



Enhanced Federated Flood Forecasting Model With 2D CNN For Improved Prediction

Dr. MUVVA VENKATESWARA RAO 1, YENUGADHATI ROKESH SAI VARMA 2,
GONUGUNTA MAHESWARI 3, SHAIK AJIS 4

#1 Professor in Department of Information Technology, NRI Institute Of Technology, Agiripalli.

#2#3#4 B.Tech with Specialization of Information Technology in NRI Institute Of Technology, Agiripalli.

Abstract: One of the most frequent natural catastrophes that causes significant harm to property, crops, the economy, and human lives is flooding. For scholars who have been trying to forecast floods for a long time, flood prediction presents a significant difficulty. This article proposes a flood forecasting model that makes use of the federated learning approach. By preventing data from being exchanged over the network for model training, Federated Learning, the most sophisticated machine learning (ML) approach, addresses network latency issues that arise in flood prediction while guaranteeing data privacy, availability, and security. Instead of transferring large data sets to a central server for local model aggregation and global data model training at the central server, the Federated Learning approach encourages onsite training of local data models and concentrates on transmitting these local models across the network. The suggested model in this article combines locally trained models of eighteen clients, determines which station is likely to experience flooding, and provides a flood alarm with a five-day lead time for a particular client. At the client station where the flood is anticipated, a local feed forward neural network (FFNN) model is developed. The local FFNN model's flood forecasting module uses a number of regional factors as input to estimate the anticipated water level. From 2015 to 2021, a dataset of five distinct rivers and barrages was gathered, taking into account four factors: hydrodynamics, flow

routing, rainfall-runoff, and snow melting. The suggested flood forecasting model has an 84% accuracy rate in predicting past floods that occurred in the chosen area between 2010 and 2015. As an addition, we employed sophisticated algorithms such as Convolution2D Neural Networks, which are well-liked across all fields due to their successful and accurate prediction accuracy of over 90%. Therefore, we have extended flood forecasting with CNN2D to increase accuracy.

Index Terms - Feed-forward neural networks, federated learning, flood forecasting systems, hydrology, and meteorology Convolutional Two-D Neural System

1. INTRODUCTION

Natural and man-made disasters have become more common worldwide in recent years [1]. Global warming, growing urbanisation, and hydrological extremes have all increased the danger of flooding worldwide [2]. Devastating natural catastrophes like floods cause a great deal of damage to crops and infrastructure, a large number of fatalities, and the collapse of a nation's whole socioeconomic structure. Although they occur often worldwide, floods differ in strength depending on the location [3]. Every year, floods in developing nations claim numerous lives, trigger harsh economic crises, and increase financial difficulties [4].

Floods are becoming more frequent and severe as a result of the increased rate of precipitation and snowmelt brought on by global warming and climate change [5]. Floods occur more often in Pakistan than other natural disasters, as Figure 1 illustrates [6]. In the South Asian nations, floods have been found to outnumber all other disasters in 2021 [7].

Governments urgently require accurate prediction tools to provide prompt and efficient responses in response to the growing risks that floods pose to both human life and economic infrastructure [8]. The intrinsic complexity of floods has prevented significant gains in accuracy, despite the fact that several regional and international approaches, models, and methods have been offered for flood prediction [9]. Complex mathematical expressions have been used to depict flood-causing physical processes via well-known statistical techniques as the climatology average method (CLIM), flood frequency analysis (FFA), Bayesian forecasting models (BFM), and artificial neural networks (ANN) [11–14].

Flood prediction systems have greatly improved with the introduction of machine learning (ML), which provides more efficient and economical solutions. In their quest for more precise and effective prediction models through the development of new ML approaches and the hybridisation of preexisting ones, hydrologists are favouring machine learning techniques more and more [15–16]. ML's reliance on large amounts of data for model training, however, presents difficulties since data sharing amongst authorities is hampered by issues with data security, privacy, and regulations [17–18]. Flood forecasting systems have historically used centralised configurations, focussing the data and prediction model in one place for training before distributing them to all customers. This method presents latency, connection problems, and possible security and privacy hazards despite its ease [19–20].

2. LITERATURE REVIEW

[16] For precise urban reservoir flood forecasting, they provide a hybrid recurrent neural network that incorporates multivariate autoregressive integrated moving average,

convolution kernel smoothing, and time series attention. This model has been verified in Ankang Reservoir. Accurate real-time forecasting in Ankang Reservoir shows that the suggested hybrid model outperforms other machine learning networks and conventional hydrological models in predicting urban reservoir floods. The system's resilience in many settings may be limited by its reliance on precise upstream data, possible susceptibility to threshold alterations, and computing complexity. The suggested flood forecasting system has difficulties in accommodating various geographic situations, guaranteeing real-time data availability, and resolving rainfall prediction uncertainty. Due to Ankang Reservoir's unique geographic features, generalisability to other areas could necessitate further calibration and validation procedures.

[21] Back-propagation networks with self-organising map cluster analysis are used in this paper's real-time flood forecasting model to improve accuracy by generating ensemble forecasts during typhoon occurrences in Taiwan's Wu River Basin. In the Wu River Basin, the ensemble model, which combines self-organising map clustering and back-propagation networks, performs better than individual models in reliably predicting flood amounts and intervals during typhoon occurrences. The suggested system's resilience may be impacted by issues including computational complexity, possible susceptibility to cluster selection, and reliance on precise typhoon data. The suggested real-time flood forecasting system has difficulties in adjusting to various typhoon characteristics, guaranteeing real-time data availability, and resolving uncertainty in rainfall forecasts. For wider geographical application, the model's usefulness may be restricted to the unique features of the Wu River Basin, necessitating thorough evaluation and validation.

[23] By merging the Monotone Composite Quantile Regression Neural Network with the Xinanjiang conceptual model, an integrated XAJ-MCQRNN model improves short-term flood probability density forecasting while tackling issues with error accumulation and transmission. In short-term flood probability density forecasting, the XAJ-MCQRNN model performs better than the MCQRNN, showing increased accuracy and dependability that helps

early warning systems and flood prevention. Deploying the suggested system may be difficult due to possible difficulties merging conceptual and machine learning models, sensitivity to input data quality, and computing needs. The proposed short-term flood probability density forecasting system has problems in resolving errors in rainfall-runoff modelling, adapting to a variety of hydrological circumstances, and guaranteeing timely rainfall data availability. Applicability could be restricted to Jianxi River-like catchments, requiring thorough evaluation of various hydrological situations and prolonged validation.

[24] In order to achieve reliable flood prediction in Xi County, China, they created a ConvLSTM model that combines CNN and LSTM processes spatial-temporal hydrological data. This model outperformed previous methods. ConvLSTM showed excellent performance in forecasting peak discharge and flood arrival time, providing a viable substitute for precise and prompt flood forecasting. Computational complexity, resource-intensive training, and sensitivity to the quantity and quality of input data are possible disadvantages. Optimising model parameters, guaranteeing data quality, and incorporating real-time monitoring for dynamic flood prediction might provide implementation issues. The model's generalisability may be restricted by certain environmental circumstances and data availability, and its efficacy may change depending on the geographic location.

[25] By filling in the holes in existing techniques for improved accuracy, they are launching a flood prediction and extent mapping model that makes use of multispectral, radar, and LIDAR remote sensing technologies. Although there are still issues and restrictions with their use, remote sensing technologies are essential for flood prediction and provide insightful information for disaster management. High equipment expenditures, restricted access to cutting-edge technology, and possible errors in data interpretation that compromise the accuracy of flood predictions are some of the difficulties. Integrating various remote sensing data, guaranteeing real-time data collection, and resolving problems with data resolution and accuracy for reliable flood prediction are implementation hurdles. Notwithstanding

developments, the model's overall efficacy in flood prediction scenarios may be impacted by remote sensing technology constraints, such as weather dependence and errors in data interpretation.

3. METHODOLOGY

i) Proposed Work:

Federated Learning (FL) and a Feed Forward Neural Network (FFNN) are used in the proposed Flood Forecasting Model (FFM) to improve flood forecast accuracy while protecting data privacy. In the first stage, data is sent to a central server for aggregation while several clients work together to build local models. A global model that forecasts flood events at particular client stations with a 5-day lead time is then trained using local models. The last phase entails notifying authorities for flood preparedness, assessing water levels, and training a local FFNN on the designated station. The humanitarian goal of the research is to stop fatalities and extensive harm. In addition to the conventional FFNN, an extension uses Convolutional Neural Network [24] (CNN2D) to increase accuracy. Performance is measured by accuracy, Mean Square Error (MSE), and Root Mean Square Error (RMSE). This integrated method offers a fresh and all-encompassing approach to flood prediction.

ii) System Architecture:

Three layers of hidden nodes make up the suggested Flood Forecasting Neural Network (FFNN) model, which is intended to forecast floods five days ahead of time. The system design incorporates input from many models, such as the hydrodynamic model, flow routing model, rainfall runoff model, and snow melting model. When combined, these models produce crucial data for flood prediction.

The architecture's initial stage combines the findings from the flow routing, rainfall runoff, and snow melting models. The FFNN model uses this combined data as input, especially to fill in the inner hidden nodes. This input is converted into flood forecast predictions via the FFNN model, which has three layers of hidden nodes. To further improve the model's capacity to represent intricate interactions in the hydrological

system, the second stage is integrating the hydrodynamic model findings into the FFNN.

The third stage uses an adaptation of the Convolutional Neural Network [24] (CNN2D) method to optimise the spatial-temporal elements of the input data. The FFNN model can efficiently handle the multidimensional input data thanks to this technique, which also extracts spatial information that are essential for precise flood forecasts. By combining these several models and utilising CNN2D's FFNN architecture, the suggested method seeks to improve flood prediction accuracy by taking into account a number of hydrological elements and their intricate relationships.

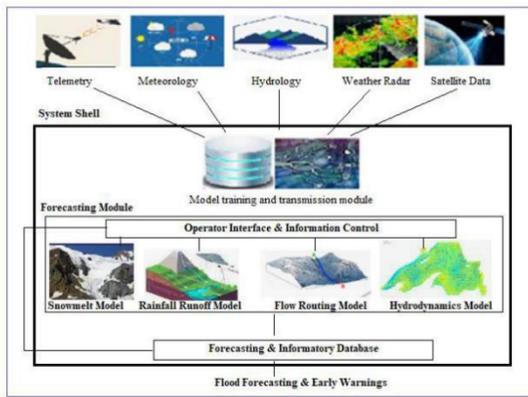


Fig 1 Proposed Architecture

iii) Dataset:

The Feed Forward Neural Network (FFNN) algorithm is locally trained using a dataset collected from 18 stations or rivers in the proposed flood prediction research. The FFNN model is trained using this dataset, which contains crucial information, enabling it to learn and forecast water levels at different places. The trained FFNN model is sent to a centralised server for worldwide updates following the local training stage. Although the original dataset has not been made publicly available online by the author, an alternative is sought, which results in the use of the KERALA flood dataset [25] available on the Kaggle website. Monthly rainfall records are included in this Kaggle dataset; column names are specified in the first row, and dataset values are found in the following rows. Importantly, the dataset's last column contains vital data on water level measurements,

which are essential for anticipating and controlling flood hazards. Based on the predicted water levels, the predictions made by the trained FFNN model let authorities warn the public about possible flood situations.

The study acknowledges the possibility of additional accuracy improvement by integrating more sophisticated approaches, even though the conventional Feed Forward Neural Network (FFNN) methods used in the suggested work are effective. Convolutional Neural Network [24] (CNN2D) was not used in the initial study, but this sophisticated method is now incorporated into the model expansion. Due to its higher accuracy than FFNN, CNN2D has become more and more popular in a variety of fields. By identifying complex geographical linkages and patterns in the dataset, this addition seeks to increase prediction accuracy and demonstrates a progressive approach to flood forecasting that is in line with recent developments in neural network techniques.

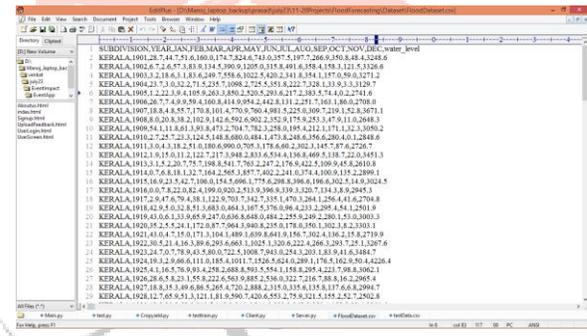


Fig 2 Dataset

The top row of the dataset screen above shows the names of the dataset's columns, while the other rows show the dataset's values. The dataset includes monthly rainfall records, and the last column shows the water level. Authorities will notify the public about flooding based on the expected water level.

iv) Pre – processing Dataset:

The construction of a hybrid predictive model in this project, which combines conventional Feed Forward Neural Network (FFNN) techniques with an extension using Convolutional Neural Network with 2D layers [24] (CNN2D), is made possible in large part by the "Pre-process Dataset" module. The main goal is to handle the dataset efficiently prior to

model training in order to improve flood forecast accuracy and resilience.

Three essential procedures are included in the pre-processing module. First and foremost, maintaining the dataset's integrity depends on locating and eliminating missing values. If left unchecked, missing values may cause bias and irregularities while training the model. The dataset then goes through a normalisation procedure, in which the characteristics are scaled to a predetermined range. Normalisation promotes impartial and equitable learning by preventing some features from unduly impacting the model because of different scales.

The dataset is shuffled in the last stage, which adds randomisation to the data instances' order. In order to reduce the danger of overfitting and improve generalisation to new data, this phase is very important since it stops the model from learning patterns based on the sequence of incoming data.

A hybrid model that combines the advantages of a conventional FFNN with the spatial-awareness capabilities of CNN2D is then trained using the pre-processed dataset. This novel method seeks to provide a more precise and contextually aware flood prediction model by capturing intricate spatial correlations within the dataset. The use of CNN2D as an expansion is in line with recent developments in neural network techniques and demonstrates a progressive approach to raising the accuracy of flood predictions.

v) Training & Testing:

To make model training and assessment easier, the dataset is split into two subsets in the project's "Train & Test Split" module: a training set and a testing set. This module's main goal is to evaluate the model's performance on unobserved data in order to make sure it can generalise outside of the training set.

To enable the predictive model to identify patterns, correlations, and trends in the data, 80% of the dataset must be set aside for training. For the model to capture the underlying traits and changes required for precise predictions,

this significant amount of the information is essential. The trained model will be tested with the remaining 20%. By simulating a real-world situation, this distinct test set enables the evaluation of the model's prediction ability on data that was not seen during training.

An 80-20 split is frequently used in machine learning to strike a compromise between preserving a sizeable amount for thorough review and having enough data for efficient model training. This section aids in identifying possible problems like overfitting, which occurs when a model works well on training data but has trouble with fresh, untested data. The project contributes to the resilience and dependability of the flood forecasting system by testing the model on the test set, which guarantees a more accurate evaluation of its generalisation performance and predictive accuracy.

descent and backpropagation to find the best filters.

4. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is determined by how well it distinguishes between sick and healthy instances. The percentage of true positive and true negative in each assessed instance should be determined in order to measure a test's accuracy. This may be expressed mathematically as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

S. NO	ALGORITHM	ACCURACY%
1	FFNN	82.76%
2	CNN2D	87.86%

MSE: A prediction or estimate aim and a predictor or estimator—which is referred to as the function of the provided data—are required for the mean squared error

measure. The average of squares of the "errors" is known as the MSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- MSE = mean squared error
- n = number of data points
- Y_i = observed values
- \hat{Y}_i = predicted values

S. NO	ALGORITHM	MSE%
1	FFNN	296.96%
2	CNN2D	147.22%

RMSE: One often used metric to quantify the discrepancies between the actual observed values and the values predicted by an estimate or model is the root mean square error, or RMSE. The square root of the variations between expected and actual values is known as the root mean square error (RMSE). In this computation, the individual variations are referred to as "residuals." The size of the mistakes is estimated by the RMSE. Because it is scale-dependent, this accuracy metric is used to compare predicting errors from various estimators for a single variable, but not among the variables.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

$\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ are predicted values

y_1, y_2, \dots, y_n are observed values

n is the number of observations

S. NO	ALGORITHM	RMSE%
1	FFNN	17.23%
2	CNN2D	12.13%

Performance Evaluation Table:

S. NO	ALGORITHM	ACCURACY%	MSE%	RMSE%
1	FFNN	82.76%	296.96%	17.23%
2	CNN2D	87.86%	147.22%	12.13%

Comparison graph:

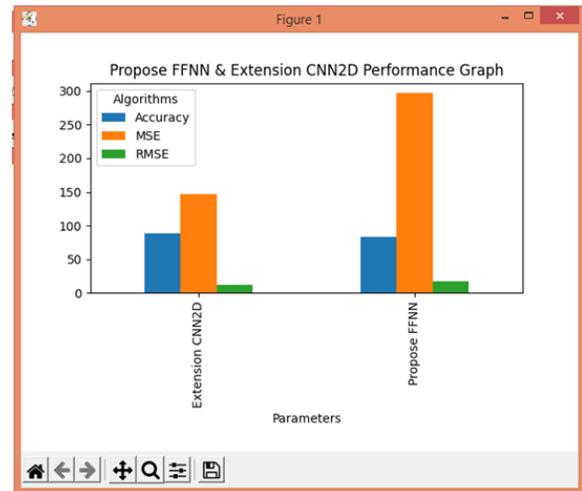


Fig 5 Comparison graph of FFNN & CNN2D algorithm



The test data appears on the screen above before the arrow symbol, and the projected water level appears after the arrow sign →

5. CONCLUSION

In summary, a two-module strategy for improved flood prediction and mitigation is demonstrated by the Flood Forecasting Model (FFM) that is being presented. A network of eighteen local monitoring stations is established by the first module, which also trains and sends data models to a central server. By examining various factors from the local models, this central server then builds a global model that can predict floods over the course of the following five days. The second module estimates the anticipated rise in water levels by using a Feed Forward Neural Network at the anticipated flood area. Hydraulic and meteorological data are processed locally to satisfy privacy, security, and data availability concerns. By providing the flood mitigation department with timely flood alarms, the FFM helps with proactive disaster prevention and response, proving its effectiveness. The analysis of past floods from 2010 to 2015 demonstrates an impressive 82.76% accuracy rate. Moreover, accuracy is greatly increased to 87.86% by extending the model using CNN2D.

REFERENCES

- [1] Floods World Health Organization. Accessed: Feb. 3, 2022. [Online]. Available: <https://www.who.int/news-room/questions-and-answers/item/how-do-i-protect-my-health-in-a-flood>
- [2] S. Patro, C. Chatterjee, R. Singh, and N. S. Raghuvanshi, "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 Commun. 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.

[16] B. Cai and Y. Yu, "Flood forecasting in urban reservoir using hybrid recurrent neural network," *Urban Climate*, vol. 42, Mar. 2022, Art. no. 101086.

Author's Profiles



Dr. MUVVA VENKATESWARA RAO is presently working as a Professor in the Department of Information Technology at NRI Institute of Technology, Vijayawada, specializing in Artificial Intelligence and Machine Learning. With over 20 years of teaching experience, he has been instrumental in shaping the careers of numerous students in AI and ML. He holds a strong academic background and has contributed significantly to the field through his expertise and mentorship. His dedication to research and teaching has made a lasting impact in academia, inspiring students to excel in the ever-evolving field of technology.



YENUGADHATI ROKESH SAI VARMA is a B.Tech student specializing in Information Technology at NRI Institute of Technology. He is passionate about exploring emerging technologies and continuously enhancing his skills through hands-on challenges. He has completed internships in DevOps and Data Science, gaining practical experience. Additionally, he has earned NPTEL certification in Cloud Computing and a Java Full Stack certification from Wipro, further strengthening his technical expertise. He is skilled in problem-solving, Artificial Intelligence and Full Stack Development and is committed to continuous learning.



GONUGUNTA MAHESWARI is a B.Tech Student specialising in Information Technology at NRI Institute of Technology. She has completed in Cloud computing and The joy of computing using python (NPTEL). She gained practical experience during her DevOps internship at Advaita Global - IT Lavs Pvt.LTd where she acquired knowledge about Core Java. She is skilled in Artificial Intelligence, problem-solving and Python programming.



SHAIK AJIS is a B. Tech student specializing in Information Technology at NRI Institute of Technology. He has completed certification in Java Programming (NPTEL). He gained practical experience during his DevOps internship at Advaita Global - IT Labs Pvt. Ltd., where he acquired knowledge about DevOps. He is skilled in Python programming, Artificial Intelligence, and problem-solving and is committed to continuous learning.

