FL-FPM: Federated Learning Based Flood Prediction Model

A. Akhila[1], Dr. K. Santhi Sree[2]
[1]*Student, M.Tech, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad*
[2]*Professor of CSE and Head of the Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad*

***Abstract*—This paper proposes a novel strategy that utilizations federated learning to improve flood prediction accuracy because of the developing danger of floods while resolving significant issues with information security and organization inertness. The concentrated information gathering limitations that have tormented customary flood prediction approaches have encouraged an interest for a decentralized model that can precisely expect floods with a lead time. This study presents a decentralized flood forecasting model that consolidates Feedforward Neural Network (FFNN) models from a few clients that have been prepared locally. This technique tackles protection issues by utilizing unified figuring out how to total bits of knowledge from a few datasets without risking the security of individual information. The program shows how viable it is at recognizing mind boggling flood examples and conveying convenient notices that are modified for specific regions, further developing readiness and response times for debacles. The model's ability to extraordinarily increment forecast accuracy when contrasted with conventional techniques is exhibited by approval preliminaries, highlighting the progressive capability of state of the art approaches in calamity aversion. This work sets a norm for involving federated learning in ecological gauging, with the potential for more extensive applications in lessening the impacts of regular calamities on a worldwide scale.**

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

I. INTRODUCTION

Floods are perhaps of the most terrible catastrophic event, annihilating networks, economy, and environments. Flood recurrence and power are ascending due to hydrological limits, developing urbanization, and environmental change. Flood prediction and early admonition frameworks are vital for diminishing these dangers and expenses.

Because of natural intricacy, flood prediction stays troublesome regardless of meteorological and hydrology progresses [2]. Customary procedures, which utilize verifiable information and actual models, battle to reflect flood occasions' dynamic and nonlinear person. Because of this imperative, complex computational strategies, for example, Machine Learning (ML), are being utilized to further develop flood forecast accuracy and cost [3].

ANNs, SVMs, and gathering procedures perform well for displaying muddled hydrological cycles and understanding flood elements. These techniques utilize gigantic datasets to figure out examples and connections from earlier flood information for additional accurate predictions. Notwithstanding, preparing information accessibility and quality extraordinarily influence ML model viability.

Enormous and different datasets are a significant hindrance for ML-based flood prediction. Training accurate models requires hydrological information such stream release, precipitation designs, land use information, and geography. Protection requirements and strategic issues limit information securing, restricting model turn of events and approval datasets.

This study offers a Federated Learning (FL) - based Flood Forecasting Model (FFM) to defeat these issues. Federated Learning permits decentralized model training across various information sources (e.g., areas or foundations) while safeguarding information [6]. FL lets information proprietors train neighborhood models utilizing their datasets as opposed to pooling information into a focal server. Just model changes are shipped off a focal aggregator, safeguarding touchy information and diminishing correspondence costs [7].

The proposed FFM utilizes FL to increment flood forecast accuracy without compromising information security. Confined models prepared on territorial information can consolidate nearby hydrological framework subtleties, further developing the unified model's abilities to guaging. FL additionally diminishes information break and administrative consistence concerns, which are basic for handling delicate natural and hydrological insights.

This exploration matches current advances in federated learning applications in medical services, money, and presently ecological sciences. This work adds to the field on decentralized and protection safeguarding ML model development by taking on FL to flood prediction [8].

This presentation plans for Federated Learning-based flood forecasting model exploration. It underlines the requirement for better flood prediction strategies as environmental change and urbanization increment flood dangers. The methodology, execution structure, trial discoveries, and results of the FL-based FFM will be analyzed in the accompanying segments to evaluate further developing flood prediction while safeguarding information protection and security potential.

II. LITERATURE REVIEW

Floods compromise worldwide wellbeing, framework, and economy [1]. Accurate prediction and early admonition innovations are being utilized to lessen these risks. Conventional methodologies battle to catch flood elements, empowering the utilization of hydrodynamic demonstrating and ML.

In information unfortunate regions, hydrodynamic displaying is fundamental for flood prediction. Hydrodynamic demonstrating was displayed to chip away at a significant flood-inclined stream framework in India by Patro et al. regardless of insignificant information [2]. Flood reenactment models utilize water driven standards and geological information to gauge risk and get ready for calamities.

Financial factors influence flood opposition and weakness. Rahman and Shaw analyzed Hindu Kush floods, zeroing in on financial variables [3]. Rahman and Khan inspected the 2010 Khyber Pakhtunkhwa floods, noticing how ecological causes and financial weaknesses demolished the misfortune [4]. Understanding these collaborations is fundamental for making natural and social flood risk control arrangements.

Flood prediction and impact appraisal are confounded by unique weakness factors. Terti et al. inspected dynamic components influencing streak flood prediction and focused on the need for versatile arrangements that record for changing climatic conditions and social weaknesses [5]. Constant information and dynamic models are vital to further developing flood forecasting systems.

Authoritative and strategy structures are vital to flood risk the executives. Yearly flood reports uncover relieving adequacy and constant issues. The Yearly Flood Report of the Federal Flood Commission of Pakistan covers public flood the board plans and significant endeavors and hardships [6]. Such investigations assist with evaluating strategy adequacy and suggest flood readiness and reaction upgrades.

Development in flood risk the board involves attempting new strategies and innovation. Martinez talked about flood risk the executives advancements, focusing on how innovation works on forecast accuracy and reaction methods [7]. Remote detecting for flood checking and further developed displaying for complex gamble evaluation are instances of these progressions.

Geology additionally influences flood elements and response. Chan and Parker analyzed unique flood danger components in Peninsular Malaysia, uncovering district explicit flood reaction issues and the meaning of restricted techniques [8]. Geographic bits of knowledge empower flood control procedures fit exceptional ecological circumstances, guaranteeing compelling reaction and flexibility.

Flood risk the executives implies hydrodynamic demonstrating, financial issues, dynamic weakness factors, strategy systems, innovation advancements, and geographic impacts, as indicated by the writing. These changed elements should be incorporated to make solid flood determining models and moderation measures. Future review ought to further develop expectation models utilizing modern information examination and demonstrating, constant checking, and risk correspondence and readiness to assemble local area flexibility.

III. METHODOLOGY

Proposed Work:
Federated learning (FL) is incorporated into flood forecasting in the proposed framework to moderate worries about information security and upgrade prediction accuracy. FL keeps up with information classification while empowering neighborhood associations to make models with their information by decentralizing model training. The total Flood Forecasting Model (FFM), which can expect flood events with expanded accuracy and lead time, is then made by conglomerating these confined models. The framework takes utilization of FL's ability to deal with an assortment of datasets from different geological regions, representing differences in provincial hydrological conditions and natural factors. This strategy works on the precision of flood gauges as well as permits proactive ways to deal with catastrophe the executives that are redone for specific locales.

System Architecture:
The flood forecasting system configuration utilizes an organized information and model evaluation process. Flood information is preprocessed first. The dataset contains training and test sets for model structure and validation. A Feed Forward Neural Network (FFNN) and 2D Convolutional Neural Network (CNN2D) are prepared to deal with fluctuated flood information spatial and worldly properties.
These models are consolidated into a federated model and sent to a focal server in the wake of preparing. Test the united model's presentation utilizing the assigned test set. Exactness, MSE, and RMSE are utilized to assess the model's prescient power.
This plan utilizes the advantages of FFNN and CNN2D models to further develop flood forecasting accuracy while guaranteeing strong execution through careful appraisal measures. The methodology permits cooperative model preparation without forfeiting information protection through federated learning, making it suitable for scale sending across flood-inclined areas. This strategy upgrades forecasting accuracy, illuminates navigation, and decreases flood gambles.

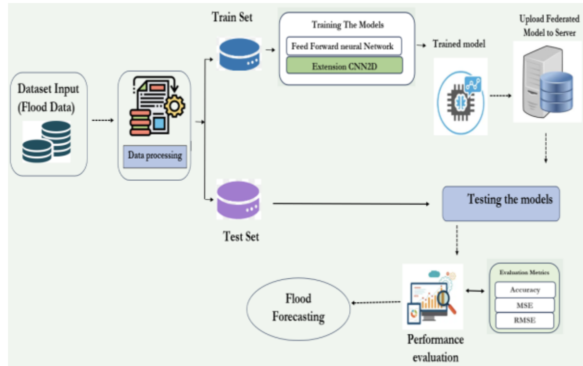


Fig.1 Proposed Architecture

Dataset:
The KERALA flood dataset on Kaggle was utilized in this work to expect floods utilizing accounts from a few stations or streams. Month to month precipitation and water level perceptions are expected to prepare and approve ML models in each dataset passage. Segment headings for factors like precipitation and water level improve on information planning such missing qualities, standardization, and rearranging. These information planning exercises set up the dataset for FFNN and CNN2D preparing. The assortment helps specialists control and relieve floods by precisely foreseeing water levels.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SUBDIVISI	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	water_level
2	KERALA	1901	28.7	44.7	51.6	160	174.7	824.6	743	357.5	197.7	266.9	350.8	48.4	3248.6
3	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205	315.8	491.6	358.4	158.3	121.5	3326.6
4	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157	59	3271.2
5	KERALA	1904	23.7	3	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7
6	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6
7	KERALA	1906	26.7	7.4	9.9	59.4	160.8	414.9	954.2	442.8	131.2	251.7	163.1	86	2708
8	KERALA	1907	18.8	4.8	55.7	170.8	101.4	770.9	760.4	981.5	225	309.7	219.1	52.8	3671.1
9	KERALA	1908	8	20.8	38.2	102.9	142.6	592.6	902.2	352.9	175.9	253.3	47.9	11	2648.3
10	KERALA	1909	54.1	11.8	61.3	93.8	473.2	704.7	782.3	258	195.4	212.1	171.1	32.3	3050.2
11	KERALA	1910	2.7	25.7	23.3	124.5	148.8	680	484.1	473.8	248.6	356.6	280.4	0.1	2848.6
12	KERALA	1911	3	4.3	18.2	51	180.6	990	705.3	178.6	60.2	302.3	145.7	87.6	2726.7
13	KERALA	1912	1.9	15	11.2	122.7	217.3	948.2	833.6	534.4	136.8	469.5	138.7	22	3451.3
14	KERALA	1913	3.1	5.2	20.7	75.7	198.8	541.7	763.2	247.2	176.9	422.5	109.9	45.8	2610.8
15	KERALA	1914	0.7	6.8	18.1	32.7	164.2	565.3	857.7	402.2	241	374.4	100.9	135.2	2899.1
16	KERALA	1915	16.9	23.5	42.7	106	154.5	696.1	775.6	298.8	396.6	196.6	302.5	14.9	3024.5
17	KERALA	1916	0	7.8	22	82.4	199	920.2	513.9	396.9	339.3	320.7	134.3	8.9	2945.3
18	KERALA	1917	2.9	47.6	79.4	38.1	122.9	703.7	342.7	335.1	470.3	264.1	256.4	41.6	2704.8
19	KERALA	1918	42.9	5	32.8	51.3	683	464.3	167.5	376	96.4	233.2	295.4	54.1	2501.9
20	KERALA	1919	43	6.1	33.9	65.9	247	636.8	648	484.2	255.9	249.2	280.1	53	3003.3
21	KERALA	1920	35.2	5.5	24.1	172	87.7	964.3	940.8	235	178	350.1	302.3	8.2	3303.1
22	KERALA	1921	43	4.7	15	171.3	104.1	489.1	639.8	641.9	156.7	302.4	136.2	15.8	2719.9
23	KERALA	1922	30.5	21.4	16.3	89.6	293.6	663.1	1025.1	320.6	222.4	266.3	293.7	25.1	3267.6
24	KERALA	1923	24.7	0.7	78.9	43.5	80	722.5	1008.7	943	254.3	203.1	83.9	41.6	3484.7
25	KERALA	1924	19.3	2.9	66.6	111	185.4	1011.7	1526.5	624	289.1	176.5	162.9	50.4	4226.4

Fig.2 Dataset
The top line of the dataset screen above shows the names of the dataset sections, while the resulting columns show the dataset values. The dataset incorporates month to month precipitation records, and the last segment shows the water level. In view of the expected water level, specialists will advise occupants about flooding.

Pre-Process Dataset:
Preprocessing the dataset guarantees information quality and reasonableness for flood forecasting machine learning model training. Month to month precipitation and water level data from KERALA stations and streams are expected to successfully assess flood conditions.
Missing qualities are credited utilizing measurable methodologies such segment mean or middle qualities or disposed of on the off chance that they are far and wide and can't be with certainty determined. Then, standardization scales the mathematical highlights to a standard reach, generally somewhere in the range of 0 and 1 or - 1 and 1, so all factors contribute similarly to show preparing without one overwhelming inferable from its size.
Rearranging the dataset is one more significant stage to randomize information passage request. This forestalls information request or patterns from influencing machine learning algorithm training. Revamping serious areas of strength for makes that sum up to new information.
These pre-handling steps — dealing with missing qualities, normalization, and rearranging — advance the dataset for FFNN and CNN2D model preparation. This upgraded dataset helps calculations gain and distinguish critical examples from the information, improving flood forecasts and supporting convenient flood control choices.

Training & Testing:
Training and testing are basic cycles in creating and assessing ML models for flood forecasting. In this exploration, 80% of the pre-handled dataset is utilized to prepare the models, which incorporate the Feed Forward Neural Network (FFNN) and the Convolutional 2D Neural Network (CNN2D). To decrease prediction blunders, the training methodology contains changing model boundaries by means of slope plummet and backpropagation. The excess 20% of the dataset is then used for testing, which permits model execution to be assessed on already obscure information. This gap ensures that the models sum up well and can dependably gauge water levels, which is basic for effective flood the board and moderation procedures.

Algorithms:
Feed Forward Neural Network: A feedforward neural network is one of the most essential types of artificial neural networks created. In this organization, data heads just in one path: forward — from the information hubs to the result hubs, passing through any secret hubs that exist. There are no cycles or circles in the organization. Feedforward neural networks were the primary type of artificial neural network developed, and they are easier than their counterparts, for example, repetitive and convolutional neural networks.

Architecture of Feed Forward Neural Networks: A feed forward neural network has three sorts of layers: input, stowed away, and yield. Each layer is comprised of neurons, which are connected by loads.

Input Layer: Neurons get and advance contributions to the following layer. The components of the info information impact the number of neurons that are utilized in the information layer.

Hidden Layers: Hidden layers act as the neural networks computational motor, as they are not presented to information or result. Each secret layer's neurons compute the weighted amount of the former layer's results, apply an enactment capability, and send the outcome to the following layer. The organization might have at least zero secret levels.

Output Layer: The output layer creates the ideal result in light of the data sources. The quantity of neurons in the result not entirely settled by the number of potential results the network that is wanted to deliver. Every neuron in one layer associates with each neuron in the accompanying layer, bringing about a totally connected network. Loads mirror the strength of the association among neurons, and learning in a brain network incorporates changing these loads relying upon yield blunder.

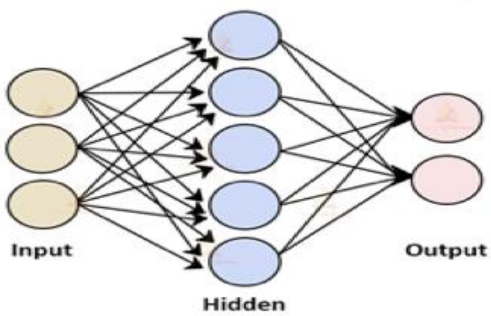


Fig.3 Architecture of Feed Forward Neural Network

Extension CNN2D Algorithm:
The Convolutional Neural Network (CNN or ConvNet) is a kind of deep neural network that is especially great at handling and assessing visual data. The Convolution2D layer is a key structure component of a CNN that performs convolutional procedure on input information.
Convolutional Neural Network (CNN) is an extended type of artificial neural networks (ANN) that is regularly used to extricate features from network like matrix datasets. For instance, consider visual datasets, for example, photographs or motion pictures, in which information designs assume a huge part.
CNN architecture: The info layer, pooling layer, convolutional layer, and completely associated layers are a portion of the layers that make up a convolutional neural network.

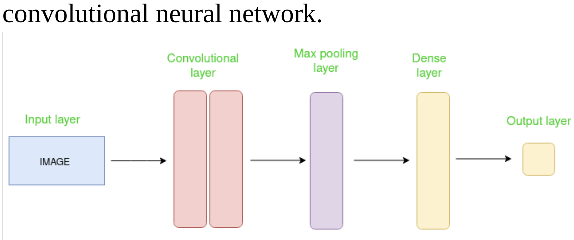


Fig.4 CNN Architecture

The info picture is handled by the Convolutional layer to remove features, the Pooling layer lessens calculation by down sampling the picture, and the completely associated layer creates the last prediction. The organization utilizes angle plummet and backpropagation to find the best channels.

EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is its ability to recognize debilitated from sound cases. To quantify test accuracy, figure the small part of true positive and true negative in completely broke down cases. Numerically, this is:

Accuracy = 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 predictor or assessor, at times alluded to as the capability of the gave information, is expected notwithstanding an objective of prediction or gauge to compute the mean squared error. The mean squared of the "errors" is known as the MSE.

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

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: The root mean square error (RMSE) is a typical proportion of the distinction between assessor or model predictions and noticed values. RMSE is the square root of anticipated noticed inconsistencies. This computation's inconsistencies are classified "residuals". The RMSE gauges error size. This scale-subordinate exactness metric is utilized to look at anticipating mistakes from various assessors for a given variable, yet not among factors.

Performance Evaluation Table:

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

$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values
$y_1, y_2, \ldots, y_n$ are observed values
n is the number of observations

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

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

Comparison graph:

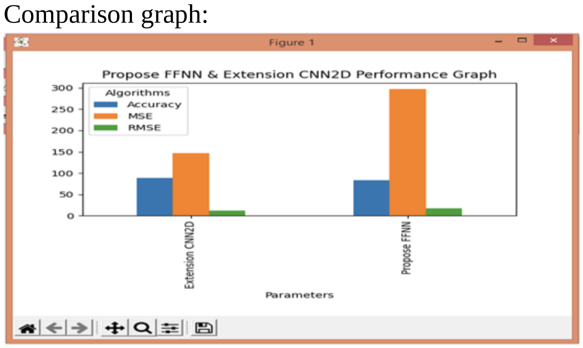
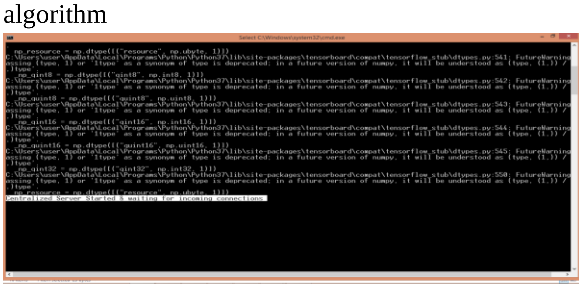
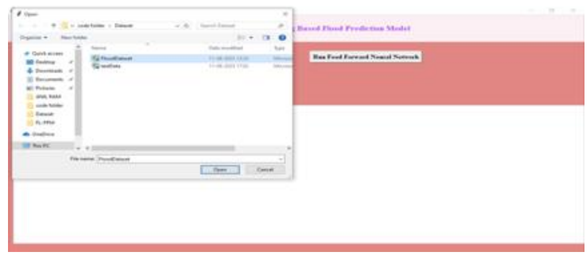


Fig.5 Comparison graph of FFNN & CNN2D algorithm

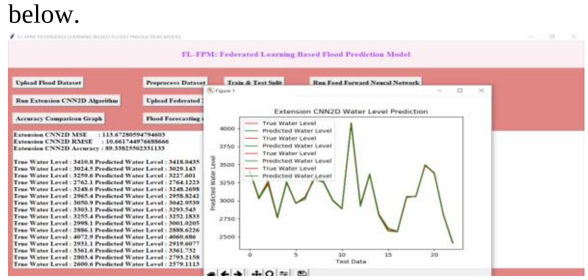


Double-click the "runServer.bat" file to launch the centralized server and see the output shown below.

The above screen shows that the centralized server has begun. Leave it running, and then double-click the "run.bat" file to launch the client, which will train the model locally by importing a local dataset and provide the output shown below.



To load the dataset and see the screen below, click the "Upload Flood Dataset" button in the above screen.
To load the dataset, choose and upload the "Flood Dataset" on the above page, then click the "Open" button.
The dataset is loaded in the above screen; click the "Pre-process Dataset" button to process the dataset and obtain the output shown below.
After completing the pre-processing steps for the dataset (normalization and shuffling) on the above page, click the "Train & Test Split" button to divide the dataset and obtain the output shown below.
The dataset size is displayed in the screen above, followed by the train and test sizes. Click the "Run Feed Forward Neural Network" button to train the suggested FFNN method and obtain the output shown below.



The FFNN training is finished in the screen above. The graph above shows the number of days, the water level, and the predicted and true water levels. Both lines fully overlap with very little space between them, indicating that the FFNN is producing the best predictions. We can now close the above graph to obtain the output below.

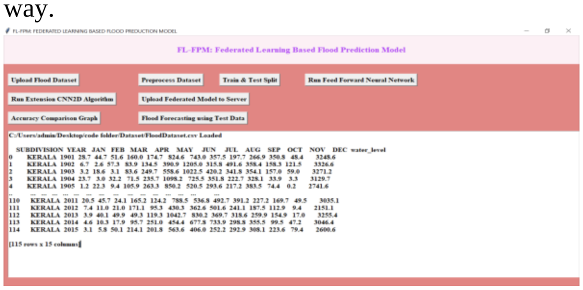
The FFNN algorithm's MSE, RMSE, and accuracy values are displayed in the first three lines of the screen above. Next, the actual and forecast water levels for the next few days are displayed. To train the extension algorithm, click the "Run Extension CNN2D Algorithm" button.

Close the above graph, then click the "Upload Federated Model to Server" button to upload the trained model to the server and receive the output below. In the above screen with the extension, we can see both predicted and true, which means reads and green lines are fully overlapping, so we can say extension model is better than propose. We can also see MSE and RMSE also lower compare to propose, and accuracy is high for extension algorithm.

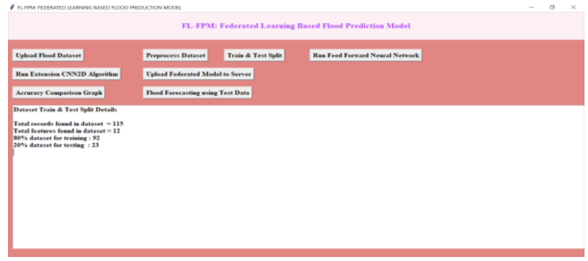
Simply input the name of the station in the above page, click OK, and the model will be uploaded to the server and produced as seen below.

The server responded with the message "model uploaded" on the screen above, and the received model data are shown in the server screen below.

The server output about model saving is seen in the white text on the screen above. The "Assam" model is stored in the server's "received" folder, and the server will save models for all stations in the same way.



The algorithm names are represented on the x-axis in the comparison graph above, and the accuracy and MSE values are represented on the y-axis. As we can see, the extension algorithm has higher accuracy and lower MSE and RMSE error when compared to the proposed FFNN algorithm. Now that the graph has been closed, click the "Flood Forecasting using Test Data" button to upload test data and predict the water level.



Upload test data on the top panel, then select "Open" to see the outcome below.

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

CONCLUSION

To summarize, the Flood Forecasting Model (FFM) that has been portrayed shows a two-module strategy to further develop flood prediction and moderation. An organization of one neighborhood checking station is laid out by the main module, which likewise prepares and sends information models to a focal server. By inspecting many variables from the nearby models, this focal server then forms a worldwide model that can foresee floods inside. To survey the expected ascent in water levels, the subsequent module utilizes a Feed Forward Neural Network at the area of the expected flood. Water driven and meteorological information are handled locally to fulfill protection, security, and information accessibility concerns. By furnishing the flood relief office with opportune flood alerts, the FFM really adds to proactive fiasco avoidance and reaction. The examination of past floods from 2010 to 2015 exhibits an incredible accuracy pace of 82.76%. Moreover, the precision of the model is tremendously expanded to 87.86% by broadening it with CNN2D.

With the coordination of datasets from different spots, the Flood Forecasting Model (FFM) expects to anticipate floods overall later on. The framework is a promising instrument for proactive flood forecasting for a bigger scope as a result of its demonstrated ability to adjust to local information. The FFM can possibly make a significant commitment to overall debacle the board endeavors through continuous turn of events, collaboration, and combination of worldwide datasets. It can give convenient estimates and experiences to numerous locales that are inclined to floods.

REFERENCE

[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.