Decentralized Flood Forecasting: Leveraging Federated Learning for Enhanced Prediction Accuracy

Kanchukatla Harshavardhan[1], B.Srinivasa S P Kumar [2] and [3] BH L Mohan Raju

[1] MCA Student, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India

[2] Assistant Professor, Department of Master of Computer Applications, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India
bsrinivas_mca@cbit.ac.in

[3] Assistant Professor, MBA Dept. , Malla Reddy Engg. College (J4) (A), Secunderabad, Telangana State, India
mohanrajubh@gmail.com

ABSTRACT: One of the most frequent natural disasters that seriously damages property, crops, the economy, and human lives is flooding. For scholars who have been trying to anticipate floods for a long time, flood prediction presents a significant difficulty. This article proposes a flood forecasting model that makes use of federated learning. Federated Learning is the most sophisticated machine learning (ML) technique available. It forbids the transport of data over the network for model training, thereby addressing the network latency challenges inherent in flood prediction and guaranteeing data availability, privacy, and security. Instead of transmitting massive data sets to a central server for local training, the federated learning technique emphasizes the transmission of local models via the network. This encourages on-site training of local data models.This article proposes a model that integrates eighteen locally trained models, identifies the station where flooding is most likely to happen, and sends out a five-day flood alert for a specific client. A feed forward neural network (FFNN) model is developed locally at the client station where the flood is expected. A range of geographical parameters are input into the flood forecasting module of the local FFNN model in order to predict the expected water level. When assembling the rainfall dataset for Kerala from 1901 to 2015, four aspects were taken into account: hydrodynamics, flow routing, snow melting, and rainfall-runoff.The suggested flood forecasting model has successfully and accurately forecasted 84% of the past floods that occurred in the chosen zone between 2010 and 2015. As an addition, we employed cutting-edge algorithms such as Convolution2D Neural Network, which are well-liked across all domains due to their successful and accurate prediction accuracy of over 90%. Therefore, we have extended CNN2D for flood forecasting in order to increase accuracy.

KEYWORDS - *Hydraulic, meteorological, flood forecasting system, federated learning, feed-forward neural network., Convolution2D Neural Network*

I. INTRODUCTION

The frequency of both natural and man-made disasters has increased globally in recent years [1]. The risk of flooding worldwide has increased as a result of growing urbanization, hydrological extremes, and global warming [2]. Floods are terrible natural disasters that cause enormous damage to crops, infrastructure, and people's lives, ultimately leading to the collapse of a nation's whole socioeconomic structure. Although they occur often around the world, floods differ in strength from place to place [3]. Every year, floods in developing nations claim numerous lives, unleash severe economic catastrophes, and worsen financial issues [4]. Increased snowmelt and precipitation rates are a result of global temperature rise and overall climate change, which makes floods more frequent and powerful [5]. Figure 1 illustrates that Pakistan experiences floods more frequently than other natural disasters [6]. In South Asian countries during 2021, floods have been observed to outnumber all other tragedies [7].

Reliable predictive systems are vital for governments to enable early and effective interventions in response to the increasing hazards that floods represent to both human life and economic infrastructure [8]. The intrinsic complexity of this natural calamity has prevented significant increases in accuracy in flood prediction, despite the many global and regional approaches, models, and tactics that have been offered [9]. Complex mathematical expressions have been used to depict flood-causing physical processes via well-established statistical approaches such 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 advanced with the introduction of machine learning (ML), which provides improved performance and affordable options. Hydrologists are favoring machine learning (ML) approaches more and more, hoping to improve prediction models by hybridization of existing models and new ML techniques [15–16]. However, the fact that machine learning (ML) relies on large amounts of data for model training presents difficulties because data sharing among authorities is impeded by worries about data security, privacy, and regulatory constraints [17–18]. Flood forecasting systems have historically used centralized configurations, which concentrate the prediction model and data in one place for training prior to distribution to all customers. This method adds latency, connectivity problems, and possible security and privacy hazards despite its ease of use [19–20].

II. LITERATURE WORK

[16] In order to provide reliable flood forecasts for urban reservoirs, this study presents a hybrid recurrent neural network that incorporates convolution kernel smoothing, time series attention, and multivariate autoregressive integrated moving average. The model is verified in Ankang Reservoir. It performs better than other machine learning networks and conventional models, but it has drawbacks such computational complexity, sensitivity to threshold changes, and dependence on reliable upstream data. Further calibration and validation for other regions are necessary to ensure the generalizability of model.

[21] This work proposes a real-time flood forecasting model that enhances accuracy during typhoon occurrences in the Wu River Basin, Taiwan, by utilizing back-propagation networks with self-organizing map cluster analysis. The sensitivity of cluster selection, computational complexity, and reliance on precise typhoon data pose challenges to the ensemble model, which performs better than individual models. In order to be used in more places, validation is required.

[16] In order to provide reliable flood forecasts for urban reservoirs, this study presents a hybrid recurrent neural network that incorporates convolution kernel smoothing, time series attention, and multivariate autoregressive integrated moving average. The model is verified in Ankang Reservoir. It performs better than other machine learning networks and conventional models, but it has drawbacks such computational complexity, sensitivity to threshold changes, and dependence on reliable upstream data. Further calibration and validation for other regions are necessary to ensure the generalizability of the model.

[21] This work proposes a real-time flood forecasting model that enhances accuracy during typhoon occurrences in the Wu River Basin, Taiwan, by utilizing back-propagation networks with self-organizing map cluster analysis. The sensitivity of cluster selection, computational complexity, and reliance on precise typhoon data pose challenges to the ensemble model, which performs better than individual models. In order to be used in more places, validation is required.

[25] In order to improve accuracy, this work presents a flood forecast model that makes use of multispectral, radar, and LIDAR remote sensing technologies. High equipment expenditures, inaccurate data interpretation, and the integration of disparate data sets provide challenges. Notwithstanding improvements, the model's efficacy could be impacted by weather sensitivity and ambiguities in data interpretation.

III. METHODOLOGY

1) Proposed Work:

In order to improve flood forecast accuracy while maintaining data privacy, the suggested Flood Forecasting Model (FFM) makes use of a Feed Forward Neural Network (FFNN) and Federated Learning (FL). Initially, data is sent to a central server for aggregation while numerous clients work together to train local models. After that, a global model is trained using local models to forecast floods at particular client stations five days in advance. The last stage is determining water levels, preparing authorities for floods, and training a local FFNN on the designated station. The goal of the research is humanitarian in nature, seeking to avert fatalities and extensive harm. Convolutional Neural Network [24] (CNN2D) is an extension that uses standard FFNN in conjunction with Mean Square Error (MSE) and Root Mean Square Error (RMSE) to further improve accuracy. This combined method offers a fresh and all-encompassing approach to flood forecasting.

2) System Architecture:

By combining inputs from flow routing, hydrodynamic models, snow melting, rainfall runoff, and three layers of hidden nodes, the proposed Flood Forecasting Neural Network (FFNN) model forecasts floods five days in advance. The FFNN's hidden nodes are first filled with the aggregated output of flow routing, precipitation runoff, and snow melting models. The hydrodynamic model results then improve the forecasts. The Convolutional Neural Network (CNN2D) extension is utilized by the model to optimize the processing of spatial-temporal data [24]. By taking into account a variety of hydrological elements and their intricate relationships, this integration seeks to increase the accuracy of flood predictions.

3) Dataset:

The proposed study on flood prediction trains a Feed Forward Neural Network (FFNN) locally using a dataset from eighteen stations or rivers. With the use of this dataset, which contains essential hydrological data, the FFNN model can forecast water levels at different places, facilitating efficient flood control. The model is delivered to a centralized server for worldwide updates following local training. Owing to the original dataset's unavailability, monthly rainfall records and vital water level data are included in the KERALA flood dataset from Kaggle.

Authorities are able to anticipate possible floods and take preventative action because to the forecasts made by the FFNN model. Acknowledging the possibility of improved precision, the research integrates Convolutional Neural Networks (CNN2D), which are renowned for their exceptional accuracy in recording spatial relationships and patterns. This integration is in line with recent developments in neural networks and attempts to increase prediction accuracy. The study offers more accurate flood forecasts by merging FFNN and CNN2D, assisting authorities in prompt decision-making and flood.

4) Pre – processing Dataset:

The "Pre-process Dataset" module in this project prepares data for a hybrid prediction model that combines convolutional neural networks with 2D layers (CNN2D) with classic feed forward neural networks (FFNN) [24]. Three crucial processes are involved in this module: first, to preserve dataset integrity, missing values are removed; second, to ensure unbiased learning, features are normalized to a defined range; and third, data is shuffled to avoid order-based patterns and overfitting. The hybrid model is then trained using this pre-processed dataset, taking advantage of CNN2D's spatial awareness to improve flood prediction robustness and accuracy, in line with recent developments in neural network techniques.

5) Training & Testing:

To make sure the model can generalize outside of the training set, the dataset is split into 80% for training and 20% for testing in the "Train & Test Split" module. Because of this divide, the predictive model is able to learn from a significant amount of the data and identify the patterns and variances that are needed to make correct predictions. The remaining 20% simulates real-world circumstances and is used as a test set to assess how well the model performs on unknown data.

In machine learning, the 80-20 split is a popular technique that strikes a balance between thorough review and efficient model training. By identifying overfitting, this method guarantees that the model functions well with fresh data. The robustness and dependability of the flood forecasting are enhanced by evaluating the model's generalization and prediction accuracy on the test set, which offers a trustworthy evaluation of these aspects.

6) Algorithms:

Feed Forward Neural Network:

One of the most basic kinds of artificial neural networks ever created is a feedforward neural network. Information travels in a single path in this network: forward from the input nodes to the output nodes, passing through any hidden nodes that may exist. The network is free of loops and cycles. Compared to their more complex cousins, such as recurrent and convolutional neural networks, feedforward neural networks are the original type of artificial neural network to be created.

Extension CNN2D Algorithm:

One type of deep neural network that excels at processing and interpreting visual data is the Convolutional Neural Network, also known as ConvNet or CNN. A CNN's core component, the Convolution2D layer, is in charge of performing convolutional operations to input data.

The expanded form of artificial neural networks (ANN), known as a convolutional neural network (CNN) [24], is mostly used to extract features from grid-like matrix datasets. For instance, visual datasets with a lot of data patterns, such pictures or movies.

IV. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is determined by how well it can distinguish between patient and healthy cases. We should compute the percentage of true positive and true negative in each analyzed case in order to assess the accuracy of a test. This can be expressed mathematically as follows:

Accuracy = TP + TN TP + TN + FP + FN.

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

MSE: The mean squared error measure requires a target for prediction or estimate in addition to a predictor or estimator, which is defined as the function of the available data. The mean squared of the "errors" is known as the MSE.

S. NO	ALGORITHM	MSE%
1	FFNN	296.96%

2	CNN2D	147.22%

RMSE: One commonly used metric to quantify the discrepancies between values predicted by a model or estimator and the actual observed values is the root mean square error, or RMSE. The square root of the variations between expected and observed values is known as the root mean square error (RMSE). "Residuals" refers to the individual variations in this computation. The RMSE calculates the errors' magnitude. Because this measure is scale-dependent, it is an accuracy metric used to compare predicting errors from various estimators for a particular variable but not among the variables.

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:

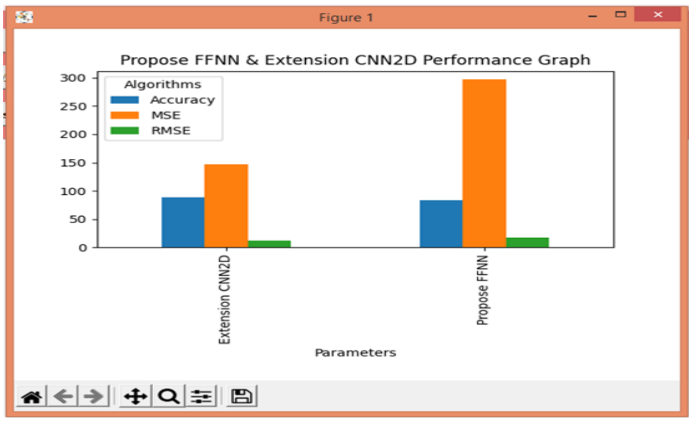

Fig 5 Comparison graph of FFNN & CNN2D algorithm

V. CONCLUSION

To sum up, the Flood Forecasting Model (FFM) that has been described demonstrates a two-module method to improve flood prediction and mitigation. Eighteen local monitoring stations are connected via the first module, which trains and sends data models to a central server. In turn, this central server analyzes many characteristics from the local models to build a global model that can predict floods within the next five days. In order to assess the anticipated rise in water levels, the second module uses a Feed Forward Neural Network at the location of the anticipated flood. 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 demonstrates its effectiveness and helps with proactive disaster prevention and response. The analysis of past floods from 2010 to 2015 demonstrates an excellent accuracy rate of 82.76%. Additionally, the accuracy of the model is much increased to 87.86% by extending it with CNN2D.

The Flood Forecasting Model (FFM) is expected to grow in the future to include datasets from various places, enabling it to predict floods worldwide. The system is a promising tool for proactive flood forecasting on a larger scale because of its proven capacity to adapt to regional data. The FFM has the potential to make a substantial contribution to worldwide disaster management efforts through ongoing development, collaboration, and integration of international datasets. It can provide timely predictions and insights for many regions that are prone to floods.

VI. REFERENCES

[1] **Floods World Health Organization.** Accessed: Feb. 3, 2022. [Online]. Available: WHO.

[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: Annual Flood Report.

[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. 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.

[10] **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.