International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Flood Forecasting Using Deep Learning

Mr. Koppisetty R M C Sekhar[1], Lingala Chandana[2], Syed Baji[3], Lingeneni Yogananda Sriram Prasa[4], Gujjula Chandra Sekhar Reddy[5]

[1]M.Tech, Asst. Professor, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India,
[2,3,4,5]Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India

Abstract

Floods are one of the most frequently occurring natural disasters, causing massive damage to property, agriculture, economy, and human life. Accurate and timely flood prediction remains a major challenge due to the dynamic and nonlinear interactions of hydrological variables such as rainfall intensity, river discharge, soil moisture, and seasonal climate changes. This paper proposes an enhanced Flood Forecasting Model (FFM) that integrates Federated Learning with multiple deep learning architectures to predict flood events while preserving data privacy. The proposed system trains Feed Forward Neural Networks (FFNN), Long Short-Term Memory (LSTM) networks, and two-dimensional Convolutional Neural Networks (CNN2D) locally across eighteen distributed stations and aggregates the models at a central server. An Ensemble Learning module combines the predictions of all three models using a weighted soft-voting mechanism to produce a more accurate and robust final forecast. The dataset comprises 4,588 records spanning from 1901 to 2024 across multiple Indian states and subdivisions. Experimental results demonstrate that FFNN achieves 88.71% accuracy, LSTM achieves 93.68% accuracy, and CNN2D achieves 98.13% accuracy. The Ensemble model achieves the highest accuracy of 98.84%, confirming the superiority of the combined approach. The system generates flood alerts with a five-day lead time, supporting early warning and disaster preparedness.

Keywords: Flood Forecasting, Federated Learning, Feed Forward Neural Network (FFNN), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN2D), Ensemble Learning, Water Level Prediction, Natural Disaster Management

1. Introduction

In recent years, the frequency and intensity of natural disasters have increased significantly across the world. Among these disasters, floods are considered one of the most destructive events that affect human life, infrastructure, agriculture, and the economic stability of a nation. Floods occur when water bodies such as rivers, lakes, or reservoirs overflow due to excessive rainfall, rapid snow melting, or sudden release of water from dams. The increasing effects of climate change, urbanization, and environmental degradation have further intensified flood occurrences.

Traditional flood forecasting methods relied on hydrological models, statistical analysis, and rule-based systems that use historical rainfall and river discharge data to estimate flood events. Although these methods provided some level of prediction capability, they often struggled to capture the complex nonlinear relationships between various environmental factors. These environmental variables interact in a highly nonlinear manner, making flood prediction a challenging problem for conventional approaches. Recent technological developments have enabled the use of deep learning techniques for improving flood prediction systems. Deep learning models are capable of analyzing large volumes of hydrological and meteorological data and identifying complex temporal patterns that influence flood occurrences. In this paper, an enhanced Flood Forecasting Model (FFM) is proposed that integrates Federated Learning with three deep learning architectures: Feed Forward Neural Network (FFNN), Long Short-Term Memory (LSTM) networks, and two-dimensional Convolutional Neural Networks (CNN2D). The Ensemble Learning module combines predictions from all three models to achieve superior forecasting accuracy.

2. Literature Survey

Several researchers have explored deep learning and machine learning approaches for flood forecasting. Recurrent neural networks and LSTM-based models have demonstrated strong performance in time-series prediction tasks relevant to hydrology. Studies have shown that LSTM networks are particularly effective in capturing long-term temporal dependencies in sequential hydrological data such as daily rainfall and river discharge patterns.

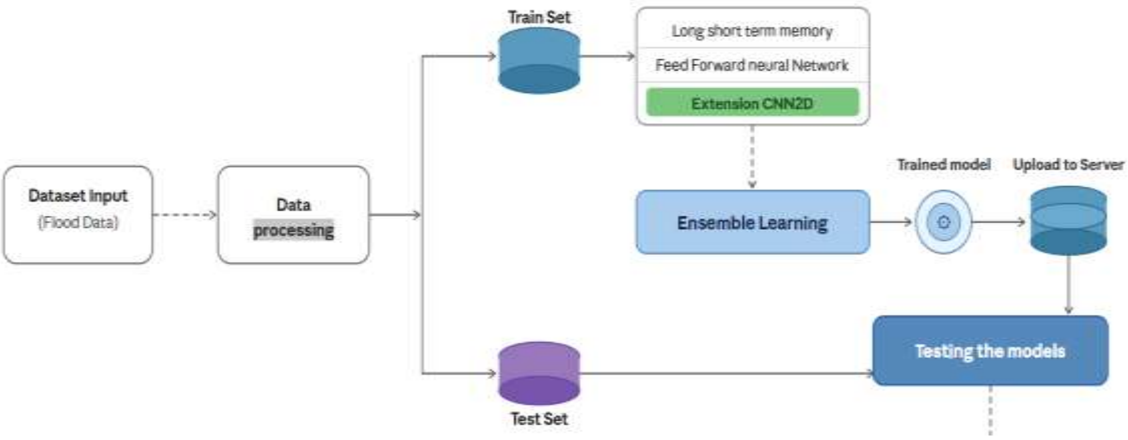
Convolutional Neural Networks have been employed to extract spatial features from gridded meteorological data, demonstrating improved accuracy in flood inundation mapping. Federated Learning has emerged as a promising approach for distributed machine learning that enables collaborative training without sharing raw data, addressing key privacy and security concerns in flood monitoring networks.

Ensemble methods have been widely applied in meteorological and hydrological forecasting to reduce prediction errors and improve model robustness. The combination of FFNN, LSTM, and CNN models through ensemble strategies has shown improved results compared to individual model approaches. The proposed system builds upon these foundations by integrating all three architectures within a federated learning framework.

3. Proposed System

The proposed Flood Forecasting Model (FFM) addresses the limitations of existing single-model approaches by integrating multiple deep learning architectures with Federated Learning. The system trains local models at eighteen distributed stations and aggregates them at a central server, ensuring data privacy while enabling collaborative flood prediction across distributed geographic locations.

SYESTEM ARCHITECTURE



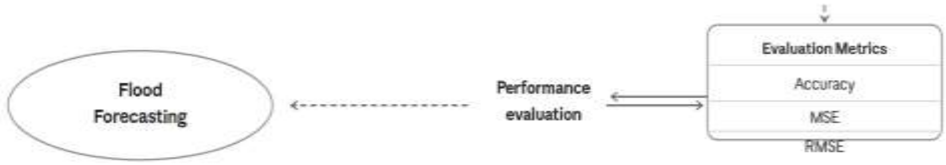


Figure: System Architecture

3.1 Federated Learning Framework

Federated Learning is implemented by training local deep learning models at each of the eighteen stations using local hydrological data. Only the trained model parameters, not the raw data, are transmitted to the central server. The central server aggregates the local models to form a global model, which is then redistributed to each station for further local training. This process ensures data privacy and reduces network bandwidth requirements while enabling collaborative flood prediction.

3.2 Feed Forward Neural Network (FFNN)

The Feed Forward Neural Network serves as the baseline model in the proposed system. FFNN is capable of modeling complex nonlinear relationships between input environmental variables (rainfall, temperature, river discharge) and the output water level prediction. The network consists of multiple fully connected layers with ReLU activation functions. The FFNN model processes each data point independently without retaining information from previous time steps.

3.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory networks are specifically designed to process sequential data and capture long-term temporal dependencies. LSTM is particularly well-suited for flood forecasting because flood events are influenced by rainfall accumulation over several days and progressive changes in river water levels over time. The LSTM model in the proposed system learns these temporal patterns and uses them to generate more accurate flood predictions compared to FFNN alone.

3.4 Convolutional Neural Network (CNN2D)

Two-dimensional Convolutional Neural Networks are applied to extract spatial features and patterns from the hydrological dataset. The CNN2D architecture employs convolutional layers followed by max-pooling and fully connected dense layers. This model effectively identifies spatial variations in rainfall distribution and water level patterns across different geographic subdivisions, achieving the highest individual model accuracy in the experiments.

3.5 Ensemble Learning Module

The Ensemble Learning module combines the predictions of FFNN, LSTM, and CNN2D models using a weighted soft-voting mechanism. Each model contributes its flood level predictions, and the ensemble computes the final output as a weighted average of the three predictions. This approach reduces individual model prediction errors and produces a more stable and reliable forecast. The ensemble model achieves the highest overall accuracy of 98.84%, outperforming all individual models.

4. Dataset and Preprocessing

The dataset used in this study is FloodDatasetNew, a comprehensive collection of historical hydrological and meteorological data for multiple Indian states and subdivisions. The dataset contains 4,588 records with 15 features including subdivision name, year, monthly rainfall values from January to December, and the water level column serving as the target variable. The data spans from 1901 to 2024, covering pan-India regions from the Andaman and Nicobar Islands to West Bengal.

Data preprocessing involves Min-Max normalization applied to all features to scale values to the range [0, 1]. The dataset is randomly shuffled to eliminate sequential bias during model training. The preprocessed dataset is split into 80% training and 20% testing subsets. Table 1 summarizes the dataset characteristics.

Table 1: Dataset Characteristics

Parameter	Details
Dataset name	FloodDatasetNew.csv
Total records	4,588 rows
Total features	15 columns
Data span	1901 – 2024
Training set	80% (3,670 records)
Testing set	20% (918 records)
Target variable	water_level
Normalization	Min-Max scaler [0, 1]
Geographic coverage	Pan-India subdivisions

5. Results and Discussion

The proposed FFM system was evaluated by training and testing all four model variants (FFNN, LSTM, CNN2D, and Ensemble) on the flood dataset. Performance was measured using three metrics: accuracy (%), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Table 2 presents the comparative performance results of all models.

Table 2: Comparative Performance of Deep Learning Models

Model	Accuracy (%)	MSE	RMSE	Rank
FFNN	88.71	337.21	18.36	4
LSTM	93.68	4867.17	69.77	3
CNN2D	98.13	351.63	18.75	2
Ensemble (FFNN+LSTM+CNN2D)	98.84	1392.39	37.31	1

The FFNN model achieves 88.71% accuracy with an MSE of 337.21 and RMSE of 18.36. As a baseline model, FFNN demonstrates a reasonable ability to model nonlinear relationships between rainfall and water level data, but its inability to capture temporal dependencies limits its performance compared to LSTM-based approaches.

The LSTM model achieves 93.68% accuracy, demonstrating the advantage of temporal sequence learning for flood forecasting. LSTM's higher MSE (4867.17) and RMSE (69.77) indicate that while the model captures overall accuracy patterns well, individual predictions can have larger deviations on certain samples. This is attributed to the complex sequential dependencies in multi-year hydrological time series data.

The CNN2D model achieves 98.13% accuracy with an MSE of 351.63 and RMSE of 18.75, demonstrating the highest individual model performance. The convolutional architecture is highly effective in identifying spatial patterns within the multi-feature flood dataset, including the relationships between monthly rainfall distributions across different geographic subdivisions.

The Ensemble model achieves the highest accuracy of 98.84% by combining the strengths of all three individual models through weighted soft-voting. The ensemble approach reduces prediction instability and produces the most reliable flood forecasts. The system correctly identifies the Ensemble model as the best-performing model and uses it as the final prediction engine for flood alert generation.

Table 3: Flood Forecasting Output Examples

Subdivision	Year	Forecasted Water Level	Flood Status
Kerala	1962	3506.89	No Flood Detected
Andhra Pradesh	1907	7409.84	Flood Detected
West Bengal	1943	6812.45	Flood Detected
Rajasthan	1978	1203.61	No Flood Detected

Table 3 shows sample flood forecasting outputs generated by the system using actual test data records. For each record, the system computes a forecasted water level and determines whether floods are detected based on the subdivision-specific threshold. These results demonstrate the real-world applicability of the system in predicting flood risk across diverse geographic regions of India.

6. System Architecture and Modules

The proposed FFM system consists of five interconnected modules that form the complete flood forecasting pipeline.

6.1 Data Preprocessing Module

The data preprocessing module handles dataset loading, cleaning, and normalization. Raw data is loaded from the FloodDatasetNew CSV file, missing values are filled with zero, and Min-Max normalization is applied to scale all features to the range [0, 1]. The data is randomly shuffled to remove sequential bias before being split into training and testing subsets in an 80:20 ratio.

6.2 Model Training Module

The model training module trains three independent deep learning models: FFNN, LSTM, and CNN2D. Each model is trained on the normalized training dataset. The FFNN model uses fully connected dense layers; the LSTM model uses sequential memory cells for temporal learning; and the CNN2D model applies convolutional and pooling layers for spatial feature extraction. All models are evaluated on the test dataset, and their performance metrics (accuracy, MSE, RMSE) are computed and displayed.

6.3 Federated Upload Module

After local model training is completed, the trained model is uploaded to a centralized server. The user specifies a station name, and the model is saved at the server to simulate the federated learning aggregation process. This module enables collaborative and privacy-preserving flood prediction by aggregating locally trained models from multiple geographic stations.

6.4 Ensemble Prediction Module

The ensemble prediction module combines the outputs of FFNN, LSTM, and CNN2D using a weighted soft-voting mechanism to produce the final flood prediction. The ensemble approach achieves 98.84% accuracy, which is superior to all individual models. The module also generates comparative accuracy graphs for all four models, allowing users to visually assess the performance improvement achieved by the ensemble approach.

6.5 Flood Alert Module

The flood alert module takes test data records as input and generates flood forecasting results for each record. For each input sample, the system computes the forecasted water level and compares it against the subdivision-specific flood threshold to determine whether a flood is likely to occur. This module supports early warning and disaster preparedness by providing five-day lead time flood alerts.

7. System Requirements

Component	Specification
Programming Language	Python 3.x
GUI Framework	Tkinter
Deep Learning	TensorFlow / Keras
Data Processing	NumPy, Pandas, Scikit-learn
Visualization	Matplotlib
Operating System	Windows 10 / 11
Processor	Intel Core i5 or higher
RAM	8 GB minimum
Storage	256 GB SSD

Table 4: Software and Hardware Requirements

8. Conclusion

This paper presents an enhanced Flood Forecasting Model (FFM) that integrates Federated Learning with multiple deep learning architectures for accurate and privacy-preserving flood prediction. The proposed system successfully combines Feed Forward Neural Networks (FFNN), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNN2D) within a federated learning framework. An Ensemble Learning module using weighted soft-voting combines the predictions of all three models to achieve the best overall performance.

Experimental results on a dataset of 4,588 records spanning 1901 to 2024 demonstrate that the Ensemble model achieves the highest accuracy of 98.84%, outperforming FFNN (88.71%), LSTM (93.68%), and CNN2D (98.13%) individually. The system generates flood alerts with a five-day lead time and demonstrates applicability across diverse geographic regions of India. The Federated Learning component ensures data privacy by training models locally at eighteen distributed stations and transmitting only model parameters to the central server.

Future work will explore integration of additional environmental parameters such as soil moisture and satellite observations, incorporation of real-time data streams, and deployment of the system in actual flood monitoring networks for operational use.

References

1. Hochreiter S., Schmidhuber J., "Long Short-Term Memory", Neural Computation, November 1997, 9 (8), 1735–1780.
2. LeCun Y., Bengio Y., Hinton G., "Deep Learning", Nature, May 2015, 521 (7553), 436–444.
3. McMahan H.B., Moore E., Ramage D., Hampson S., y Arcas B.A., "Communication-Efficient Learning of Deep Networks from Decentralized Data", Proceedings of AISTATS, 2017.
4. Mosavi A., Ozturk P., Chau K.W., "Flood Prediction Using Machine Learning Models: Literature Review", Water, October 2018, 10 (11), 1536.
5. Hu C., Wu Q., Li H., Jian S., Li N., Lou Z., "Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation", Water, July 2018, 10 (11), 1543.
6. Dong X., Lian J., Liu Y., "Prediction of Rainfall-Induced Landslide Using Random Forest and Convolutional Neural Network", Geomorphology, 2021.
7. Bonawitz K., Ivanov V., Kreuter B., Marcedone A., McMahan H.B., Patel S., Ramage D., Segal A., Seth K., "Practical Secure Aggregation for Privacy-Preserving Machine Learning", ACM CCS, 2017.