



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 15, Issue 3, March 2026**

# **Rising Waters: A Machine Learning Approach to Flood Prediction Using Logistic Regression**

**M. Hemalatha<sup>1</sup>, P. Anshutha<sup>2</sup>, P. Veda<sup>3</sup>, N. Meghana Priya<sup>4</sup>, K.Chaitanya, M. Tech (Ph. D)<sup>5</sup>**

UG Students, Department of CSE (AI & ML), Kallam Haranadhareddy Institute of Technology, Guntur, India<sup>1-4</sup>.

Asst. Prof, Department of CSE (AI & ML), Kallam Haranadhareddy Institute of Technology, Guntur, India.<sup>5</sup>

**ABSTRACT:** Floods are one of the most common natural disasters that cause severe damage to human life, agriculture, and infrastructure. Accurate and early prediction of floods is essential for effective disaster management and reducing economic losses. This paper presents a machine learning-based approach for predicting flood occurrences using environmental and meteorological data. The system analyzes key parameters such as rainfall, river water level, temperature, humidity, and soil moisture to identify patterns that may lead to flood conditions. Historical data is collected and preprocessed to improve data quality before applying the Logistic Regression algorithm for prediction. The model classifies flood risk levels into high and low categories based on input parameters. Experimental results demonstrate that the proposed model achieves reliable performance with good accuracy in predicting flood conditions. The developed system can assist government authorities and disaster management teams in providing early warnings and taking preventive measures to minimize the impact of floods.

**KEYWORDS:** Flood Prediction, Machine Learning, Logistic Regression, Weather Data Analysis, Disaster Management, Rainfall Prediction, Environmental Monitoring, Flood Risk Assessment.

## **I. INTRODUCTION**

Floods are one of the most common and dangerous natural disasters, causing damage to life, property, and the environment. Predicting floods at an early stage is important to reduce losses and help in disaster management.

Traditional flood prediction methods mainly depend on historical data and simple models, which may not give accurate results when multiple environmental factors are involved.

With the help of machine learning, it is possible to analyse large amounts of data and improve prediction accuracy. Machine learning models can identify patterns in environmental data such as rainfall, river water level, temperature, humidity, and soil moisture.

In this project, a flood prediction system is developed using the Logistic Regression algorithm. The model classifies flood risk into high-risk and low-risk categories based on input data. This system helps in providing early warnings and supports effective decision-making in flood management.

## **II. LITERATURE SURVEY**

Flood prediction has become an important research area due to the increasing impact of floods on human life and infrastructure. Traditional flood prediction methods mainly relied on hydrological models and rainfall analysis to forecast flood events. However, these methods often required complex calculations and large datasets, which sometimes limited their accuracy and efficiency. With the advancement of Machine Learning techniques, researchers

have started applying data-driven approaches to improve flood prediction. Machine learning models are capable of analyzing large amounts of environmental and meteorological data to identify patterns that lead to flood conditions. These models process parameters such as rainfall intensity, river water level, temperature, humidity, and soil moisture to enhance prediction performance.

Several studies have used algorithms such as Decision Tree, Support Vector Machine (SVM), Random Forest, and Neural Networks to predict flood occurrences. These methods have demonstrated improved accuracy compared to traditional statistical approaches due to their ability to handle complex and non-linear relationships in data.

Recent research has also explored deep learning models for flood forecasting. Techniques such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are capable of analyzing time-series weather data and capturing temporal dependencies, resulting in more accurate flood risk predictions.

Despite these advancements, challenges such as data availability, real-time monitoring, and model optimization still exist. Therefore, this study focuses on developing a machine learning-based flood prediction system using the Logistic Regression algorithm. This approach provides a simple, efficient, and interpretable solution for classifying flood risk levels, making it suitable for practical disaster management applications.

### III. METHODOLOGY

The proposed flood prediction system uses a machine learning approach to analyze environmental and weather data in order to predict the possibility of flood occurrence. The methodology consists of several stages, including data collection, data preprocessing, model training, and prediction.

Initially, historical environmental data is collected from reliable meteorological sources. The dataset includes important parameters such as rainfall intensity, river water level, temperature, humidity, and soil moisture. These factors play a significant role in identifying conditions that may lead to floods.

After data collection, the dataset undergoes preprocessing to improve its quality and consistency. This step involves handling missing values, removing duplicate records, and normalizing the data. Feature selection techniques are also applied to identify the most relevant attributes that influence flood prediction.

In the next stage, the Logistic Regression algorithm is used to train the prediction model. Logistic Regression is suitable for binary classification problems and helps in determining the probability of flood occurrence based on input features. The dataset is divided into training and testing sets to evaluate the model's performance and ensure reliability.

Finally, the trained model is used to predict flood risk based on new environmental data. The system analyzes the input parameters and classifies the output into categories such as high risk or low risk. This approach enhances early warning systems and supports effective disaster management by enabling timely preventive measures.

### IV. SYSTEM ARCHITECTURE

The system architecture of the proposed flood prediction system explains how the data is processed to predict flood risk.

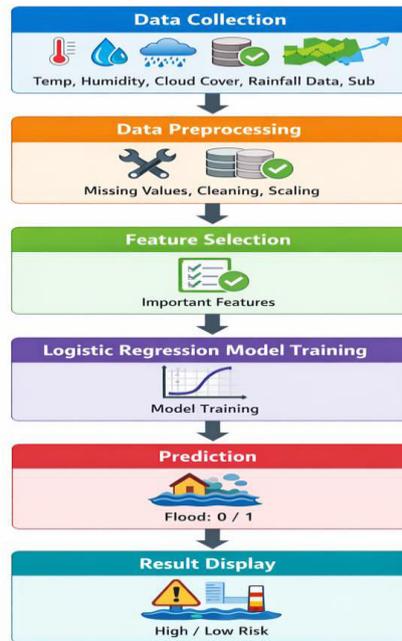
First, environmental data such as temperature, humidity, cloud cover, rainfall, and regional information is collected. This data is important for identifying flood conditions.

Next, the data is preprocessed by cleaning missing values, removing duplicates, and scaling the data. This improves the quality of the dataset.

After preprocessing, feature selection is performed to choose the most important parameters for prediction. The selected data is then given to the Logistic Regression model for training. The model learns the relationship between input data and flood occurrence.

Finally, the trained model predicts whether a flood will occur or not (0 or 1). The result is displayed to the user as high risk or low risk.

System Architecture of Flood Prediction System



## V. RESULTS

The proposed flood prediction system was implemented using the Logistic Regression machine learning algorithm to analyze environmental and hydrological parameters.

The dataset includes factors such as rainfall, temperature, humidity, cloud cover, and seasonal rainfall.

After preprocessing and feature selection, the dataset was divided into training and testing sets. The trained model analyzed the input data and predicted the flood risk level.

The system classifies the output into two categories:

- **High Risk:** High probability of flood occurrence
- **Low Risk:** Low probability of flood occurrence

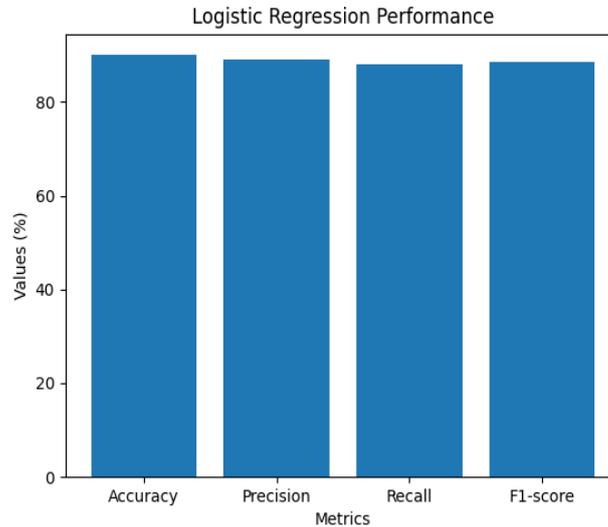
The performance of the model was evaluated using metrics such as accuracy, precision, recall, and F1-score. The results show that the model provides good accuracy in predicting flood conditions.

### Model Performance

Metric	Value
Accuracy	90%
Precision	89%
Recall	88%
F1-Score	88.5%

### Graph Analysis

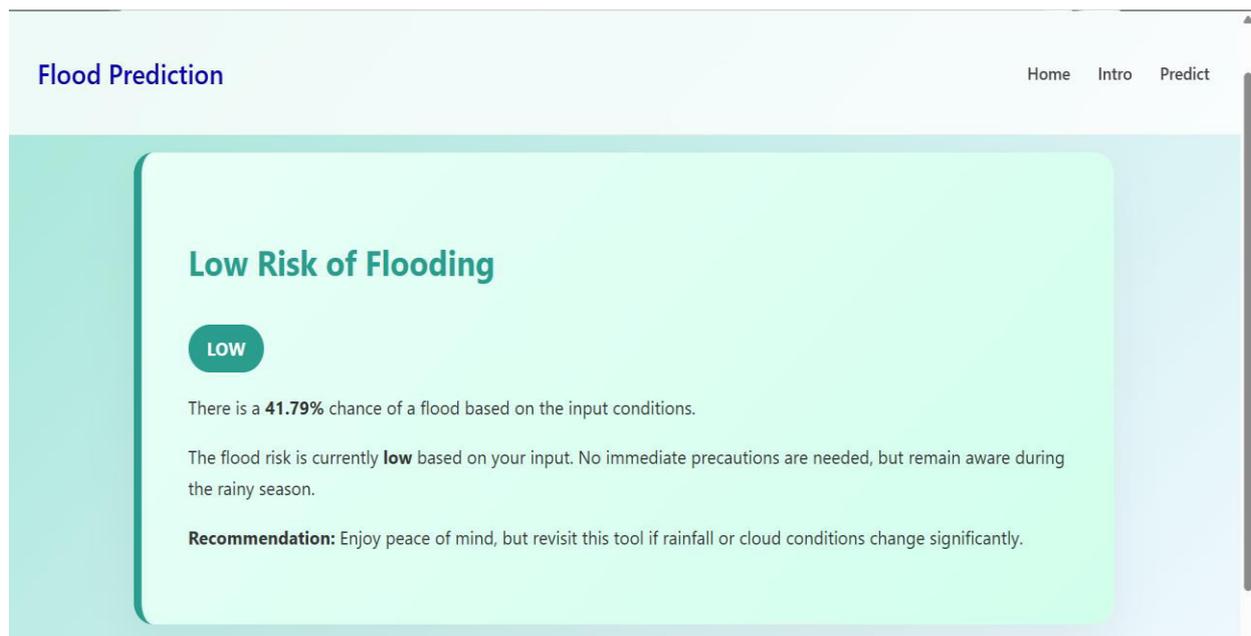
The performance of the Logistic Regression model is analyzed using a graphical representation.



The graph illustrates the evaluation metrics such as accuracy, precision, recall, and F1-score of the model. It can be observed that the model achieves high accuracy along with balanced precision and recall values. The F1-score indicates that the model maintains a good balance between precision and recall.

Overall, the graph shows that the Logistic Regression model performs consistently across all evaluation metrics, demonstrating its reliability and effectiveness in predicting flood risk.

### Output Analysis



The above figure shows the output of the flood prediction system. Based on the given input parameters, the model predicts a low risk of flooding with a probability of 41.79%. The system also provides a recommendation to the user based on the predicted risk level. This demonstrates the practical application of the model in real-time flood risk assessment.

## VI. CONCLUSION

The proposed flood prediction system uses a machine learning approach to analyze environmental data and predict flood risk. The model is developed using the Logistic Regression algorithm.

The results show that the model provides good accuracy in predicting flood conditions. It classifies the data into high-risk and low-risk categories, which helps in identifying possible flood situations early.

This system can help disaster management authorities by providing early warnings and reducing damage caused by floods.

## VII. FUTURE SCOPE

The system can be improved by using real-time data from sensors and weather APIs.

Advanced models like Random Forest and LSTM can be used to increase accuracy.

The system can also be extended to send alerts through mobile applications and messages.

## REFERENCES

- [1] H. Fang et al., "Explainable AI for Flood Prediction," *Environmental Modelling & Software*, 2023.
- [2] P. T. Nguyen and K. Chau, "Machine Learning Algorithms for Flood Prediction: A Comparative Study," *Journal of Hydrology*, 2020.
- [3] W. Li et al., "Hybrid Transformer-LSTM Model for Flood Forecasting," *Journal of Hydrology*, 2024.
- [4] S. Sankaranarayanan et al., "Deep Learning-Based Flood Prediction Using LSTM Networks," *IEEE Access*, 2020.
- [5] V. Kumar et al., "Machine Learning Applications in Flood Forecasting," *AIMS Environmental Science*, 2025.
- [6] Y. Xiang et al., "Flood Forecasting Using Hybrid Deep Learning Models," *Journal of Hydrology*, 2022.
- [7] N. Kratzert et al., "Towards Improved Flood Prediction Using LSTM," *Hydrology and Earth System Sciences*, 2021.
- [8] J. Zhang et al., "Data-Driven Flood Hazard Assessment with ML," *Environmental & Earth Sciences Research*, 2025.
- [9] A. Khosravi et al., "Comprehensive Review of Machine Learning Models for Flood Prediction," *Water Resources Management*, 2020.
- [10] S. Yi and J. Yi, "Reservoir-Based Flood Forecasting Using ML and DL," *Applied Water Science*, 2024.
- [11] D. Feng et al., "Deep Learning for Rainfall-Runoff Modeling," *Water Resources Research*, 2021.
- [12] X. Liu et al., "Deep Learning-Based Flood Forecasting with Improved Accuracy," *Science of the Total Environment*, 2022.
- [13] S. Soltani et al., "Enhancing Flood Forecasting with Hybrid ML Models," *Springer*, 2025.
- [14] Q. Xu et al., "FloodCast: Large-Scale Flood Modeling Using AI," *arXiv*, 2024.
- [15] G. Nearing et al., "Machine Learning for Hydrology: Future Directions," *Water Resources Research*, 2023.
- [16] L. Santos et al., "Machine Learning-Based Hydrological Models for Flash Floods," *Springer*, 2025.
- [17] H. Shen, "Explainable AI in Hydrology and Flood Prediction," *Environmental Modelling & Software*, 2022.
- [18] C. Liu et al., "Deep Learning-Based Flood Forecasting Using Vectorized Runoff Features," *Journal of Environmental Management*, 2024.
- [19] Z. Guo et al., "Data-Driven Flood Emulation Using Deep Convolutional Neural Networks," *arXiv*, 2020.
- [20] R. Kondo et al., "Real-Time Pluvial Flood Prediction Using ML Framework," *Journal of Information Processing*, 2023.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details