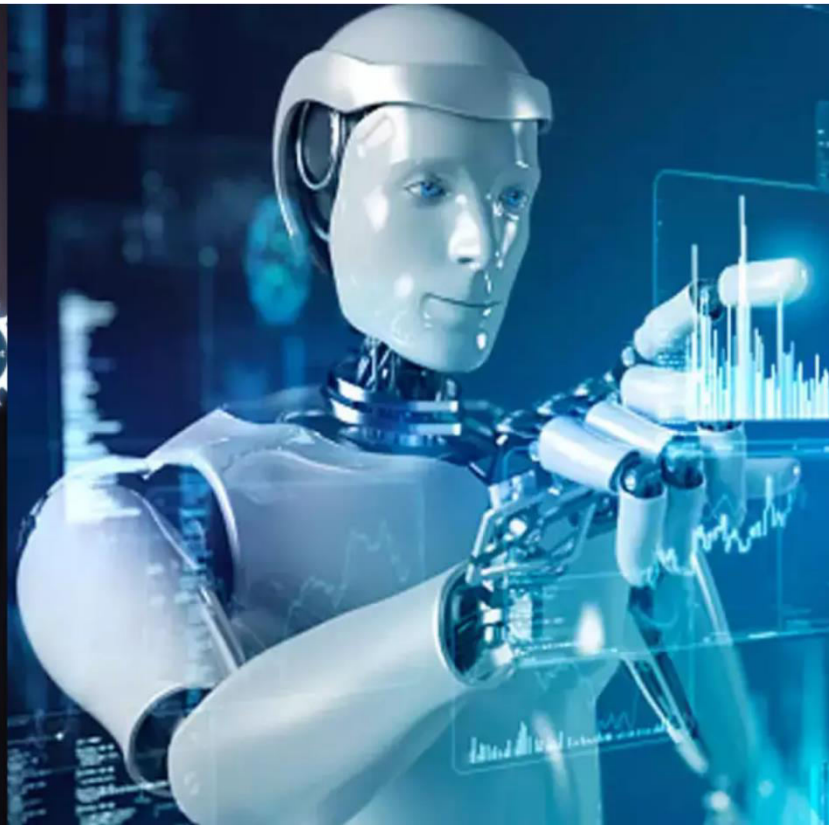
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

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

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

Rising Waters: A Machine Learning Approach to Flood Prediction using Random Forest Classification

Chintapalli Aarya[1], Madhu Kiran Reddy[1], Ch Abhinav[1], Kanaparthi Bhaskar[1], M. Ashok Naga Sai[2]

UG Students, Department of CSE (AI & ML), Kallam Haranadha Reddy Institute of Technology, Guntur, India[1]

Assistant Professor, Department of CSE (AI & ML), Kallam Haranadha Reddy Institute of Technology, Guntur, India[2]

ABSTRACT: Floods are among the most destructive natural disasters affecting communities across the world. Predicting flood conditions in advance can significantly help authorities reduce damage and improve disaster preparedness. This research proposes a machine learning-based flood prediction system that analyzes environmental parameters such as temperature, humidity, cloud cover, and rainfall to estimate flood risk. A Random Forest classification model is trained using historical weather data to identify patterns that are commonly associated with flood occurrences. The general workflow of the proposed system is illustrated in Fig. 1. The trained model is further integrated with a Flask-based web application that allows users to input weather conditions and obtain real-time predictions. The study demonstrates how machine learning techniques can be applied to environmental datasets to support intelligent flood prediction systems.

KEYWORDS: Flood Prediction, Machine Learning, Random Forest Algorithm, Climate Data Analysis, Disaster Management, Flood Risk Assessment, Weather Parameters, Environmental Data Modeling.

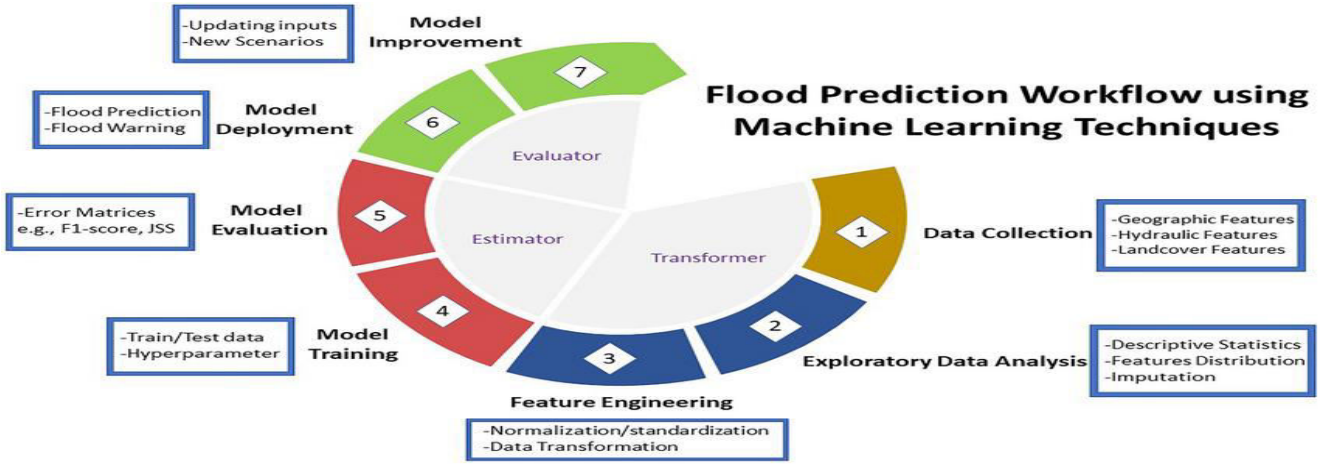


Fig. 1. Workflow of the Machine Learning Flood Prediction System

I. INTRODUCTION

Floods are among the most common and destructive natural disasters that affect communities across the world. They cause significant damage to infrastructure, agriculture, and human life, especially in regions that experience heavy rainfall and changing climatic conditions. In recent years, the increasing frequency of extreme weather events has made flood prediction an important area of research. Accurate and timely prediction of floods can help governments and disaster management authorities take preventive actions, reduce economic losses, and protect human lives.Traditional flood prediction methods mainly rely on hydrological models that analyze rainfall levels, river flow, and geographical conditions. Although these models are useful, they often require complex calculations and large amounts of

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

environmental data, making them difficult to implement in real-time systems. With the advancement of data science and artificial intelligence, machine learning techniques have emerged as a powerful alternative for analyzing environmental data and identifying patterns related to flood occurrences.

II. LITERATURE SURVEY

Flood prediction has been widely investigated by researchers using both traditional hydrological methods and modern data-driven techniques. Earlier systems relied mainly on mathematical hydrological models that simulated rainfall accumulation and river flow patterns. Although these models provide useful insights, they often require detailed environmental measurements and complex calculations.In recent years, machine learning algorithms have been applied to environmental datasets to identify patterns related to flood occurrences. Techniques such as Support Vector Machines, Decision Trees, Artificial Neural Networks, and Random Forest models have demonstrated encouraging results in analyzing climatic variables and predicting flood risk. The literature review presented in Table 1 summarizes several representative studies that applied machine learning techniques for flood prediction.

Table 1. Summary of important research works on flood prediction.

Paper Title	Authors	Method Used	Key Contribution	Limitation
Machine Learning Flood Forecasting	Smith et al.	Random Forest	Improved prediction accuracy	Needs large datasets
Rainfall Pattern Analysis	Kumar & Patel	Regression	Studied rainfall impact	Limited parameters
Flood Classification System	Chen et al.	SVM	Classified flood events	Sensitive to tuning
Neural Network Flood Study	Rahman et al.	ANN	Captured complex patterns	High computation
Hybrid Flood Model	Zhao et al.	Hybrid ML	Combined climatic variables	Model complexity

• Many recent studies emphasize the importance of using environmental datasets such as rainfall records, temperature variations, and humidity levels to predict flood events.
• Machine learning algorithms have shown strong potential in identifying hidden relationships between climatic variables and flood occurrence.
• Compared with traditional hydrological models, data-driven methods can process large datasets more efficiently and produce faster predictions.
• Researchers have also explored hybrid models that combine multiple algorithms to improve prediction accuracy.

III. DATASET DESCRIPTION

The performance of a machine learning model largely depends on the quality and relevance of the dataset used for training and evaluation. In this study, a dataset containing weather and environmental parameters related to flood conditions was utilized to train the flood prediction model. The dataset includes several important climatic attributes such as temperature, humidity, cloud cover, annual rainfall, and seasonal rainfall indicators. These environmental parameters play a significant role in influencing flood conditions.Before training the model, the dataset was carefully examined and preprocessed to ensure data quality and consistency. Missing values were handled using appropriate preprocessing techniques, and the numerical attributes were prepared for machine learning analysis. This preprocessing step improves the reliability of the trained model and ensures that the environmental variables are represented consistently. As illustrated in (**Fig. 2)**, the dataset first provides the **weather parameters**, which represent the key environmental factors affecting flood occurrence. These parameters are then subjected to a **feature selection process**, where the most relevant attributes are identified and selected for model training. Feature selection helps in reducing unnecessary data and improves the efficiency and accuracy of the prediction model.Finally, the selected features are used as inputs to the **machine learning model**, which learns the relationships between environmental conditions and

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

flood occurrences. The dataset is divided into training and testing subsets to evaluate the model's performance. The training dataset allows the model to learn patterns associated with flood risk, while the testing dataset is used to assess the model's ability to generalize to unseen data.

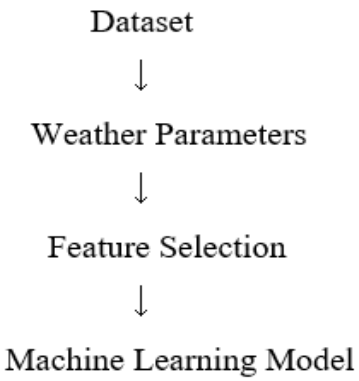


Fig. 2. **Dataset structure used for flood prediction model training**

- The dataset contains multiple environmental attributes that influence flood conditions.
- Proper preprocessing ensures that the dataset is suitable for machine learning training.
- Splitting the dataset into training and testing sets allows accurate evaluation of model performance.
- Environmental variables such as rainfall and humidity provide important indicators for flood prediction.

IV. SYSTEM ARCHITECTURE

The system architecture describes how different components of the flood prediction system interact to generate flood risk predictions. The proposed system integrates environmental data analysis, machine learning techniques, and a web-based interface to provide an efficient flood prediction solution.

As illustrated in (**Fig. 3)**, the architecture begins with the **data collection stage**, where environmental and weather-related parameters such as temperature, humidity, cloud cover, and rainfall indicators are gathered from the dataset. These parameters represent the fundamental inputs required for predicting flood occurrences.The collected data then undergoes a **data preprocessing stage**, where the dataset is cleaned and prepared for machine learning analysis. During this stage, missing values are handled, irrelevant data is removed, and the input variables are formatted to ensure consistency and reliability.

After preprocessing, the next stage is **building the machine learning model**. In this system, the **Random Forest algorithm** is used to construct the flood prediction model. The model is designed to analyze relationships between weather parameters and flood occurrences.

Once the model structure is built, the system proceeds to the **training phase**, where the machine learning model learns patterns and relationships from the training dataset. This process allows the model to understand how different environmental factors contribute to flood risk.

Finally, the system performs **testing**, where the trained model is evaluated using unseen data to measure its prediction accuracy and performance. The testing stage ensures that the model can effectively generalize and provide reliable flood predictions.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

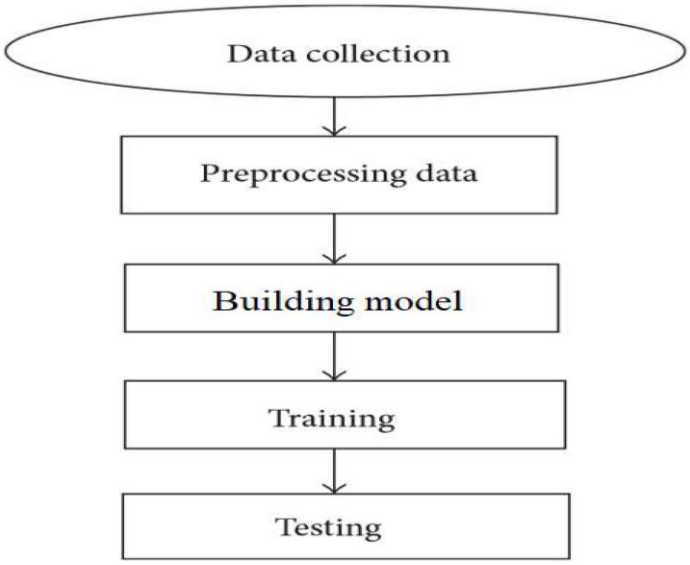


Fig. 3. System architecture of the machine learning-based flood prediction system.

V. METHODOLOGY

The proposed flood prediction system follows a structured workflow consisting of model training, model saving, deployment, and user interaction through a web application. The overall deployment architecture of the system is illustrated in **Fig. 4**.

Initially, the machine learning model is trained using environmental and weather-related parameters such as temperature, humidity, cloud cover, and rainfall indicators. As shown in **Fig. 4**, the training process is performed using the **Scikit-learn library within a Jupyter Notebook environment**, where the Random Forest classification algorithm learns patterns associated with flood occurrences from the dataset.After the training process is completed, the trained model is **serialized and saved as a .pkl (pickle) file**. This file stores the learned parameters of the model so that it can be reused for prediction without retraining the model.

The saved model is then integrated into a **Flask-based web application**, where a **REST API** is created to handle prediction requests. The Flask application receives input data from the user, processes the input parameters, and passes them to the trained machine learning model for prediction.

To enable efficient communication between the web application and the server, the Flask application runs on a **WSGI HTTP server (Gunicorn)**, which manages incoming HTTP requests and sends prediction responses back to the user.

Finally, the system is deployed on a **cloud platform (Heroku PaaS)**, allowing users to access the flood prediction service through a web interface. Users submit weather parameters through the application, and the system returns predictions indicating whether flood conditions are likely to occur.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

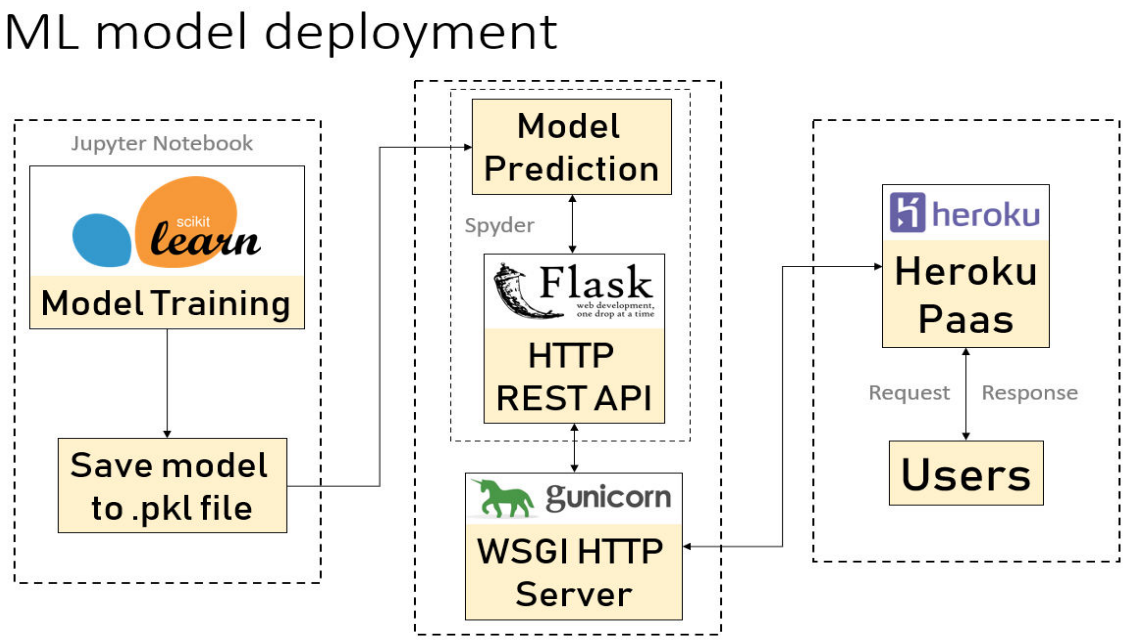


Fig. 4. Architecture of the Flood Prediction Web Application

• Data normalization and feature selection techniques are applied to improve the quality of the dataset before training the model.
• The dataset is divided into training and testing subsets to ensure that the model can be evaluated objectively.
• During training, the Random Forest algorithm generates multiple decision trees that collectively determine the final prediction result.
• Ensemble learning helps improve model stability and reduces the possibility of overfitting.

VI. MODEL EXPLANATION (RANDOM FOREST)

Random Forest is a supervised machine learning algorithm widely used for classification and prediction tasks. It belongs to the ensemble learning family, where multiple decision trees are constructed and combined to produce a more accurate and stable prediction.

During the training process, the Random Forest algorithm generates several decision trees using different subsets of the dataset and randomly selected feature sets. Each tree learns patterns from the input features and independently predicts an output class. This randomness in data sampling and feature selection improves the diversity among trees and reduces the risk of overfitting.

As illustrated in **Fig. 5**, the input dataset is divided into multiple subsets of features. Each subset is used to train an individual decision tree (Tree #1, Tree #2, Tree #3, Tree #4). Every tree produces its own prediction based on the learned relationships between the input variables and the target variable.After all trees generate their predictions, the Random Forest algorithm combines the outputs using a **majority voting mechanism**. The class that receives the highest number of votes from the individual decision trees is selected as the **final predicted class**.This ensemble approach improves prediction accuracy and robustness compared to a single decision tree. By aggregating the predictions of multiple trees, Random Forest effectively handles complex datasets and nonlinear relationships between environmental variables, making it highly suitable for flood prediction tasks.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

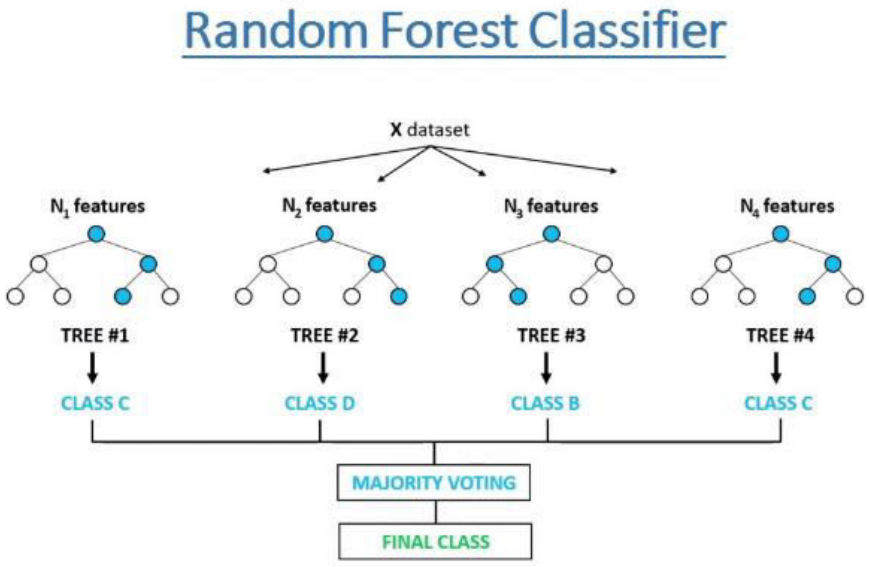


Fig. 5. Working principle of the Random Forest classification model used for flood prediction.

VII. RESULTS AND EXPERIMENTS

The flood prediction system provides a user-friendly web interface that allows users to input environmental parameters and obtain flood risk predictions. The interface is designed using a web-based framework and is connected to the trained machine learning model through a Flask backend.

As shown in Fig. 6.1, the system initially displays an input form where users can enter important weather parameters such as temperature, humidity, and cloud cover. These input fields allow users to provide real-time environmental conditions that influence flood occurrence. At this stage, the system waits for the user to submit the data by clicking the "Predict Flood Risk" button.

Once the user submits the input parameters, the data is sent to the backend server where the trained Random Forest model processes the values and calculates the probability of flood occurrence.

The prediction result is then displayed on the interface, as illustrated in Fig. 6.2. In this stage, the system shows the prediction outcome along with the calculated risk score. If the model determines that the environmental conditions indicate a potential flood, the system displays a warning message such as "Flood Expected" along with the corresponding risk probability.

This interactive interface allows users to easily provide environmental inputs and quickly obtain flood risk predictions. By integrating the machine learning model with a web-based interface, the system enables real-time flood risk assessment and supports early warning for potential flood conditions.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

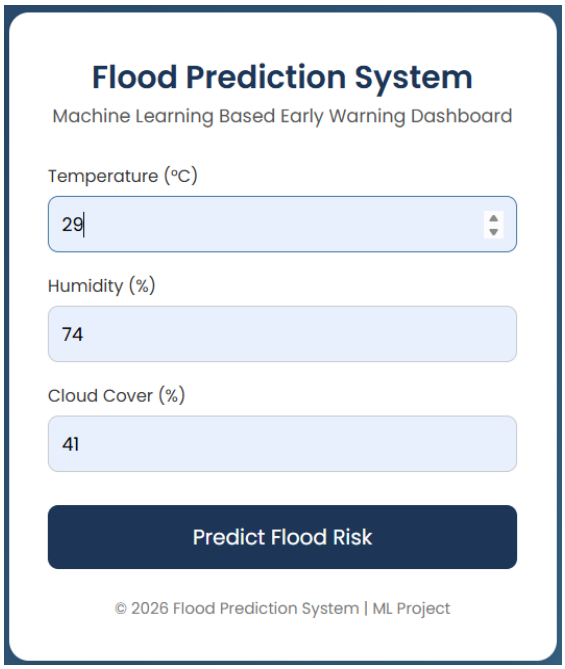


Fig6.1 Before Prediction

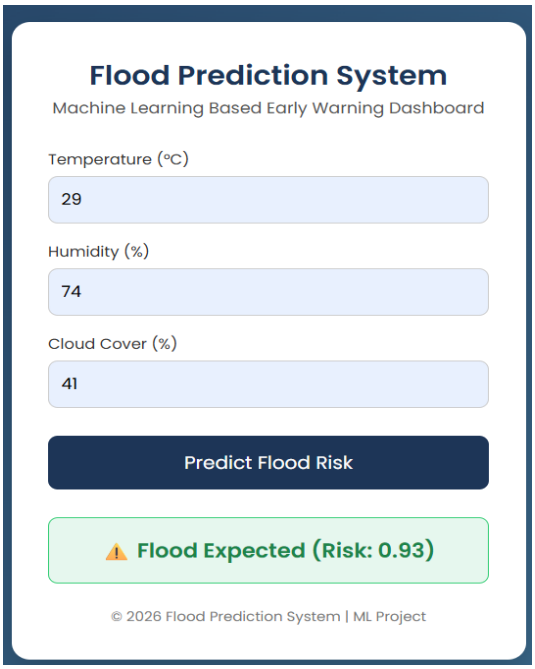


fig6.2 After Prediction

VIII. EVALUATION METRICS

To evaluate the effectiveness of the flood prediction model, a **confusion matrix** is used to analyze the classification performance of the trained machine learning algorithm. The confusion matrix provides a detailed comparison between the **actual values** and the **predicted values** generated by the model.As illustrated in **Fig. 7**, the confusion matrix contains four important components: **True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)**. These values help measure how accurately the model predicts flood and non-flood events.

In **Fig. 7(a)**, the confusion matrix represents the evaluation results obtained from a larger dataset. The model correctly predicted **485 flood events (True Positive)** and **490 non-flood events (True Negative)**. Additionally, **15 cases were incorrectly predicted as floods (False Positive)**, while **10 flood cases were incorrectly classified as non-flood events (False Negative)**. These results indicate that the model performs with high reliability in distinguishing between flood and non-flood conditions.

Similarly, **Fig. 7(b)** shows the confusion matrix results for another test evaluation of the model. In this case, the model correctly predicted **95 flood events (True Positive)** and **97 non-flood events (True Negative)**. Only **5 instances were misclassified as floods (False Positive)** and **3 instances were incorrectly classified as non-flood events (False Negative)**.

The results shown in **Fig. 7 demonstrate that the Random Forest model achieves high prediction accuracy and maintains a balanced performance in identifying both flood and non-flood conditions**. The confusion matrix therefore confirms the effectiveness of the proposed machine learning approach for flood prediction.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

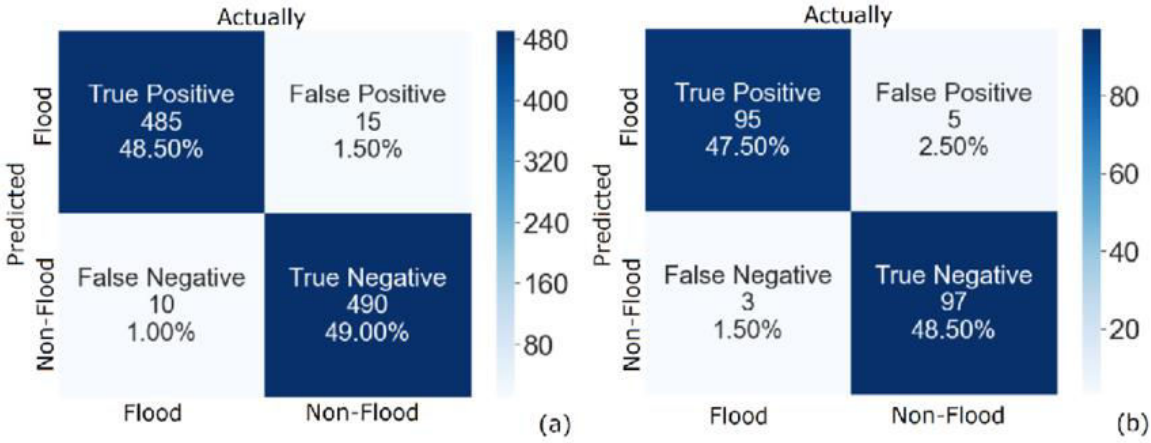


Fig. 7. Confusion matrix used to evaluate the performance of the flood prediction model: (a) training dataset evaluation and (b) testing dataset evaluation.

IX. CONCLUSION

This study presented a machine learning-based flood prediction system that analyzes environmental data to identify potential flood conditions. The Random Forest algorithm demonstrated reliable performance when applied to historical weather datasets.By integrating the trained model with a web application interface, the system becomes accessible and easy to use for prediction purposes. In the future, the model can be further improved by incorporating real-time weather data and exploring advanced machine learning or deep learning techniques.

• The study confirms that environmental parameters play an important role in predicting flood conditions.
• Machine learning techniques provide an efficient alternative to traditional simulation-based flood prediction models.
• The developed system demonstrates how predictive analytics can support disaster management strategies.
• Future research may explore advanced algorithms such as deep learning models to further improve prediction accuracy.

REFERENCES

- **Mobley et al. (2021)** — *Quantification of Continuous Flood Hazard Using Random Forest Classification*
This research applies Random Forest classification to identify flood-prone areas using geographical and hydrological datasets. The study demonstrates how machine learning techniques can effectively support flood hazard assessment and risk management.
- **Fang et al. (2021)** — *Flood Susceptibility Prediction Using Long Short-Term Memory Neural Networks*
This study proposes a deep learning approach using LSTM networks to analyze time-series climate data such as rainfall and water levels. The results show that deep learning models can capture complex temporal relationships and improve flood prediction performance.
- **Tang et al. (2023)** — *Flood Forecasting Based on Machine Learning Pattern Recognition*
This work focuses on identifying flood patterns using machine learning models combined with pattern recognition techniques. The study shows that analyzing rainfall distribution and environmental indicators can enhance flood forecasting accuracy.

1. Mosavi, A., Ozturk, P., & Chau, K. (2018). *Flood Prediction Using Machine Learning Models: Literature Review*. **Water Journal (MDPI)**. https://www.mdpi.com/2073-4441/10/11/1536
2. Mobley, W., et al. (2021). *Quantification of Continuous Flood Hazard Using Random Forest Classification*. **Natural Hazards and Earth System Sciences**. https://nhess.copernicus.org/articles/21/807/2021/
3. Fang, Z., et al. (2021). *Predicting Flood Susceptibility Using Long Short-Term Memory Neural Networks*. **Journal of Hydrology**. https://www.sciencedirect.com/science/article/pii/S0022169420311951

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503245|

4. Tang, Y., et al. (2023). *Flood Forecasting Based on Machine Learning Pattern Recognition*. **Environmental Modelling & Software**. https://www.sciencedirect.com/science/article/pii/S2214581823000939
5. Wahba, M., et al. (2024). *Flash Flood Susceptibility Mapping Using Machine Learning Techniques*. **Heliyon Journal**.
https://www.sciencedirect.com/science/article/pii/S2405844024100138
6. Dahal, D., et al. (2024). *Analyzing Climate Dynamics and Developing Machine Learning Models for Flood Prediction*. **Environmental Data Science**. https://www.cambridge.org/core/journals/environmental-data-science
7. Jang, S. D., et al. (2025). *Flood Prediction in Urban Areas Based on Random Forest Machine Learning Model*. **Environmental Engineering Research**. https://www.eeer.org

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details