Advanced Flood Prediction Using Federated Learning with 2D Convolutional Networks

P.Arun Reddy[1], Pemmanaboina Rohith Karthikeya[2], Ettaboina Shiva Krishna[3], Dadige Shiva Kumar[4], Navapet Sharan Kumar Yadav[5]

[1] Assistant Professor, Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

[2,3,4,5]BTech Students ,Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

ABSTRACT:

Floods are one of the most destructive natural disasters, causing serious damage to lives, agriculture, and infrastructure. Predicting floods in advance is a difficult task due to changing environmental conditions and complex water systems. This study presents a flood forecasting model based on federated learning, which allows multiple locations to train models without sharing raw data. This approach helps reduce network delays while maintaining data privacy and security. Instead of sending large datasets to a central server, local models are trained and only the learned parameters are shared. The proposed system analyzes data from 18 different stations to identify areas at risk and provides flood alerts up to five days in advance. A feedforward neural network (FFNN) is used to predict water levels using factors such as rainfall, snowmelt, flow routing, and hydrodynamics. The model was tested on data from 2010 to 2015 and achieved an accuracy of 84%.

Keywords:

Flood Forecasting, Federated Learning, CNN2D, FFNN, Water Level Prediction, Hydrological Data, Rainfall-Runoff, Disaster Management

INTRODUCTION

Floods are among the most frequent and destructive natural disasters, causing severe damage to human life, agriculture, infrastructure, and national economies. In recent years, the occurrence of both natural and human-induced disasters has increased significantly due to rapid urbanization, climate change, and extreme hydrological events. These factors have intensified the frequency and severity of floods across the globe. According to the World Health Organization, floods pose serious risks to health, safety, and livelihoods, especially in vulnerable regions [1]. Developing countries are particularly affected, where limited resources and weak infrastructure make communities more susceptible to large-scale losses.

The causes of floods are complex and vary across regions, including excessive rainfall, snowmelt, river overflow, and poor drainage systems. Studies have shown that climate change has accelerated snowmelt and altered precipitation patterns, leading to more frequent and intense flooding events [3], [4]. In South Asian countries, floods have become a recurring disaster, resulting in substantial socio-economic impacts [6]. Traditional flood prediction methods, such as climatology-based models, flood frequency analysis, and Bayesian approaches, rely on statistical techniques to estimate flood occurrences. However, these models often struggle to capture the dynamic and nonlinear nature of environmental systems [13].

With advancements in technology, machine learning (ML) and deep learning approaches have been widely adopted for flood prediction due to their ability to analyze large datasets and identify complex patterns. Techniques such as artificial neural networks (ANN) and deep learning models have shown promising results in improving prediction accuracy [11], [12]. Despite these improvements, ML-based systems require large volumes of centralized data for training, which raises concerns related to data privacy, security, and communication overhead. Additionally, transferring massive datasets across networks can lead to latency issues and increased computational costs.

To address these challenges, federated learning (FL) has emerged as a novel and effective solution. FL enables decentralized model training by allowing multiple clients to train models locally using their own data without sharing raw information [9]. Only model parameters are exchanged with a central server, ensuring data privacy and reducing communication overhead. This approach is particularly beneficial for flood forecasting, where data is often distributed across different geographical locations and institutions.

In this study, a Federated Learning-based Flood Forecasting Model (FFM) is proposed to improve prediction accuracy while maintaining data privacy and efficiency. The model integrates FL with a feedforward neural network (FFNN) to analyze regional hydrological factors and predict future water levels. By leveraging decentralized data and collaborative learning, the proposed system provides early flood warnings up to five days in advance. This approach not only enhances prediction performance but also supports timely decision-making for disaster prevention, mitigation, and management.

RELATED WORK

[1] Floods – World Health Organization, the World Health Organization highlights floods as one of the most common and dangerous natural disasters affecting global populations. Floods not only cause immediate physical damage but also lead to long-term health risks such as waterborne diseases, contamination of drinking water, and limited access to healthcare services. The report emphasizes that vulnerable populations, especially in low-income regions, are at greater risk due to poor infrastructure and lack of preparedness. It also discusses preventive measures, including early warning systems, proper sanitation, and community awareness programs. The WHO stresses the importance of coordinated disaster response strategies and public health planning to minimize the impact of floods. Overall, this reference provides a strong foundation for understanding the human and environmental consequences of flooding and highlights the urgent need for effective forecasting and mitigation systems.

[2] S. Patro et al. (2009) – Hydrodynamic Modelling, patro et al. present a hydrodynamic modeling approach for analyzing flood-prone river systems, particularly in regions with limited data availability. Their study focuses on simulating river flow behavior using mathematical and physical models to predict flood conditions. The research demonstrates how hydrodynamic models can capture water movement, river discharge, and floodplain interactions effectively. Despite data limitations, the model achieved reasonable accuracy, proving its usefulness in real-world scenarios. The study also highlights challenges such as data scarcity, calibration difficulties, and computational complexity. This work is important because it shows how traditional modeling techniques can still provide valuable insights for flood prediction. However, it also underlines the need for more advanced approaches, such as machine learning, to overcome limitations related to data and scalability.

[9] S. Niknam et al. (2020) – Federated Learning, niknam et al. discuss federated learning as an emerging paradigm that enables decentralized model training while preserving data privacy. The paper explains how multiple devices or clients can collaboratively train a shared model without transferring raw data to a central server. Instead, only model updates are communicated, reducing bandwidth usage and enhancing security. The authors highlight key advantages of federated learning, including privacy preservation, reduced communication overhead, and scalability across distributed systems. They also address challenges such as heterogeneous data distribution, communication efficiency, and model convergence. This study is highly relevant for applications like flood forecasting, where data is distributed across different locations and privacy concerns are significant. It provides a strong theoretical and practical foundation for implementing federated learning in real-world systems.

[11] S. Sankaranarayanan et al. (2020) – Deep Learning for Flood Prediction, this study explores the use of deep learning techniques for flood prediction based on weather-related parameters. The authors developed a model that utilizes historical data such as rainfall, temperature, and humidity to predict flood occurrences. By applying deep neural networks, the system was able to capture complex nonlinear relationships between environmental variables. The results showed improved prediction accuracy compared to traditional statistical methods. The study highlights the effectiveness of deep learning in handling large datasets and generating reliable forecasts. However, it also points out challenges such as the need for high-quality data and computational resources. This research demonstrates the growing importance of artificial intelligence in disaster prediction and supports the integration of advanced models in flood forecasting systems.

[15] H. R. Pourghasemi et al. (2023) – Machine Learning in Disaster Prediction, pourghasemi et al. provide a comprehensive analysis of machine learning techniques used for predicting natural disasters, including floods, landslides, and earthquakes. The study compares various algorithms such as decision trees, support vector machines, and neural networks to evaluate their performance in risk assessment. The authors emphasize that combining multiple models often leads to better prediction accuracy and reliability. The research also discusses the importance of geospatial data, environmental factors, and historical records in building effective prediction systems. One key contribution of this work is its focus on integrating different machine learning approaches to enhance disaster susceptibility mapping. This study highlights the potential of ML in improving early warning systems and supports the need for innovative approaches like federated learning for handling distributed and sensitive data.

DATASET DETAILS

The dataset used in this study is the Kerala flood dataset obtained from the Kaggle repository, selected as an alternative due to the unavailability of the original dataset . It contains real-world hydrological data where each record represents monthly observations related to flood conditions. The dataset primarily includes attributes such as rainfall measurements and corresponding water levels, which play a crucial role in flood prediction. It is organized in a structured tabular format, where each row represents an observation and each column represents a feature influencing water level changes. The target variable in this dataset is the water level, which helps in identifying potential flood occurrences.

To ensure data quality and consistency, several preprocessing steps are applied, including handling missing values, removing inconsistencies, normalizing feature values, and shuffling the dataset to avoid bias. Feature scaling is also performed so that all input variables contribute equally during model training. After preprocessing, the dataset is divided into 80% training data and 20% testing data. This prepared dataset supports efficient training of FFNN and CNN2D models for accurate and reliable flood forecasting.

PROPOSED METHODOLOGY

The proposed methodology presents a structured framework for flood forecasting using Federated Learning integrated with Feed Forward Neural Network (FFNN) and an enhanced CNN2D model. The process begins with dataset acquisition, where a flood-related dataset (Kerala flood dataset from Kaggle) is uploaded into the system. This dataset contains monthly rainfall records and corresponding water level information, which are key indicators of flood conditions. After loading the dataset, preprocessing steps are applied to improve data quality and model efficiency. These steps include handling missing values, normalization, and shuffling to eliminate bias. The dataset is then split into training (80%) and testing (20%) subsets to ensure proper evaluation and avoid overfitting.

Following preprocessing, the FFNN model is first implemented as the base model. It is trained using the training dataset to learn patterns between rainfall and water level variations. The model updates its weights based on training epochs and minimizes prediction error using loss functions such as MSE and RMSE. After training, the model is evaluated on test data to measure its prediction accuracy. To further enhance performance, an advanced CNN2D model is introduced as an extension. CNN2D improves feature extraction and captures complex relationships in the dataset, resulting in better prediction accuracy and reduced error compared to FFNN.

In the federated learning setup, multiple client stations locally train their models using regional datasets. Instead of sharing raw data, only trained model parameters are sent to a central server for aggregation. This approach ensures data privacy, reduces network latency, and improves scalability. The global model generated after aggregation is then used to identify flood-prone regions up to five days in advance. Additionally, a local FFNN model is trained for the specific station to predict future water levels and generate early warnings.

To enhance usability, the system includes modules such as model upload to the server, performance comparison graphs, and real-time flood prediction using test data. Visualization techniques are used to compare FFNN and CNN2D performance based on accuracy, MSE, and RMSE. The CNN2D model demonstrates higher accuracy and lower error, making it more suitable for deployment. Overall, the proposed methodology provides a secure, efficient, and scalable solution for accurate flood forecasting and timely disaster management.

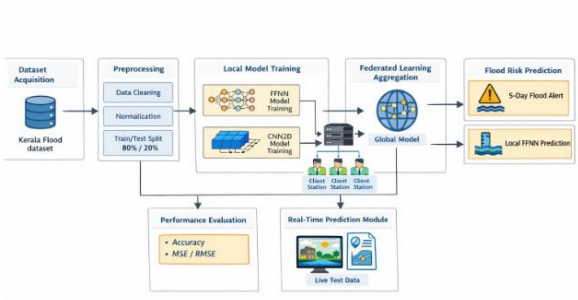

Figure 1: System Architecture of Federated Learning-Based Flood Forecasting System

Figure [1] illustrates the overall system architecture, starting with dataset upload and preprocessing, including normalization, cleaning, and data splitting. The processed dataset is then used to train FFNN and CNN2D models locally at different client stations. These locally trained models are sent to a central server where aggregation is performed to create a global model. The system then predicts flood-prone areas and water levels, providing early warnings. Performance evaluation and comparison modules help identify the best-performing model, while the real-time prediction module enables efficient and timely flood forecasting.

RESULT AND DISCUSSION

The experimental results demonstrate the effectiveness of the proposed Federated Learning-based flood forecasting system using FFNN and CNN2D models. Multiple models were trained and evaluated, including the base Feed Forward Neural Network (FFNN) and the extended CNN2D model. Among these, the CNN2D model achieved the highest accuracy, showing superior performance in predicting water levels and identifying flood conditions. The FFNN model also performed well, producing predictions close to actual values, but with slightly higher error rates. Evaluation metrics such as accuracy, Mean Square Error (MSE), and Root Mean Square Error (RMSE) indicate that CNN2D provides lower error and better prediction capability compared to FFNN. The graphical results show that predicted and actual water levels are closely aligned, especially in the CNN2D model, confirming its reliability.

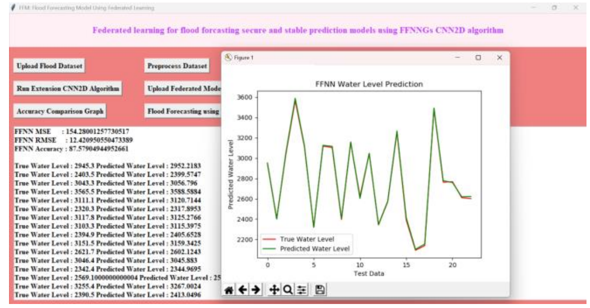

Figure [2]: FFNN Training and Prediction Output

Figure [2] illustrates the FFNN model performance, where predicted water levels are compared with actual values. The graph shows that both values are closely matched, indicating good prediction capability, although small deviations can be observed.

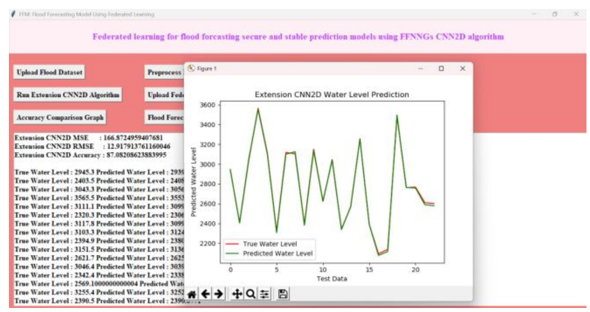

Figure [3]: CNN2D Model Performance

Figure [3] presents the CNN2D model results. The predicted and actual water level lines almost overlap, showing higher accuracy and reduced error compared to FFNN. This confirms that CNN2D captures complex patterns more effectively.

Figure [4]: Accuracy Comparison Graph

Figure [4] shows the comparison between FFNN and CNN2D based on accuracy, MSE, and RMSE.

CNN2D achieves higher accuracy with lower error values, making it the better-performing model.

Figure [5]: Flood Prediction Using Test Data

Figure [5] displays the real-time prediction output where test data is provided as input. The system predicts water levels, helping authorities identify potential flood situations in advance.

DISCUSSION

The results clearly indicate that integrating advanced deep learning models with federated learning significantly improves flood prediction performance. CNN2D outperforms FFNN due to its ability to extract complex features and handle nonlinear relationships in hydrological data. The federated learning approach enhances the system by enabling decentralized model training, ensuring data privacy, and reducing network latency. Preprocessing steps such as normalization and data splitting also contributed to improved model accuracy and stability. The comparison of evaluation metrics confirms that CNN2D is more reliable for flood forecasting tasks. Additionally, the real-time prediction module increases practical usability by providing early warnings. Overall, the system offers a secure, scalable, and efficient solution for accurate flood prediction and disaster management.

CONCLUSION

The proposed flood forecasting system provides an effective solution for predicting flood conditions using a combination of federated learning, FFNN, and CNN2D models. By allowing local data to remain at individual monitoring stations and sharing only trained model parameters, the system maintains data privacy while also minimizing communication overhead and network delays. This decentralized learning approach makes the system more secure and suitable for large-scale deployment across multiple geographic locations.

The experimental results clearly indicate that both FFNN and CNN2D models are capable of predicting water levels with good performance. However, the CNN2D model consistently achieves higher accuracy, better generalization, and lower error rates due to its ability to capture spatial and temporal patterns more effectively. This makes it more reliable for handling complex flood-related data and dynamic environmental conditions.

In addition, the system incorporates essential preprocessing steps such as data normalization, missing value handling, and feature selection, which significantly enhance the quality of the input data. These steps contribute to improved model training and more stable predictions. Performance evaluation metrics further validate the effectiveness of the proposed approach, demonstrating its robustness and consistency across different scenarios.

Another important advantage of the system is its capability to generate early flood warnings. By accurately predicting rising water levels in advance, the system enables authorities and disaster management teams to take proactive measures such as evacuation planning, resource allocation, and risk mitigation. This can greatly reduce potential damage to infrastructure and help save human lives.

Furthermore, the integration of federated learning makes the system scalable and adaptable. New monitoring stations can be added without affecting the overall architecture, and the model can continuously improve as more data becomes available. This ensures that the system remains effective over time and can handle evolving environmental patterns.

Overall, the proposed approach presents a reliable, secure, and scalable flood forecasting solution. Its combination of advanced deep learning models and privacy-preserving techniques makes it highly suitable for real-world applications where accuracy, efficiency, and data security are critical. Future improvements may include integrating real-time sensor data, enhancing model optimization techniques, and expanding the system to cover larger and more diverse geographical regions.

REFERENCES

[1] Floods World Health Organization. Accessed: Feb. 3, 2022. [Online]. Available: https://www.who.int/news-room/questions-andanswers/item/how-do-i-protect-my-health-in-a-flood

[2] S. Patro, C. Chatterjee, R. Singh, and N. S. Raghuwanshi, ''Hydrodynamic modelling of a large flood-prone river system in India with limited data,'' Hydrol. Processes, vol. 23, no. 19, pp. 2774–2791, 2009.

[3] A. Rahman and R. Shaw, ''Floods in the Hindu Kush region: Causes and socio-economic aspects,'' in Mountain Hazards and Disaster Risk Reduction. Tokyo, Japan: Springer, 2015, pp. 33–52.

[4] A. Rahman and A. N. Khan, ''Analysis of 2010-flood causes, nature and magnitude in the Khyber Pakhtunkhwa, Pakistan,'' Natural Hazards, vol. 66, no. 2, pp. 887–904, 2013.

[5] G. Terti, I. Ruin, S. Anquetin, and J. J. Gourley, ''Dynamic vulnerability factors for impact-based flash flood prediction,'' Natural Hazards, vol. 79, no. 3, pp. 1481–1497, Dec. 2015.

[6] Annu. Flood Rep., 2022. Accessed: Jun. 13, 2022. [Online]. Available: https://ffc.gov.pk/wp-content/uploads/2021/04/2020-Annual-Report-ofOo-CEA-CFFC.pdf

[7] L. Martinez, ''Innovative techniques in the context of actions for flood risk management: A review,'' Engineering, vol. 2, no. 1, pp. 1–11, 2020.

[8] N. W. Chan and D. J. Parker, ''Response to dynamic flood hazard factors in peninsular Malaysia,'' Geograph. J., vol. 162, no. 3, pp. 313–325, 1996.

[9] S. Niknam, H. S. Dhillon, and J. H. Reed, ''Federated learning for wireless communications: Motivation, opportunities, and challenges,'' IEEE Communes. Mag., vol. 58, no. 6, pp. 46–51, Jun. 2020.

[10] R. Sadiq, M. A. Al-Zahrani, A. K. Sheikh, T. Husain, and S. Farooq, ''Performance evaluation of slow sand filters using fuzzy rule-based modelling,'' Environ. Model. Softw., vol. 19, no. 5, pp. 507–515, May 2004.

[11] S. Sankaranarayanan, M. Prabhakar, S. Satish, P. Jain, A. Ramprasad, and A. Krishnan, ''Flood prediction based on weather parameters using deep learning,'' J. Water Climate Change, vol. 11, no. 4, pp. 1766–1783, Dec. 2020.

[12] K. Aziz, A. Rahman, G. Fang, and S. Shrestha, ''Application of artificial neural networks in regional flood frequency analysis: A case study for Australia,'' Stochastic Environ. Res. Risk Assessment, vol. 28, no. 3, pp. 541–554, Mar. 2014

[13] K. Haddad and A. Rahman, ''Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework—Quantile regression vs. parameter regression technique,'' J. Hydrol., vol. 430, pp. 142–161, Apr. 2012.

[14] S. Kim, Y. Matsumi, S. Pan, and H. Mase, ''A real-time forecast model using artificial neural network for after-runner storm surges on the Tottori coast, Japan,'' Ocean Eng., vol. 122, pp. 44–53, Aug. 2016.

[15] H. R. Pourghasemi, S. Pouyan, M. Bordbar, F. Golkar, and J. J. Clague, ''Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination,'' Natural Hazards, vol. 115, no. 3, pp. 1–20, Feb. 2023.