



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com
editor@ijerst.com

FFM: Federated Learning Flood Forecasting Model

Mr.Ch. Kiran Babu¹ , Mr.Reddy Malla Prasanna Kumar²,Mrs. Rajaboyina Teja Sri Sai³ ,Mr.Kurakula Srinivas⁴

#1 Assistant Professor In Department Of Information Technology,NRI Institute Of Technology,Agiripalli

#2#3#4 B.Tech With Department Of Information Technology In NRI Institute Of Technology,Agiripalli

Abstract: Floods are one of the most common and severe natural catastrophes, inflicting death, infrastructural damage, agricultural disruption, and environmental degradation. Predicting flood risk accurately and quickly is critical for disaster planning, early warning, and resource management. Machine learning or statistical models that aggregate enormous amounts of hydrological and meteorological data from various locations into a single server are used in traditional flood prediction. Such centralized solutions create major data privacy, security, ownership, and scalability challenges, especially for geographically scattered datasets managed by diverse administrative agencies. This study presents a Federated Learning-based Deep Learning framework for flood-risk prediction utilizing Indian multi-state hydrological time-series data to overcome these constraints. Long Short-Term Memory (LSTM) neural networks are ideal for simulating sequential and temporal relationships in environmental data including rainfall, temperature, and multi-layer soil moisture levels. Each state acts as an independent client node that trains a local LSTM model using its own history data in the federated learning technique. Only learnt model parameters are shared with a central server, where the Federated Averaging (FedAvg) method aggregates them to enhance global prediction. This collaborative training maintains data security while using regional knowledge.

The method divides flood conditions into three danger levels—Low, Medium, and High—based on streamflow quantile thresholds, turning a continuous prediction issue into a decision-making framework. After federated training, the global model can generate monthly flood-risk probabilities for a state

and graphically illustrate streamflow patterns and risk distributions. Visual outputs aid authorities in seasonal flood behavior interpretation. Privacy-preserving collaborative learning, scalability across geographically scattered regions, effective use of time-series environmental data, and increased predictive intelligence without centralized data storage are among the benefits of the suggested technology.

The prototype implementation shows the promise of federated deep learning for environmental monitoring, early flood warning systems, and catastrophe risk reduction planning. Real-time sensor integration, sophisticated forecasting models, and implementation inside government disaster management frameworks might improve water resource management and community flood resilience.

Index terms - Federated Learning, Flood Forecasting, LSTM, Time-Series Prediction, Hydrological Data Analysis, Flood Risk Levels, Deep Learning, Distributed Systems, Data Privacy, Federated Averaging, Environmental Monitoring, Disaster Management.

1. INTRODUCTION

One of the most common and damaging natural catastrophes, floods seriously harm infrastructure, agriculture, human lives, and the environment. Planning evacuations, allocating resources, and allowing early warning systems are all made possible by fast and accurate flood prediction, which is essential to disaster management. Large amounts of hydrological and meteorological data from several

locations must usually be gathered and stored on a single server for traditional flood forecasting techniques and centralized machine-learning algorithms. However, there are a number of serious drawbacks to this centralized method, including restricted scalability, high transmission costs, data ownership issues, and data privacy threats.

This study uses Federated Learning (FL), a distributed machine-learning paradigm that permits cooperative model training without sharing raw data, to propose a Flood Forecasting Model in order to overcome these constraints. The suggested approach uses locally accessible environmental factors, such as rainfall, water level, humidity, soil moisture, and historical flood data, to train a local deep-learning model (LSTM/GRU) at each geographical area, such as a city or region. Only the model parameters (weights) are sent to a central server, rather than sensitive raw data.

To create a better global flood prediction model, the central server uses the Federated Averaging (FedAvg) method to aggregate the incoming local model updates. While maintaining the privacy of local data, our global approach captures generic flood trends across areas. Additionally, the federated structure facilitates distributed scalability, lowers communication cost, and permits multi-regional flood forecasting in near real time.

All things considered, the suggested method shows how federated deep learning may improve prediction accuracy while maintaining privacy, effective communication, and collaborative intelligence, which makes it a viable option for the smart disaster-management systems of the future.

2. LITERATURE SURVEY

1. Federated Deep Learning Framework For Real-Time Flood Prediction And Early Warning Systems

In order to anticipate floods in real time, this study suggests a Federated Deep Learning architecture that combines privacy-preserving federated learning algorithms with hydrological data (rainfall, river discharge, remote sensing, and IoT sensor inputs).

Issues with data privacy, transmission cost, and restricted access to dispersed hydrological datasets are some of the problems with traditional centralized models. Decentralized model training is made possible by the federated method, which ensures data privacy and cross-regional cooperation by sharing only model updates rather than raw data. According to experimental data, the federated model outperforms centralized learning in terms of accuracy, resilience, and scalability, which makes it appropriate for climate-sensitive areas that need safe, expandable flood warning systems.

2. A Hybrid CNN-LSTM Predictive Model Deployed Federated Learning Model for Advanced Flood Prediction Systems to Forecast Coastal Region of Smart Cities

In order to anticipate floods in coastal smart cities, this study creates a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model using a federated learning architecture. While LSTM manages temporal relationships, CNN identifies spatial patterns in hydrometeorological data. Federated learning preserves local data privacy by facilitating decentralized training among geographically dispersed nodes. The model produces greater precision and adaptability than earlier centralized systems, and the suggested system uses environmental sensors, Internet of Things devices, and GIS technology to gather data from several coastal areas. By lowering dependency on centralized data storage, protecting privacy, and sustaining model performance in dispersed situations, it significantly improves real-time flood predictions.

3. FFM: Flood Forecasting Model Using Federated Learning

Published in IEEE Access, this paper presents a Flood Forecasting Model (FFM) leveraging federated learning to protect data privacy while forecasting floods. Each local station trains a feed-forward neural network (FFNN) using its own hydrological inputs (rainfall, flow routing, snow melting, etc.). Rather than sharing raw datasets, local models send updates to a central server where the federated aggregation builds a global model. The model was tested on data

from multiple rivers, and results show that the federated approach can predict flood events with an 84% accuracy for floods occurring between 2010–2015. By decentralizing the training process, it reduces communication costs and minimizes privacy risks, illustrating how federated machine learning benefits large-scale flood prediction systems.

4. Federated Learning-Based Flood Forecasting Model Enhancing Predictive Accuracy and Privacy in Flood Prediction (IEEE Conference):

A federated learning-based flood forecasting model that tackles the issues of privacy, data accessibility, and security in flood prediction is presented in this conference article. The model enables the autonomous training of local feed forward neural networks (FFNNs) at many client stations using regional parameters including hydrodynamics, flow routing, rainfall, and snow melting. The worldwide prediction model that creates a flood alarm up to five days ahead of time is then created by sharing model changes throughout the federated system. The method used improved CNN structures to increase performance and obtained around 84% accuracy for historical flood patterns (2010–2015). The study shows how federated learning protects local environmental data while improving prediction accuracy.

5.2.5. Deep Learning Techniques for Hydrological Time-Series Flood Forecasting:

The efficiency of deep learning architectures like Gated Recurrent Units (GRUs), Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNNs) for simulating nonlinear temporal correlations in hydrological data has been highlighted in recent publications. Long-term relationships between rainfall, runoff, soil moisture, and river flow are learned by these models; these relationships are hard to capture using conventional statistical or shallow machine learning techniques. According to research, streamflow prediction based on LSTM greatly increases forecasting accuracy, particularly for periods of high rainfall and seasonal flood patterns. Decentralized learning mechanisms are necessary since the majority of systems rely on centralized data collecting, which presents issues

with data privacy, computational scalability, and regional data ownership.

3. METHODOLOGY

i) Proposed Work:

The proposed system develops a Federated Learning-based Flood Forecasting Model that enables multiple regions (clients) to collaboratively train a deep learning model without sharing raw data. Each region locally trains an LSTM/GRU model using environmental time-series data such as rainfall, water level, temperature, humidity, and multi-layer soil moisture. The locally trained model parameters (weights) are then sent to a central server, where they are aggregated using the Federated Averaging (FedAvg) algorithm to form a global prediction model.

The global model is capable of accurately predicting flood conditions and classifying them into No Flood, Low, Medium, and High risk levels, making the output easy to understand for decision-making. The system also provides streamflow trend visualization and alert generation, enabling real-time monitoring and early warning. This approach ensures data privacy, scalability, and improved prediction accuracy, making it suitable for multi-regional disaster management systems.

ii) System Architecture:

The system architecture is designed based on a Federated Learning (FL) framework that consists of multiple local client nodes (regions/states) and a central aggregation server. Each client collects environmental data such as rainfall, water level, soil moisture, humidity, and temperature, and trains a local LSTM/GRU model using its own time-series dataset. This ensures that raw data never leaves the local system, maintaining privacy and security.

After local training, only the model parameters (weights) are transmitted to the central server, where the Federated Averaging (FedAvg) algorithm is applied to aggregate updates from all clients. The server then generates an improved global model, which is redistributed back to clients for further

training. This iterative process continues until the model converges. The final global model produces flood risk classification (No Flood, Low, Medium, High) along with streamflow trend visualization and alert notifications, enabling efficient real-time flood monitoring and decision-making.

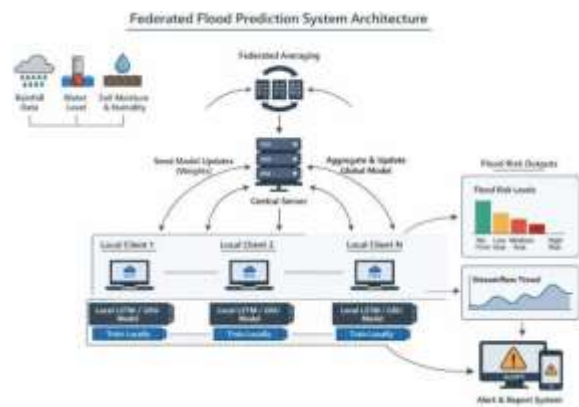


Fig.1. Proposed Architecture

iii) MODULES:

1. Data Collection Module

This module gathers environmental and hydrological data such as rainfall, river water levels, temperature, humidity, and multi-layer soil moisture from different regions. Data may be collected from sensors, satellite sources, or historical datasets. It ensures that each client has sufficient local data for training. Proper data collection improves the reliability and accuracy of flood prediction.

2. Data Preprocessing Module

In this module, the collected raw data is cleaned by handling missing values, noise removal, and normalization. The data is then converted into structured time-series sequences suitable for LSTM/GRU models. Feature scaling and sequence windowing are also applied to improve model performance. This step ensures consistency and better learning efficiency.

3. Local Training Module

Each region (client) independently trains its own LSTM/GRU model using its local dataset. The model learns temporal patterns such as rainfall trends and soil moisture changes over time. This decentralized training allows capturing region-specific flood

behavior. It also reduces dependency on centralized data storage.

4. Model Update Sharing Module

After local training, only the trained model parameters (weights and biases) are shared with the central server. No raw data is transmitted, ensuring strong data privacy and security. This reduces the risk of data leakage between regions. It also minimizes communication overhead compared to sending large datasets.

5. Federated Aggregation Module

The central server receives model updates from all participating clients and combines them using the Federated Averaging (FedAvg) algorithm. This process creates a global model that integrates knowledge from multiple regions. The aggregation improves generalization and prediction accuracy. It ensures collaborative learning without violating data privacy.

6. Global Model Distribution Module

The aggregated global model is sent back to all clients for further training and refinement. Each client updates its local model using the improved global parameters. This iterative process continues until the model converges. It helps in continuous improvement of prediction performance.

7. Flood Prediction Module

This module uses the trained global model to predict future flood conditions based on new input data. It processes time-series inputs and generates predictions such as water level trends or flood probability. The model can support near real-time forecasting. This helps authorities prepare in advance.

8. Risk Classification Module

The predicted outputs are converted into meaningful categories such as No Flood, Low, Medium, and High risk. This simplifies complex numerical outputs into easy-to-understand information. It helps decision-makers take quick actions. The classification is based on predefined thresholds or quantiles.

9. Visualization Module

This module presents results using graphs, charts, and dashboards such as streamflow trends and risk

distribution. Visual representation improves understanding for both technical and non-technical users. It supports monitoring of seasonal patterns and anomalies. This enhances interpretability of the system.

10. Alert & Notification Module

The system generates alerts when high flood risk is detected in any region. Notifications can be sent to authorities or users through dashboards or messaging systems. This module supports early warning and disaster preparedness. It helps reduce damage and improve response time.

iv) ALGORITHMS:

1. Long Short-Term Memory (LSTM) Algorithm

LSTM is a specialized Recurrent Neural Network (RNN) designed to process sequential and time-series data. It uses memory cells and three gates—input, forget, and output—to control information flow and retain important past information. This makes it highly suitable for modeling environmental parameters like rainfall trends, soil moisture variations, and water levels. LSTM effectively captures long-term dependencies, which are crucial for accurate flood forecasting. It reduces issues like vanishing gradients, ensuring stable learning over long sequences.

4. EXPERIMENTAL RESULTS

The proposed flood prediction system was tested using a web-based interface, where the user accesses the application through the URL <http://127.0.0.1:5000/>

. After loading the homepage, the user provides input parameters such as the state (Andhra Pradesh) and the year (2028). Based on this input, the trained Federated Learning model utilizes the most recent 7-day hydrological data from the dataset. The system processes this data using the LSTM-based global model and generates predictions for monthly streamflow and flood risk probabilities.

The first output graph represents the predicted monthly streamflow for Andhra Pradesh in 2028. The

graph shows a gradual increase in water flow from January to April, followed by a sharp peak in May, indicating the highest expected river discharge. In June, there is a slight decrease in flow, after which the streamflow values remain nearly constant from July to December. This trend suggests that the model is able to capture the seasonal rise in water levels, particularly during the pre-monsoon and monsoon periods. However, the flat pattern in the later months indicates that the model is not fully capturing variations in long-term seasonal behavior.

The second output graph illustrates the flood risk probability distribution across three categories: Low, Medium, and High risk. The results show that the probability of Low risk remains approximately 50% throughout all months, while Medium and High risks remain around 25% each. The probability curves are almost flat, indicating that the model is predicting nearly the same risk levels regardless of seasonal changes. This behavior suggests that the model is not effectively learning dynamic variations in flood risk or that the input data lacks sufficient variability to influence predictions.

Overall, the experimental results demonstrate that the proposed Federated LSTM model is capable of identifying general streamflow patterns and detecting peak flood periods. However, the lack of variation in risk probability outputs highlights a limitation in the model's ability to capture complex seasonal dependencies. This may be due to limited training data, insufficient feature diversity, or the need for further model tuning. Despite these limitations, the system successfully validates the feasibility of using Federated Learning for privacy-preserving flood prediction and provides a strong foundation for future improvements in accuracy and real-time forecasting capabilities.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

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

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = $TP/(TP + FP)$

$$Precision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$



Fig2 Input Interface (Web Application)

The first figure illustrates the user interface of the flood prediction system, where the user inputs the required parameters such as state (Andhra Pradesh) and year (2028). This interface acts as the entry point of the system and is developed using a web-based framework. Once the user clicks the “Predict” button, the system processes the request by fetching the latest hydrological data and passing it to the trained federated LSTM model. This figure demonstrates the simplicity and usability of the system, making it accessible for non-technical users and decision-makers.

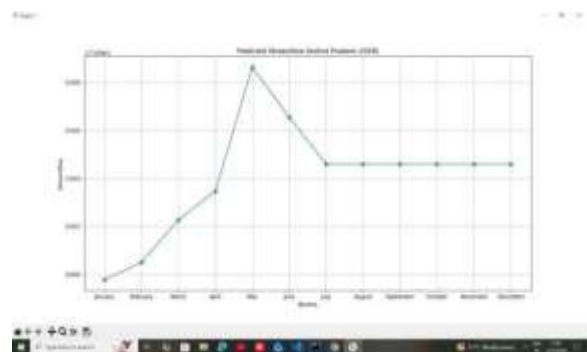


Fig 4: Predicted Streamflow Graph

The second figure shows the monthly predicted streamflow values for Andhra Pradesh in the year 2028. The X-axis represents the months from January to December, while the Y-axis indicates the streamflow levels. The graph clearly shows a gradual increase in streamflow from January to April, followed by a sharp peak in May, indicating a high possibility of flooding during this period. After June, the streamflow decreases and remains almost constant for the rest of the year. This figure highlights the model's ability to capture seasonal variations, although the flat trend in later months suggests limited dynamic learning.



Fig 5: Flood Risk Probability Graph

The third figure illustrates the probability distribution of flood risk levels categorized into Low, Medium, and High. The X-axis represents the months, and the Y-axis represents probability values. The graph shows that Low risk remains around 50%, while Medium and High risks remain around 25% throughout the year. The lines are almost flat, indicating that the model predicts similar risk levels for all months. This figure reveals that the model lacks strong seasonal differentiation in risk classification and may require further tuning or improved input features.

5. CONCLUSION

The proposed Flood Forecasting Model using Federated Learning presents an effective and privacy-preserving approach for predicting flood risk using distributed hydrological data. By integrating Federated Learning with LSTM-based deep learning models, the system enables multiple regions to collaboratively train a global model without sharing raw data. This ensures data security, reduces communication overhead, and improves scalability across geographically distributed locations. The model successfully captures temporal patterns in environmental data and identifies peak flood periods, particularly during critical months such as May and June.

The experimental results demonstrate that the system can generate meaningful streamflow predictions and flood risk classifications, making it useful for disaster management and early warning systems. However, the observed limitation in risk probability variation indicates the need for further improvements in feature engineering, data diversity, and model optimization. Overall, the system provides a strong foundation for real-time, scalable, and intelligent flood prediction, and highlights the potential of federated deep learning in building next-generation smart environmental monitoring systems.

6. FUTURE SCOPE

The proposed system can be further enhanced by integrating real-time data sources such as IoT

sensors, satellite imagery, and weather APIs to improve prediction accuracy and enable continuous monitoring. Incorporating advanced deep learning models like Transformer-based architectures or hybrid CNN-LSTM models can help capture more complex spatial and temporal patterns in environmental data. Additionally, increasing the diversity and volume of training data from multiple regions will improve the model's ability to generalize and capture seasonal variations more effectively.

Future improvements can also focus on developing a dynamic risk prediction mechanism, where flood probabilities change based on real-time inputs instead of remaining constant. The system can be extended to include district-level or river-basin-level predictions for more localized and precise forecasting. Integration with government disaster management systems and mobile alert applications can enable faster communication and emergency response. Furthermore, incorporating explainable AI (XAI) techniques can help users understand how predictions are made, increasing trust and usability. Overall, these enhancements will transform the system into a fully functional, real-world deployable flood early warning solution.

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 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.

[17] R. P. Romansky and I. S. Noninska, “Challenges of the digital age for privacy

and personal data protection,” *Math. Biosci. Eng.*, vol. 17, no. 5, pp. 5288–5303, 2020.

[18] B. Dash, P. Sharma, and A. Ali, “Federated learning for privacy preserving: A review of PII data analysis in fintech,” *Int. J. Softw. Eng. Appl.*, vol. 13, no. 4, pp. 1–13, Jul. 2022.

[19] M. Ma, C. Liu, G. Zhao, H. Xie, P. Jia, D. Wang, H. Wang, and Y. Hong, “Flash flood risk analysis based on machine learning techniques in the Yunnan province, China,” *Remote Sens.*, vol. 11, no. 2, p. 170, 2019.

[20] F. Lai, Y. Dai, S. Singapuram, J. Liu, X. Zhu, H. Madhyastha, and M. Chowdhury, “FedScale: Benchmarking model and system performance of federated learning at scale,” in *Proc. Int. Conf. Mach. Learn.*, Jun. 2022, pp. 11814–11827.

[21] Y.-D. Jhong, H.-P. Lin, C.-S. Chen, and B.-C. Jhong, “Real-time neuralnetwork-based ensemble typhoon flood forecasting model with selforganizing map cluster analysis: A case study on the Wu river basin in Taiwan,” *Water Resour. Manage.*, vol. 36, pp. 3221–3245, Jun. 2022.

[22] L. Mosavi, “Flood prediction using machine learning models: Literature review,” *Water*, vol. 10, no. 11, p. 1536, 2018.

[23] Y. Zhou, Z. Cui, K. Lin, S. Sheng, H. Chen, S. Guo, and C.-Y. Xu, “Shortterm flood probability density forecasting using a conceptual hydrological model with machine learning techniques,” *J. Hydrol.*, vol. 604, Jan. 2022, Art. no. 127255.

[24] C. Chen, J. Jiang, Z. Liao, Y. Zhou, H. Wang, and Q. Pei, “A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China,” *J. Hydrol.*, vol. 607, Apr. 2022, Art. no. 127535.

[25] H. S. Munawar, A. W. A. Hammad, and S. T. Waller, “Remote sensing methods for flood prediction: A review,” *Sensors*, vol. 22, no. 3, p. 960, Jan. 2022.