Deep Learning Approaches for Hydrological Forecasting: A Systematic Review

Leishangthem Sashikumar Singh[1], Ksh. Nilakanta Singh[1],
Khumukcham Robindro Singh[1,*], Nazrul Hoque[1] and Dhruba K Bhattacharyya[2]

[1] Department of Computer Science, Manipur University, Imphal, 795003, India

[2] Department of Computer Science & Engineering, Tezpur University, Tezpur, 784028, India

INFORMATION

Keywords:
Flood prediction
recurrent neural network
long short-term memory
gated recurrent unit
convolutional neural network
generative adversarial network

DOI: 10.23967/j.rimni.2025.10.67112

Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Deep Learning Approaches for Hydrological Forecasting: A Systematic Review

Leishangthem Sashikumar Singh[1], Ksh. Nilakanta Singh[1], Khumukcham Robindro Singh[1,*], Nazrul Hoque[1] and Dhruba K Bhattacharyya[2]

[1]Department of Computer Science, Manipur University, Imphal, 795003, India

[2]Department of Computer Science & Engineering, Tezpur University, Tezpur, 784028, India

ABSTRACT

Machine learning (ML) techniques, especially Deep Learning (DL) techniques, have been applied for flood risk analysis and prediction on spatial historical data to minimize the risk of the loss of lives and properties associated with floods. In recent years, various studies have established DL as an effective approach for building potential flood prediction models. However, a thorough, systematic integration of recent advancements in DL for hydrological forecasting, their reported performance, and challenges is essential for guiding future work effectively. This paper presents a systematic survey of various DL models applied to flood prediction. This systematic review integrates studies published between 2018 and 2025 that used DL models in hydrological forecasting, such as flood prediction, streamflow forecasting, and runoff modeling. A systematic search of major electronic databases was performed using pre-defined inclusion and exclusion criteria. The overall search provided 647 records, which were narrowed down to a final collection of 45 studies after screening and full-text examination based on factors like study design, hydrological forecasting relevance, and application of DL. The review quantitatively summarizes the reported performances of various DL models, including RNN variants (LSTM, GRU), CNN, GAN, and hybrid architectures, across different hydrological forecasting tasks and datasets. Findings indicate that DL models consistently achieve high performance metrics such as NSE of (0.99), RMSE of (24.61) and MAPE of (1.73) for certain applications. Despite these advancements, significant research gaps remain, particularly concerning the scarcity of high-quality, publicly available datasets with detailed spatial information, the need for more robust real-time prediction systems with minimized false alarms, and the development of more generalized models applicable across diverse geographical regions. this review highlights the significant potential of DL in hydrological prediction as well as clearly stating the fundamental challenges that need to be overcome in order to achieve more robust and generalizable flood prediction systems.

OPEN ACCESS

Received: 25/04/2025

Accepted: 22/07/2025

Published: 27/10/2025

DOI
10.23967/j.rimni.2025.10.67112

Keywords:
Flood prediction
recurrent neural network
long short-term memory
gated recurrent unit
convolutional neural network
generative adversarial network

*Correspondence: Khumukcham Robindro Singh (rbkh@manipuruniv.ac.in).

1 Introduction

Floods are widely recognized as among the most devastating natural calamities. They cause significant loss of life, displacement, and damage to infrastructure and agriculture worldwide. The frequency and severity of flood events have increased in recent years due to global climate change. Rising temperatures, erratic rainfall, and extreme weather patterns, intensified by global warming and human activities such as deforestation and urbanization, have made flood occurrences more unpredictable and widespread [1]. Although it is difficult to prevent floods, flood prediction systems can be used to estimate the hazards and extent of flooding, helping to mitigate their impact to a certain degree [2,3]. The development of reliable and accurate prediction systems is essential not only for evacuation strategies, disaster management, analysis, and policymaking, but also for the sustainable management of water resources, agriculture, and infrastructure [4]. Traditional flood prediction methods rely on hydrological models [5] which are primarily based on river basin monitoring [6] and water level estimation [7]. While these models have been useful, their accuracy is often limited due to the inherent complexity of hydrological systems, inconsistent data availability, and the highly dynamic nature of extreme weather events. These challenges are especially pronounced in regions outside the coverage of the World Meteorological Organization's global data networks, where the lack of reliable data creates significant barriers to developing accurate flood forecasting systems [8,9].

Hydrological prediction systems are increasingly recognized as essential tools for mitigating the risks associated with loss of life, property damage, and environmental degradation. These systems typically focus on both short-term and long-term forecasting, using historical data to improve predictive accuracy. Short-term prediction models are primarily employed in early warning systems and provide forecasts with real-time, hourly, daily, or weekly lead-times to facilitate timely responses. In contrast, long-term prediction models support policy development and the implementation of preventive strategies by providing forecasts with monthly to yearly lead-times [10]. Moreover, different lead-time predictions can be examined using sequence data. Predicting flood lead-times and identifying flood-prone areas has become increasingly challenging due to the volatile and shifting nature of climate patterns, which introduces a high degree of uncertainty into the process. Therefore, the major flood forecasting models are mainly data-specific and require various simplified assumptions [11]. Development of a flood prediction system can be organized in accordance with water resource parameters [12], including water level, river flood, soil moisture, rainfall-discharge, precipitation, river inflow, peak flow, river flow, rainfall-runoff, flash flood, rainfall, streamflow, seasonal streamflow, flood peak discharge, urban flood, plain flood, groundwater level, rainfall stage, flood frequency analysis, flood quantiles, surge level, extreme flow, storm surge, and so on.

Historically, hydrological forecasting relied on physical and conceptual models, and was subsequently complemented by statistical methods including Multiple Linear Regression (MLR), Autoregressive Integrated Moving Average (ARIMA), Climatology Average Method (CLIM), Quantile Regression Techniques (QRT), and Bayesian forecasting as the established physical and statistical models used for flood prediction [10,13–16]. Although these approaches contributed significantly and provide valuable insights, they often face challenges such as limited accuracy in short-term predictions, the need for region-specific parameters, and reduced adaptability to the complex or changing flood dynamics [17]. Moreover, physical models simulate water flow using complex physical laws and differential equations [18,19], leading to high computational demands due to the need for detailed spatiotemporal resolution. These limitations make it difficult for such models to meet the timeliness requirements for real-time forecasting and early warning [19–21]. In response, the use of machine learning-based surrogate models has emerged as a significant advancement, enabling the emulation of complex physical systems with substantially lower computational demands [20–22]. Building on this progress, the field has increasingly turned to Machine Learning (ML) and its subfield, Deep Learning (DL), which are recognized as potent, data-driven approaches for developing more accurate and robust systems suitable for both short-term and long-term flood prediction [23–27].

Artificial Intelligence (AI) methods, particularly ML and its subfield DL play a crucial role in advancing prediction systems. ML and DL contribute to achieving better performance and cost-effective solutions. ML involves computer algorithms that learn and improve from past experiences without explicit programming [28]. To effectively harness machine intelligence, understanding algorithm characteristics is essential. High-quality training data is pivotal for optimal model performance. However, inadequate, irrelevant, or noisy data can hinder the direct application of ML models. Therefore preprocessing and data analysis before model application facilitate the development of successful prediction systems. ML and its subset DL encompass diverse learning tasks, including supervised, unsupervised, and reinforcement learning [29]. Supervised learning, where input datasets are labeled with desired outputs, is commonly employed when there is an ample amount of labeled data. On the other hand, unsupervised learning gleans insights directly from data for decision-making. Reinforcement learning involves trial-and-error decision-making. Notably, numerous single and hybrid ML models for flood prediction are documented [10].

DL employs ANN with multiple layers and mimics human brain's learning mechanism using input data, weights, and bias. DL techniques are increasingly popular, especially in data science, due to their computational prowess [30–32]. This has led to substantial success across various fields. In implementing DL flood prediction models, understanding the problem and relevant data, assessing feasibility, and selecting appropriate algorithms are crucial. Historical flood records, encompassing rainfall, temperature, humidity, water flow, and water level, are essential inputs obtained from sources like remote sensing technologies [33]. DL models can effectively utilize rich data from remote sensing [34]. Specifically, radar-based datasets are valuable input sources; their utility is enhanced by extensive network coverage [35] and advancements in data correction methodologies that improve the accuracy of derived products like rainfall estimates [36]. For instance, these improved radar datasets are leveraged by DL approaches for enhanced precipitation estimation and nowcasting [37], which are crucial for flood prediction and contribute to better overall predictive capabilities in hydrological applications. Practical DL applications has become common and can easily adept at handling large sequence data, due to advancements in AI and the availability of powerful graphics processing units (GPUs) [31,38]. Deep Neural Network (DNN) or DL models excels at automatically learning intricate patterns and hierarchical representations from data, overcoming some limitations of traditional methods [28,39–41]. Its ability to handle complex, high-dimensional, and sequential data makes it particularly well-suited for time series forecasting problems in hydrology [28,42]. Open-source tools, educational resources, and generalized algorithms further enhance DL's applicability [43].

Recent studies have demonstrated the effectiveness of various DL architectures in improving the accuracy and robustness of hydrological predictions across different lead-times and spatial scales [44–46]. These studies showcase the methodological contributions of advanced ML/DL in handling complex data patterns and non-linear relationships, reflecting a significant shift towards more sophisticated data-driven approaches for flood forecasting and management. Comparisons of recent ML models further highlight their performance in various hydrological contexts [46]. This advancement has motivated us to conduct an exhaustive survey of various DL methods for flood risk analysis. Several review papers have explored aspects of flood prediction and the application of different modeling techniques, including some focusing on ML and DL in hydrology [10,47–49]. However, a comprehensive systematic review specifically dedicated to synthesizing the diverse range of DL architectures applied to various hydrological forecasting tasks, analyzing their reported performances using quantitative metrics, and identifying key research gaps and future directions based on recent literature remains valuable to guide future research efforts in this rapidly evolving field.

This systematic review is intended to present a thorough and current overview of DL methods for hydrological forecasting. Specifically, the study emphasize on identifying and classifying the various DL architectures utilized, comparing the types of hydrological forecasting issues dealt with (e.g., flood forecasting, streamflow forecasting, and runoff modeling), noting the datasets and quantitatively summarizing reported performance levels of various models and applications. One of the novelties in this review is its systematic approach to quantitatively assessing the reported performance across a wide range of studies and providing an detailed analysis of the state-of-the-art DL architectures used. Further, the review is presented with a new taxonomy of DL techniques in this field and proposes a generic framework based on the synthesis of effective methods. The importance of this work stems from the necessity for precise and reliable hydrological forecasting systems in the face of increasing climate variability and extreme events. Integrating the state of the art of DL for this field, this review is an invaluable guide for researchers who aim to identify promising avenues of future research, practitioners seeking to utilize effective forecasting tools, and policymakers in need of evidence-based information for disaster risk reduction and water resource planning. The main contributions of this review work are mentioned below:

1. A taxonomy of the DL methods for flood prediction.
2. A detailed comparative analysis of the existing survey works.
3. Discussion on a set of prominent DL methods for flood prediction with their architectures, pros and cons.
4. A comparative study on some prominent DL methods for flood prediction with reference to a set of essential parameters.
5. Some issues and research challenges of flood prediction using DL.
6. Recommendation of an experimental framework for robust DL-based flood prediction to address early warning challenges.

The rest of the paper is organized as follows. Section 2 describes a detailed background study that has been performed with this review. Section 3 presents the methodology used in this study. Section 4 briefly explains a comprehensive literature review of the current state of the art of DL methods for flood prediction. Section 5 presents the flood-related dataset descriptions and preprocessing techniques. Section 6 discussed the comparative studies of DL models utilized for early warning flood prediction system. Section 7 presents the results and discussion, which include the critical research issues and challenges, and the proposed experimental framework. The last Section 8 presents the concluding remarks.

2 Background Study

To provide context for this systematic review and position it within the existing body of literature, relevant previous work was examined. This included a focus on existing review papers discussing various methods used in flood forecasting and management. We specifically examined 17 review papers that analyzed flood prediction models from diverse methodological perspectives. It covers a wide range of techniques applied in flood forecasting and management, including statistical approaches, ML algorithms, DL techniques, and hydrological modeling frameworks. By synthesizing insights from these previous works, this review seeks to identify valuable insights into the strengths, weaknesses, research gaps, and challenges associated with various techniques employed in flood prediction models. It also aims to highlight prevailing trends, suggest potential future directions, and situate the present review within the broader development of flood prediction methodologies. Several review papers have been studied to explore the applications of diverse approaches used in flood prediction and management. These approaches include Remote Sensing (RS) and Geographical Information System (GIS) [48,50–53]; Bayesian methods [54]; empirical, hydrological, conceptual and physical models [15,16,55,56]; Light Detection and Ranging (LiDAR) and Digital Elevation Model (DEM) [57]; ML, ANN and DL [10,47,49,58]; image processing and computer vision [59,60]. Some of these articles also provide valuable insight and practical guidance on the implementation of flood information management systems. The summary of the findings from the surveyed works is presented in Table A1.

Recent advancements in DL architectures have shown promising potential in addressing the urgent need to minimize property damage and human loss caused by floods. This summary reveals that researchers are increasingly focusing on the development and deployment of DL models for the prediction of key hydrological indicators such as runoff, river water levels, streamflow patterns, and flood forecasts. These models aim to enhance early warning systems, thereby supporting timely preparedness, prevention, and response strategies. Several studies have explored both conventional and advanced DL approaches in flood prediction. Traditional neural networks continue to be widely applied for runoff and streamflow prediction, while more recent work has shifted toward leveraging the capabilities of specialized architectures such as Convolutional Neural Networks (CNN) and Generative Adversarial Network (GAN). These models have been particularly effective in tasks like flood susceptibility mapping and high-resolution forecasting, offering improved spatial and temporal accuracy. Moreover, there is a noticeable trend toward optimizing DL performance through the integration of preprocessing techniques and parameter tuning. For instance, reference [47] presents a method that shows reducing the complexity of rainfall and runoff input parameters prior to model training significantly improves prediction accuracy. This indicates a growing awareness of the importance of data quality and feature selection in DL-based hydrological modeling. In addition to architectural innovations, many researchers are also experimenting with hybrid models and ensemble methods to enhance robustness and generalizability. Such approaches effectively manage the uncertainty and variability inherent in flood-related data. The following section of this review presents a technical summary of these DL methodologies and offers a concise evaluation of their practical applications in flood prediction and management. This synthesis highlights current trends and identifies future research directions, particularly in improving model interpretability and operational deployment in real-time systems.

3 Methodology

This systematic review presents an overview of the use of DL techniques in hydrological prediction. Its primary purpose is to provide an overview of the current state of the art in DL methodologies and to determine the gaps in existing studies in order to better address their contribution to hydrological modeling. The review was conducted in line with standard systematic review protocols to present exhaustive, transparent, and unbiased literature review of the use of DL in this domain. The review was guided by the following specific research questions: What DL approaches have been applied to hydrological forecasting (including flood prediction, streamflow forecasting, and rainfall-runoff modeling)? What datasets and evaluation metrics have been commonly used in these studies? What are the reported performance levels of different DL models in hydrological forecasting tasks? What are the key challenges and future directions identified in this research area?

A comprehensive search was carried out to identify relevant studies published between 2018 and 2025, from major electronic databases including IEEE Xplore, Google Scholar, Elsevier (ScienceDirect), SpringerLink, and Wiley Online Library. The search strategy was developed with a combination of keywords and, where feasible, controlled vocabulary, including DL and hydrological forecasting related keywords. Key search terms were "Deep Learning," "DL," "hydrological forecasting," "flood prediction," "streamflow forecasting," "rainfall-runoff model," "water level prediction." These terms were combined using Boolean operators (AND, OR) to produce long search strings for every database. Screening of the reference lists of the included studies and relevant review articles was performed to identify further relevant studies. The initial database search yielded a total of 647 records. Following the removal of 93 duplicate records, 554 unique records remained for screening. We then collectively screened the titles and abstracts of these records based on predefined inclusion and exclusion criteria. These criteria were as follows:

Inclusion Criteria:

1. Original research articles.
2. Studies focusing on hydrological forecasting tasks (including flood prediction, streamflow forecasting, rainfall-runoff modeling, and water level prediction) using DL models.
3. Studies published between 2018, and 2025.
4. Studies published in English-language articles.
5. Empirical research reporting the application of DL models on hydrological and meteorological data.

Exclusion Criteria:

1. Review articles, survey papers (except for background or reference screening) specifically those published before 2013.
2. Theoretical papers with no empirical use or findings.
3. Studies applying ML techniques aside from DL (unless as a baseline for comparison).
4. Hydrological studies on subject other than prediction (e.g., water quality, observing drought without prediction component).
5. Editorials, book chapters, reports, or conference abstracts (unless full papers available and language other than English).
6. Studies not published within the specified timeframe for primary studies (i.e., outside 2018, and 2025).

Studies were screened for eligibility if they were original research articles that used DL models for hydrological forecasting and provided empirical results; studies were excluded if they were not original research, did not use DL, or were not relevant to hydrological forecasting. This initial screening yielded 554 potentially relevant articles. The full text of these papers was then retrieved and independently reviewed using the same inclusion and exclusion criteria. Disagreement was resolved by discussion. The study selection process is summarized in Fig. 1. This full-text screening led to the exclusion of 509 papers. The primary reasons for exclusion at the full-text stage were: not focusing on hydrological forecasting (136 studies), not using DL (167 studies), not being original research (e.g., reviews other than those included in the background) (61 studies), lack of empirical results or insufficient detail (53 studies), not freely available (92 studies). This resulted in a final list of 45 studies included in this systematic review. Data were extracted from the included studies using a standardized form to capture relevant information on study characteristics, methodology, results, and performance metrics. The quality and risk of bias of each included study was evaluated to assist in interpreting the findings. The extracted data were synthesized narratively, and a comparative analysis of reported performance was undertaken, acknowledging the heterogeneity across studies.

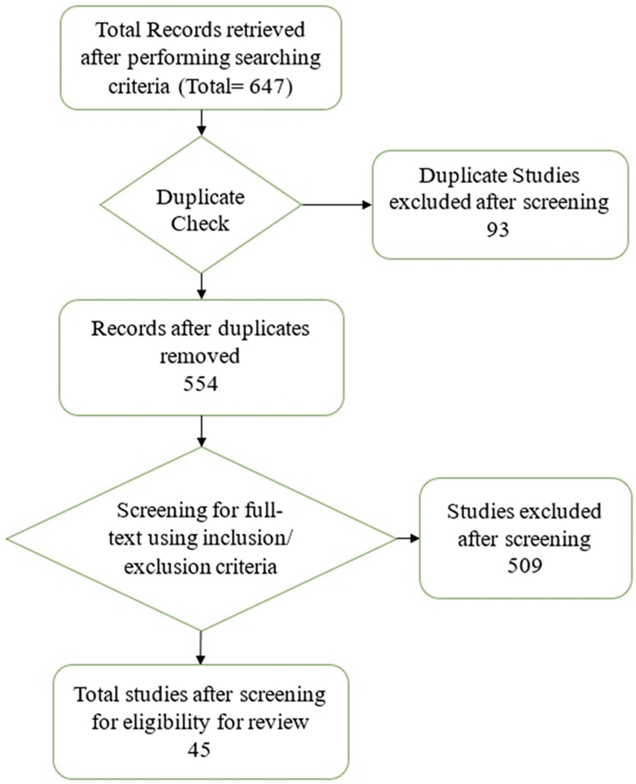


Figure 1: Flowchart illustrating the systematic study selection process for DL-based hydrological forecasting

4 DL Methods for Flood Prediction

A DNN is a distinct category of ANN, characterized by its inclusion of multiple hidden layers, enabling it to process large volumes of data and offer more parameters capable of capturing underlying patterns in the data. DNN imitates the human brain's functionality, where neurons or nodes interconnected with weighted connections process input information to produce relevant outputs. This structural similarity divides ANN into input, hidden, and output layers. A block diagram of DNN is illustrated in Fig. 2.

The application of DNN spans various domains, including AI chatbots that offer detailed responses to queries, self-driving vehicles, language recognition, text generation, and more. These methods efficiently solve complex algorithms, conserving computational resources and space. While standard neural networks might have a single hidden layer, DNN has multiple hidden layers that can learn nonlinear patterns, which is helpful in addressing intricate problems. The depth of a DNN can be adjusted to suit the specific requirements of the task at hand. DL models are trained using random weights for input data and can be supervised, unsupervised, or hybrid [43]. DL encompasses a family of techniques for representation learning, enabling machines to automatically learn features from raw data and employ them for specific tasks. With multilayer neural networks at their core, DL models excel by autonomously extracting high-level or abstract features from raw data, eliminating the need for manual feature engineering [61]. This capacity allows transfer learning, where features learned on one dataset can be transferred to a different dataset. The flexibility of DL lies in its ability to expand its layer count and neuron density, thus empowering the creation of powerful machines capable of handling complex functions [28]. The remarkable strides made by DL in solving AI problems are succinctly summarized in [31]. This paper primarily focuses on flood lead-time predictions, emphasizing early warning and flood risk reduction. Predicting lead-times, such as yearly, monthly, daily, hourly, or real-time forecasts, is key to mitigating flood damage. The discussion encompasses various types of DL methods like supervised, unsupervised, and hybrid learning applied in various lead-time prediction scenarios. The taxonomy of different DL methods employed in flood prediction is illustrated in Fig. 3.

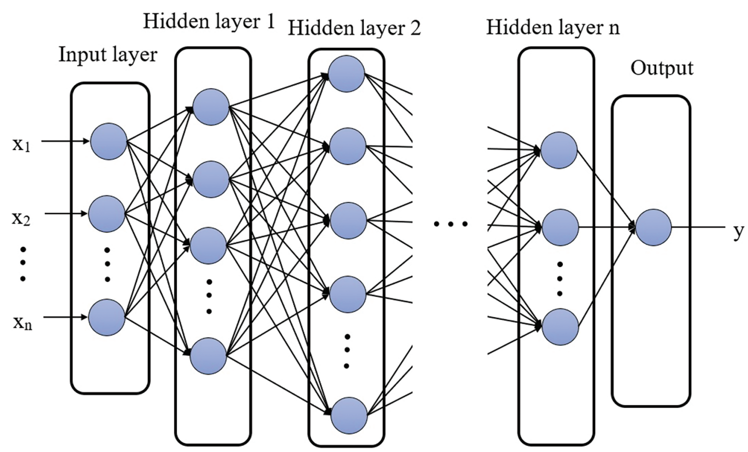


Figure 2: Architecture of densely connected ANN also known as DNN

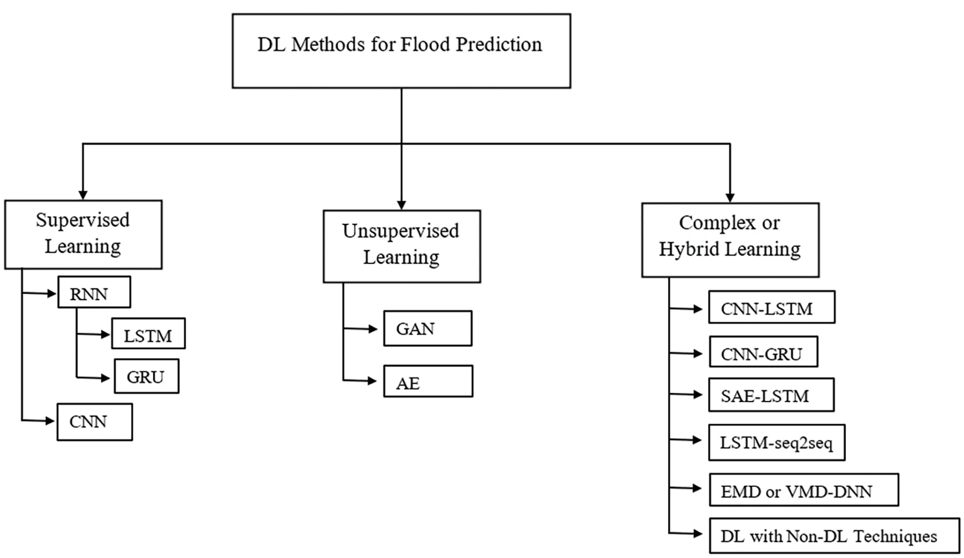


Figure 3: Taxonomy of DL methods for flood prediction

4.1 Supervised

In supervised DL, the primary objective is to achieve the intended output value using a labeled dataset. For instance, when training a model for classification, regression, or sequence prediction, a substantial labeled dataset, including corresponding categories, is necessary. Subsequently, intensive computations are conducted to adjust the tunable weights of input parameters. This adjustment aims to enable the model to attain the highest possible accuracy in generating the desired output. Several supervised DL models have been employed for flood prediction, such as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and CNN [28].

4.1.1 Recurrent Neural Networks

RNN and its variations, like LSTM and GRU, have emerged as a widely used set of networks for constructing models to predict flood occurrences based on time series data [38]. An RNN [40] is a type of ANN where hidden units possess interconnections that loop back on each other. The fundamental purpose of RNN is to effectively process sequences or time series data. The configuration of a recurrent neural network is illustrated in Fig. 4.

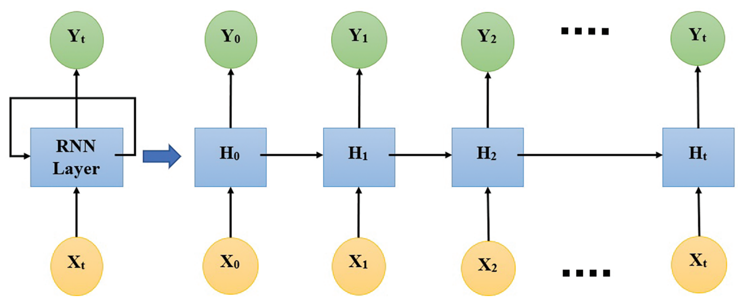


Figure 4: Architecture of RNN

The mechanism that enables the transition from one hidden unit to the next in an RNN is termed forward propagation. This network maintains a memory that captures information from previous time steps of a given data sample and strives to pass it along to subsequent time steps. This memory aids in processing sequential or time-dependent predictors by utilizing past information and the current input as inputs for the following hidden layer [28]. Mathematically, the forward propagation function of an RNN can be expressed as $h_t = f\,(x_t W_x + h_{t-1} W_h)$. In this equation, h_t signifies the output at the current time step, x_t stands for the present input, W_x represents the weight associated with the current input, h_{t-1} corresponds to the output or information from the previous time step, and W_h denotes the weight linked to the prior output. Upon completing all the time steps, an activation function is applied to the final output to compute the error, which is then propagated backward to update the weights within the RNN. The intricate derivations of the RNN's mechanisms are thoroughly expounded in the resource [38]. RNN is advantageous for processing inputs of varying lengths, maintaining a consistent model size irrespective of the sequence length. However, when dealing with extensive sequence data, both simple and deep RNN models encounter challenges such as the vanishing gradient issue, which results in slow and challenging computations [41]. Additionally, the exploding gradient problem arises, causing gradient values to become excessively large during backpropagation [28]. To address these challenges, two alternative variants of RNN networks have been introduced: LSTM [62] and GRU [63]. These variations have been developed specifically to tackle the aforementioned problems associated with traditional RNN.

LSTM: The LSTM network [62] is specifically designed to tackle the problem of extended patterns of connection within sequential datasets. This is particularly important in cases where traditional RNN struggle due to issues like vanishing gradient [64]. The LSTM architecture, illustrated in Fig. 5.

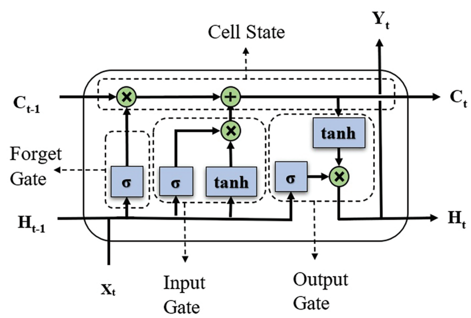


Figure 5: Architecture of LSTM Cell

Three gates are employed to manage information within the cells, including a) The forget gate eliminates unnecessary information from the cell state. b) The input gate incorporates valuable information by employing the sigmoid function to regulate and filter data into the cell state. This information is further processed using the tanh function. c) The output gate generates the output based on the current cell state information and transmits it to the subsequent cell state. LSTM neurons generate distinct output values to convey what knowledge has been gleaned from preceding sequences and then relay it to the subsequent hidden state. A novel variant of LSTM known as Entity-Aware-LSTM, introduced by [65], involves incorporating additional inputs containing static features at each time step. This architecture handles static and dynamic inputs separately: static inputs are directed to the input gate, while dynamic inputs are distributed to the forget gate, memory gate, and output gate.

GRU: Although LSTM networks have gained widespread recognition as effective variants of RNNs, successfully addressing the vanishing gradient issue and contributing to significant advancements in NLP and time-series prediction fields, they do have significant drawbacks in terms of their time complexity. In contrast, GRU networks [63] are lesser-known but equally potent alternatives. They offer quicker computation speed while maintaining effectiveness [66]. The primary purpose of GRU closely resembles that of the basic RNN, except for the internal mechanisms operating within each recurrent unit. Much like LSTM, GRU controls the information flow within its gated unit without employing separate memory cells. The output from GRU neurons serves a dual purpose: conveying valuable learned features to both the subsequent layer and the next neuron within the same layer. The architectural layout of GRU is depicted in Fig. 6.

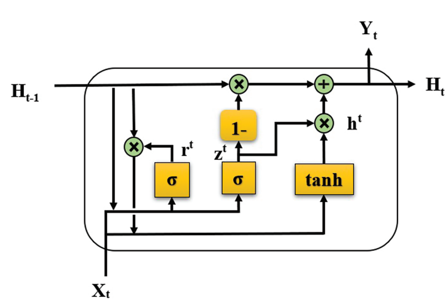


Figure 6: Architecture of GRU Cell

The authors [67] create RNN and LSTM models to predict monthly precipitation using monthly average rainfall time series data. The study suggests that LSTM exhibits better performance compared to RNN and has the potential to serve as an alternative for assessing climatic events. The application of LSTM [42] is extended to create prediction models for reservoir inflow and outflow across hourly, daily, and monthly time scales. This investigation demonstrates the superior performance of the LSTM model in comparison to Backpropagation neural networks (BPNNs) and other ML models. LSTM model [68] is proposed to simulate rainfall-runoff prediction from meteorological observations by utilizing 241 catchments of the CAMELS dataset of the USA. The results are obtained with better performance than the physical model. Reference [69] presents a streamflow forecasting framework using an LSTM model based on past meteorological and streamflow datasets for short and long-term purposes. The results show a better performance as compared to the physical model. Reference [70] constructs the autoregression and LSTM models to predict daily flow, compares the results, and demonstrates the feasibility of the models. Reference [71] explores the use of an RNN model for real-time reservoir inflow and outflow prediction. The authors [72] utilize both LSTM and GRU to predict short-term rainfall runoff for the Yutan catchment station. The study indicates that both LSTM and GRU models exhibit comparable and improved performance compared to ANN models when optimizing time steps. The findings suggest that, specifically for short-term runoff prediction, the GRU model is favored. An LSTM-based discharge prediction model [73] is designed for flash flood forecasting in mountainous regions. This model, termed LSTM flood forecasting (LSTM-FF), leverages spatial and temporal dynamics information from both observed and predicted rainfall data, along with early discharge information, as input parameters. The study showcases the model's capability to accurately predict flash floods. In the study [65], they develop an entity-aware LSTM network for regional rainfall-runoff modeling. This model incorporates meteorological time series data from 531 basins sourced from the CAMELS dataset, along with static catchment attributes, enabling the learning of catchment similarities. The outcomes demonstrate superior performance compared to previous hydrological models in terms of predictive accuracy. A study developed and evaluated RNN and LSTM models for monthly streamflow forecasting in Iran's Zarrine River basin, revealing that RNN outperforms LSTM in accuracy due to its ability to filter redundant information [74].

4.1.2 Convolutional Neural Network

CNN architectures predominantly focus on developing flood susceptibility models and innovative flood monitoring systems. They are primarily designed for processing data represented as multi-dimensional arrays [75] and are particularly adept at identifying spatial and temporal correlations found in recurring patterns in input data formats like photos, movies, and texts when appropriate filters are used. Due to these, they are extensively utilized in computer vision and image classification tasks. A CNN architecture typically consists of the following components: a) Convolutional layer-comprises multiple cascaded convolution filters, systematically generating a 3D tensor. It performs a convolution operation over the input vector to extract spatial information. b) Pooling layer-combines the convoluted information to further reduce the feature size. Common functions used in the pooling layer include maximum, average, and sum. c) A fully connected layer is usually utilized to process the feature generated by the preceding layers to generate the final output [76]. The 1D-CNN demonstrates efficiency in processing sequence or time-series data, providing a number of benefits like great noise resistance, the capacity to extract rich and insightful features irrespective of temporal correlations, and quick complex computation. The architecture of a 1-D convolutional neural network is depicted in Fig. 7.

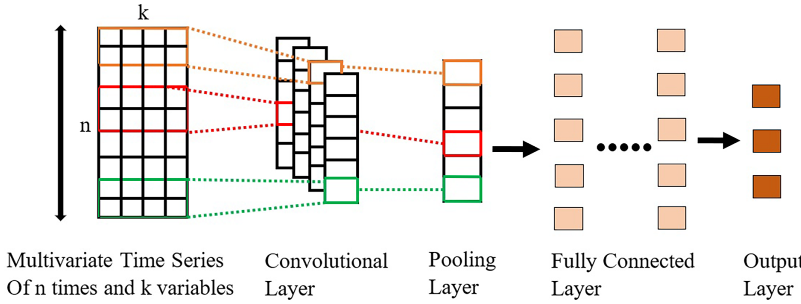


Figure 7: Architecture of CNN working on time series

Here, the input data consists of ‘n’ instances of time series data with ‘k’ variables. Various filter sizes are employed to process the input data within the convolutional layer, extracting profound features. An activation function is subsequently applied to the convolutional layer’s output. For dimensionality reduction and feature selection, the Max-pooling technique [77] is a widely used pooling technique. This involves selecting the maximum value from each output vector produced by the convolutional layers. The pooling kernel size can be determined using the formula n/p, where n represents the length of the time series and p is the pooling factor. The outcome from the pooling layer is a single vector that holds significant information for generating the desired output. This vector serves as the input for the fully connected layer. In [78], a CNN network is proposed to map flood susceptibility areas for preventing and reducing flood damage with an appropriate mapping technique. The model uses a dataset having 13 flood-triggering factors and 108 flood historical events in the study area to train different dimensional CNN architectures, such as 1D-CNN, 2D-CNN, 3D-CNN, 1D-CNN-SVM, 2D-CNN-SVM, 3D-CNN-SVM, for flood susceptibility analysis. The authors use CNN architectures for classification and feature extraction. The model shows better performance than other benchmark models like SVM. In [79], a Dilated Causal CNN (DCCNN) model is presented to provide 1-h, 3-h, and 6-h lead-time water-level forecasts on rainfall and water-level data provided by Taiwan Water Resources Agency and Taiwan Typhoon and Flood Research Institute. In this work, the performance of the proposed DCCNN model is compared with the existing models based on Multilayer Perceptron (MLP) and SVM using various performance measures. The authors conclude by observing the outstanding performance of DCCNN in forecasting real-time extreme events during typhoons. In [80], a scalable flood level monitoring approach is proposed that exploits the existing surveillance camera systems to examine the flooded area using a Deep Convolutional Neural Network (DCNN). U-net CNN [81] is used to perform flood segmentation from a surveillance camera’s viewpoint. A simple CNN model with a novel index is employed to provide the qualitative flood level trend information of the segmented images. A flood inundation forecasting model is presented in [82], which is used to estimate the water levels using computer vision algorithms by employing the CNN network and edge detection techniques. This model uses new data sources, including images and videos captured by smartphones and CCTV cameras, to identify the water surface in the images and calculate the water level from physical measurements.

4.2 Unsupervised

Unsupervised DL models are effective in the estimation of probabilities and likelihoods concerning unstructured data. This category of DL models possesses the capacity to generate novel outcomes relevant to a given problem. These models glean insights directly from the dataset itself and function by assimilating information or grouping data points to make informed decisions. They tackle problems by relying on these deductions. Various unsupervised DL models, such as GAN, Autoencoder (AE) neural networks, Self-Organizing Maps, Restricted Boltzmann Machines, and Deep Belief Networks, are explored to address real-world challenges. In the context of flood prediction models, particular attention is given to GAN and hybrid variations of AE networks. These models are leveraged to create effective solutions for flood prediction by harnessing their unsupervised learning capabilities.

4.2.1 Generative Adversarial Network

GAN model comprises two neural network models, namely the generator and the discriminator [83]. These models engage in a concurrent competition, forming a two-player minimax game.

Illustrated in Fig. 8, the GAN architecture involves the generator's task of creating lifelike synthetic samples from random input vectors. Simultaneously, the discriminator is trained to differentiate real data from artificially generated data through binary decisions. In this scenario, the generator endeavors to deceive the discriminator by crafting new fabricated data that appears authentic, while the discriminator strives to distinguish between actual and fake data [84]. The iterative competition improves both models as they adjust their weights to optimize their respective loss functions. GANs are versatile and can perform tasks such as data synthesis, time series classification and prediction [85], image super-resolution [86], and image-to-image translation [87]. In [88], a flood susceptibility and innovative flood monitoring system named FloodGAN prediction model is proposed. Utilizing a deep convolutional GAN, this model consists of a generator and a discriminator, aiming to forecast maximum inundation extent and water depth for short-duration rainfall events. It employs synthetic and actual rainfall event image data. The model demonstrates superior accuracy and speed compared to hydrodynamic models. The article [89] presents a real-time flood prediction model employing a deep convolutional GAN. This research utilizes a high-resolution MIKE surface flow dataset, collected at 10-min intervals, to develop the model. The study reveals a strong positive correlation between incoming wave data and the accuracy of flood predictions. Furthermore, the model showcases a remarkable capacity to forecast extended lead-time dynamic flow conditions effectively. In [90], a radar-driven conditional GAN model is introduced for the task of precipitation nowcasting. This model employs radar reflectivity image data specifically within the Soyang-gang Dam basin. The study reports achieving superior performance compared to other ML models and effectively tackles the need to enhance prediction accuracy in the presence of data imbalance issues.

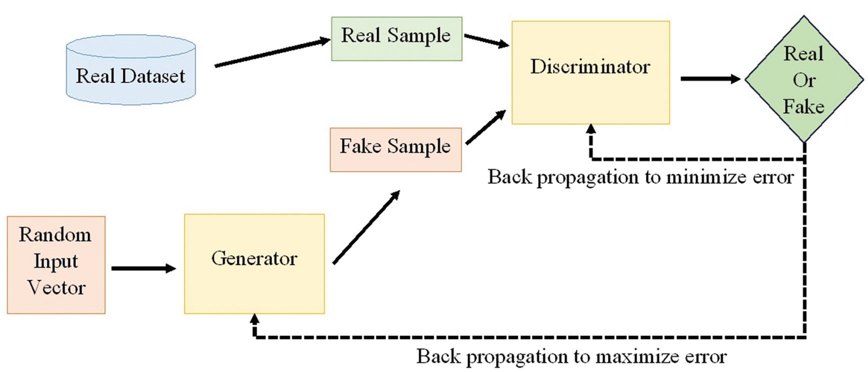


Figure 8: Architecture of GAN

4.2.2 Autoencoders

AE are a category of neural networks structured around three key components: encoder, code (or context vector), and decoder [30]. The encoder reduces input datasets into a low-dimensional representation, capturing essential features, while the decoder reconstructs outputs from this representation. The architectural layout of an AE is illustrated in Fig. 9.

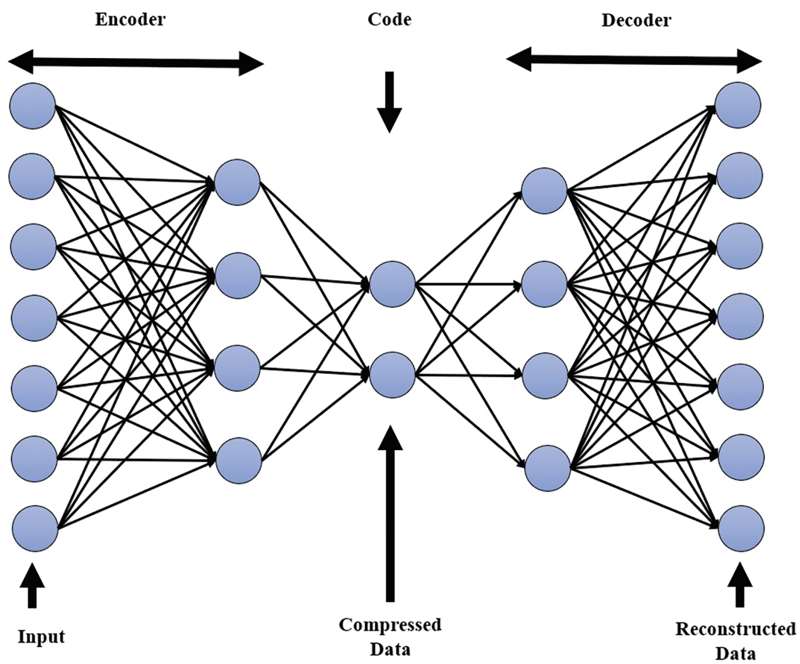


Figure 9: Architecture of Autoencoder

It creates a compressed knowledge representation, often referred to as the context vector, for the dataset. Notably, in autoencoders, the number of neurons in the output layer matches that in the input layer. Autoencoders follow an unsupervised learning approach, aiming to reconstruct the code into its original form rather than predicting specific target outputs for given inputs. Various adaptations of autoencoders are employed for constructing flood prediction models. These include Stacked Autoencoders (SAE) with multiple hidden layers, where each hidden layer's output serves as input for successive layers, as well as complex or hybrid autoencoder architectures like Encoder-Decoder (En-De) and Sequence-to-Sequence (seq2seq). The intricate structures of autoencoders, such as SAE, En-De, and seq2seq, for flood prediction purposes, are discussed in detail in Section 4.3.3.

4.3 Hybrid Model

Hybrid DL structures often comprise a combination of two or more supervised or unsupervised DL models. Diverse integrations of distinct DL models have been developed to construct intricate DL frameworks, like CNN-LSTM, CNN-GRU, SAE-LSTM, En-De-LSTM, and LSTM-seq2seq, aiming for robust flood forecasting. Notably, in multiple investigations, the performance of physical flood simulation models has demonstrated substantial improvement through the integration of cutting-edge models. The incorporation of supplementary optimization algorithms and data pretreatment techniques, namely the decomposition of rainfall and runoff prior to employing models, further bolsters the precision of flood prediction models. These hybrid models are meticulously designed to extract more insightful features, thereby enhancing the efficacy of various flood prediction models.

4.3.1 CNN-LSTM

Hybrid CNN-LSTM architectures are used to address the visual time-series prediction tasks and predict textual descriptions associated with sequences of images, videos, and activities, like activity recognition. In the context of flood prediction, a composite CNN-LSTM architecture is constructed, employing both CNN and LSTM components. The schematic representation of the CNN-LSTM hybrid model can be found in Fig. 10.

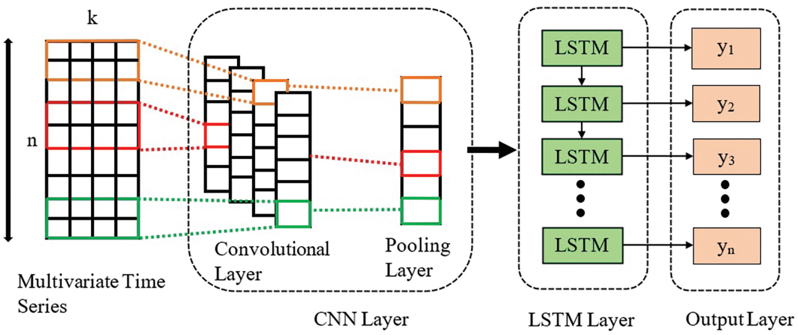


Figure 10: Architecture of CNN-LSTM hybrid model

Here, CNN is utilized to capture temporal features from the input data prior to its conveyance to the LSTM network. The features extracted are then funneled into the LSTM network. There are scenarios where CNN struggles to effectively handle input subsequences related to flood events and their pertinent characteristics, which RNN handles more adeptly. However, RNN might encounter challenges when dealing with long temporal sequences. The hybrid model offers a solution by employing CNN to extract more salient features, thus shortening lengthy temporal sequences. In this hybrid framework, CNN functions as a feature extractor for LSTM, enhancing the efficiency of predictions by condensing and enhancing the lengthy sequences. In [91], a hybrid CNN-LSTM model, encompassing LSTM and CNN, is introduced for forecasting monthly streamflow and rainfall. The application of convolutional layers aids in the extraction of temporal features. The outcomes of the CNN-LSTM model surpass those of both MLP and basic LSTM. Another study [92] applies the CNN-LSTM approach to a daily runoff forecasting model using data from the Feilaixia catchment, demonstrating its superiority over the LSTM model. Likewise, a rainfall-runoff model utilizing CNN-LSTM is proposed by [93], showcasing the network's efficacy in estimating flood alerts. Additionally, reference [25] explores the enhancement of flood forecasting model accuracy through the use of fundamental LSTM, CNN-LSTM, and LSTM augmented with spatiotemporal attention mechanisms. A study on the Mahanadi River basin demonstrated that hybrid models, particularly CNN-LSTM, outperform standalone models in short-term streamflow forecasting, especially up to a 2-day lead-time [94]. This study highlighted the importance of upstream discharge and meteorological inputs, and suggested that using a constant basin lag time improves performance under limited computational resources. It also noted limitations due to manual hyperparameter tuning and recommended future integration of land use and climate change analysis for improved forecasting. A Python-based DL framework called FlowDyn is presented for global daily streamflow prediction using models like LSTM, GRU, and CNN-LSTM, integrated into a web-based visualization tool in [95]. These models are trained on data from 183 catchments, and the CNN-LSTM model achieved an average accuracy of 0.83, demonstrating strong adaptability to diverse hydrological behaviors. A study demonstrates that CNN-LSTM and LSTM, effectively improve flood forecasting accuracy in data-scarce regions like the Ouémé basin in Africa, achieving high performance [45]. The integration of DL with remote sensing offers valuable tools for proactive flood risk management, although challenges remain in model generalization and uncertainty handling.

4.3.2 CNN-GRU

The hybrid CNN-GRU is modeled by employing a CNN layer followed by a GRU layer, and a dense layer. Fig. 11 shows a general CNN-GRU model structure.

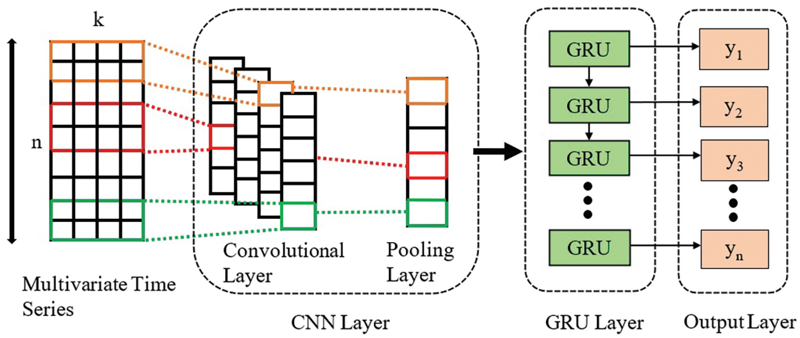


Figure 11: Architectural diagram of CNN-GRU hybrid model

In this architecture, the CNN layer works as a feature extractor of the model to obtain more prominent features of the input variables for further processing by the GRU layer, and finally connected to the fully connected dense layer. Reference [96] proposes a CNN-GRU neural network model that combines CNN and GRU to predict river water level and its anomaly. Here, the CNN layer is applied to extract the temporal features from the dataset at the water level station and then map to the GRU layer for analyzing the information and the time series flood prediction. This architecture is more convenient for problems with spatial and temporal structure in their input, such as the 1D structure of sequence data, a pattern of images in a video, words in a text, etc. The study compares the performance of the CNN-GRU model with four other models and demonstrates its superior predictive capabilities. Reference [97] proposes another CNN-GRU model for monthly streamflow prediction, applying it to many watersheds with global hydroclimatic characteristics to examine the effects of spatial and temporal analysis on predictive performance. This study shows that the CNN-GRU model is more suitable for monthly streamflow prediction with large drainage areas. In [98], CNN-GRU and CNN-LSTM models are built to simulate daily, weekly, and monthly streamflow of two different climatic regions using time series hydrological and meteorological data. They compared their results with MLP, LSTM, and GRU. The performance of CNN-GRU is found to be better for simulating the daily, weekly, and monthly streamflow.

4.3.3 SAE-LSTM

The SAE-LSTM model combines SAE with LSTM. The SAE neural network serves as a crucial feature representation layer, and the LSTM further processes these features to produce the output. The architecture is visualized in Fig. 12.

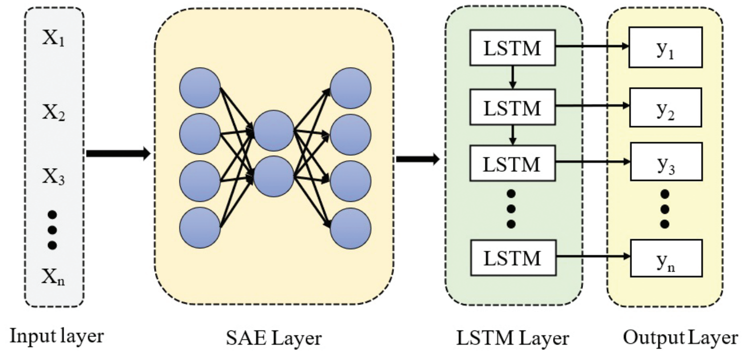


Figure 12: Architectural diagram of SAE-LSTM hybrid model

The training process of this model includes training the model on raw data to derive meaningful features in an unsupervised manner. This process involves extracting improved features from the input data, which are less susceptible to errors. The enhanced features yielded by the SAE network are then passed to the LSTM network, comprising higher-order features aimed at simulating the desired output predictions. The architecture of the LSTM network is significantly influenced by the output features obtained from the SAE network. Additionally, SAE-LSTM architectures can encompass encoder and decoder components, featuring hidden LSTM layers that together form the architecture of LSTM-seq2seq. This variant is primarily employed for tasks such as language translation. The arrangement of the LSTM-seq2seq model is depicted in Fig. 13.

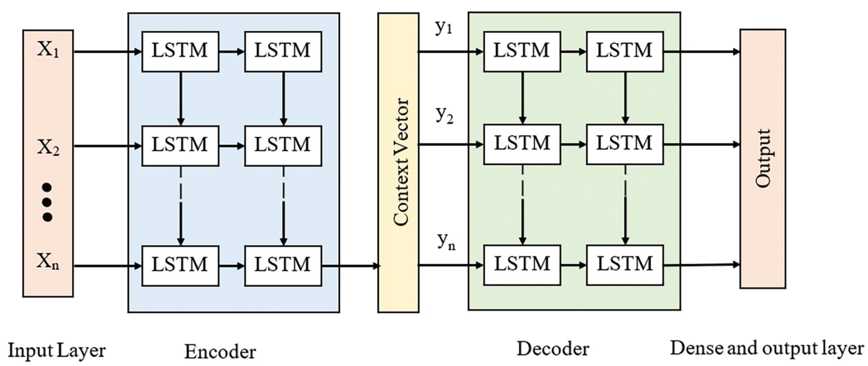


Figure 13: Architectural diagram of LSTM-seq2seq hybrid model

This model takes the predictor's input variables in the encoder layer of the LSTM. The ultimate output of the encoder LSTM is retained in a context vector (essentially the network's hidden state or encoded vector), which is subsequently fed into the decoder LSTM layer. In the decoder layer, the model learns the characteristics present in the original input to generate multiple outputs, which are then passed through a fully connected dense layer to produce the final output.

Some studies emphasize the construction of AE, En-De, and seq2seq. An LSTM-based seq2seq model [99] is proposed for predicting hourly runoff over a 24-h span using data from rainfall, runoff, and monthly evapotranspiration in two distinct watersheds. This model demonstrates superior forecast accuracy in comparison to linear regression, lasso regression, Ridge regression, support vector regression, Gaussian processes regression, and LSTM for short-term flood prediction. Likewise, reference [100] presents an LSTM-driven architecture known as the En-De LSTM model, tailored for forecasting runoff in multi-step flood prediction scenarios. Their approach involves leveraging hourly hydrological data derived from 23 typhoon events, with prediction horizons extending up to 6 h. The En-De LSTM model is shown to exhibit enhanced precision and reliability when applied to multi-step flood forecasting tasks. In [101], a novel feature-enhanced regression model is introduced, designed through the integration of an SAE with LSTM, leveraging historical daily discharge data. This model operates by taking one week's worth of data as input and predicting the corresponding daily discharge values. Significantly, it surpasses the performance of CNN, yielding more favorable outcomes. Furthermore, reference [102] adopts a distinctive approach by combining Empirical Mode Decomposition (EMD) with the En-De LSTM architecture. This fusion is employed to devise a streamflow forecasting model for the Yangtze River, utilizing monthly streamflow data sourced from the Hankou Hydrological Station. The model's performance excels, particularly in the context of minimum periods spanning 6 months and 12 months.

4.3.4 EMD or VMD DNN

This hybrid model integrates EMD or variational mode decomposition (VMD) techniques with a DNN. EMD [103] is a powerful analysis method specifically suited for nonlinear and non-stationary data. It decomposes the original input signal based on the intrinsic time series characteristics of the data, all within the time domain. Conversely, VMD [104] is a non-recursive adaptive signal processing technique employed to disassemble non-stationary complex input signals into finite bandwidth frequency components. The primary objective of these techniques is to acquire high-frequency data, thereby extracting latent informative features for predictive modeling. The architectural framework of this hybrid model is illustrated in Fig. 14.

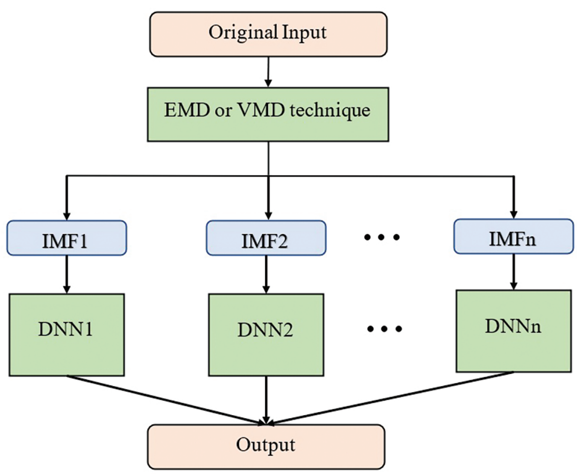


Figure 14: Architectural diagram of EMD or VMD-DNN hybrid model

The initial step involves feeding the original input data into either the EMD or VMD algorithm, generating Intrinsic Mode Functions (IMF), a finite set of components. Subsequently, the decomposed IMF is normalized within a 0 to 1 range, serving as input for the DNN model. The DNN model is then trained with optimal hyperparameters to forecast future outcomes based on each decomposed IMF. In [105], a single forecasting model anchored in LSTM is presented for predicting daily streamflow up to 7 days in advance. Notably, it incorporates three distinct data preprocessing techniques—ensemble

EMD, VMD, and Discrete Wavelet Transform (DWT) to determine effective model parameter configurations for forecasting daily streamflow series. Among these approaches, the SF-VMD-LSTM model demonstrates robustness and efficiency in its forecasting capabilities. Furthermore, reference [106] presents a hybrid model that combines VMD and DNN. This model employs VMD to decompose daily runoff series data into multiple IMF components. The results highlight superior prediction performance compared to five other hybrid models under consideration.

4.3.5 DL Based Hybrid Models with Non DL Techniques

This section presents several notable hybrid models that integrate LSTM with non-DL techniques, such as hydrological or physical models, optimization algorithms, and data pretreatment techniques. GHM-LSTM [107] is a physics-guided hybrid model [108] that combines the Global Hydrological Model (GHM) with LSTM. In this model, the output generated by the GHM model is utilized as the input for the LSTM. This integration allows the combined capabilities of physical insights from GHM with the predictive capabilities of LSTM. The model is employed with a dataset comprising basin-wide daily average air temperature, rainfall, wind speed, and simulated daily discharge from the Global Hydrological Model-Catchment-based Macro-scale Floodplain model (GHM-CaMa-Flood) model spanning the period from 1971 to 2010. The outcomes, as indicated by metrics such as Nash-Sutcliffe efficiency (NSE) and Relative Error (RE), highlight a substantial enhancement in the performance of LSTM across all basins in comparison to streamflow simulation models driven by the Variable Infiltration Capacity (VIC) method in the CaMa-Flood framework. These findings underscore the model's proficiency in hybrid LSTM with the GHM model, resulting in notable improvements in global flood simulation accuracy and providing increased confidence in flood risk assessment. Two advanced models, VIC-CaMa-Flood-RNN and VIC-CaMa-Flood-LSTM [109] are developed to predict the daily streamflow of the Xijiang River in China. The study compares the performance of these models with baseline models, including physics-based models. The findings revealed that VIC-CaMa-Flood-LSTM outperformed the other models in predicting daily streamflow. A multi-output DNN model [110] is introduced for the prediction of Flow-Duration Curve (FDC). The model leverages daily mean streamflow data spanning six decades (from 1950 to 2009) obtained from the "United States Geological Survey National Water Information System." The dataset encompasses 15 quantiles of the FDC, effectively capturing the comprehensive shape of these curves. Through the application of the Local Interpretable Model-agnostic Explanations (LIME) method, the authors demonstrate that the connection between basin characteristics and anticipated FDC can be inferred using localized surrogate models. This study underscores the potential of employing models as surrogate representations for physical models or curves. Reference [111] proposes a model based on an LSTM network and develop a coupled 1D/2D (1D conduit network model and 2D overland flow model) urban hydrological model to predict water levels and inundation phenomena in an urban catchment. This model uses the Quantitative Precipitation Forecasts (QPF) of the McGill Algorithm Precipitation nowcasting by Lagrangian Extrapolation (MAPLE) system to make 3-h Mean Arial Precipitation (MAP) forecasts. A database of 24 heavy rainfall events, 22 grid points from the MAPLE system, and the observed MAP values estimated from five ground rain gauges are used to train and test the model. This model is found to have superior correction capability with higher performance than simple models (LR, MLP) and indicates the improvement in generating the 3-h MAP forecasts and urban inundation forecasts from the original MAPLE. In [112], the authors introduce an innovative data-driven method for predicting urban floods and supporting emergency decision-making. This approach combines LSTM, Bayesian optimization, and transfer learning techniques, leveraging spatial and temporal drainage as well as rainfall data. The study demonstrates that this proposed method can efficiently forecast maximum water depths and generate time series flood maps with significantly reduced computational expenses. LSTM-ALO is a hybrid model that utilizes the Ant Lion Optimizer (ALO) algorithm [113] to optimize the key parameters, including the number of hidden layers and the learning rate of the LSTM model. ALO has two elements: ants (use different random walks to move over the search space) and antlions (hunt the position of ants and trap them in antlions' pits). A hybrid LSTM-ALO model for monthly runoff prediction [114] is proposed using the Ant Lion Optimisation (ALO) algorithm to optimize the essential LSTM model parameters. This model is contrasted with standard LSTM and LSTM-PSO (Particle Swarm Optimisation). With a few time lags, the LSTM-ALO model outperforms in terms of its ability to predict monthly runoff. The LSTM-HetGP [115] model combines the LSTM network with the heteroscedastic Gaussian (HetGP) for probabilistic prediction. This hybrid model comprises four layers: input, recurrent hidden, GP [116] (a nonparametric and kernel machine) layer, and output layer. In this model, instead of connecting the output of the recurrent hidden layer of LSTM to the output layer, it is estimated with the kernel function of the HetGP and then connected to the output layer. Reference [115] offers a probabilistic streamflow forecasting model based on an LSTM network by connecting a heteroscedastic Gaussian process to the internal structure of the LSTM to deal with probabilistic daily streamflow forecasting. HetGP addresses the interdependencies among the input variables in this model and embeds them into a single hidden vector. In this work, prediction accuracy is observed to be higher than with the standard LSTM model, and the adaptive prediction interval is provided. LSTM-ROM [117] is a hybrid model of an LSTM network coupled with a Reduced-Order Model (ROM) [118]. LSTM networks may take hours or days to train massive spatial datasets and have high computational complexity. ROM based on data-driven techniques is used in this hybrid model to reduce the size of the datasets, significantly speed up the simulation, and reduce computational complexity. Proper Orthogonal Decomposition (POD) and Singular Value Decomposition (SVD) [117] are commonly used ROM-based data reduction methods to reduce the dataset size. The reduced dataset is then fed to the LSTM network to predict the level. These studies show that data preprocessing may improve the accuracy of models. A DeepGR4J model, a hybrid approach that integrates the GR4J (Génie Rural á 4 paramètres Journalier) conceptual rainfall-runoff model with DL architectures like LSTM and CNN, outperforms both the standalone GR4J model and individual DL models [119]. By replacing the routing storage component of GR4J with neural networks and incorporating antecedent streamflow data, DeepGR4J achieves superior predictive performance, particularly in arid regions, and demonstrates stronger generalization across 222 Australian catchments. This highlights the model's effectiveness in leveraging both physical and data-driven insights for enhanced hydrological forecasting. A hybrid DL framework—CEEMDAN-VMD-TCN-GRU&RF—for accurate daily streamflow prediction at 11 stations in the Jialing River Basin, China is presented in [120]. This model integrates multi-source data from the Global Land Data Assimilation System (GLDAS), hydro-meteorological records, and streamflow data. Maximum Information Coefficient (MIC) is used for feature selection, while Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Variational Mode Decomposition (VMD) extract key features. A combined Temporal Convolutional Network (TCN) and GRU model, further enhanced by a Random Forest (RF) ensemble, is employed for prediction. This model outperforms traditional approaches in flood forecasting and long-range streamflow predictions (7, 14, and 28 day leads). A study develops two hybrid flood forecasting models—CNN–RF (Convolutional Neural Network–Random Forest) and CNN–SVR (Convolutional Neural Network-Support Vector Regression)—for the Mahanadi River Basin, where CNN serves as a feature extractor and RF/SVR as forecasting models [46]. Among various data splits and input configurations tested at Kantamal and Kesinga stations, CNN–RF consistently outperformed other models, demonstrating its effectiveness in improving flood prediction accuracy. A hybrid LSTM-MMSCS runoff simulation model is proposed that combines a Long Short-Term Memory (LSTM) neural network with the Modified Michel Soil Conservation Service (MMSCS) algorithm to enhance runoff prediction accuracy in the Songhua River Basin, China [121]. By integrating satellite meteorological data and hydrological mechanisms, the model significantly reduced runoff prediction errors and peak runoff error outperforming standalone LSTM models, especially during extreme runoff events. A study introduces Stream-LSTM, an adaptive runoff prediction model combining DL with data stream mining to handle non-stationary rainfall-runoff relationships [122]. Using dynamic threshold adjustment and selective network updates, the model achieved high accuracy (NSE up to 0.91) in the Yellow River basin and showed strong generalization in inter-basin tests, outperforming traditional models.

Table A2 summarizes the existing DL methods for flood prediction studied in this review. The summarized performance metrics provided in Table A3 present either the mean or maximum values for NSE, R^2, WI, LMI, and ρ, as well as the mean or minimum values for RMSE, RSR, MAE, RE, MARE, MSLE, and MAPE. These metrics are focused on monthly, daily, and hourly prediction scenarios. The reviewed literature shows significant advancements applying DL models in hydrological forecasting but also reveals some of the common issues and challenges that would be overcome. Among these, one of the issue is data quality and availability, since DL models require large, consistent, and high-quality datasets, which in hydrological contexts are not readily available due to missing values, noise, and sparse spatial or temporal coverage. Effective feature selection and engineering are also difficult, since the selection of good and no-redundant inputs is domain dependent and requires trail-and-error experiments. Hyperparameter tuning is required for achieving good model performance but is not usually reported, and replication and comparison thus are not straightforward. Overfitting is still a problem, especially with complex models and limited data, and needs robust validation strategies. Additionally, real-world deployment faces problems such as high computational demands, model interpretability issues, and the difficulty of developing generalizable models across multiple basins.

5 Data Preprocessing & Flood Dataset

The flood prediction data may comprise information on rainfall, water level, MIKE surface flow, radar imagery data, streamflow, runoff, evapotranspiration, watershed, discharge, radar map, catchment flow, etc. [10,47]. Many studies concentrate on the spatial and temporal data gathered by governmental entities in order to construct a reliable lead-time flood prediction model utilising DL [65,99]. The Global Runoff Data Centre, the National River Flow Archive, the United Kingdom, the National Water Information System, the United States, and the African database of hydrometric indices are examples of entities that facilitate access to ongoing records of flood-related data on national and international levels [123–125]. However, flood-related data can suffer from issues such as missing records, outliers, noise, and inconsistent scaling of variables, often due to recording errors or data corruption at specified time steps [126,127]. In order to enhance the performance of the model, it is crucial to perform preprocessing techniques such as handling missing data, eliminating outliers and noise, scaling the data, etc. [12,55,128]. Fig. 15 shows the taxonomy of the various methods used in time series data preprocessing.

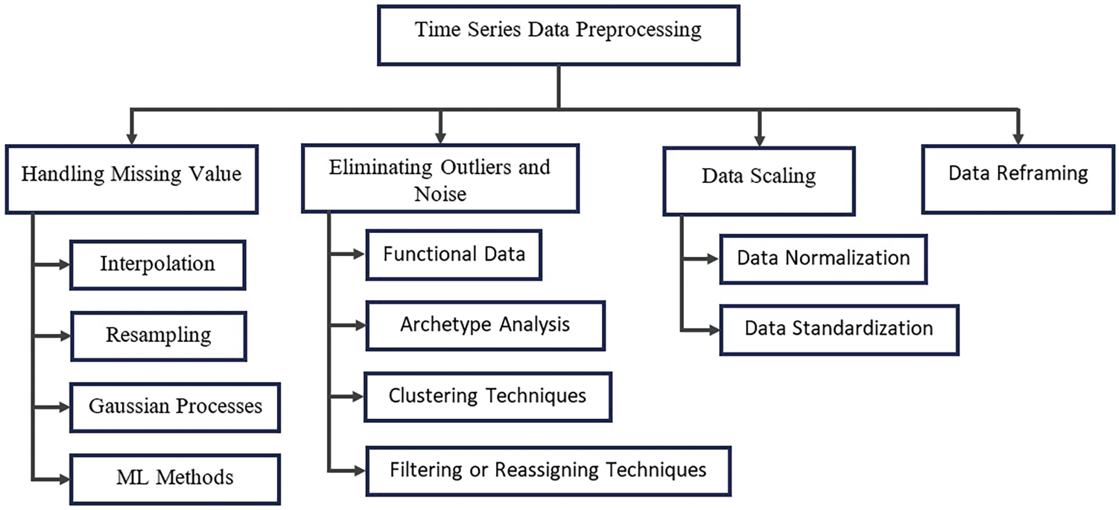


Figure 15: Taxonomy for time series data preprocessing techniques

5.1 Handling Missing Value

Multiple techniques are employed to address the challenge of managing missing data, encompassing approaches like interpolation [129], resampling [130], autoregression [131], Gaussian processes [132], as well as ML methods. While these strategies demonstrate efficacy and yield favorable outcomes for datasets featuring a limited volume of missing information, they prove inefficient in situations where datasets are sparsely populated with a substantial proportion of missing values [133]. In such instances, conventional methods struggle to grasp the temporal patterns and valuable connections between observed and absent data points. However, the combination of certain ML techniques with RNN has exhibited a capacity to generate intricate outcomes in managing missing data [132,134–136].

5.2 Eliminating Outliers and Noises

Outliers and noises disrupt the inherent patterns or trends embedded within time series data, exerting an adverse influence on forecasting models, ultimately leading to inaccuracies in outcomes and a degradation in the reliability and effectiveness of future predictions. Anomalies or significant deviations, are classified as outliers, whereas inaccuracies or errors in the data, such as those arising from misclassified instances or faulty measurements, are regarded as noise [137]. Detecting outliers within multivariate time series datasets is addressed through a functional data approach, as proposed by [138]. Another strategy, outlined in [139], involves subspace outlier detection through archetype analysis coupled with the nearest neighbor algorithm. As discussed in [140], clustering techniques are explored for identifying outliers within multidimensional datasets. The realm of statistical outlier detection and imputation is also explored in [141]. Enhancing forecasting model performance can be achieved by mitigating outliers and noise by means of filtering or reassigning mislabeled data points. A study in [142] delves into the utilization of and visualization algorithms to identify and eliminate noise within time series data.

5.3 Data Scaling

Data normalization and standardization are extensively employed techniques for feature scaling. The selection and application of appropriate data scaling or reformation techniques can significantly impact machine learning model performance by ensuring equitable feature contribution and addressing potential biases, as explored in studies like [143] on data reformation for streamflow extremes and demonstrated in practice by [144] who applied standard-scaler normalization for consistent streamflow prediction. Normalization is particularly valuable as it prevents overlooking data attributes characterized by smaller numerical ranges. Therefore, it becomes imperative to normalize both the substantial and diminutive numerical data patterns and attributes to establish a coherent range. The generalized formula utilized for normalizing input multivariate time series data is provided below:

$$O_{scale} = \frac{O_t - O_{min}}{O_{max} - O_{min}} \tag{1}$$

Here O_t and O_{scale} are the observed and rescaled value of multivariate flood-related series data at time step t; O_{min} and O_{max} are the minimum and maximum of the observed multivariate time series flood-related data, respectively. Standardization can be helpful when the features follow a Normal or Gaussian distribution. It rescales the data to ensure that the mean and standard deviation of the data would become 0 and 1, respectively, and does not have a predefined range. The generalized formula used for standardization is given below.

$$O_{scale} = \frac{(O_t - mean)}{\sigma} \tag{2}$$

Here O_t and O_{scale} are the observed and rescaled value of multivariate flood-related series data at time step t; *mean* and σ are the mean and standard deviation of the multivariate flood-related series data, respectively.

5.4 Data Reframing

Reframing time series data into a supervised learning framework is a crucial step that unlocks valuable insights and simplifies the task of forecasting. This transformation involves converting a multivariate time series into a collection of input features and corresponding output sequences. The fundamental concept underlying this data reframing is the application of the sliding window technique. This method takes into account previous data as input and potential outcomes as output. In the studies by authors in [145,146], the utility of sliding windows is explored for predicting stock prices and future trends in the context of the coronavirus disease pandemic. The dimension of the sliding window can be tailored to suit the specific challenges of prediction. To illustrate, imagine a dataset spanning ten days focused on flood-related multivariate daily reports. The dataset incorporates attributes such as rainfall and river water levels. Consider the scenario where the aim is to predict the river water level for the following day using data from the preceding six days. This concept is illustrated in Fig. 16, depicting the sliding window technique in action.

In this instance, the window size is set to six. The initial window encompasses data from time t_1 to t_6, which is employed to predict the water level for the subsequent day. As the window advances, the second window encompasses data from t_2 to t_7, predicting the water level for the next day. This iterative process continues until the complete series is utilized, resulting in a fresh dataset comprising numerous input sequences matched with their corresponding output predictions. In essence, by employing the sliding window technique, the time series forecasting problem is transformed into a structured supervised learning problem, enabling the application of established machine learning methodologies to tackle the task effectively. The observation is that data preprocessing plays a key role in the time series predictive model, which can reduce the computational and time complexity of the model without compromising flood prediction accuracy.

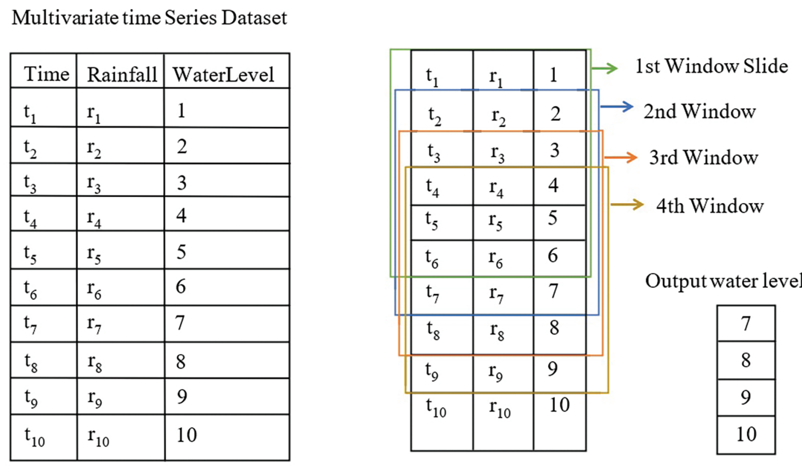


Figure 16: Reframing time series as supervised learning using sliding window

6 Comparative Study of DL Models for Flood Prediction

This section presents the key findings from the systematic review and a comparative analysis. We first synthesize the reported performances and applications of DL models from the reviewed literature (summarized in Table A3). Following this synthesis, we present the design and results of our own comparative analysis employing various cutting-edge DL architectures on the publicly available Waterbench-Iowa dataset, aiming to predict hourly streamflow and provide a more controlled performance comparison.

6.1 Datasets

This dataset, meticulously gathered from 125 distinct stations in the heart of Iowa, USA, served as the foundation of this study [125]. Specifically, we narrow our focus to data from five specific stations identified by station IDs 519, 521, 522, 523, and 525 for this comparative analysis. The analysis primarily concentrates on essential parameters like precipitation, evapotranspiration, and discharge. This analysis revolves around the complex task of predicting streamflow, emphasizing the detailed examination of these specific variables.

6.2 Methods

The exploration of this study leads us through a diverse array of DL models. These models encompass various architectures, including LSTM, GRU, 1D-CNN, and innovative hybrid variations like CNN-LSTM and CNN-GRU. Each of these architectures is meticulously designed to extract meaningful patterns and has been specifically implemented for the comparative analysis presented in this section using the selected dataset from the Waterbench-Iowa collection [125]. We have verified the accuracy of these implemented architectural details for technical correctness and reproducibility. The detailed compositions of these models are outlined in Table 1.

Table 1: Architectural overview of designed models

Designed model	Architecture details
LSTM	LSTM (16, return sequences = False), Dense (16), Activation = (ReLu), Dense (1)
LSTM	LSTM (16, return sequences = False), Dense (16), Activation = (ReLu), Dense (1)
GRU	GRU (32, return sequences = False), Dense (16), Activation = (ReLu), Dense (1)
1D-CNN	Conv1D (filters = 16, kernel size = 3), MaxPooling1D(2), Flatten(), Dense(32), Activation = (ReLu), Dense(1)
CNN-LSTM	Conv1D (filters = 16, kernel size = 3), LSTM (8, return sequences = False), Dense(16), Activation = (ReLu), Dense(1)
CNN-GRU	Conv1D (filters = 16, kernel size = 3), GRU(8, return sequences = False), Dense(16), Activation = (ReLu), Dense(1)

6.3 Evaluation Metrics

Evaluation of flood prediction models' performance relies on various essential metrics, as outlined in Table 2. These metrics are commonly used in the reviewed literature for assessing hydrological model performance.

Table 2: Formulae and descriptions of the metrics used for evaluating flood prediction models

Metrics	Formulae	Descriptions
NSE	$1-\left\{\frac{\sum_{t=1}^{T}(y_t-p_t)^2}{\sum_{t=1}^{T}(y_t-\bar{y})^2}\right\}$	Here, y_t and p_t are the observed and predicted water level at the time step t, $\bar{y}$ and $\bar{p}$ are the mean of the observed and predicted water level.
R^2	$\frac{\sum_{t=1}^{T}(y_t-\bar{y})(p_t-\bar{p})}{\sqrt{\left\{\sum_{t=1}^{T}(y_t-\bar{y})^2\right\}\left\{\sum_{t=1}^{T}(p_t-\bar{p})^2\right\}}}$	
ρ	$\frac{cov(y,p)}{\sigma_y.\sigma_p}$	Here, $cov(y, p)$ is the covariance, σ_y, and σ_p are the standard deviation of observed and predicted water level.
$RMSE$	$\sqrt{\frac{\sum_{t=1}^{T}(y_t-p_t)^2}{T}}$	
MAE	$\frac{\sum_{t=1}^{T}\lVert y_t-p_t\rVert}{T}$	
RSR	$\frac{RMSE}{STDEV_y}=\left[\frac{\sum_{t=1}^{T}(y_t-p_t)^2}{\sqrt{\sum_{t=1}^{T}(y_i-p)^2}}\right]$	Here, y_t and p_t are the observed and predicted water level at the time step t.

(Continued)

Table 2 (continued)

Metrics	Formulae	Descriptions
RE	$\frac{\Vert y_t - p_t \Vert}{y_t}$	
MARE	$\frac{1}{T}\sum_{t=1}^{T} \Vert \frac{y_t - p_t}{y_t} \Vert$	
MSLE	$\frac{1}{T}\sum_{t=1}^{T} (\log y_t - \log p_t)^2$	
MAPE	$\frac{1}{T}\sum_{t=1}^{T} \left(\Vert \frac{y_t - p_t}{y_t} \Vert \right) 100\%$	
WI	$1 - \left[\frac{\sum_{t=1}^{T} (y_t - p_t)^2}{\sum_{t=1}^{T} (\Vert p_t - \bar{y}_t \Vert + \Vert y_t - \bar{y}_t \Vert)^2} \right]$	Here, y_t and p_t are the observed and predicted water level at the time step t, $\bar{y}$ is the mean of the observed water level.
LMI	$1 - \left[\frac{\sum_{t=1}^{T} \Vert y_t - p_t \Vert}{\sum_{t=1}^{T} \Vert y_t - \bar{y}_t \Vert} \right]$	Here, y_t and p_t are the observed and predicted water level at the time step t, $\bar{y}$ is the mean of the observed water level.
KGE	$1 - \sqrt{(\rho - 1)^2 + (\beta - 1)^2 + (\alpha - 1)^2}$	Here ρ is the pearson correlation coefficient, $\beta = \frac{\mu p}{\mu y}$ are the mean of y observe and p prediction water level, and $\alpha = \frac{\sigma p}{\sigma y}$ are the standard deviation of the y observe and p prediction water level.

As discussed in the synthesis of literature, widely used metrics include the NSE, which measures the predictive capability of a simulation model. It spans from 0 to 1, where higher values indicate superior predictive performance. The coefficient of determination (R^2), which measures the correlation between observed and predicted water levels. Pearson correlation coefficient (ρ) also assesses this correlation. Additional metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), RMSE-observation standard deviation ratio (RSR), RE, Mean Absolute Relative Error (MARE), Mean Squared Logarithmic Error (MSLE), and Mean Absolute Percentage Error (MAPE) quantify the discrepancies between observed and predicted values, providing insights into model accuracy. The Willmott's Index of agreement (WI) evaluates prediction errors relative to potential errors, yielding a value between 0 and 1. Legates-McCabe's Index (LMI) serves a similar purpose, ranging from 0 to 1 as well. A higher value indicates better model performance for both WI and LMI. The Kling-Gupta efficiency (KGE) considers error, bias, and relative variability, providing a comprehensive measure of model accuracy.

7 Results and Discussion

7.1 Results

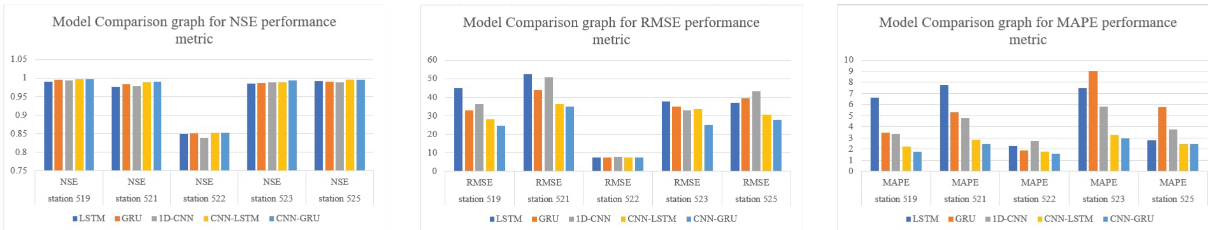
Table A3 presents a condensed summary of performance metrics across various forecasting scenarios (monthly, daily, and hourly), based on the studies reviewed in Section 4. It includes either the mean or maximum values for NSE, R^2, WI, LMI, and ρ, and the mean or minimum values for RMSE, RSR, MAE, RE, MARE, MSLE, and MAPE. The table also provides an overview of the DL models applied, the hydrological tasks addressed, datasets used, and the associated performance outcomes. In this analysis, we prioritize the assessment of the implemented streamflow forecasting models by concentrating on three vital performance metrics: NSE, RMSE, and MAPE. This deliberate focus on these specific metrics ensures a thorough and intricate comparative analysis of the designed forecasting models. By utilizing these metrics, we gain valuable insights into the accuracy and precision of our predictions.

The outcomes of the analysis, presented in Table 3, clearly demonstrate the effectiveness of the designed models for each dataset associated with the five stations. These results provide a controlled comparison of the performance of selected DL models on a specific, publicly available dataset, complementing the broader but more varied findings synthesized from the literature review and Table A3.

Table 3: The NSE, RMSE, and MAPE metrics are used to assess the performance of hourly streamflow prediction models at each of the five stations

Model	Station 519			Station 521			Station 522			Station 523			Station 525		
	NSE	RMSE	MAPE	NSE	RMSE	MAPE	NSE	RMSE	MAPE	NSE	RMSE	MAPE	NSE	RMSE	MAPE
LSTM	0.990	44.87	6.61	0.976	52.37	7.75	0.849	7.56	2.30	0.984	37.79	7.50	0.991	37.13	2.81
GRU	0.995	32.90	3.49	0.983	43.76	5.31	0.850	7.54	1.90	0.987	34.90	9.03	0.990	39.52	5.74
1D-CNN	0.995	36.21	3.39	0.978	50.66	4.79	0.839	7.80	2.71	0.988	33.01	5.82	0.988	43.33	3.75
CNN-LSTM	0.996	28.16	2.21	0.989	36.16	2.84	0.852	7.48	1.76	0.988	33.47	3.24	0.994	30.39	2.46
CNN-GRU	**0.997**	**24.61**	**1.73**	**0.989**	**34.92**	**2.43**	**0.853**	**7.46**	**1.59**	**0.993**	**24.98**	**2.96**	**0.995**	**27.64**	**2.44**

Among the evaluated models, the hybrid CNN-GRU model stands out with superior performance values in all three metrics: NSE, RMSE, and MAPE. These results signify the model's ability to accurately capture the complex patterns inherent within the studied datasets. The strong performance of hybrid models, particularly CNN-GRU, in this analysis (achieving NSE values frequently above 0.99 and as high as 0.997, with MAPE values for some stations around 24%–27% as shown in Table 3 aligns with the observations in the reviewed literature that hybrid approaches can leverage the strengths of different architectures to improve prediction accuracy for complex hydrological data [100,101]. While direct quantitative comparison to studies in Table A3 is limited due to dataset differences, our findings on the Iowa dataset support the potential benefits of combining CNN and GRU for streamflow forecasting. Although, the high NSE values (often >0.98) demonstrated by our standalone LSTM and GRU implementations for hourly streamflow are consistent with the performance levels reported in literature, such as [72], who achieved NSE of 0.98 for LSTM and 0.99 for GRU in hourly flow discharge predictions. These comparisons, despite variations in datasets and specific experimental setups, suggest that our findings on the Waterbench-Iowa dataset are well-supported by the broader literature and further underscore the potential benefits of combining CNN and GRU components for robust streamflow forecasting. The model comparison graph in Fig. 17 provides additional clarity, using the y-axis to represent NSE, RMSE, and MAPE values, and the x-axis indicating the model performance for each station.



(**a**) (**b**) (**c**)

Figure 17: Comparison of LSTM, GRU, 1D-CNN, CNN-LSTM, and CNN-GRU models for predicting hourly streamflow. Sub-figures (**a**), (**b**), and (**c**) display their respective performances in terms of NSE, RMSE, and MAPE

The graph clearly shows that the hybrid CNN-GRU and CNN-LSTM models closely align in terms of NSE performance across all five stations. However, when it comes to RMSE and MAPE performance, the CNN-GRU models consistently outperform the others.

7.2 Discussion

The key findings from the systematic review and comparative analysis are examined for their significance in the context of the current literature, the novel contributions of this study are highlighted, and major research gaps and future research directions are identified. Based on the synthesis of the reviewed literature, DL models have shown significant potential for hydrological forecasting, successfully applied to tasks such as flood prediction, streamflow forecasting, and rainfall-runoff modeling across various locations and datasets. Dominance of RNN variants, particularly LSTM and GRU, highlights their capacity to learn temporal patterns in hydrological time series, as evidenced through frequently reported high performance metrics [42,67,71,72,98]. CNN have shown strength in utilizing spatial data [96]. The increasing exploration of hybrid models also indicates the acknowledgment that utilizing architectural strengths through combination can address the complexity of hydrological processes [92,93,98–100]. However, the high variability in reported performance across studies (Table A2) indicates the significant impact of dataset characteristics, prediction task specifics, lead-time, and methodological choices on model effectiveness.

The comparative analysis on the publicly available Waterbench-Iowa dataset (Table 3) provides a controlled evaluation that complements these findings. The superior performance of the hybrid CNN-GRU model's prediction of hourly streamflow on this dataset shows that models capable of integrating both spatial (potentially represented by relationships between stations or features if CNN is used for feature extraction across inputs) and temporal information are highly effective for this task. The consistent performance of LSTM and GRU in this analysis is in line with their overall successful application reported in the literature synthesis for time series prediction.

It is challenging to directly compare results with the broad range of findings in the reviewed literature due to the heterogeneity of datasets, hydrological regimes, and assessment protocols. However, relative model performance ranks and overall trends can be compared. The robust performance of LSTM and GRU in the controlled experiment aligns with their common co-emergence as top performers in many of the studies within the systematic review of time series forecasting [42,67,71,72,98]. The findings on the performance of the CNN-GRU hybrid model for hourly streamflow forecasting on the Iowa dataset supports the literature trend towards the development of hybrid architectures to potentially achieve improved accuracy than single models [92,98–100,102]. The absence of consistency in performance across the studies in Table A2, even for the same model types, is an expression of the site-specific and data-dependent nature of hydrological forecasting and places the need for model selection and tuning according to local conditions.

Beyond predictive performance, a major consideration for the practical application of such DL models, especially in the context of real-time or near real-time flood forecasting that this review addresses, is their computational cost. Many DL architectures, particularly the increasingly popular hybrid models explored in this review, can be computationally intensive. This intensity is both in the long training phase, which can be time consuming and deployment of specialized hardware like GPUs, and more significantly, in the inference phase. In time-sensitive for flood warnings, high inference time can be a significant bottleneck, potentially undermining the 'real-time' applicability of an otherwise accurate model. The computation and memory requirements for executing complex models can also be a challenge to use in resource-constrained operational environments. Therefore, assessing the trade-offs between model complexity, predictive accuracy, and computational efficiency is paramount. Future research and implementation should increasingly focus on strategies like model optimization (e.g., pruning, quantization), the deployment of inherently faster architectures where applicable (e.g., GRUs which can offer faster computation), and the exploitation of available hardware advancements to ensure that DL-based flood forecasting systems are not only robust but deployable in operational, time-sensitive environments. This consideration has a direct bearing on the practical utility and wider applicability of the reviewed methodologies.

Drawing on prior review efforts, this systematic review and comparative analysis present several novel contributions to the field of DL for hydrological forecasting. First, through systematic synthesis of literature from 2018 to 2025, we provide an overview of the state of the art in the DL architectures and their applications, extending the scope of prior reviews. Second, despite recognition of the heterogeneity, this review provides quantitative synthesis of reported performance metrics (Table A3), offering insights into overall trend in performance and relative effectiveness of different model types across various hydrological tasks as reported in practice. Third, the development of a taxonomy of DL methods provides an ordered structured for understanding various models employed. Importantly, the comparative analysis on the publicly available Waterbench-Iowa dataset (Table 3, Fig. 17) provides a useful and valuable contribution by presenting a controlled comparison of leading DL and hybrid models under standardized conditions, addressing the data availability and comparability limitation that characterises the literature and offering a benchmark for future research. Finally, this study presents an effective analysis of the current research gaps and outlines a recommended framework, directly informing future research and implementation activities.

7.2.1 Key Issues and Challenges

There are several key issues and challenges in developing a successful hydrological prediction system. Based on the literature, some of the major challenges include dynamic environmental variability, the unavailability of adequate benchmark datasets, the need for real-time or near real-time response generation with minimal false alarms, limited spatial information, data labeling difficulties, and the development of a generic prediction method [127]. These challenges are briefly discussed in the subsequent subsection.

Dynamic Environmental Cardinality

Because of the climate's frequent volatility, irregularity, dynamic character, and frequency of flood episodes, it is challenging to develop a reliable flood prediction model. It might be difficult to analyze flood features because the frequency of flood episodes varies over time (month by month and year by year). Environmental cardinality is impacted by changing weather patterns, human land use changes, and growing urbanization. Therefore, choosing dynamically the best and possible subset of features that ensures flood prediction with high accuracy is an important issue.

Lack of Dataset

Unlike other fields, this field has a shortage of curated, high-quality, labeled, or benchmark datasets. Most studies in this literature acquire government datasets from different agencies, depending on their requirements. Sometimes, the acquired datasets may be (*i*) inconsistent, (*ii*) for limited time, (*iii*) with missing values, (*iv*) contain gaps, (*v*) uncertain because of measurement errors, or (*vi*) not publicly available. The datasets available on governmental websites are still scarce, require strict access policies and licensing restrictions, and need time to make a globally usable dataset. So, collecting raw data from relevant sources and designing an adequate data preprocessing technique would significantly help in developing appropriate flood prediction models.

Minimal Spatial Information

Flood-associated datasets containing both temporal and spatial details are notably scarce. On a global scale, the availability of river water level observations, along with records of hydrological and flood incidents for specific locations, remains limited. There are variations in the densities of hydrological stations, which can impede experts' efforts in comprehensively analyzing spatial attributes. During the collection of spatial data, there's a possibility of encountering temporal discrepancies in overlapping data records. This arises due to differing agency-specific needs for their spatiotemporal analyses. Addressing this challenge represents a contemporary research avenue within this domain.

Development of Generic Model for Prediction

In order to get the desired outcomes, training a DL model requires huge datasets with pertinent characteristics. If the DL model is trained with insufficient and inadequate instances, it will perform poorly and produce ambiguous results, rendering its performance useless. Researchers encounter difficulties in validating their models since generally recognized data is not publicly available. Therefore, the creation of a general approach that can effectively manage data from different basins throughout the world is the need of the hour.

Real Time Detection with High Accuracy

Creating an accurate real-time flood detection system presents a formidable challenge, primarily stemming from the limited availability of spatial data. This constraint is attributed to the prevalence of floods across both monitored regions and inadequately or unmonitored areas. When striving to devise a real-time flood detection model for such unmonitored regions, essential datasets encompassing insights from hydrological and social sciences become crucial. These datasets should encompass comprehensive details such as land use patterns, hydro-meteorological analyses, hazard intensity assessments, and historical event records. Adopting this approach holds the potential to significantly enhance the precision of real-time flood detection models.

Computational Cost and Real-Time Feasibility

A key practical challenge in deploying DL models for time-sensitive tasks like real-time flood forecasting is managing their computational cost. Deep architectures such as DNNs, RNNs, LSTMs, GRUs, CNNs, and hybrid models (e.g., CNN-LSTM, CNN-GRU) often demand significant computational resources and time, especially during training on large hydrological datasets [38,133]. Although inference is typically faster, complex models can still introduce latency that may hinder timely flood warnings. Factors such as model depth, parameter count, and spatio-temporal data resolution contribute to this computational burden. While advancements in GPU technology have improved feasibility [38], access and affordability remain concerns for operational deployment. Moreover, hybrid models, despite offering improved accuracy, tend to compound the computational demands of their components. Hence, a trade-off arises between model complexity and computational efficiency. Real-time systems require models that balance predictive accuracy with low-latency execution, which may involve choosing simpler models like GRUs over LSTMs in some cases, employing optimization techniques (e.g., pruning, quantization), or enhancing infrastructure. Future work should focus on optimizing this balance to meet the practical demands of flood early warning systems.

7.2.2 Recommendation

This study investigates the importance of analyzing time series data for predicting river water levels using various DL models. The analysis highlights several key factors essential for improving prediction accuracy, including effective preprocessing, time series-based feature extraction, well-designed DL model architectures, and appropriate training strategies. With these considerations in mind, we propose a general experimental framework—illustrated in Fig. 18 to develop a robust and generalized DL-based flood prediction model.

This framework consists of the following steps:

1. **Data Collection & Preprocessing for Time Series Data:** Since most flood forecasting models rely on historical data from various variables, it is essential to examine this data through a time series perspective, considering aspects such as time granularity, feature generation, and extraction. These components fall under the broader scope of data collection and preprocessing. This stage involves gathering historical spatial time series data related to floods from diverse sources and compiling it into a centralized repository. A critical part of the process is analyzing and preprocessing the collected time series data. This includes handling missing values, outliers, and noise, as well as exploring temporal interdependencies within the data. Additionally, preprocessing may involve rescaling or normalizing the data. To make the dataset suitable for DL models, the time series is ultimately transformed into a supervised learning format.
2. **Data Distribution:** Processed data might encompass a substantial volume of spatial time series data for various gauge stations. This data is distributed separately for each gauge station.
3. **Hybrid DL-Based Prediction:** Hybrid DL techniques are employed to train on the distributed time series data individually, facilitating water level prediction for each station.
4. **Compilation of Results:** The outcomes generated by the hybrid DL models are consolidated to showcase their predictive performances.
5. **Consensus Prediction:** Predicted water levels are derived using weighted majority voting for each gauge station, and these predictions are subsequently compared. Weights are assigned to each hybrid DL model based on its previous performance.

6. **Comparative Analysis:** Predicted water levels are compared with observed warning, flood, and high flood levels. This comparison aids in determining whether the projected water levels fall within the categories of normal, warning, flood, or high flood. The outcomes of this comparison contribute to early alarm mechanisms.
7. **Feedback-Loop Learning for Predictive Accuracy Enhancement:** The framework incorporates a continuous feedback loop for model improvement. This begins with Data Archiving and Outcome Verification, where all predictions and relevant operational data are stored, and critically, incoming actual observed outcomes are aligned with these predictions. Following this, Performance Analysis and Data Update Strategy involves systematically comparing these archived predictions against the verified outcomes. This analysis identifies biases and performance issues, informing how datasets will be updated for retraining. Finally, Model Refinement and Incremental Training with Error Management uses these insights to periodically retrain models. To prevent cumulative errors, this retraining primarily uses new, verified ground truth data, while the analysis of past prediction errors guides model adjustments, ensuring the system learns and adapts robustly.

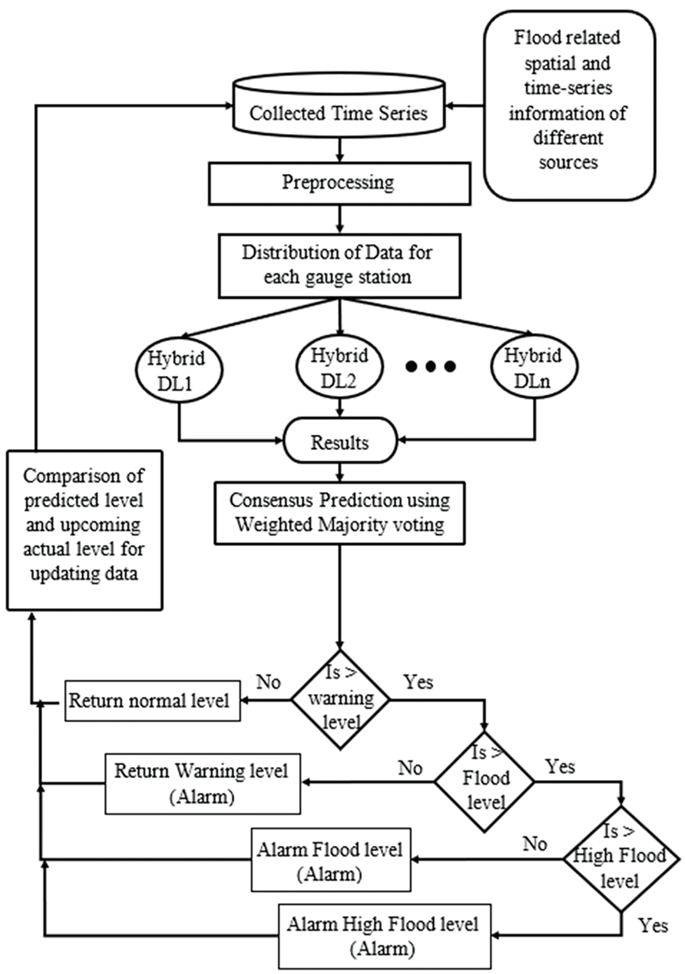


Figure 18: Framework of the recommended model

The innovative aspect of the recommended framework (Fig. 18) is not being focused on the individual application of widely used techniques like multiple model architectures or standard pre-processing operation. Instead, its main contribution and innovative value lie in the systematic and unique combination of these components in a coherent, adaptable, and end-to-end process specifically designed for forecasting river water levels and generating actionable flood alerts. Although such components such as data preprocessing and the use of hybrid models are well-established in time series modeling, this framework provides a clear, structured approach that leverages these components to effectively address key challenges in hydrological forecasting. These challenges include managing spatial-temporal data complexities, improving prediction robustness through diverse model integration, and translating complex model outputs into clear, operational early warnings. The key elements that, in their specific combination and application within this cohesive structure, underscore the framework's novelty include:

i) *Strategic and Distributed Integration of Hybrid DL Models*: Although the use of hybrid models is an developing field, this framework innovatively suggests their systematic deployment in a distributed manner. Specifically, varied hybrid DL models are trained on time series data tailored to each gauge station. The innovative lies in this customized deployment of several advanced hybrid models to different spatial locations to capture localized hydrological details, followed by a consensus mechanism to enhance overall predictive accuracy and reliability.

ii) *Preprocessing*: While time series data analysis and preprocessing are standard initial steps, the framework highlights their rigorous and thorough application (as outlined in Section 4) as an integral, foundational stage. These comprise systematic handling of missing data, outliers, noise, temporal interdependencies, and proper data scaling and reframing. The unique feature of the framework is its explicit emphasis on elaborate preprocessing as a prerequisite step, followed by high-quality, distributed hybrid modeling tailored to each gauge station. Furthermore, the recommendation to utilize CPU-GPU platforms for such preprocessing steps highlights a practical consideration for operational efficiency when dealing with large-scale hydrological datasets.

iii) *Individualized Prediction through Spatially Distributed Data Handling*: A key innovative aspect is the explicit distribution of preprocessed spatial time series data for individualized model training and prediction at each gauge station. This approach addresses spatial heterogeneity systematically through enabling hybrid DL models to specialize in the specific hydrological characteristics of individual locations. This is different from more generalizable modeling approaches and is designed to capture localized details, hence improving the accuracy of water level predictions across a network of stations.

iv) *Actionable Alerts via Consensus Prediction and Multi-Level Threshold Comparison*: The system incorporates a novel decision-making layer where water level predictions from potentially multiple hybrid DL models for each station are aggregated using a weighted majority voting mechanism, regarding model performance. Crucially, such consensus predictions are then directly compared against predefined operational thresholds for 'warning', 'flood', and 'high flood' levels. This quantitative-to-qualitative translation of predictions via systematic comparison yields actionable alert levels as an explicit early warning mechanism within the framework's architecture. While ensemble methods are well-documented, their application herein with a multi-level alert system specifically for flood management presents a unique contribution to operational forecasting.

v) *Holistic End-to-End Framework for Flood Prediction*: The primary novelty of the recommended approach is its conceptualization as an end-to-end, holistic framework. This framework incorporates all key steps in a systematically manner: from careful data collection and comprehensive, customized preprocessing, through to distributed hybrid DL model execution, results compilation, consensus-based prediction, and ending in a formalized, multi-level flood alert system. It is this consolidation of established and specialized methods into a unified, cohesive, and reproducible system for flood forecasting that forms its main contribution, providing a clear pathway from raw data to actionable intelligence.

vi) *Embedded Mechanism for Iterative Improvement and Long-Term Adaptation*: The framework's design incorporates the systematic archiving of predicted water levels and alert outcomes into the data repository. Although data archiving is normal, its explicit inclusion here is to support a feedback loop for subsequent model improvement, long-term performance analysis, trend identification, and continuous adaptation of the hybrid models and weighting schemes in the consensus mechanism. This provides the framework not as a static system, but one that is designed for continuous learning and improvement.

In summary, though individual models such as specific DL architectures or preprocessing methods are part of the general ML toolkit, the novelty of this recommended framework emerges from the unique, synergistic combination of these elements. It suggests a unified, multi-stage approach—from spatially-sensitive individualized modeling to consensus-driven actionable alert generation—that is customized to solve the complex challenges in river water level prediction and flood early warning. This unification aims to provide a more robust, accurate, and operationally relevant system than ad-hoc combinations of methods might achieve.

7.2.3 Practical Implications and Implementation

The findings and methodologies presented in this systematic review have significant practical implications for enhancing real-time flood forecasting and supporting robust modeling solutions for surface water management. This section presents actionable strategies for the deployment of DL models, especially hybrid architectures such as CNN-GRU in real-time flood forecasting systems.

The identification of effective DL architectures, particularly LSTM, GRU, and hybrid CNN variants, like CNN-GRU (as synthesized from the literature and demonstrated in the comparative analysis), provides actionable guidance for practitioners in selecting models suited to capture complex temporal and potentially spatial patterns in hydrological data. This review also underscores the importance of time series-based data preprocessing. For operational deployment, rigorous preprocessing, such as missing value imputation, outlier filtering, normalization, and temporal feature engineering, should be implemented using automated, scalable data pipelines. These pipelines can run continuously on real-time data streams, ensuring that data are properly formatted for model prediction. The recommended framework incorporates the spatially distributed deployment of hybrid DL models across gauge stations. Each model is independently trained on time series data from a specific station, allowing for specialization and higher accuracy in localized forecasting. These models can be efficiently scaled and managed across multiple monitoring locations. The integration of a consensus prediction mechanism further enhances operational reliability by aggregating the quantitative model outputs and mapping them to predefined flood alert thresholds. This process produces tiered warning levels that provide interpretable and actionable information for emergency response units and decision-makers. Moreover, the embedded feedback-loop mechanism transforms the framework from a static prediction system into a dynamic, adaptive learning system. By incorporating new observations and analyzing forecast errors over time, the system ensures continuous learning and adaptability to seasonal or climate-induced changes in hydrological behavior. The integration of preprocessing, distributed modeling, consensus alerting, and adaptive learning into a unified workflow provides a practical and reproducible blueprint for real-world flood prediction. For surface water management, the use of hybrid approaches can leverage the strengths of different models to enhance accuracy and robustness, which is crucial for managing diverse hydrological conditions. The concept of distributed data handling and consensus prediction also provides viable strategies for managing large-scale, multi-station monitoring systems. Ultimately, the integration of effective DL techniques with a structured model development framework can support water management agencies in developing more accurate, timely, and dependable flood forecasting systems. This, in turn, can enhance disaster preparedness, minimize flood-related damages, and promote more efficient and informed management of surface water resources.

8 Conclusion

This systematic review thoroughly examines the recent applications of various DL models for flood prediction. Although ML models have been frequently applied in flood prediction systems, they often lack effectiveness in the presence of sparse, irrelevant, redundant, and noisy data. To address these limitations, this review discusses several DL models, analyzes their performance, and presents a comprehensive summary using various standard performance measures. The survey investigates recent flood prediction models, considering multiple aspects, including: a) Commonly used DL-based models for streamflow prediction and flood forecasting, such as RNN, LSTM, and GRU. b) CNN and GAN for assessing flood susceptibility and predictive capability. c) Hybrid DL architectures for flood prediction, including combinations like CNN-LSTM, CNN-GRU, SAE-LSTM, LSTM-seq2seq, and En-De-LSTM. d) Physical models enhanced with DL and coupled with GHM for execution of flood simulations. e) Various optimization algorithms and data preprocessing techniques to improve the performance and quality of DL models, such as VMD-DNN, LSTM-ALO, and LSTM-PSO. Our survey identifies DL-based models as the most widely used approach for flood prediction. Among them, RNNs and their variants, particularly LSTM and GRU are the most commonly applied architectures. Their inherent ability to handle sequential data makes them well-suited for hydrology-based flood prediction, which relies on time-series analysis of parameters such as water level, rainfall, and other related meteorological data. For instance, LSTM has been extensively used in flood modeling and has demonstrated its effectiveness in managing sequential data, yielding accurate forecasts of flood events on hourly, daily, and monthly lead-time.

This study presents new contributions by conducting a systematic review of recent literature 2018–2025, in a quantitative overview of reported trends in performance across various studies despite data heterogeneity. In addition, the review presents a taxonomy of DL approaches and a recommended framework. Most importantly, the comparison on the publicly available Waterbench-Iowa dataset provides a unique controlled benchmark of major DL and hybrid models, overcoming the problem of data comparability in the literature.

In this survey, the critical importance of thorough data preprocessing techniques is highlighted, including handling missing data, outliers, and scaling. The necessity of transforming time series data into a supervised learning format using techniques like the sliding window method to enhance predictive accuracy is also discussed. The review conducts an extensive comparative analysis of LSTM, GRU, 1D-CNN, CNN-LSTM, and CNN-GRU models applied to hourly streamflow prediction using the rich Waterbench-Iowa dataset from specific Iowa stations. The study emphasizes the impressive capabilities of these diverse architectures. Among the models evaluated, the hybrid CNN-GRU model proved extremely effective, demonstrating superior performance in capturing intricate dataset patterns as measured by the NSE, RMSE, and MAPE metrics. These findings offer a clear roadmap for real-world adoption. For instance, a water management agency can directly implement the high-performance CNN-GRU architecture identified in this study to enhance the lead time and accuracy of streamflow forecasts. Moreover, the proposed framework (Fig. 18) provides an operational blueprint that integrates data preprocessing, distributed modeling, and consensus-based alert generation. Adopting this framework would allow practitioners to systematically translate raw hydrological data into actionable multilevel flood warnings, thus strengthening emergency preparedness and resource allocation in flood-prone regions.

While this systematic review provides a comprehensive overview of hydrological flood prediction systems, it also has certain limitations that future researchers with a strong interest in this domain may address. One significant limitation is the absence of direct quantitative comparisons between different flood prediction models. This is primarily due to the heterogeneity of datasets and methodologies used across the literature, which makes such comparisons challenging. Although a comparative analysis was conducted in this study, the scope was constrained by the limited number of models and datasets available. Specifically, due to the scarcity of standardized datasets for hydrological research, our experimental evaluation was restricted to the publicly available Iowa dataset. Since hydrological systems are highly influenced by local climate, topography, and land-use characteristics, a model that performs well on Iowa data may not be directly transferable to a region with vastly different environmental conditions, such as Manipur, India, which experiences recurring and unpredictable flood during monsoon season. Furthermore, we propose a generalized framework for implementing flood prediction systems. However, this framework may not be suitable for certain specific scenarios, such as uni-station systems. Addressing these limitations is a key objective of our ongoing research. Future research directions include the creation and publication of new hydrological datasets, particularly focusing on flood-prone regions such as Manipur in India. This region frequently experiences flooding due to river overflow, and we are in the process of collecting relevant data, which we intend to make publicly available. Another promising direction is the development of generic flood prediction models that are transferable, interpretable, and capable of uncertainty estimation. Explainable AI (XAI), a state-of-the-art approach in deep learning, should be embraced and leveraged to enhance the transparency and trustworthiness of predictive models. Additionally, there is a need to explore DL models that can effectively process and learn from spatially rich datasets, as modern hydrological data often span multiple dimensions and contain latent features that can significantly improve prediction accuracy when properly extracted. Conducting case studies on real-world and practical flood prediction systems in regions like Manipur is also part of our ongoing research work. In this context, collaboration among multidisciplinary researchers and government agencies is essential to develop more robust and actionable flood forecasting solutions, which will ultimately enhance preparedness and response capabilities.

Acknowledgement: The authors sincerely acknowledge the Department of Computer Science, Manipur University, for providing a supportive and conducive research environment, which played a vital role in the successful completion of this work.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization: Leishangthem Sashikumar Singh and Ksh. Nilakanta Singh; validation: Nazrul Hoque, Khumukcham Robindro Singh and Dhruba K Bhattacharyya; formal analysis: Leishangthem Sashikumar Singh and Ksh. Nilakanta Singh; investigation: Leishangthem Sashikumar Singh; writing—original draft preparation: Leishangthem Sashikumar Singh; writing—review and editing: Leishangthem Sashikumar Singh, Ksh. Nilakanta Singh, Nazrul Hoque, Khumukcham Robindro Singh and Dhruba K Bhattacharyya; visualization: Leishangthem Sashikumar Singh and Ksh. Nilakanta Singh; supervision: Nazrul Hoque, Khumukcham Robindro Singh and Dhruba K Bhattacharyya. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data utilized in the empirical analysis of this survey is openly accessible on Zenodo at the following doi:https://doi.org/10.5281/zenodo.7087806 (accessed on 21 July 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Supplementary Materials: The supplementary material is available online at https://doi.org/10.23967/j.rimni.2025.067112.

References

1. Bongaarts J. IPCC, 2023: climate change 2023: synthesis report. IPCC, 184 p. Popul Develop Rev. 2024;50(2):577–80. doi:10.59327/IPCC/AR6-9789291691647.
2. Field CB, Barros V, Stocker TF, Dahe Q. Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. Cambridge, UK and New York, NY, USA: Cambridge University Press; 2012.
3. Ahmadi SM, Balahang S, Abolfathi S. Predicting the hydraulic response of critical transport infrastructures during extreme flood events. Eng Appl Artif Intell. 2024;133:108573. doi:10.1016/j.engappai.2024.108573.
4. Xie K, Ozbay K, Zhu Y, Yang H. Evacuation zone modeling under climate change: a data-driven method. J Infrastruct Syst. 2017;23:360. doi:10.1061/(asce)is.1943-555x.0000369.
5. Arsenault R, Martel JL, Brunet F, Brissette F, Mai J. Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models. Hydrol Earth Syst Sci. 2023;27(1):139–57. doi:10.5194/hess-27-139-2023.
6. Ávila L, Silveira R, Campos A, Rogiski N, Gonçalves J, Scortegagna A, et al. Comparative evaluation of five hydrological models in a large-scale and tropical river basin. Water. 2022;14(19):3013. doi:10.3390/w14193013.
7. Pandi D, Kothandaraman S, Kuppusamy M. Hydrological models: a review. International Journal of Hydrol Sci Technol. 2021;12(3):223–42. doi:10.1504/ijhst.2021.117540.
8. Yu Y, Zhang H, Singh VP. Forward prediction of runoff data in data-scarce basins with an improved ensemble empirical mode decomposition (EEMD) model. Water. 2018;10(4):388. doi:10.3390/w10040388.
9. Dembélé M, Oriani F, Tumbulto J, Mariéthoz G, Schaefli B. Gap-filling of daily streamflow time series using Direct Sampling in various hydroclimatic settings. J Hydrol. 2019;569:573–86. doi:10.1016/j.jhydrol.2018.11.076.
10. Mosavi A, Ozturk P, Chau KW. Flood prediction using machine learning models: literature review. Water. 2018;10(11):1536. doi:10.3390/w10111536.

11. Lohani AK, Goel NK, Bhatia K. Improving real time flood forecasting using fuzzy inference system. J Hydrol. 2014;509:25–41. doi:10.1016/j.jhydrol.2013.11.021.
12. Maier HR, Jain A, Dandy GC, Sudheer KP. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Environ Model Softw. 2010;25(8):891–909. doi:10.1016/j.envsoft.2010.02.003.
13. Kim B, Sanders BF, Famiglietti JS, Guinot V. Urban flood modeling with porous shallow-water equations: a case study of model errors in the presence of anisotropic porosity. J Hydrol. 2015;523:680–92. doi:10.1016/j.jhydrol.2015.01.059.
14. Valipour M, Banihabib ME, Behbahani SMR. Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting. J Math Stat. 2012;8(3):330–8. doi:10.3844/jmssp.2012.330.338.
15. Jain SK, Mani P, Jain SK, Prakash P, Singh VP, Tullos D, et al. A Brief review of flood forecasting techniques and their applications. Int J River Basin Manage. 2018;16(3):329–44. doi:10.1080/15715124.2017.1411920.
16. Chourushi S, Lodha P, Prakash I. A critical review of hydrological modeling practices for flood management. Pramana Res J. 2019;9:352–62.
17. Singh S, Mishra K, Chavan R, Tiwari HL. Advancements and challenges in hydrological modeling: a comprehensive review. In: Pandey M, Umamahesh NV, Das J, Pu JH, editors. Hydrology and hydrologic modelling. Singapore: Springer Nature Singapore; 2025. p. 423–42. doi:10.1007/978-981-97-7474-6_32.
18. Freeze RA, Harlan RL. Blueprint for a physically-based, digitally-simulated hydrologic response model. J Hydrol. 1969;9(3):237–58. doi:10.1016/0022-1694(69)90020-1.
19. Paniconi C, Putti M. Physically based modeling in catchment hydrology at 50: survey and outlook. Water Resour Res. 2015;51(9):7090–129. doi:10.1002/2015wr017780.
20. Zhao X, Wang H, Bai M, Xu Y, Dong S, Rao H, et al. A comprehensive review of methods for hydrological forecasting based on deep learning. Water. 2024;16(10):1407. doi:10.3390/w16101407.
21. Zhou J, Chen L, Hu T, Lu H, Shi Y, Chen L. The comparative study of machine learning agent models in flood forecasting for tidal river reaches. Sci Rep. 2025 May;15(1):19130. doi:10.1038/s41598-025-04633-w.
22. Samantaray S, Sahoo A, Agnihotri A. Prediction of flood discharge using hybrid PSO-SVM algorithm in barak river basin. MethodsX. 2023;10:102060. doi:10.1016/j.mex.2023.102060.
23. Huang PC. An effective alternative for predicting coastal floodplain inundation by considering rainfall, storm surge, and downstream topographic characteristics. J Hydrol. 2022;607:127544. doi:10.1016/j.jhydrol.2022.127544.
24. Gu H, Xu YP, Ma D, Xie J, Liu L, Bai Z. A surrogate model for the variable infiltration capacity model using deep learning artificial neural network. J Hydrol. 2020;588:125019. doi:10.1016/j.jhydrol.2020.125019.
25. Zhang Y, Gu Z, Thé JVG, Yang SX, Gharabaghi B. The discharge forecasting of multiple monitoring station for humber river by hybrid LSTM models. Water. 2022;14(11):1794. doi:10.3390/w14111794.
26. Zhang Y, Zhou Z, Van Griensven Thé J, Yang SX, Gharabaghi B. Flood forecasting using hybrid LSTM and GRU models with lag time preprocessing. Water. 2023;15(22):3982. doi:10.3390/w15223982.
27. do Lago CAF, Giacomoni MH, Bentivoglio R, Taormina R, Gomes MN, Mendiondo EM. Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. J Hydrol. 2023;618:129276. doi:10.1016/j.jhydrol.2023.129276.
28. Goodfellow I, Bengio Y, Courville A. Deep learning. In: Adaptive computation and machine learning series. Cambridge, MA, USA: MIT Press; 2016.
29. Stinis P. Enforcing constraints for time series prediction in supervised, unsupervised and reinforcement learning. In: *2020 AAAI Spring Symposium on Combining Artificial Intelligence and Machine Learning with Physical Sciences AAAI-MLPS*; 2020 Mar 23–25; Stanford, CA, USA.

30. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Parallel distributed processing, volume 1: explorations in the microstructure of cognition: foundations. Cambridge, MA, USA: MIT Press; 1986. doi:10.7551/mitpress/5236.003.0012.
31. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. doi:10.1038/nature14539.
32. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput. 2006;18(7):1527–54 doi:10.1162/neco.2006.18.7.1527.
33. Seo DJ, Breidenbach J. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. J Hydrometeorol. 2002;3(2):93–111.
34. Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J Photogram Remote Sens. 2019;152:166–77. doi:10.1016/j.isprsjprs.2019.04.015.
35. Maddox RA, Zhang J, Gourley JJ, Howard KW. Weather radar coverage over the contiguous United States. Weather Forecast. 2002;17(4):927–34.
36. Gude V, Corns S, Long S. Flood prediction and uncertainty estimation using deep learning. Water. 2020;12(3):884. doi:10.3390/w12030884.
37. Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. In: Proceedings of the 29th International Conference on Neural Information Processing Systems-Volume 1, NIPS'15. Cambridge, MA, USA: MIT Press; 2015. p. 802–10.
38. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D Nonlinear Phenomena. 2020;404:132306. doi:10.1016/j.physd.2019.132306.
39. Sengupta S, Basak S, Saikia P, Paul S, Tsalavoutis V, Atiah F, et al. A review of deep learning with special emphasis on architectures, applications and recent trends. Knowl Based Syst. 2020;194:105596. doi:10.1016/j.knosys.2020.105596.
40. Pollack JB. Recursive distributed representations. Artif Intell. 1990;46(1–2):77–105. doi:10.1016/0004-3702(90)90005-k.
41. Bengio Y, Frasconi P, Simard P. The problem of learning long-term dependencies in recurrent networks. In: IEEE International Conference on Neural Networks. San Francisco, CA, USA: IEEE; 1993. Vol. 3, 1183–8.
42. Zhang D, Lin J, Peng Q, Wang D, Yang T, Sorooshian S, et al. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J Hydrol. 2018;565:720–36. doi:10.1016/j.jhydrol.2018.08.050.
43. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci. 2021;2(6):1–20.
44. Hadi FAA, Mohd Sidek L, Ahmed Salih GH, Basri H, Sammen SS, Mohd Dom N, et al. Machine learning techniques for flood forecasting. J Hydroinform. 2024;26(4):779–99. doi:10.2166/hydro.2024.208.
45. Obada E, Biao EI, Zohou PJ, Yarou H, Hounnondaho FZ, Alamou EA. Using machine learning and satellite data to improve flood forecasting: the case of the Ouémé basin at the Bétérou outlet. Hydrol Res. 2025;56(2):153–66. doi:10.2166/nh.2025.133.
46. Sharma S, Kumari S. Comparison of machine learning models for flood forecasting in the Mahanadi River Basin, India. J Water Climate Change. 2024;15(4):1629–52. doi:10.2166/wcc.2024.517.
47. Sit M, Demiray BZ, Xiang Z, Ewing GJ, Sermet Y, Demir I. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci Technol. 2020;82(12):2635–70. doi:10.2166/wst.2020.369.
48. Rehman S, Sahana M, Hong H, Sajjad H, Ahmed BB. A systematic review on approaches and methods used for flood vulnerability assessment: framework for future research. Nat Hazards. 2019;96(2):975–98. doi:10.1007/s11069-018-03567-z.

49. Maspo NA, Harun ANB, Goto M, Cheros F, Haron NA, Nawi MNM. Evaluation of machine learning approach in flood prediction scenarios and its input parameters: a systematic review. In: IOP Conference Series: Earth and Environmental Science. Kuala Lumpur, Malaysia: IOP Publishing; 2020. Vol. 479.
50. Opolot E. Application of remote sensing and geographical information systems in flood management: a review. Res J Appl Sci Eng Technol. 2013;6(10):1884–94.
51. Ahmad N, Hussain M, Riaz N, Subhani F, Haider S, Alamgir KS, et al. Flood prediction and disaster risk analysis using GIS based wireless sensor networks, a review. J Basic Appl Sci Res. 2013;3(8):632–43.
52. Wang X, Xie H. A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management. Water. 2018;10(5):608. doi:10.3390/w10050608.
53. Munawar HS, Hammad AW, Waller ST. Remote sensing methods for flood prediction: a review. Sensors. 2022;22(3):960. doi:10.3390/s22030960.
54. Han S, Coulibaly P. Bayesian flood forecasting methods: a review. J Hydrol. 2017;551:340–51. doi:10.1016/j.jhydrol.2017.06.004.
55. Teng J, Jakeman AJ, Vaze J, Croke BFW, Dutta D, Kim S. Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. Environ Modell Softw. 2017;90:201–16.
56. Piadeh F, Behzadian K, Alani A. A critical review of real-time modelling of flood forecasting in urban drainage systems. J Hydrol. 2022;607:127476. doi:10.1016/j.jhydrol.2022.127476.
57. Muhadi NA, Abdullah AF, Bejo SK, Mahadi MR, Mijic A. The use of LiDAR-derived DEM in flood applications: a review. Remote Sens. 2020;12(14):2308. doi:10.3390/rs12142308.
58. Jaafar K, Ismail N, Tajjudin M, Adnan R, Rahiman MHF. A review on flood modelling and rainfall-runoff relationships. In: 2015 IEEE 6th Control and System Graduate Research Colloquium (ICSGRC). Shah Alam, Malaysia: IEEE; 2015. p. 158–62. doi:10.1109/icsgrc.2015.7412484.
59. Iqbal U, Perez P, Li W, Barthelemy J. How computer vision can facilitate flood management: a systematic review. Int J Disaster Risk Reduct. 2021;53:102030. doi:10.1016/j.ijdrr.2020.102030.
60. Munawar HS, Hammad AW, Waller ST. A review on flood management technologies related to image processing and machine learning. Autom Constr. 2021;132:103916. doi:10.1016/j.autcon.2021.103916.
61. Shen C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour Res. 2018;54(11):8558–93. doi:10.1029/2018wr022643.
62. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
63. Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder-decoder approaches. In: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, Qatar: Association for Computational Linguistics; 2014. p. 103–11.
64. Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model. Artif Intell Rev. 2020;53(8):5929–55. doi:10.1007/s10462-020-09838-1.
65. Kratzert F, Klotz D, Shalev G, Klambauer G, Hochreiter S, Nearing G. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol Earth Syst Sci. 2019;23(12):5089–110. doi:10.5194/hess-23-5089-2019.
66. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555. 2014.
67. Kumar D, Singh A, Samui P, Jha RK. Forecasting monthly precipitation using sequential modelling. Hydrol Sci J. 2019;64(6):690–700. doi:10.1080/02626667.2019.1595624.
68. Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M. Rainfall-runoff modelling using long short-term memory (LSTM) networks. Hydrol Earth Syst Sci. 2018;22(11):6005–22. doi:10.5194/hess-22-6005-2018.

69. Damavandi HG, Shah R, Stampoulis D, Wei Y, Boscovic D, Sabo J. Accurate prediction of streamflow using long short-term memory network: a case study in the Brazos River Basin in Texas. Int J Environ Sci Develop. 2019;10(10):294–300. doi:10.18178/ijesd.2019.10.10.1190.
70. Qin J, Liang J, Chen T, Lei X, Kang A. Simulating and predicting of hydrological time series based on TensorFlow deep learning. Polish J Environ Stud. 2019;28(2):795–802. doi:10.15244/pjoes/81557.
71. Yang S, Yang D, Chen J, Zhao B. Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. J Hydrol. 2019;579:124229. doi:10.1016/j.jhydrol.2019.124229.
72. Gao S, Huang Y, Zhang S, Han J, Wang G, Zhang M, et al. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. J Hydrol. 2020;589:125188. doi:10.1016/j.jhydrol.2020.125188.
73. Song T, Ding W, Wu J, Liu H, Zhou H, Chu J. Flash flood forecasting based on long short-term memory networks. Water. 2019 Dec;12(1):109. doi:10.3390/w12010109.
74. Nakhaei M, Zanjanian H, Nakhaei P, Gheibi M, Moezzi R, Behzadian K, et al. Comparative evaluation of deep learning techniques in streamflow monthly prediction of the Zarrine River Basin. Water. 2024;16(2):208. doi:10.3390/w16020208.
75. LeCun Y. Generalization and network design strategies. Connectionism Perspect. 1989;19(143–155):18.
76. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems-Volume 1, NIPS'12. Red Hook, NY, USA: Curran Associates Inc.; 2012. p. 1097–1105.
77. Zhou YT, Chellappa R. Computation of optical flow using a neural network. In: IEEE 1988 International Conference on Neural Networks. San Diego, CA, USA: IEEE; 1988. Vol. 2, p. 71–8.
78. Wang Y, Fang Z, Hong H, Peng L. Flood susceptibility mapping using convolutional neural network frameworks. J Hydrol. 2020;582:124482. doi:10.1016/j.jhydrol.2019.124482.
79. Wang JH, Lin GF, Chang MJ, Huang I, Chen YR, et al. Real-time water-level forecasting using dilated causal convolutional neural networks. Water Resources Manage. 2019;33(11):3759–80. doi:10.1007/s11269-019-02342-4.
80. de Vitry MM, Kramer S, Wegner JD, Leitão JP. Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network. Hydrol Earth Syst Sci. 2019;23(11):4621–34. doi:10.5194/hess-23-4621-2019.
81. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer; 2015. Vol. 9351, 234–41.
82. Bhola PK, Nair BB, Leandro J, Rao SN, Disse M. Flood inundation forecasts using validation data generated with the assistance of computer vision. J Hydroinform. 2019;21(2):240–56. doi:10.2166/hydro.2018.044.
83. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–44. doi:10.1145/3422622.
84. Gautam A, Sit M, Demir I. Realistic river image synthesis using deep generative adversarial networks. Front Water. 2022;4:784441. doi:10.3389/frwa.2022.784441.
85. Brophy E, Wang Z, She Q, Ward T. Generative adversarial networks in time series: a systematic literature review. ACM Comput Surv. 2022;55(10):199. doi:10.1145/3559540.
86. Demiray BZ, Sit M, Demir I. D-SRGAN: DEM super-resolution with generative adversarial networks. SN Comput Sci. 2021;2(1):1–11.
87. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE; 2017. p. 5967–76.

88. Hofmann J, Schüttrumpf H. floodGAN: using deep adversarial learning to predict pluvial flooding in real time. Water. 2021;13(16):2255. doi:10.3390/w13162255.
89. Cheng M, Fang F, Navon I, Pain C. A real-time flow forecasting with deep convolutional generative adversarial network: application to flooding event in Denmark. Phys Fluids. 2021;33(5):056602. doi:10.1063/5.0051213.
90. Choi S, Kim Y. Rad-cGAN v1.0: radar-based precipitation nowcasting model with conditional generative adversarial networks for multiple dam domains. Geosci Model Develop. 2022;15(15):5967–85. doi:10.5194/gmd-15-5967-2022.
91. Ni L, Wang D, Singh VP, Wu J, Wang Y, Tao Y, et al. Streamflow and rainfall forecasting by two long short-term memory-based models. J Hydrol. 2020;583:124296. doi:10.1016/j.jhydrol.2019.124296.
92. Deng H, Chen W, Huang G. Deep insight into daily runoff forecasting based on a CNN-LSTM model. Nat Hazards. 2022;113:1675–96. doi:10.1007/s11069-022-05363-2.
93. Li P, Zhang J, Krebs P. Prediction of flow based on a CNN-LSTM combined deep learning approach. Water. 2022;14(6):993. doi:10.3390/w14060993.
94. Khatun A, Nisha MN, Chatterjee S, Sridhar V. A novel insight on input variable and time lag selection in daily streamflow forecasting using deep learning models. Environ Modell Softw. 2024;179:106126. doi:10.1016/j.envsoft.2024.106126.
95. Sadeghi Tabas S, Humaira N, Samadi S, Hubig NC. FlowDyn: a daily streamflow prediction pipeline for dynamical deep neural network applications. Environ Modell Softw. 2023;170:105854.
96. Miau S, Hung WH. River flooding forecasting and anomaly detection based on deep learning. IEEE Access. 2020;8:198384–402. doi:10.1109/access.2020.3034875.
97. Xu W, Chen J, Zhang XJ. Scale effects of the monthly streamflow prediction using a state-of-the-art deep learning model. Water Resour Manage. 2022;36:3609–25. doi:10.1007/s11269-022-03216-y.
98. Wegayehu EB, Muluneh FB. Multivariate streamflow simulation using hybrid deep learning models. Comput Intell Neurosci. 2021;2021:16. doi:10.1155/2021/5172658.
99. Xiang Z, Yan J, Demir I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resour Res. 2020;56(1):e2019WR025326.
100. Kao IF, Zhou Y, Chang LC, Chang FJ. Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting. J Hydrol. 2020;583:124631. doi:10.1016/j.jhydrol.2020.124631.
101. Bai Y, Bezak N, Sapač K, Klun M, Zhang J. Short-term streamflow forecasting using the feature-enhanced regression model. Water Resources Manage. 2019;33(14):4783–97. doi:10.1007/s11269-019-02399-1.
102. Liu D, Jiang W, Mu L, Wang S. Streamflow prediction using deep learning neural network: case study of Yangtze River. IEEE Access. 2020;8:90069–86. doi:10.1109/access.2020.2993874.
103. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc London Series A Math Phys Eng Sci. 1998;454(1971):903–95. doi:10.1098/rspa.1998.0193.
104. Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process. 2013;62(3):531–44. doi:10.1109/tsp.2013.2288675.
105. Zuo G, Luo J, Wang N, Lian Y, He X. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. J Hydrol. 2020;585:124776. doi:10.1016/j.jhydrol.2020.124776.
106. He X, Luo J, Zuo G, Xie J. Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks. Water Resources Manage. 2019;33(4):1571–90. doi:10.1007/s11269-019-2183-x.

107. Yang T, Sun F, Gentine P, Liu W, Wang H, Yin J, et al. Evaluation and machine learning improvement of global hydrological model-based flood simulations. Environ Res Lett. 2019;14(11):114027. doi:10.1088/1748-9326/ab4d5e.
108. Wang J, Li Y, Zhao R, Gao RX. Physics guided neural network for machining tool wear prediction. J Manuf Syst. 2020;57:298–310. doi:10.1016/j.jmsy.2020.09.005.
109. Zhong M, Zhang H, Jiang T, Guo J, Zhu J, Wang D, et al. A hybrid model combining the cama-flood model and deep learning methods for streamflow prediction. Water Resour Manage. 2023 Sep;37(12):4841–59. doi:10.1007/s11269-023-03583-0.
110. Worland SC, Steinschneider S, Asquith W, Knight R, Wieczorek M. Prediction and inference of flow duration curves using multioutput neural networks. Water Resources Res. 2019;55(8):6850–68. doi:10.1029/2018wr024463.
111. Nguyen DH, Bae DH. Correcting mean areal precipitation forecasts to improve urban flooding predictions by using long short-term memory network. J Hydrol. 2020;584:124710. doi:10.1016/j.jhydrol.2020.124710.
112. Zhou Q, Teng S, Situ Z, Liao X, Feng J, Chen G, et al. A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions. Hydrol Earth Syst Sci. 2023;27(9):1791–808. doi:10.5194/hess-27-1791-2023.
113. Mirjalili S. The ant lion optimizer. Adv Eng Softw. 2015;83:80–98. doi:10.1016/j.advengsoft.2015.01.010.
114. Yuan X, Chen C, Lei X, Yuan Y, Muhammad Adnan R. Monthly runoff forecasting based on LSTM-ALO model. Stochastic Environ Res Risk Assess. 2018;32(8):2199–212. doi:10.1007/s00477-018-1560-y.
115. Zhu S, Luo X, Yuan X, Xu Z. An improved long short-term memory network for streamflow forecasting in the upper Yangtze River. Stochastic Environ Res Risk Assess. 2020;34(9):1313–29. doi:10.1007/s00477-020-01766-4.
116. Rasmussen CE. Gaussian processes in machine learning. In: Advanced lectures on machine learning: ML summer schools 2003. Canberra, Australia, 2003 Feb 2–14, Tübingen, Germany, 2003 Aug 4–16, Revised Lectures. Berlin/Heidelberg, Germany: Springer; 2004. p. 63–71. doi:10.1007/978-3-540-28650-9_4.
117. Hu R, Fang F, Pain C, Navon I. Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. J Hydrol. 2019;575:911–20. doi:10.1016/j.jhydrol.2019.05.087.
118. Schilders W. Introduction to model order reduction. In: Model order reduction: theory, research aspects and applications. Vol. 13. Berlin/Heidelberg, Germany: Springer; 2008. p. 3–32.
119. Kapoor A, Pathiraja S, Marshall L, Chandra R. DeepGR4J: a deep learning hybridization approach for conceptual rainfall-runoff modelling. Environ Modell Softw. 2023;169:105831. doi:10.1016/j.envsoft.2023.105831.
120. Wang Z, Xu N, Bao X, Wu J, Cui X. Spatio-temporal deep learning model for accurate streamflow prediction with multi-source data fusion. Environ Model Softw. 2024;178:106091.
121. Yu C, Hu D, Shao H, Dai X, Liu G, Wu S. Runoff simulation driven by multi-source satellite data based on hydrological mechanism algorithm and deep learning network. J Hydrol Reg Stud. 2024;52:101720. doi:10.1016/j.ejrh.2024.101720.
122. Hu F, Yang Q, Yang J, Shao J, Wang G. An adaptive rainfall-runoff model for daily runoff prediction under the changing environment: stream-LSTM. Environ Model Softw. 2025;191:106524. doi:10.1016/j.envsoft.2025.106524.
123. Hannah DM, Demuth S, van Lanen HAJ, Looser U, Prudhomme C, Rees G, et al. Large-scale river flow archives: importance, current status and future needs. Hydrol Process. 2011;25(7):1191–200. doi:10.1002/hyp.7794.
124. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017;37(12):4302–15. doi:10.1002/joc.5086.
125. Demir I, Xiang Z, Demiray B, Sit M. WaterBench-Iowa: a large-scale benchmark dataset for data-driven streamflow forecasting. Earth Syst Sci Data. 2022;14(12):5605–16. doi:10.5194/essd-14-5605-2022.

126. Teegavarapu RSV, Chandramouli V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. J Hydrol. 2005;312(1):191–206. doi:10.1016/j.jhydrol.2005.02.015.
127. Brunner MI, Slater L, Tallaksen LM, Clark M. Challenges in modeling and predicting floods and droughts: a review. WIREs Water. 2021;8(3):e1520. doi:10.1002/wat2.1520.
128. Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3rd ed. Boston, MA, USA: Morgan Kaufmann; 2012.
129. Lepot M, Aubin JB, Clemens FH. Interpolation in time series: an introductive overview of existing methods, their performance criteria and uncertainty assessment. Water. 2017;9(10):796. doi:10.3390/w9100796.
130. Cismondi F, Fialho AS, Vieira SM, Reti SR, Sousa JM, Finkelstein SN. Missing data in medical databases: impute, delete or classify? Artif Intell Med. 2013;58(1):63–72. doi:10.1016/j.artmed.2013.01.003.
131. Ansley CF, Kohn R. On the estimation of ARIMA models with missing values. In: Time series analysis of irregularly observed data. New York, NY, USA: Springer; 1984. p. 9–37. doi:10.1007/978-1-4684-9403-7_2.
132. Fortuin V, Baranchuk D, Rätsch G, Mandt S. Gp-vae: deep probabilistic time series imputation. In: International Conference on Artificial Intelligence and Statistics. Palermo, Italy: PMLR; 2020. Vol. 108, 1651–61.
133. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. Sci Rep. 2018;8(1):1–12. doi:10.1038/s41598-018-24271-9.
134. Weerakody PB, Wong KW, Wang G, Ela W. A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing. 2021;441:161–78. doi:10.1016/j.neucom.2021.02.046.
135. Luo Y, Cai X, Zhang Y, Xu J, Yuan XJ. Multivariate time series imputation with generative adversarial networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2018. Vol. 31, p. 1603–14.
136. Nguyen M, Sun N, Alexander DC, Feng J, Yeo BT. Modeling Alzheimer's disease progression using deep recurrent neural networks. In: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI). Singapore: IEEE; 2018. p. 1–4. doi:10.1109/prni.2018.8423955.
137. Salgado CM, Azevedo C, Proença H, Vieira SM. Noise versus outliers. Second Anal Electron Health Rec. 2016;14:163–83. doi:10.1007/978-3-319-43742-2_14.
138. López-Oriona Á, Vilar JA. Outlier detection for multivariate time series: a functional data approach. Knowl Based Syst. 2021;233:107527. doi:10.1016/j.knosys.2021.107527.
139. Cabero I, Epifanio I, Piérola A, Ballester A. Archetype analysis: a new subspace outlier detection approach. Knowl Based Syst. 2021;217:106830. doi:10.1016/j.knosys.2021.106830.
140. Koufakou A, Georgiopoulos M. A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes. Data Min Knowl Discov. 2010;20(2):259–89. doi:10.1007/s10618-009-0148-z.
141. Akouemo HN, Povinelli RJ. Time series outlier detection and imputation. In: 2014 IEEE PES General Meeting Conference & Exposition. National Harbor, MD, USA: IEEE; 2014. p. 1–5.
142. Atkinson G, Metsis V. Identifying label noise in time-series datasets. In: Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers. New York, NY, USA: Association for Computing Machinery; 2020. p. 238–43.
143. Tran VN, Ivanov VY, Kim J. Data reformation—a novel data processing technique enhancing machine learning applicability for predicting streamflow extremes. Adv Water Resour. 2023;182:104569. doi:10.1016/j.advwatres.2023.104569.
144. Sharma RK, Kumar S, Padmalal D, Roy A. Streamflow prediction using machine learning models in selected rivers of Southern India. Int J River Basin Manage. 2024;22(4):529–55. doi:10.1080/15715124.2023.2196635.

145. Vafaeipour M, Rahbari O, Rosen MA, Fazelpour F, Ansarirad P. Application of sliding window technique for prediction of wind velocity time series. Int J Energy Environ Eng. 2014;5(2):1–7. doi:10.1007/s40095-014-0105-5.

146. Norwawi NM. 29-Sliding window time series forecasting with multilayer perceptron and multiregression of COVID-19 outbreak in Malaysia. In: Data science for COVID-19. Puchong, Malaysia: Academic Press; 2021. p. 547–64. doi:10.1016/b978-0-12-824536-1.00025-3.

Appendix A

Table A1: Summarized reports of the survey works studied in this review

Ref.	Year	Author	Title	Method used	Research issues
[50]	2013	Opolot	Application of Remote Sensing and Geographical Information Systems in Flood Management: A Review	RS and GIS method	Yes
[51]	2013	Ahmad et al.	Flood Prediction and Disaster Risk Analysis using GIS-based Wireless Sensor Networks, A Review	GIS method	No
[58]	2015	Jaafar et al.	A Review on Flood Modelling and Rainfall-Runoff Relationships	ANN Method	No
[54]	2017	Han and Coulibaly	Bayesian Flood Forecasting Methods: A Review	Bayesian method	Yes
[55]	2017	Teng et al.	Flood inundation modelling: A review of methods, recent advances and uncertainty analysis	Empirical methods, Hydrological methods, and Simplified conceptual methods	Yes
[52]	2018	Wang and Xie	A Review on Applications of Remote Sensing and Geographic Information Systems (GIS) in Water Resources and Flood Risk Management	RS, GIS, and Hydrological models	Yes
[10]	2018	Mosavi et al.	Flood Prediction Using Machine Learning Models: Literature Review	ML methods	No
[15]	2018	Jain et al.	A Brief review of flood forecasting techniques and their applications	Hydrological methods	Yes
[48]	2019	Rehman et al.	A systematic review on approaches and methods used for flood vulnerability assessment: framework for future research	RS, GIS, hydrological and ML methods	No
[16]	2019	Chourushi et al.	A Critical Review of Hydrological Modeling Practices for Flood Management	Hydrological methods	No

(Continued)

Table A1 (continued)

Ref.	Year	Author	Title	Method used	Research issues
[57]	2020	Muhadi et al.	The Use of LiDAR-Derived DEM in Flood Applications: A Review	LiDAR with DEMs	Yes
[49]	2020	Maspo et al.	Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review	ML methods	No
[47]	2020	Sit et al.	A comprehensive review of DL applications in hydrology and water resources	DL methods	Yes
[59]	2021	Iqbal et al.	How computer vision can facilitate flood management: A systematic review	Computer vision method	Yes
[60]	2021	Munawar et al.	A review on flood management technologies related to image processing and machine learning	Image processing and ML methods	Yes
[56]	2022	Piadeh et al.	A critical review of real-time modelling of flood forecasting in urban drainage systems	Empirical, conceptual, and physical methods	No
[53]	2022	Munawar et al.	Remote Sensing Methods for Flood Prediction: A Review	RS methods	Yes

Table A2: DL models for different types of flood predictions

Ref.	DL models	Data types	Prediction types	Location
[42]	BPNN, LSTM	Sequential reservoir data	Monthly, Daily and Hourly	China
[68]	LSTM	Sequential Meteorological and streamflow data	Daily	USA
[69]	LSTM	Sequential Meteorological and streamflow data	Both short and long term	USA
[70]	AR, LSTM	Hydrological time series data	Daily	China
[67]	RNN, LSTM	Meteorological time series data	Monthly	India

(Continued)

Table A2 (continued)

Ref.	DL models	Data types	Prediction types	Location
[71]	RNNs, LSTM	Sequential Reservoir inflow, outflow, geographical and meteorological data	Daily	Thailand
[73]	LSTM	Observed or Predicted Rainfall and Early Discharge	Hourly	China
[72]	GRU, LSTM	Hourly flow discharge and precipitation data	Hourly	China
[79]	CNN	Hourly Rainfall and water-level data	Hourly	Taiwan
[80]	CNN	Image and video data	Image Classification	Europe
[82]	CNN	Image and Water-level data	Estimate inundation	Germany
[88]	GAN	Synthetic and Historical rainfall events	Real time	Germany
[89]	GAN	real time high fidelity-MIKE surface flow data	Hourly	Denmark
[90]	GAN	Radar reflectivity Imagery data	Real time	South Korea
[107]	GHM-LSTM, LSTM	Daily streamflow and meteorological data	Daily	Global
[110]	DNN	Daily mean streamflow data	Real time	USA
[111]	LSTM	Three-hour QPF's data	Hourly	South Korea
[99]	LSTM, LSTM-seq2seq	Hourly rainfall, runoff, and Monthly evapotranspiration of two watersheds	Hourly	USA
[100]	En-De-LSTM, En-De-FFNN	Hourly hydrological data	Hourly	Taiwan
[96]	CNN-GRU, CNN, LSTM, Seq2Seq	Hourly river water level data	Real time	Taiwan
[97]	CNN-GRU	Monthly streamflow data	Monthly	Global
[98]	CNN-GRU, CNN-LSTM, LSTM, GRU	Time series hydrological and meteorological data of two different location	Monthly, weekly, and daily	Italy, Ethiopia
[101]	SAE-LSTM, LSTM	Historical daily discharge data	Daily	China
[102]	EMD-En-De-LSTM	Monthly streamflow data	Monthly, yearly	China
[65]	EA-LSTM	Meteorological time series and catchment data	Streamflow prediction	USA

(Continued)

Table A2 (continued)

Ref.	DL models	Data types	Prediction types	Location
[114]	LSTM-ALO	Monthly runoff series data	Monthly	Pakistan
[91]	WLSTM, CNN-LSTM	Monthly streamflow and rainfall series data	Monthly	China
[92]	CNN-LSTM	Daily runoff, rainfall and temperature data	Daily	China
[93]	CNN-LSTM	Hourly radar maps, rainfall and catchment flow data	Daily	Germany
[25]	LSTM, CNN-LSTM	Hourly rainfall and watershed data	Hourly	Canada
[115]	LSTM-HetGP, LSTM	Daily streamflow time series data	Daily	China
[105]	LSTM	Daily streamflow data	Daily	China
[117]	LSTM-ROM	Free surface height of induced wave of Okushiri tsunami	Real-time	Japan
[106]	VMD-DNN	Daily runoff series data	Daily	China
[109]	VIC-CaMa-Flood -RNN, VIC-CaMa-Flood -LSTM	Geographical, Meteorological and streamflow data	Daily	China
[94]	CNN-LSTM, CNN-GRU	Meteorological data containing Gridded rainfall and temperature data	Daily	India
[119]	DeepGR4J	CAMELS-AUS dataset with 222 catchments	Daily	Australia
[120]	CVTGR	Hydrological and meteorological data	Daily	China
[46]	CNN-RF, CNN-SVR	Daily Rainfall, discharge, and water level measurement	Daily	India
[121]	LSTM-MMSCS	Satellite data for precipitation temperature, and evapotranspiration	Daily	China
[122]	Stream-LSTM	CAMELS dataset	Daily	China
[95]	CNN-LSTM	CAMELS, NCDC, and GRDC streamflow data	Daily	North and South America
[45]	CNN-LSTM, LSTM	Historical climate data	Hourly	Benin
[74]	RNN, LSTM	Monthly streamflow of Zarrine River	Monthly	Iran

Table A3: Performances of flood prediction models survey

DL models	Ref.	Performance metrics	Prediction type	Datasets
BPNN	[42]	RMSE = 1122.827	Monthly	Sequential Reservoir data
		RSR = 0.1411	Monthly	
		NSE = 0.9799	Monthly	
		RMSE = 1073.402	Daily	
		RSR = 0.1253	Daily	
		NSE = 0.9843	Daily	
		RMSE = 1596.052	Hourly	
		RSR = 0.1885	Hourly	
		NSE = 0.9645	Hourly	
LSTM	[42]	RMSE = 161.712	Monthly	Sequential Reservoir data
		RSR = 0.0203	Monthly	
		NSE = 0.9996	Monthly	
		RSME = 663.697	Daily	
		RSR = 0.023	Daily	
		NSE = 0.9935	Daily	
		RMSE = 488.121	Hourly	
		RSR = 0.0577	Hourly	
		NSE = 0.9967	Hourly	
	[67]	$R^2 = 0.960$	Monthly	Meteorological time series data
		RMSE = 263.89	Monthly	
		NSE = 0.916	Monthly	
		MAE = 177.14	Monthly	
	[91]	RMSE = 88.53	Monthly	Monthly streamflow and rainfall series data
		NSE = 0.81	Monthly	
		MARE = 0.18	Monthly	
	[98]	RMSE = 5.55	Monthly	Time series hydrological and meteorological daily data of two different location
		MAE = 3.49	Monthly	
		$R^2 = 0.91$	Monthly	
		RMSE = 7.93	Daily	
		MAE = 3.91	Daily	
		$R^2 = 0.84$	Daily	
	[68]	NSE = 0.68	Daily	Meteorological and Daily discharge data

(Continued)

Table A3 (continued)

DL models	Ref.	Performance metrics	Prediction type	Datasets
	[71]	$R^2 = 0.96$	Daily	Sequential Reservoir inflow, outflow, geographical and meteorological data
		RMSE = 37.71	Daily	
		NSE = 0.91	Daily	
	[115]	RMSE = 1311	Daily	Daily streamflow time series data
		MRE = 0.17	Daily	
		MSLE = 0.06	Daily	
		NSE = 0.86	Daily	
	[101]	RMSE = 2.95	Daily	Historical daily discharge data
		$R^2 = 0.93$	Daily	
		NSE = 0.83	Daily	
LSTM	[70]	$R^2 = 0.4543$	Daily	Hydrological time series data
		RMSE = 122.01	Daily	
		MAE = 115.82	Daily	
	[69]	$\rho = 0.95$	Daily	Sequential meteorological and streamflow data
		NSE = 0.88	Daily	
	[92]	NSE = 0.97	Daily	Daily runoff, rainfall and temperature data
		RMSE = 212	Daily	
	[99]	$R^2 = 0.97$	Hourly	Hourly rainfall, runoff, and Monthly evapotranspiration of two watersheds
		NSE = 0.93	Hourly	
		RMSE =0.21	Hourly	
	[96]	RMSE = 1.032	Hourly	Hourly river water level data
		MAPE = 31.035	Hourly	
		MAE = 0.620	Hourly	
	[72]	NSE = 0.98	Hourly	Hourly flow discharge and precipitation data
		MAE = 6.92	Hourly	
GRU	[72]	NSE = 0.99	Hourly	Hourly flow discharge and precipitation data
		MAE = 6.00	Hourly	
	[98]	RMSE = 5.22	Monthly	Time series hydrological and meteorological daily data of two different location
		MAE = 3.06	Monthly	
		$R^2 = 0.92$	Monthly	

(Continued)

Table A3 (continued)

DL models	Ref.	Performance metrics	Prediction type	Datasets
		RMSE = 7.94	Daily	
		MAE = 3.64	Daily	
		R^2 = 0.85	Daily	
GAN	[89]	RMSE = 0.15	Hourly	Real time high fidelity-MIKE surface flow data
		MAE = 0.08	Hourly	
		NSE = 0.97	Hourly	
		ρ = 0.98	Hourly	
EMD-En-De -LSTM	[102]	WI = 0.9868	Monthly	Monthly streamflow data
		LMI = 0.7611	Monthly	
		RMSE = 2133.748	Monthly	
		R^2 = 0.9384	Monthly	
LSTM-ALO	[114]	R^2 = 0.9484	Monthly	Monthly runoff series data
		RMSE = 0.074	Monthly	
		MAE = 0.040	Monthly	
		NSE = 0.867	Monthly	
	[91]	RMSE = 88.42	Monthly	Monthly streamflow and rainfall series data
		NSE = 0.81	Monthly	
		MARE = 0.20	Monthly	
		RMSE = 5.36	Monthly	
		MAE = 3.17	Monthly	
		R^2 = 0.92	Monthly	Daily hydrological and meteorological data of two different location
	[98]	RMSE = 7.99	Daily	
CNN-LSTM		MAE = 4.09	Daily	
		R^2 = 0.85	Daily	
	[92]	NSE = 0.98	Daily	Daily runoff, rainfall and temperature data
		RMSE = 203	Daily	
	[93]	R^2 = 0.91	Daily	Hourly radar maps, rainfall and catchment flow data
		NSE = 0.91	Daily	
		KGE = 0.92	Daily	
		ρ = 0.97	Daily	
	[94]	NSE = 0.95	Daily	Meteorological data containing Gridded rainfall and temperature data
		RSR = 0.23	Daily	

(Continued)

Table A3 (continued)

DL models	Ref.	Performance metrics	Prediction type	Datasets
		MAE = 496.36	Daily	
	[95]	NSE = 83	Daily	CAMELS, NCDC, and GRDC streamflow data
	[45]	NSE = 0.95	Hourly	Historical climate data
		R^2 = 0.96	Hourly	
		RMSE = 0.0024	Hourly	
SAE-LSTM	[101]	RMSE = 1.76	Daily	Historical daily discharge data
		R^2= 0.98	Daily	
		NSE = 0.94	Daily	
LSTM-HetGP	[115]	RMSE = 1298	Daily	Daily streamflow time series data
		MARE = 0.16	Daily	
		NSE = 0.87	Daily	
		MSLE = 0.05	Daily	
VMD-DNN	[106]	NSE = 0.95	Daily	Daily runoff series data
		RMSE = 9.92	Daily	
		MAE = 3.82	Daily	
LSTM-seq2seq	[99]	R^2 = 0.98	Hourly	Hourly rainfall, runoff, and Monthly evapotranspiration of two watersheds
		NSE = 0.93	Hourly	
		RMSE = 0.21	Hourly	
EnDe-LSTM	[100]	R^2 = 0.99	Hourly	Hourly hydrological data
		NSE = 0.99	Hourly	
		RMSE = 25	Hourly	
EnDe-FFNN		R^2 = 0.99	Hourly	
		NSE = 0.96	Hourly	
		RMSE = 43	Hourly	
CNN-GRU	[98]	RMSE = 5.15	Monthly	Time series hydrological and meteorological daily data of two different location
		MAE = 3.18	Monthly	
		R^2 = 0.92	Monthly	
		RMSE = 7.94	Daily	
		MAE = 3.66	Daily	
		R^2 = 0.85	Daily	
		RMSE = 0.774	Hourly	
		MAPE = 30.684	Hourly	
		MAE = 0.567	Hourly	
CNN	[96]	RMSE = 1.144	Hourly	Hourly river water level data
		MAPE = 37.154	Hourly	

(Continued)

Table A3 (continued)

DL models	Ref.	Performance metrics	Prediction type	Datasets
		MAE = 0.745	Hourly	
DeepGR4J	[119]	NSE = 0.9262	Daily	CAMELS-AUS dataset with 222 catchments
CVTGR	[120]	NSE = 0.943	Daily	Hydrological and meteorological data
		RMSE = 23.746	Daily	
		MAE = 13.670	Daily	
CNN-RF	[46]	$R^2 = 0.95$	Daily	Daily Rainfall, discharge, and water level measurement
		MAE = 54.19	Daily	
		RMSE = 172.83	Daily	
		NSE = 0.95	Daily	
LSTM-MMSCS	[121]	$R^2 = 0.90$	Daily	Satellite data for precipitation temperature, and evapotranspiration
		NSE = 0.82	Daily	
		KGE = 0.85	Daily	
		RMSE = 24.47	Daily	
		RE = 13.39	Daily	
Stream-LSTM	[122]	NSE = 0.91	Daily	CAMELS dataset
		RMSE = 147.167	Daily	
RNN	[74]	NSE = 0.85	Monthly	Monthly streamflow of Zarrine River
		R2 = 0.87	Monthly	
		RMSE = 13.65	Monthly	