INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

Flood Risk Prediction Using Reservoir and Rainfall Data Based on Machine Learning Techniques

[1]D Hemavathi, [2]Akkinapalli Nirmal Kumar, [2]Jeevan Sarasram SS

[1]Associate Professor Department of Data Science and Business Systems, School of Computing SRM Institute of Science and, Technology Kattankulathur, Chennai

[2]Department of Data Science and Business Systems, SRM Institute of Science and Technology Kattankulathur, Chennai

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150800015

Received: 16 August 2026; Accepted: 21 August 2026; Published: 03 September 2026

ABSTRACT

In the recent years, the use of machine learning approaches has shown promising results in enhancing the processes of environmental risk assessment and disaster prediction models. The main of this aim of this research study is to develop an automated flood risk prediction model with the help of historical rainfall data and the level of the reservoir storage. The proposed system will use machine learning models such as Logistic Regression, Decision Trees, and Random Forest to predict the flood risk levels based on hydrological and meteorological parameters. A formatted dataset containing the historical rainfall and water level of the reservoir data was employed to train and test the prediction models, and data preprocessing and feature extraction methods were employed to improve the accuracy of the prediction models. The experimental outcome reveals that tree-based ensemble models are more accurate in predicting the flood risk levels compared to the existing statistical models, especially when dealing with imbalanced datasets. The proposed system validates the use of machine learning analytics in disaster management by developing an automated decision support system for the early detection of flood-prone situations.

Index Terms: Flood risk prediction, machine learning, rainfall data, reservoir levels, classification models, disaster management.

INTRODUCTION

N recent years, the rising number of flood disasters that occurrences with increased intensity has created serious concerns for public safety, infrastructure, and economic security, and because of these serious concerns, early flood risk prediction has become an integral part of efficient disaster management. Floods caused by high rainfall intensity and high water levels in the reservoir are some of the most common cause for triggering natural hazards like floods, especially in areas where monsoon climate conditions prevail. Without early risk prediction, these situations can cause massive damage to property, essential services, and human life, thus making it very crucial to have accurate and reliable flood prediction systems. Rainfall intensity and water levels in the reservoir are one among the most important factors that determine flood occurrence, hence monitoring and analysing these factors are essential in minimizing flood effects.

Conventional techniques that are being used for flood risk evaluation are mainly based on hydrological models, past data, and manual analysis of rainfall levels and reservoir levels. Even though these techniques have been proven to be useful, they are always considered to be very time consuming, needs higher amount of resources and they are not easily scalable for larger or remotely located areas where data availability may be a constraint. Due to the increase in need for an effective and flexible solution for flood monitoring, the machine learning algorithms have started to become a rising and promising area of research due to their ability to process the complex environmental data and detect the risk patterns present in the data.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

In this research work, machine learning methods are used to predict and classify flood risk levels using past data of rainfall and reservoir water levels. The models that were used in this study are Logistic Regression, Decision Tree and Random Forest. Logistic Regression is taken as the base for the model because it is simple and easy to understand. Decision Tree is used to handle non linear relationships between different environmental factors. Random Forest is selected as the main model since it can handle the data with better and works well with structured datasets. The dataset was created with the help of collecting historical rainfall levels and reservoir level information. Before training the models, data preprocessing and feature engineering were performed like splitting the dataset into training and testing subsets to improve the overall performance of the models. In between the normalization of the rainfall level and reservoir level values are done. After training, the models were evaluated using standard evaluation metrics. The final models were then connected to a Flask backend, which allows the integration of a web interface. Through this interface, users can enter environmental inputs and get instant flood risk predictions. This system shows how machine learning can be practically used to support disaster management by helping experts assess flood risks faster and more efficiently.

LITERATURE REVIEW

The use of data driver approaches for flood risk prediction have been significantly increased in recent years, the main reason for this is especially due to the increase in extreme rainfall events which cause floods near civilized areas. It causes serious damage to people and the public properties and infrastructure. Floods are among the most destructive natural disasters, and the early warning of floods are mostly inaccurate as the behavior of the nature is unpredictable, the absence of high quality of data, and the time taken by traditional monitoring systems. In this respect, several studies have been made to explore the use of machine learning algorithms as an alternative approach to improve the accuracy of flood risk prediction based on the available environmental variables like rainfall intensity and water levels in the reservoir.[1], [2].

In the beginning or initial stages of predicting the forecast floods, scientists mainly relied on the physical and statistical approaches to relate rainfall intensities and response of reservoirs.[3] Also the earlier versions of flood predictions are very difficult to adapt by only using the traditional methods alone. It was a good foundation, but it had its flaws, such as oversimplification, a higher reliance on the parameters and they are not flexibile enough so they were painful when conditions changed. As climate change and land use patterns began to alter the flood process in several ways and due to this the traditional methods didnt respond well with their assumptions, scientists turned to data models that could recognize complex patterns instead.

Certain studies have also emphasized the significance of joint use of rainfall and reservoir-related variables for flood forecasting [4], [5]. Machine learning-based models have also been found to be able to learn patterns from multi-variable datasets, especially when the patterns are non-linear in nature, and when traditional models are unable to learn these patterns. Certain review articles have also mentioned that machine learning-based models are generally able to provide better flexibility and forecasting capabilities than traditional flood forecasting models, if proper feature selection and model design are employed [6], [7]. However, these studies also mention that the performance of these models is still highly dependent on data.

However, aside from the accuracy of the predictions, the use of decision support systems is to manage the flood risks and that has also been widely studied in the current literature. The combination of predictive results and advisory data has been found to provide value to planning and response operations by providing the opportunity for more informed decision-making during flood events [8], [9]. On the other hand, when it comes to the relation to the effect of climate change, there are findings of increased rainfall variability and flood risk, thus this assures the need for adaptive and data driven predictive systems that can adapt to the changing environment [10].

However, most of the current flood prediction models are based on complex modeling paradigms or it will require a lot of data, which may not be very useful in regions that do not have the required infrastructure. It is with this background that this paper proposes to develop a flood risk prediction model based on rainfall and reservoir data using machine learning models that are accurate and simple. This paper hopes to fill the gap that currently exists between the existing research and the practical management of flood risks.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

METHODOLOGY

This research work is purely based on an analytical model framework that uses machine learning concepts to determine the flood risks based on the historical rainfall data and water levels in the reservoir. The research framework is mainly designed to address the problems of delayed warnings, lack of monitoring in flood-prone regions, and the need for accurate results in flood risk analysis. Fig. 1 represents the architecture of the research work which is categorized into stages, which include the processing of the dataset, data preprocessing, feature extraction, model creation, training, and evaluation, as well as the integration of flood risk prediction.

Fig. 1. System architecture diagram

Dataset Preparation

The dataset employed for the analysis consists of historical records of rainfall data and water level indicators of the reservoir that are linked with occurrences of flood. The dataset was grouped into separate groups based on the flood risk. The dataset was obtained from publicly accessible sources that are linked with hydrological and meteorological conditions and was structured to ensure uniformity. In consideration of the imbalanced distribution of the dataset between flood and non-flood occurrences, there was a focus on ensuring the representation of the minority class in the dataset. The dataset was split into a test and training dataset with an 80-20 composition.

Data Preprocessing and Feature Engineering

Prior to model training, raw data were subjected to preprocessing steps to improve data quality and compatibility across features. Numerical attributes such as rainfall levels and reservoir level percentage were normalized to reduce scaling related issues during learning. In addition to raw measurements, derived features including rainfall intensity categories and normalized reservoir stress indicators were generated to capture the underlying flood risk patterns more accurately. These feature transformations were designed to enhance the model sensitivity to risk related variations.

Model Development

Three supervised machine learning models were created and tested to predict the risk levels of the floods. Logistic Regression was used as the benchmark model as it is a simple model and well known for its ease of interpretation easy understanding of linear correlations. For nonlinear relationships between rainfall cycles and reservoir attributes, Decision Tree were used. Random Forest was chosen as the primary model as it is the combination of many trees and also due to its ability to easily handle data with noise and its strong performance in generalization on structured tabular data.

Training Procedure

The dataset is divided into training and testing sets. The model was then trained using the training dataset that had been prepared with the help of historical environmental data and then tested on the unseen data to determine its prediction ability of the flood risks. The parameters were selected based on experience to ensure

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

that the complexity and to also reduce the chance of over fitting. The training process also ensures that the learning behaviour is stable across the training cycles.

Performance Evaluation

The trained models were evaluated using the performance metrics like accuracy, precision, recall and F1score to provide a clean and proper assessment of the classification performance of the respective model which predicts the flood risks. The imbalanced and unpredictable nature of the flood related datasets are given special attention calculation the recall and F1score to measure the ability of the models to correctly identify flood risk cases. By doing comparative analysis we came to know that the tree based models offered better and improved sensitivity to flood conditions than when compared to baseline statistical approaches. Based on these evaluation results, Random Forest was identified as the most reliable model for integrating and deploying the prediction system.

Deployment

After the completion of the training and evaluation phase, the selected model was deployed using a Flask based application programming interface. Using the features of Flask, the setup will provide a lightweight, flexible and more stable backend for processing and handling the flood risk predictions. A web interface is created to connect to the backend so that the users can easily enter the rainfall values and reservoir water level information in the web interface. Also the flood risk prediction logic were integrated in the web interface. Based on the given inputs, the system returns the predicted flood risk level along with its confidence value. This design is made in a way so that the users can easily access the application with ease and doesn't need much of the software installation procedures.

Risk Advisory

In an attempt to make this application more relevant, this system doesn't stop after only doing prediction alone. After predicting the flood levels the results are converted into practical guidance where the users will get advisory messages and warnings on what to do in this situation. Different practical actions are suggested depending upon the different risk levels. High risk cases will mainly focus on citizen's safety and emergency readiness whereas, Moderate risks will focus on monitoring and preparation. Low risk prediction will just indicate stable conditions. Because of these features, it is easy for the authorities and flood management team to be ready for emergencies.

RESULTS AND DISCUSSIONS

Model Performance Evaluation

Three machine learning models are used mainly for the comparison of the model performance. They are Logistic Regression, Decision Tree, and Random Forest. With the help of the historical data of the rainfall measurements and reservoir water level the models are trained and evaluated carefully. The performance of each model was calculated using standard classification metrics which consists of accuracy, precision, recall and F1score to understand their how effective they are when it comes to identifying flood risk under unpredictable and variable conditions.

Table I: Model Comparison Results

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	72.31	0	0	0
Decision Tree	60.98	29.79	30.17	29.98
Random Forest	60.74	30.2	31.9	31.03

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

After testing the trained models, these are all the results that were recorded. From this table it indicates that the Logistic Regression got the highest amount of accuracy but when it comes to precision, recall and F1score it returns zero values. But when it comes to Decision Tree and Random Forest it also manages to balance other metrics and performs better than Logistic Regression. Among these Random Forest achieves the best balance between every metrics. Hence, Random Forest was selected as the most reliable model.

Flood Risk Prediction Results

The trained model was then tested using the real input scenarios using different rainfall level and reservoir level data by uploading them in the web interface. Fig. 2 presents a high risk prediction case were recorded with extreme rainfall conditions of approximately 500 mm and reservoir level of 65% and resulted in a flood risk probability of 79%. This indicates that the chance of floods are very high which requires immediate actions like evacuation.

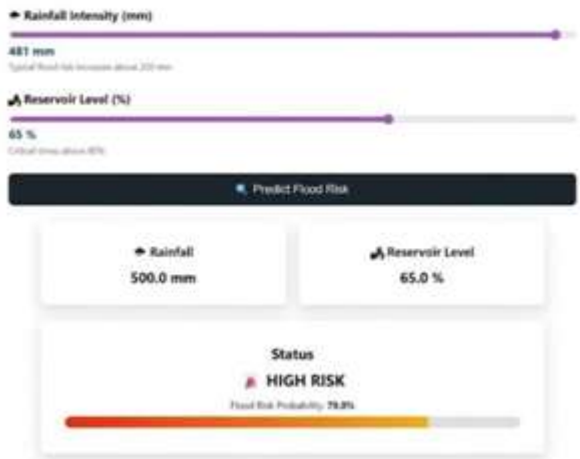

Fig. 2. High Flood Risk Prediction Output

Fig. 3 shows a low risk prediction scenario where moderate rainfall is recorded with approximately 125 mm and a low reservoir level of 17%. The model classified this case as low risk with a probability of 16%, this indicates that the hydrological conditions are stable and there wont be much of any infrastructural damage or any other emergencies.

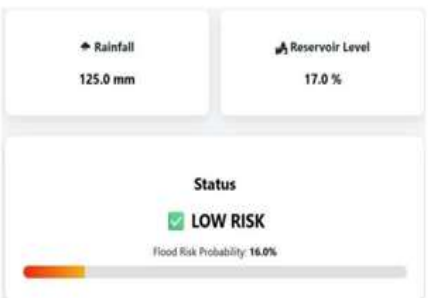

Fig. 3. Low Flood Risk Prediction Output

Fig. 4 shows a moderate risk scenario, where the rainfall is recorded with approximately 120 mm and a reservoir level of 65% produced a flood risk probability of 61%. This case indicates that there is not much of an emergency but it might lead to the cause of floods. So it requires constant monitoring.

Risk Advisory Integration

To increase and enhance the usability of the prediction system the risk advisory module is being integrated to the mechanism. Based on the predicted flood risk levels results, the system will provide the corresponding survival recommendations to the citizens present in the predicted area. High risk predictions suggests for

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

immediate evacuation. Moderate risk predictions suggest continuous monitoring of the environmental behavior, and when it comes to Low risk predictions, it suggests for normal and stable monitoring of the climate and weather. This transforms the system into

Fig. 4. Moderate Flood Risk Prediction Output a combination of both prediction and decision supporting platform.

DISCUSSION

From the experimental results it indicates that the tree based machine learning models are more effective when compared to simple statistical approaches for flood risk prediction. While the logistic regression provides a useful base level, but it is still not able to detect the minority flood cases hence limiting its practical applicability. Random Forest showed improved sensitivity to flood conditions by capturing non linear relationships between rainfall and reservoir levels which results in more reliable risk classification.

From the results it also confirms that the prediction model responds logically to changes that occur in the environmental parameters and assigns high risk to scenarios which involves intense rainfall and higher reservoir levels. The model combines probabilistic prediction with respective advisory outputs. Hence the system supports better flood management decisions.

Table II: Flood Risk Prediction Scenarios

Rain (mm)	Reservoir (%)	Prob (%)	predicted
500	65	79	High
125	17	16	Low
120	65	61	Moderate

The Random Forest model predicts the flood risk levels based on the rainfall and reservoir level conditions that often vary in different locations. When the rainfall is extremely high of (500 mm) a moderate reservoir level of(65%), the model predicts it as a high flood risk (79%) and it indicates that even intense rainfall is enough to trigger floods. But when it comes to another result where the moderate rainfall of (125 mm) with a low reservoir level (17%) it predicts and results in a low flood risk (16%) this indicates that sufficient storage capacity of reservoirs reduces the chance of floods to occur. Also when moderate rainfall (120 mm) combines with a higher reservoir level (65%), the model predicts it as a moderate flood risk (61%), this indicates and highlights the fact that the combined effect of rainfall and reservoir storage on flood occurrence.

LIMITATIONS AND FUTURE WORK

Although the proposed system of flood prediction demonstrates effectively and predicts the flood risk there are still few limitations in the system. The dataset which is being used restricts the exposure to different and rare extreme flood causing scenarios. This might affect the generalization that are under common conditions. When it comes to food related data the class imbalance can create bias towards the dominant classes.

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

Future work may mainly focus on expanding the data coverage by collecting data from multiple regions and seasons and also integrate real time rainfall and reservoir monitoring sources to know the immediate change in the trends of the climate and weather.

CONCLUSION AND FUTURE WORK

This research shows the utilization of machine learning techniques to predict flood risk by analyzing data of the reservoir water levels and rainfall level intensity. We used three classification models such as Random Forest, Decision Tree, and Logistic Regression to see how well they could find patterns in structured data about flooding. We also tried to use few tree based models. In then then we came to the con-clusion that, Random Forest worked better than other simpler statistical methods in predicting the floods more effectively.

Another outcome of this study is that the development of a user centric flood risk prediction system. Users can give the user input of the rainfall intensity levels and the level of the reservoir, and the application quickly gives them a flood risk percentage, a probability score, and some advice. When the prediction results got combined with clear instructions we can get the better results. This will help the people make smart choices about how to be aware of floods, mostly in places where it is hard to get advanced monitoring systems.

This app makes it easier to understand the results of flood risk assessments. This helps the users to be aware of their respective region's environmental changes in hydrology. A simple advisory system and machine learning can work together to link old ways of measuring water flow with new systems that help people make decisions based on facts. This gives you more information and helps you stop floods more quickly.

Although the suggested system functions well, it could be improved in a few ways:

Dataset Expansion : If we can able to get the real data from any organization it will be more useful .Also by including more regional records, real data, like last 100 years data, and infrequent extreme flood events, the model's robustness and dependability under a variety of circumstances would be improved.

Model Enhancement: By using other tree-based models also we can increase the accuracy and efficiency.

Real-time Integration: The applications would be enhanced by using the data of real-time rainfall and reservoir monitoring data, which would enable continuous flood risk assessment.

Operational Validation: By testing the application helps in many ways like to update , debugging etc.

Additional Environmental Inputs: Weather forecasts data of short-term make the model more accurate.

Scalable Deployment: By deploying the application in cloud technologies like AWS, Azure many people can use this and get the risk prediction easily.

REFERENCES

1. B. Merz, A. H. Thieken, and M. Gocht, "Flood risk mapping at the local scale: Concepts and challenges," Flood Risk Management, vol. 1, no. 4, pp. 223–236,
2. 2008.
3. E. J. Plate, "Flood risk and flood management," Journal of Hydrology, vol. 267, no. 1–2, pp. 2–11, 2002.
4. J. C. Refsgaard and J. Knudsen, "Operational flood forecasting systems," Hydrological Sciences Journal, vol. 41, no. 4, pp. 581–601, 1996.
5. D. Feng, J. Zhang, and H. Liu, "Flood prediction using rainfall and reservoir data," Water Resources Research, vol. 45, no. 8, pp. 1–12, 2009.
6. J. Abbot and J. Marohasy, "Input selection and optimization for rainfall-runoff models," Environmental Modelling & Software, vol. 25, no. 8, pp. 1005–1015, 2010.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VIII, August 2026

7. A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: A review," Water, vol. 10, no. 11, p. 1536, 2018.
8. Y. Zhou, L. Guo, and H. Chen, "Machine learning-based flood risk prediction using hydrological data," Applied Sciences, vol. 9, no. 23, p. 5125, 2019.
9. M. Sit and I. Demir, "Decision support systems for flood risk management," Environmental Modelling & Software, vol. 26, no. 11, pp. 1414–1424, 2011.
10. R. J. Dawson, "Decision-support tools for flood risk assessment," Natural Hazards, vol. 30, pp. 239–252, 2003.
11. T. Wahl, S. Jain, and J. Bender, "Increasing flood risk due to climate change," Nature Climate Change, vol. 5, pp. 109–114, 2015.