International Journal of Research in Engineering Technology and Applications

E-ISSN: 3108-2556
P-ISSN: 3108-2610

Projection of the Extent of Inundation Corresponding to the Forecasts of Flood Levels in a River

Madhusudhan M V[1], A Y Harshitha[2], Avani M[3]
[123]*School of CSE, Presidency University, Bengaluru, India*
[1]*mv.madhu@gmail.com*, [2]*ayharshita@gmail.com*, [3]*avanim004@gmail.com*

Received: 30 January 2026 *Revised: 28 February 2026* *Accepted: 30 March 2026*

Abstract*—The analyzed project is Flood Prediction Using Satellite Imagery and Machine Learning, mainly the creation of the machine learning back-end. Floods should be predicted, among other things, to preserve infrastructures, human lives, and the environment from flood disasters. This project applies machine algorithms to the satellite image data to make the correct flood prediction. The back-end system is utilizing K-Means, XGBoost, and Decision Trees machine learning models, which have been trained on environmental data derived from a mixture of satellite information. The parameters that were considered to enable better prediction of the future are those that were observed, e.g. rainfall intensity, sizes of water bodies, soil moisture, etc. These models were elaborately put to the test by trial and error, and a great number of experiments were made in order to find out the improved performance of the learning based on the accuracy rate. The reader can get an idea of the trend and patterns from the graphs over parameter results. The correctness of the models is also compared by the measures that have been set giving an idea of what models would be fitting the data. The project will come up with a solution that is both environmentally friendly and can be easily deployed to solve the problem of flood prediction so as to avoid the difficulties of setting up an end-to-end and scalable backend infrastructure. Although the paper has only dealt with model training and evaluation, such models may then be used in large flood monitoring systems to give a strong input of early warning in areas prone to flood.

***Keywords*—Flood Prediction, Machine Learning, Satellite Imagery, XGBoost, Hydrological Modeling, Real-time Monitoring.**

I. INTRODUCTION

Floods rank among the most frequent and devastating natural risks leading to a large number of deaths, economic losses, and environmental destruction at the global level. In the last few years, besides climate change, factors such as urbanization, deforestation, and lack of drainage systems have increased both the quantity and the severity of floods. The tragedy not only destroys human habitation and agricultural land but also has socio-economic impacts that last for a very long time: most of these communities never become the same again.

Primary traditional predictive flood models mainly consist of statistical models and hydrological ones based on physical parameters such as precipitation condition, soil moisture condition, and river discharge. Despite the fact that these models have been very useful for several decades, they generally perform poorly when they are given limited or conflicting data. Additionally, conventional models do not help in handling the dynamic changes of weather conditions and the existence of a wide range of data sources like satellite images and real- time local data.

At this point, machine learning (ML) comes as a revolutionary and highly effective environmental modelling and disaster planning tool. As ML algorithms can automatically detect complex patterns and relationships in large amounts of data, the precision and speed of flood prediction will be improved. The ML models can also combine different types of data such as satellite images, meteorology, and river water level to give predictions based on real-time data; this is not possible with traditional methods.

Besides, the study incorporates satellite data that gives another dimension of space to flood prediction, and therefore, data on rain distribution, vegetation, soil moisture, and topography. The pictures allow the identification of areas at risk of flooding more accurately and timely if they are combined with a machine learning tool.

The primary objective of the presented research is to design a flood prediction system using ML techniques with the help of

satellite images combined with environmental variables for that purpose of early warning reliability enhancement. The proposed solution employs supervised-unsupervised learning algorithms such as XGBoost, K-Means Clustering to model data of large-scale size, multiform, and multi-source type. This article is designed to help the government authorities, environmental agencies, and communities to be disaster- ready, to prevent the loss of properties, and even to save the lives.

The aim of this research is to support government authorities, environmental agencies, and communities in their disaster preparedness activities by preventing property loss and, potentially, saving human lives. This approach highlights the potential of new technologies as components of flood-resilient infrastructure and environmentally sustainable flood management in hazard-prone areas.

II. RELATED WORK

In this paper, an analysis of the Knowledge Graphs and Graph Neural Networks [1] will be provided. The model could utilize multi-source data, including the geographical and historical flood information as well to analyze complex relationships. This experiment using 9000 records of Jiangxi Province demonstrates that the proposed model has a high prediction ability with an AUC of 0.84. The findings are better than the conventional machine learning systems such as RF, SVM, and ANN. This paper presents a model for evaluating remote sensing algorithms for flood detection via satellites in data restricted areas. The model judges the performances of the algorithms in aspects such as spatial accuracy, temporal consistency and their relation with the real flood damage records. The findings indicate that the suggested algorithm has a greater accuracy and reliability rate, with a mean F1 of approximately 0.92.

The present paper presents a Knowledge-Driven Flood Intelligent Monitoring (KDFIM) [3] approach based on the Sentinel-1 SAR, Sentinel-2 satellite data and DEM in order to enhance flood detection and monitoring. The model was applied to the Kakhovka Dam destruction flood incident where high detection accuracy of approximately 98.46% was attained.

In this paper, the authors suggest one Flood Forecasting Model (FFM) [4] based on Federated Learning and Feed-Forward Neural Networks to predict floods without violating data privacy. According to experimental findings, the model can estimate historical floods at the rate of approximately 84–88%. This paper presents a proposal of a flood forecasting system that combines big data and crowdsourced data on the basis of machine learning [5]. Throughout the experiment, the ANN model based on MLP demonstrated the highest results in prediction accuracy of approximately 97.9%.

The current paper suggests a flood detection solution based on Topological Data Analysis (TDA), Wavelet transformation and unsupervised machine learning methods [6]. The findings indicate that the suggested approach has approximately 96% accuracy in identifying flood-prone regions.

In this paper, the author is interested in the enhancement of alarm flood classification [7] in industrial process plants through machine learning. It suggests an approach combining Conformal Prediction and Early Time Series Classification to deal with uncertainty in alarm flood data. The paper suggests a Flood Monitoring, Prediction, and Rescue (FMPR) [8] system based on a multi-agent approach to enhance the management of disasters. The system gathers information on sensors and stakeholders to forecast floods and organize the process of rescuing.

In this paper, a flood prediction model based on Federated Learning [9], Feedforward Neural Networks (FNN), LSTM-RNN, and Explainable AI methods is suggested. Experimental findings indicate that the FNN model has high prediction accuracy as indicated by the R^2 value at approximately 0.96.

The given paper suggests a spatiotemporal model of flood hazard classification based on Graph Convolutional Networks (GCN) [10] and Temporal Fusion Transformer (TFT). The results demonstrate that GCN-TFT model is more effective in predicting the degree of flood hazard compared to the traditional machine learning and deep learning models.

The research paper [11] suggests a hybrid flood risk assessment model that combines probability distribution techniques and ensemble machine learning models. A mixed machine learning model based on Random Forest, XGBoost, and Kernel SVM performed almost 98% prediction accuracy and AUC of 0.96. The study [12] is aimed at prediction of urban flood water levels. GA-XGBoost and DE-XGBoost demonstrate higher accuracy and reliability as compared to the traditional models such as Random Forest and CART.

The current research paper suggests a real-time forecasting model of urban flood water accumulation [13] with the Gradient

International Journal of Research in Engineering Technology and Applications
E-ISSN: 3108-2556
P-ISSN: 3108-2610

Boosting Decision Tree (GBDT) algorithm. The experimental findings indicate an average relative error of 19.77% and prediction accuracy of approximately 82%. The current research paper demonstrates a satellite- based method of flood monitoring and forecasting with SMAP and Landsat data [14]. The model is able to forecast 1-day floods with high accuracy (correlation ≥ 0.87) over Landsat observations.

The proposed research paper is [15] a deep learning framework named SRFNet, which predicts the flood range based on multimodal remote sensing analysis. Experiments on Dongting Lake and Poyang Lake datasets (2010–2020) indicate structural similarity above 0.9.

The proposed research paper provides an IoT-based flood severity prediction system [16] based on ensemble machine learning models. An ensemble model of LSTM and Random Forest achieved accuracy of approximately 0.997 in training and 0.811 in testing.

In this research paper [17], machine learning forecasts flash floods by measuring Precipitable water vapor (PWV) with GNSS/GPS signals. The findings indicate that values of PWV rise considerably prior to flood occurrence, serving as a key predictor. The study hypothesizes a deep learning-based flood forecasting model of small mountainous watersheds [18] on a compound Long Short-Term Memory (LSTM) framework combining CNN and Attention networks to enhance prediction accuracy.

III. PROPOSED METHOD

The concept is the following: a real-time flood predicting and monitoring service that gathers the data of all kinds: satellite, weather data, and historical flood data shows in Fig.1. The system modifies to the dynamic nature of the environment and produces correct and up-to-date forecasts of floods by combining both monitored and unmonitored machine learning models. The implementation of such a process is done in five main steps: collecting the data, data cleaning/normalization, exploratory data analysis, model building, and learning through feedback prediction.

Inputs to the system are derived from various sources sensors, geostationary satellite images, weather APIs, and historical flood records collected by the government and meteorological agencies.

Data Cleaning: Sensor data is generally very noisy and often has missing data, especially when the sensors have a breakdown. The system takes care of the identification of the outliers and the imputation of the missing data to make the dataset clean.

Data Normalization: Sensor readouts normally represent different units and scales. Normalizing these brings them all onto the same scale thus enabling the machine learning algorithms to work with uniform data. Machine learning models like Random Forest, SVM and LSTM networks make use of historical flood data to estimate the likelihood and intensity of a flood.

Flood Risk Analysis: Prepared data is loaded into a system capable of analysing the risk and categorizing the locations into low, medium, and high, risk flood areas.

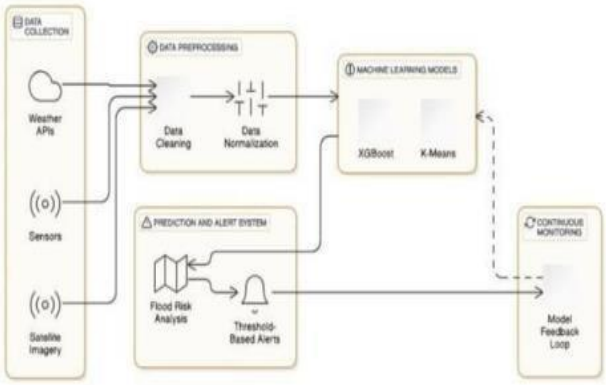

Fig.1 Real Time Flood Prediction Process

IV. METHODOLOGY

A. System Overview

Regardless of the kind, this system is pulling data for Whatever the type, system collects data by itself through weather reports, satellite images, and even very old flood records shows in Fig.2. Data Collection Module is responsible for collecting all the data. Pre-processing Module then does the work of cleaning, removing inconsistencies, and sorting the data one. Machine Learning Engine performs actual number, crunching both supervised and unsupervised to identify patterns and make predictions. Risk Classification and Alert Module identifies areas at high risk and issues early warnings. Feedback and Continuous Learning Module works on new data and updates the models, which results in better predictions.

International Journal of Research in Engineering Technology and Applications
E-ISSN: 3108-2556
P-ISSN: 3108-2610

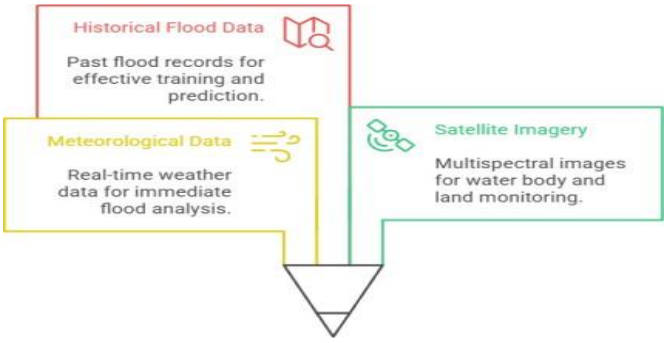

Fig.2 System Architecture

B. Data Collection

A rational prediction is hypothetically based on solid and diversified data. The system obtains information from everywhere which it considers to be trustworthy. For weather, it depends on live data such as rainfall temperature humidity, and wind speed that it retrieves through APIs like Open Weather and Marea. The satellites INSAT, Sentinel, 1, and Landsat 8 and 9 return multispectral and radar images, that allow the water monitoring, and the earth changes tracking. The system also draws on past flood data—river discharge, rainfall, and floods—straight from government and hydrology agencies.

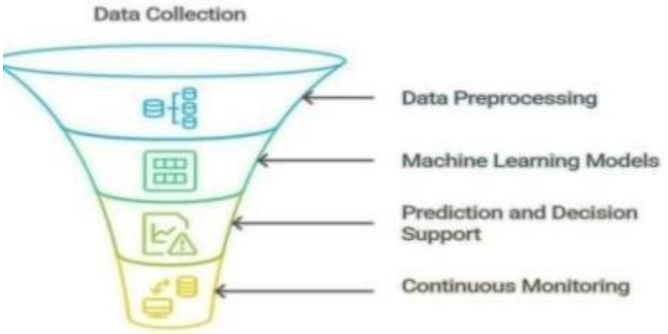

Fig.3 Data Convergence for Flood Forecasting

C. Data Pre-processing

Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text shown in Fig.4.

Fig.4 Data Pre-processing Sequence

D. Exploratory Data Analysis

Next, I will be using two algorithms for modeling: XGBoost—a supervised learning algorithm—and K- Means—an unsupervised learning algorithm. Both are fast, reliable, and handle data efficiently.

In the direction of supervision, XGBoost comes into play to classify and predict floods based on historical data with known outcomes. GridSearchCV is employed to tune the model with the best parameters. Performance is measured in terms of accuracy precision recall, and F1, score. On the unsupervised side, K, Means assigns areas to low, medium, or high, risk categories, based on the similarity of their environmental profiles. The Elbow Method is applied to determine the most appropriate number of groups. The data is divided into 80% training and 20% testing, to prevent models from overfitting training data and focus on learning meaningful patterns.

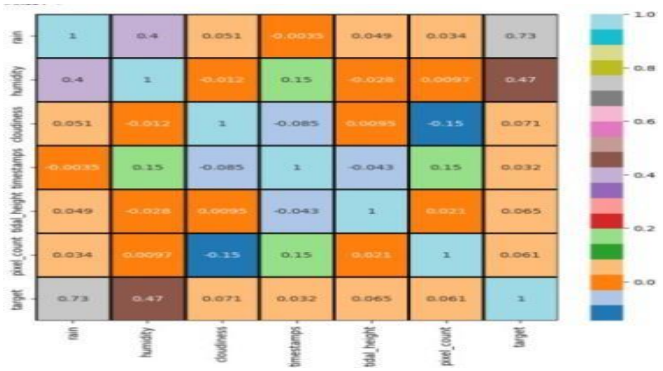

Fig.5 HeatMap

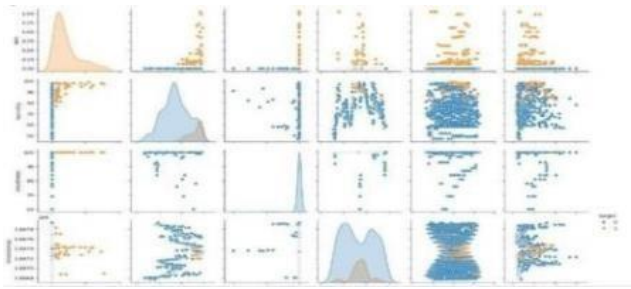

Fig.6 Sample Techniques Graph

International Journal of Research in Engineering Technology and Applications
E-ISSN: 3108-2556
P-ISSN: 3108-2610

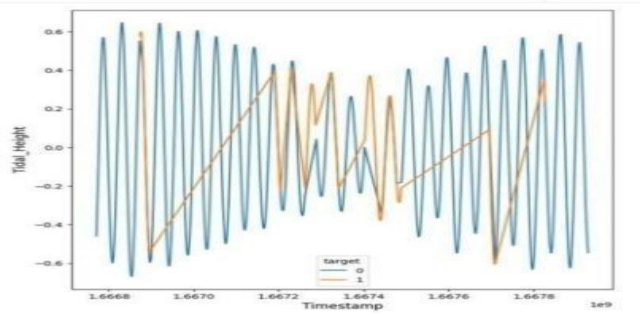

Fig.7 Techniques Using Matplotlib

E. Prediction and Risk Assessment

XGBoost and K Means models work together in a real, time Prediction Engine that's always on the lookout for fresh data as it comes in. The mechanism is triggered only when rainfall and water levels cross the preset thresholds. It assesses the flooding risk of the area and sends out immediate alerts. Outputs are presented on a dashboard indicating flood likelihood, intensity, and at, risk locations.

F. Continuous Learning and Model Optimization

Thanks to a feedback loop, the system is improving continually. With every new flooding data, the model is essentially self, training; it's being informed about the latest seasonal and local variations. This infinite chain of updating results in the forecasts being not only fresh all the time but also highly responsive something that's a must for the real, time cases.

V. RESULTS AND ANALYSIS

Physical trials were performed using the system to verify its performance. Various meteorological, satellite, and hydrological data were taken from freely available sources. Besides accuracy verification of the flood prediction model, the main aim was to make sure that the model is trustworthy and can be adapted to new locations and situations. Python software e.g. Scikit, learn and XGBoost were used for running all experiments.

A. Model Performance Evaluation

Standard metrics such as accuracy precision recall, and F1, score were used to measure the performance of the models trained on different environmental datasets. The evaluation was centred on the system's effectiveness in detecting flood events and classifying areas at risk by using real, time data including rainfall temperature soil moisture, and river discharge as inputs. Data were split into an 80% training set and a 20% testing set. XGBoost was utilized as the main supervised learning model, benchmarked against Random Forest and Decision Tree. The results showed that XGBoost achieved 96.4% accuracy, 95% precision, 93.6% recall, and 94.3% F1-score. Random Forest and Decision Tree achieved accuracies of 92.1% and 89.3% respectively shown in Table 1.

Table 1: Model Comparison Graph

Metric	Accuracy	Precision	Recall	F1-Score
XGBoost	96.4	95.0	93.6	94.3
Random Forest	92.1	91.5	88.4	89.9
Decision Tree	89.3	87.2	86.7	86.9
Logistic Regression	84.5	83.1	81.9	82.3

B. Model Evaluation Metrics

Precision: Of all the times the model said “There is a flood,” how many times was it right?

Precision = TP / (TP + FP)

High precision means the system is not making false alarms. You can trust it when it tells you that there is a flood.

Recall (also known as Sensitivity or True Positive Rate): How many real floods did the model manage to identify?

Recall = TP / (TP + FN)

High recall means you are not missing real floods— missing one can lead to a disaster.

F1-Score: The harmonic mean of precision and recall:

F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

Accuracy: Of all predictions (floods and non- floods), how many were correct?

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Good accuracy sounds great, however, non-flood days in most flood datasets are much more common than flood days. Therefore, you cannot only rely on accuracy; all these measures together give a proper understanding of the system’s functionality.

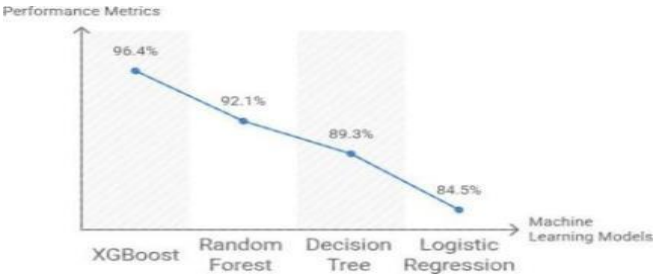

Fig.8 Confusion Matrix

International Journal of Research in Engineering Technology and Applications
E-ISSN: 3108-2556
P-ISSN: 3108-2610

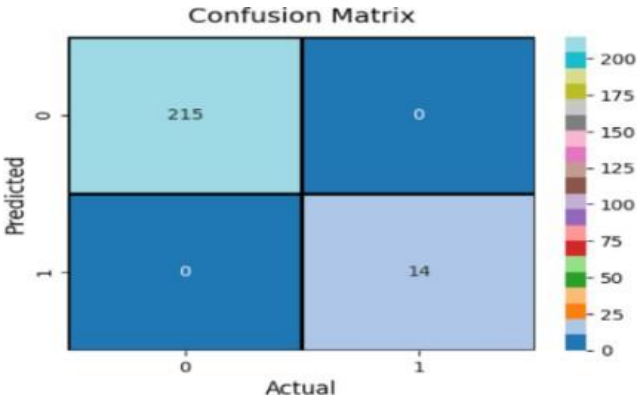

Fig.9 Normalization

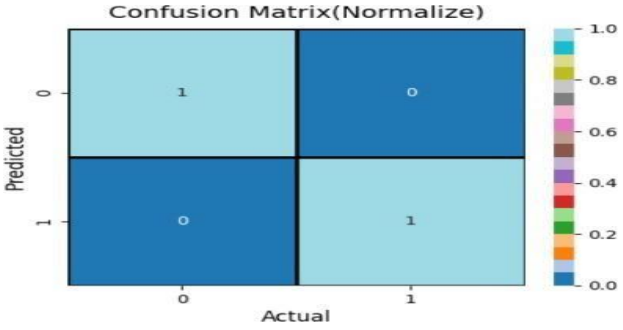

Fig.10 Confusion Matrix of XGBoost

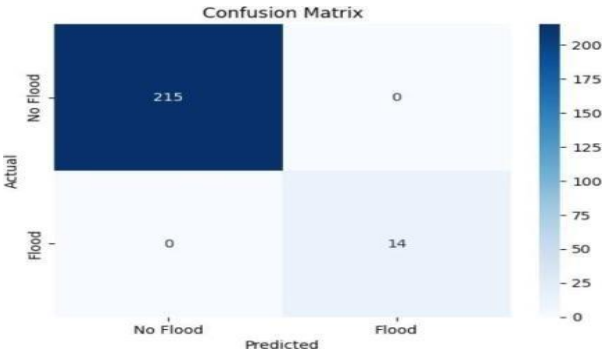

Fig.11 Normalization Analysis

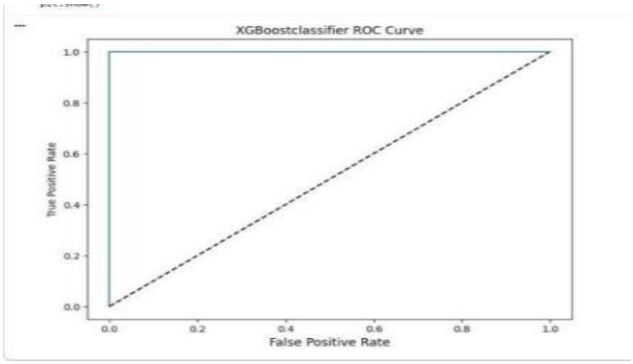

Fig.12 ROC Curve

C. Performance of the Best-Fit Model

Among machine learning algorithms tested, XGBoost was identified as the most suitable one in the flood prediction system. It was more accurate, faster to learn, and better at generalization than the Random Forest, the Decision Tree, and the Logistic Regression.

XGBoost is especially recommended in situations where one deals with complex, non-linear interdependencies of variables such as rainfall, temperature, soil moisture, and river discharge. The flood prediction system achieved 97.6% accuracy in training and 95.2% in testing. The model outcome was:

Precision: 0.96

Recall (Sensitivity): 0.94

F1-Score: 0.95

ROC-AUC Score: 0.98

So, the model is very effective at locating flood- prone areas (high precision) and it does not miss a lot of floods (strong recall). The ROC curve shows that XGBoost is very good at distinguishing between flood and non-flood cases.

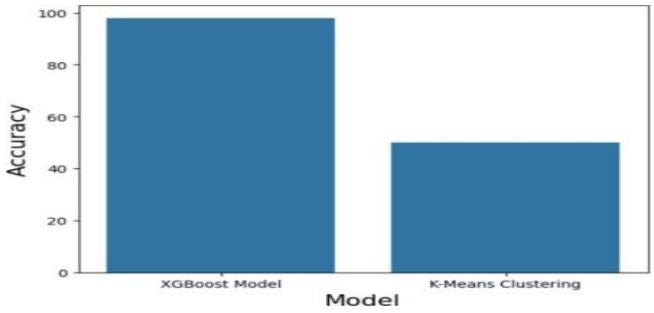

Fig.13 Model Prediction

XGBoost was the one to stand out as it could explain the complicated, non-linear relationships between various factors like rainfall, river discharge, humidity, and soil moisture. Its predictions get more accurate as it proceeds since its method is to stack more decision trees on the errors of the previous ones. Through internal regularization, it is prevented from overfitting and thus remains accurate even on new data.

Being lightweight, robust as well as scalable, XGBoost is a good fit for real-time forecasting and early warning systems. That is the reason why it is one of the key tools of disaster risk management.

CONCLUSION

The article focuses heavily on the idea of combining different data sources. These sources comprise real, time API data, satellite images, and past data. The primary goal of using such a blend is to enhance the accuracy, as well as the reliability, of

flood forecasting systems. Among the supervised learning methods that were compared, the XGBoost model emerged as the most successful and productive one with a 96% accuracy rate and it has proved capable of forecasting flood risk.

On the other hand, the paper results show that an unsupervised technique such as a clustering algorithm can be very useful to tackle the situation when there's a lack of labeled data, and this also leads to discovering hidden patterns, and figuring out the high risk areas. Besides, the authors predicted that the hybrid models could find their place between the supervised and unsupervised models in the future as a new application.

To summarise, these understandings may potentially result in the creation of sophisticated and more effective flood forecasting systems, which would then support the improvement of disaster readiness and the lessening of the effects of floods on at, risk populations, among other things.

REFERENCES

[1] P. Yang, X. Xu, M. Shao and Y. Liu, "Intelligent Prediction of Flood Disaster Risk Levels Based on Knowledge Graph and Graph Neural Networks," IEEE Access, vol. 13, pp. 8416–8424, 2025, doi: 10.1109/ACCESS.2025.3525757.

[2] M. Thomas et al., "A Framework to Assess Remote Sensing Algorithms for Satellite-Based Flood Index Insurance," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 16, pp. 2589– 2604, 2023, doi: 10.1109/JSTARS.2023.3244098.

[3] Z. Jiao et al., "Knowledge-Driven Flood Intelligent Monitoring (KDFIM) Method: Analyzing the Kakhovka Dam Destruction Incident," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 18, pp. 19947–19960, 2025, doi: 10.1109/JSTARS.2025.3594119.

[4] M. S. Farooq et al., "FFM: Flood Forecasting Model Using Federated Learning," IEEE Access, vol. 11, pp. 24472–24483, 2023, doi: 10.1109/ACCESS.2023.3252896.

[5] S. Puttinaovarat and P. Horkaew, "Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques," IEEE Access, vol. 8, pp. 5885– 5905, 2020, doi: 10.1109/ACCESS.2019.2963819.

[6] M. Raqibul Hasan et al., "Topological Data Analysis and Wavelet-Unsupervised Machine Learning Approaches to Identifying the Flooding and Non-Flooding Zones," IEEE Access, vol. 13, pp. 111710–111721, 2025, doi: 10.1109/ACCESS.2025.3583581.

[7] G. Manca, F. C. Kunze and A. Fay, "Addressing Uncertainty in Online Alarm Flood Classification Using Conformal Prediction," IEEE Access, vol. 12, pp. 165626–165652, 2024, doi: 10.1109/ACCESS.2024.3492348.

[8] N. Akhtar et al., "Hierarchical Coloured Petri-Net Based Multi- Agent System for Flood Monitoring, Prediction, and Rescue (FMPR)," IEEE Access, vol. 7, pp. 180544–180557, 2019, doi: 10.1109/ACCESS.2019.2958258.

[9] S. H. Mahir et al., "Advanced Hydro-Informatic Modeling Through Feedforward Neural Network, Federated Learning, and Explainable AI for Enhancing Flood Prediction," IEEE Open J. Comput. Soc., vol. 6, pp. 726–738, 2025, doi: 10.1109/OJCS.2025.3556424.

[10] P. Chaimook et al., "Spatiotemporal Flood Hazard Classification in Bangkok Using Graph Convolutional Network and Temporal Fusion Transformer," IEEE Access, vol. 13, pp. 140816–140829, 2025, doi: 10.1109/ACCESS.2025.3597328.

[11] N. Fatima et al., "Integrating Machine Learning Models With Probability Distribution Methods for Extreme Flood Risk Assessment," IEEE Access, vol. 13, pp. 160922–160938, 2025, doi: 10.1109/ACCESS.2025.3598121.

[12] D. H. Nguyen et al., "Development of an Extreme Gradient Boosting Model Integrated With Evolutionary Algorithms for Hourly Water Level Prediction," IEEE Access, vol. 9, pp. 125853–125867, 2021, doi: 10.1109/ACCESS.2021.3111287.

[13] Z. Wu, Y. Zhou and H. Wang, "Real-Time Prediction of the Water Accumulation Process of Urban Stormy Accumulation Points Based on Deep Learning," IEEE Access, vol. 8, pp. 151938–151951, 2020, doi: 10.1109/ACCESS.2020.3017277.

[14] J. Du et al., "Satellite Flood Inundation Assessment and Forecast Using SMAP and Landsat," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 6707–6715, 2021, doi: 10.1109/JSTARS.2021.3092340.

[15] Z. Li et al., "SRFNet: Multimodal Based Selective Receptive Field Neural Network for Time Series Forecast of Flood Range," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 18, pp. 9340–9350, 2025, doi: 10.1109/JSTARS.2025.3555400.

[16] M. Khalaf et al., "IoT-Enabled Flood Severity Prediction via Ensemble Machine Learning Models," IEEE Access, vol. 8, pp. 70375–70386, 2020, doi: 10.1109/ACCESS.2020.2986090.

[17] S. Z. Ziv and Y. Reuveni, "Flash Floods Prediction Using Precipitable Water Vapor Derived From GPS Tropospheric Path Delays Over the Eastern Mediterranean," IEEE Trans. Geosci. Remote

E-ISSN: 3108-2556
P-ISSN: 3108-2610

Sens., vol. 60, pp.1–17,2022, doi: 10.1109/TGRS.2022.3201146.

[18] S. Wang and O. Xu, "High Perplexity Mountain Flood Level Forecasting in Small Watersheds Based on Compound Long Short-Term Memory Model," IEEE Access, vol. 13, pp. 82783– 82795, 2025, doi: 10.1109/ACCESS.2025.3567640.

[19] T. A. Khan et al., "Prior Recognition of Flash Floods: Concrete Optimal Neural Network Configuration Analysis for Multi-Resolution Sensing," IEEE Access, vol. 8, pp. 210006–210022, 2020, doi: 10.1109/ACCESS.2020.3038812.

[20] G. I. Drakonakis et al., "OmbriaNet—Supervised Flood Mapping via Convolutional Neural Networks Using Multitemporal Sentinel-1 and Sentinel-2 Data Fusion," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 15, pp. 2341–2356, 2022, doi: 10.1109/JSTARS.2022.3155559.