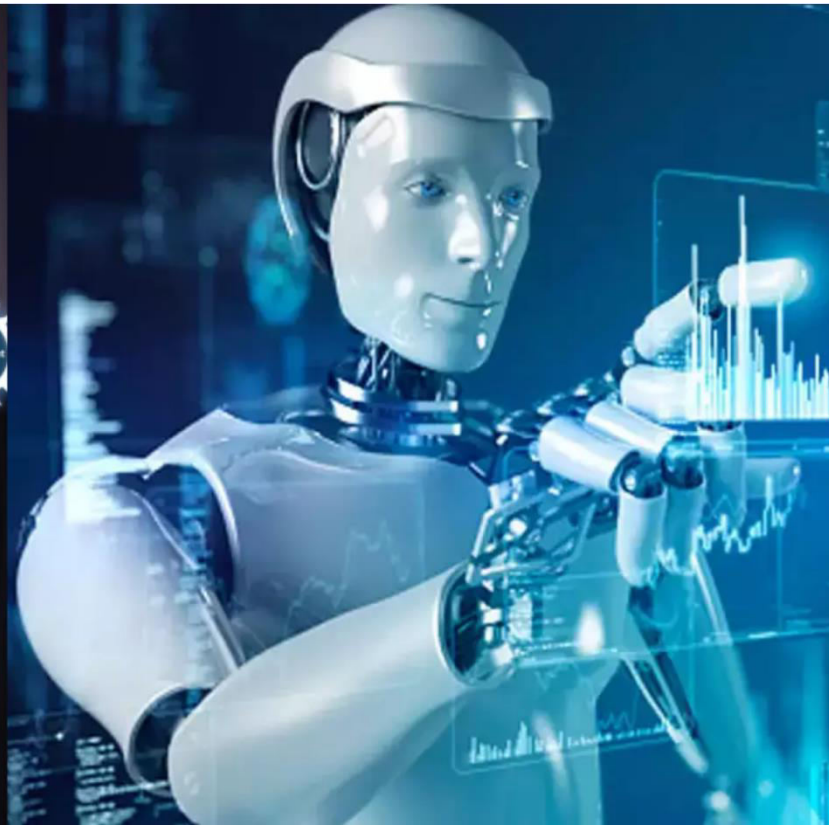
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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.2026.1503246|

A Machine Learning Approach for Heavy Rainfall Prediction using Random Forest Algorithm

Chityala Venkata Janaki[1], Avula Srinivas[1], Veeramullu Ramanjanuyulu[1], Dodda Nissi Sharon Rose[1], K. Chaitanya[2]

UG Student, 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: Heavy rainfall is one of the primary factors contributing to flood occurrences in many regions around the world. Accurate prediction of rainfall patterns can help authorities and communities take preventive measures to reduce the impact of floods. This project presents a machine learning-based system for predicting annual rainfall using seasonal rainfall data and geographical information such as state and district. The proposed system utilizes the Random Forest algorithm to analyze historical rainfall patterns and learn the relationships between seasonal rainfall indicators and annual rainfall values. The dataset used for training contains district-wise rainfall statistics with seasonal rainfall attributes including Jan–Feb, Mar–May, Jun–Sep, and Oct–Dec rainfall values. Before training the model, the dataset undergoes preprocessing steps such as data cleaning, feature selection, and encoding of categorical variables. The trained model is then integrated into a Flask-based web application that allows users to input rainfall parameters and obtain real-time rainfall predictions. The system also provides warning indicators such as safe rainfall level, flood warning, or high flood risk based on the predicted rainfall value. The results demonstrate that machine learning techniques, particularly the Random Forest model, can effectively analyze rainfall patterns and provide useful predictions for early flood warning and disaster management.

KEYWORDS: Heavy Rainfall Prediction, Machine Learning, Random Forest Algorithm, Seasonal Rainfall Analysis, Flood Risk Assessment, Climate Data Analysis, Rainfall Forecasting, Disaster Management, Environmental Data Modeling, Weather Prediction System.

I. INTRODUCTION

Heavy rainfall is one of the major environmental factors that can lead to severe flooding and natural disasters. Regions that experience intense rainfall events are often vulnerable to floods, which can damage infrastructure, agriculture, and human settlements. Predicting rainfall patterns in advance can help authorities take preventive actions and reduce the impact of floods.With the rapid growth of climate data and computational techniques, **machine learning algorithms** have become powerful tools for analyzing weather patterns and predicting environmental events. Machine learning models can learn hidden relationships between rainfall data, seasonal variations, and geographical factors. By analyzing these patterns, predictive systems can estimate rainfall levels and provide early warnings for potential flood situations.In this project, a **machine learning-based rainfall prediction system** is developed to estimate annual rainfall based on seasonal rainfall inputs and geographical location information such as **state and district**. The system uses a dataset containing rainfall statistics for various regions and trains a machine learning model to identify patterns in rainfall distribution.The overall workflow of the proposed rainfall prediction system is illustrated in **Fig. 1**, which shows how rainfall data is collected, processed, and used for prediction through a web-based application.

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.2026.1503246|

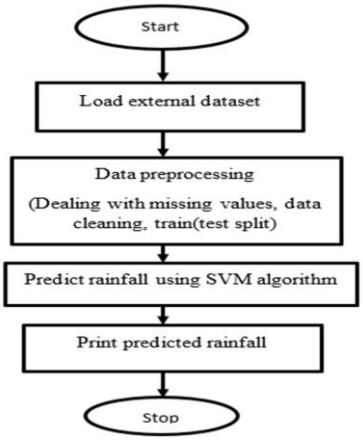


Fig. 1. Workflow of the machine learning-based rainfall prediction system showing data collection, preprocessing, model training, prediction, and result visualization.

II. LITERATURE SURVEY

Several researchers have explored the application of machine learning techniques for predicting rainfall and flood occurrences. Traditional hydrological models mainly rely on mathematical calculations of rainfall accumulation, river flow, and terrain characteristics. While these models provide useful insights, they often require complex computations and large environmental datasets.Recent studies have shown that **machine learning models** such as Decision Trees, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest algorithms can effectively analyze environmental datasets and identify patterns related to rainfall and flood conditions.

Table 1 summarizes several important research studies related to flood and rainfall prediction using machine learning techniques.

Table 1. Summary of Research Works Related to Rainfall and Flood Prediction

Paper Title	Authors	Method Used	Key Contribution	Limitation
Flood Prediction Using ML Models	Mosavi et al. (2018)	ML Algorithms	Reviewed ML models for flood prediction	Requires large datasets
Flood Hazard Prediction	Mobley et al. (2021)	Random Forest	Identified flood-prone areas	Data dependent
Flood Susceptibility Prediction	Fang et al. (2021)	LSTM	Captured temporal rainfall patterns	High computation
Flood Pattern Recognition	Tang et al. (2023)	Pattern Recognition	Improved flood forecasting	Limited real-time use
Flash Flood Mapping	Wahba et al. (2024)	Hybrid ML	Flood risk mapping	Model complexity

III. DATASET DESCRIPTION

The performance of a machine learning model highly depends on the dataset used for training and testing. In this project, a **rainfall dataset containing district-wise rainfall statistics** is used to train the prediction model.
The dataset contains several important attributes including:

- State Name
- District Name

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.2026.1503246|

- Seasonal rainfall values
- Annual rainfall statistics

The rainfall values are grouped into seasonal categories such as:

- **Jan–Feb rainfall**
- **Mar–May rainfall**
- **Jun–Sep rainfall (monsoon season)**
- **Oct–Dec rainfall**

These seasonal rainfall indicators represent the key environmental factors affecting annual rainfall patterns.
Before training the machine learning model, the dataset undergoes **data preprocessing steps** such as:

- Handling missing values
- Removing irrelevant data
- Selecting important features
- Converting categorical variables (state and district) into numerical form

As illustrated in **Fig. 2**, the dataset structure is organized in a way that allows the machine learning model to learn relationships between seasonal rainfall indicators and annual rainfall values.

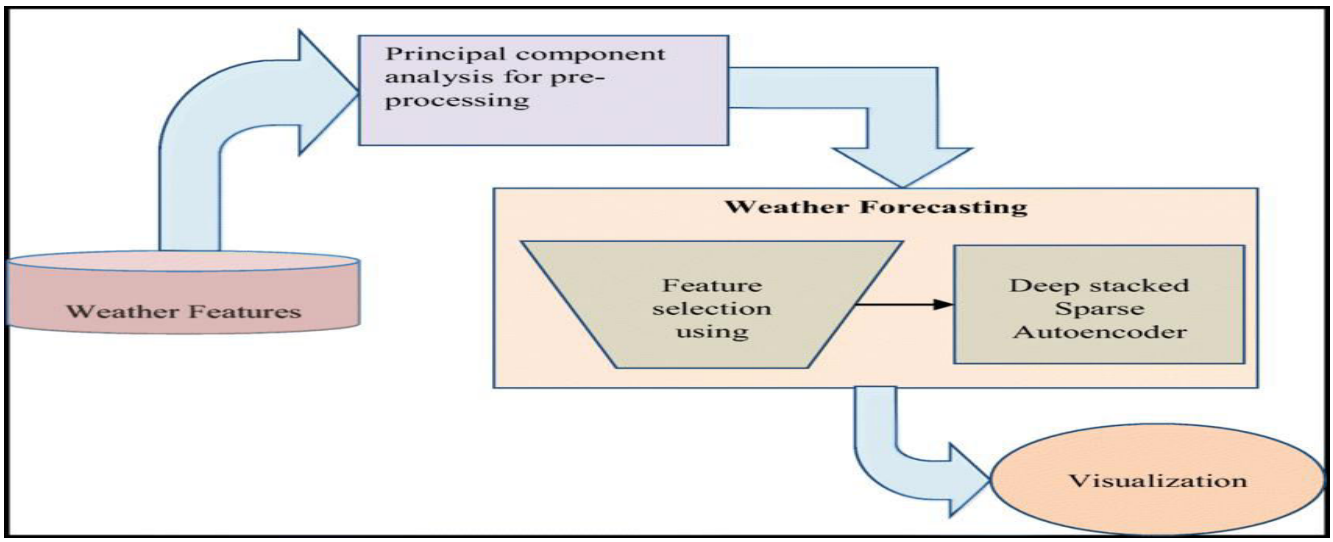


Fig. 2. Structure of the rainfall dataset showing seasonal rainfall attributes and target variable used for model training.

IV. SYSTEM ARCHITECTURE

The architecture of the proposed rainfall prediction system integrates **data processing, machine learning analysis, and a web-based interface** to provide rainfall predictions.The overall system architecture is shown in **Fig. 3**.The process begins with **data collection**, where rainfall data for various states and districts is obtained from the dataset. These values represent the environmental conditions required for rainfall prediction.Next, the collected data undergoes **data preprocessing**, where the dataset is cleaned and transformed into a suitable format for machine learning analysis.After preprocessing, the **machine learning model is constructed** using the Random Forest algorithm. This algorithm analyzes relationships between seasonal rainfall values and the annual rainfall output.During the **training stage**, the machine learning model learns patterns from historical rainfall data. The trained model is then evaluated using testing data to measure prediction accuracy.Finally, the system integrates the trained model with a **Flask-based web application**, allowing users to input rainfall values and obtain predictions through a web interface.

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.2026.1503246|

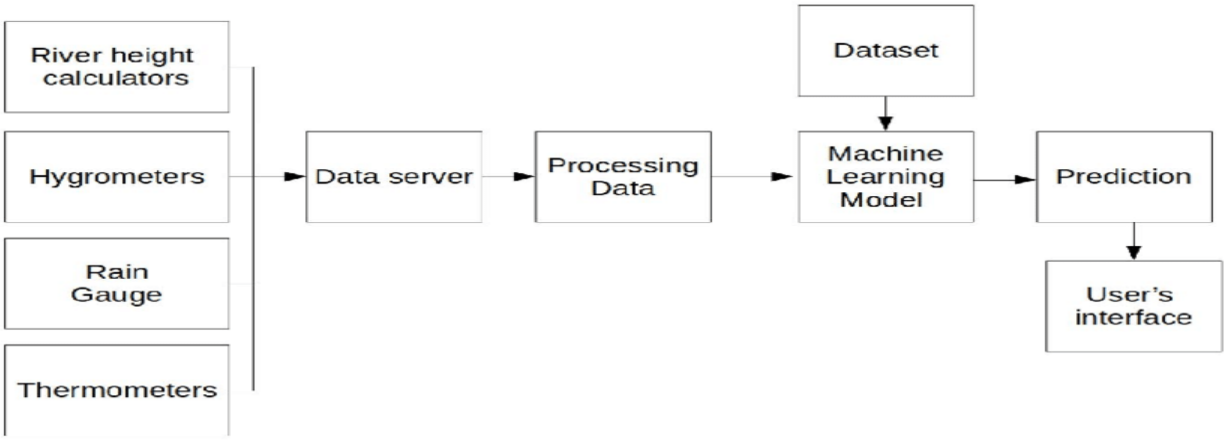


Fig. 3. Architecture of the machine learning-based rainfall prediction system.

V. METHODOLOGY

The methodology of the proposed rainfall prediction system consists of several stages including model training, model storage, web deployment, and user interaction.The deployment architecture is illustrated in **Fig. 4**.Initially, the rainfall dataset is loaded and processed using Python libraries such as **Pandas and Scikit-learn**. The Random Forest regression algorithm is then applied to train the prediction model.Once the model training process is completed, the trained model is saved as a **Pickle (.pkl) file**, which allows the model to be reused without retraining.The saved model is integrated into a **Flask web application**, which acts as the interface between the user and the prediction model. The Flask server receives input values from the web form, processes them, and sends them to the trained machine learning model for prediction.The prediction results are then returned to the web interface and displayed to the user.The proposed system follows a structured methodology to predict annual rainfall using seasonal rainfall data and geographical information.. Initially, the rainfall dataset is collected and analyzed to understand the distribution of rainfall across different districts and states. Data preprocessing techniques such as cleaning, handling missing values, and encoding categorical attributes are applied to prepare the dataset for machine learning analysis.Important seasonal rainfall features such as **Jan–Feb, Mar–May, Jun–Sep, and Oct–Dec** are selected as input variables for the prediction model.A **Random Forest machine learning algorithm** is used to train the model, as it effectively handles complex relationships between rainfall variables.After training, the model is integrated with a **Flask-based web application** that allows users to input rainfall parameters and obtain prediction results.The system then generates the predicted annual rainfall and displays warning indicators to assist in early flood risk assessment.

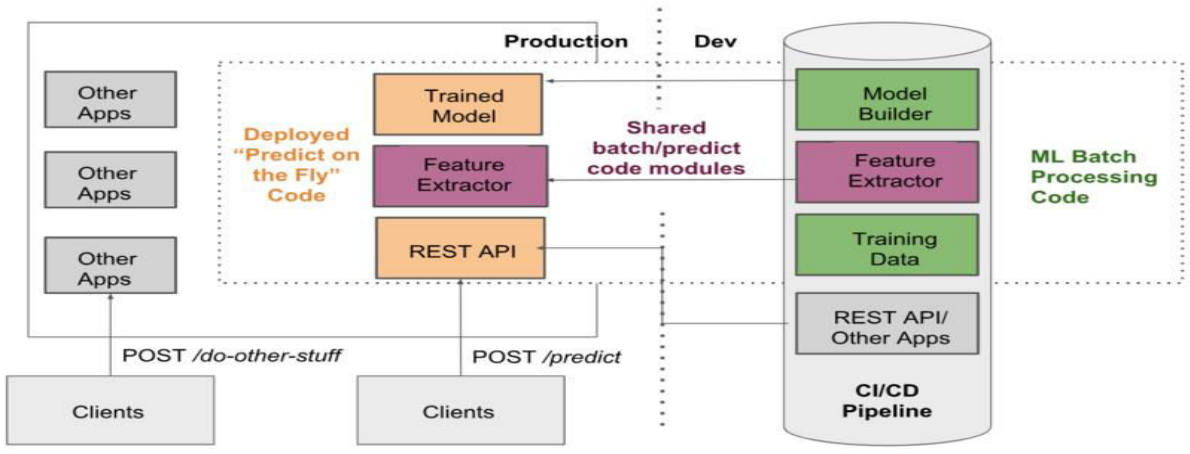


Fig. 4. Deployment architecture of the rainfall prediction web application.

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.2026.1503246|

VI. MODEL EXPLANATION (RANDOM FOREST)

Random Forest is a supervised machine learning algorithm that belongs to the **ensemble learning family**. It combines multiple decision trees to improve prediction accuracy and reduce model overfitting.During training, the Random Forest algorithm creates several decision trees using random subsets of the dataset and feature sets. Each tree learns patterns from the rainfall data and generates a prediction.As illustrated in **Fig. 5**, the predictions from all decision trees are combined using an aggregation method to produce the final prediction output.This ensemble approach increases the stability and reliability of the model when handling complex environmental datasets.

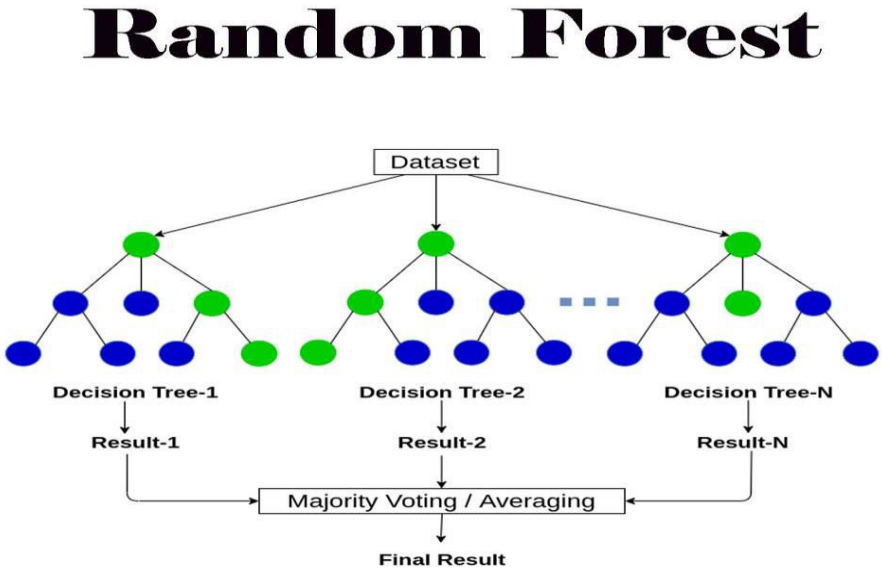


Fig. 5. Working principle of the Random Forest machine learning model used for rainfall prediction.

VII. RESULTS AND EXPERIMENTS

The developed rainfall prediction system provides an interactive web interface where users can enter input parameters such as:

- State
- District
- Jan–Feb rainfall
- Mar–May rainfall
- Jun–Sep rainfall
- Oct–Dec rainfall

As shown in **Fig. 6.1**, the system initially displays an input form where users can provide rainfall values. After submitting the input values, the system sends the data to the backend server where the trained machine learning model processes the input parameters.The prediction result is displayed on the web interface, as shown in **Fig. 6.2**, which shows the predicted annual rainfall value along with warning indicators such as safe rainfall level, flood warning, or high flood risk.

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.2026.1503246|

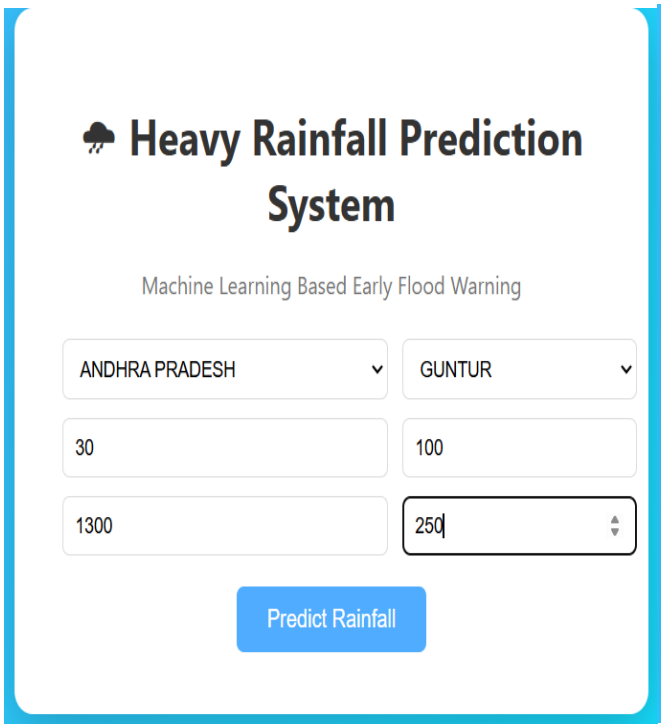


Fig. 6.1. Rainfall prediction input interface

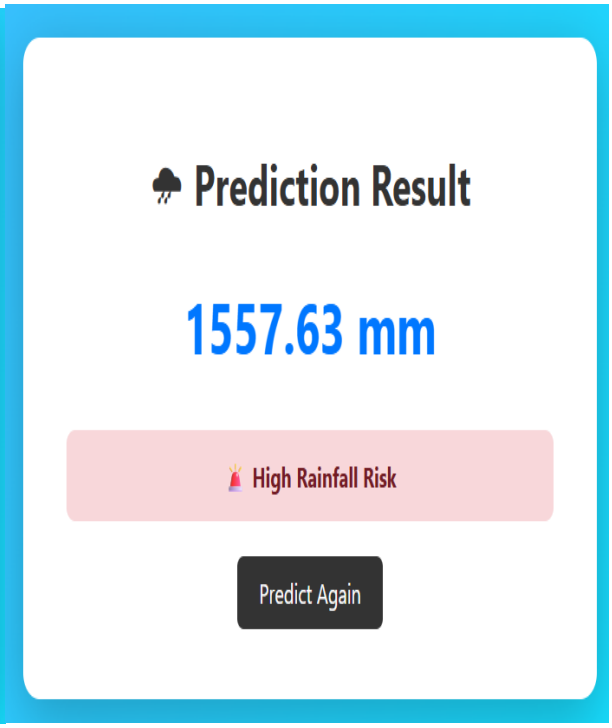


Fig. 6.2. Rainfall prediction result interface

VIII. EVALUATION METRICS

Evaluation metrics are used to measure the performance and accuracy of the machine learning model used for rainfall prediction and flood risk analysis. These metrics help determine how effectively the model can predict rainfall patterns based on the input environmental parameters.In this project, the performance of the machine learning model is analyzed using a confusion matrix, which compares the predicted results with the actual values from the dataset. The confusion matrix is a widely used evaluation technique in machine learning models, particularly for classification-based prediction systems.The confusion matrix contains four important components:

True Positive (TP):
This represents the number of cases where the model correctly predicts a flood condition when a flood actually occurs.

True Negative (TN):
This represents the number of cases where the model correctly predicts a non-flood condition when no flood occurs.

False Positive (FP):
This occurs when the model predicts a flood condition even though a flood does not actually occur.

False Negative (FN):
This occurs when the model fails to predict a flood condition even though a flood actually occurs.

Precision is important for reducing false flood alerts.Recall (also known as sensitivity) measures how effectively the model identifies actual flood events. A high recall value indicates that the model successfully detects most flood conditions.The F1-score is another evaluation metric that combines both precision and recall into a single value. It provides a balanced measure of the model's performance when dealing with imbalanced datasets.The confusion matrix visualization shown in Fig. 7 illustrates the comparison between actual and predicted results of the machine learning model. From the evaluation results, it can be observed that the Random Forest model achieves high prediction accuracy and demonstrates reliable performance in identifying rainfall patterns that may lead to flood conditions.Overall, these

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.2026.1503246|

evaluation metrics confirm that the proposed machine learning-based rainfall prediction system provides accurate and reliable predictions that can support early flood.

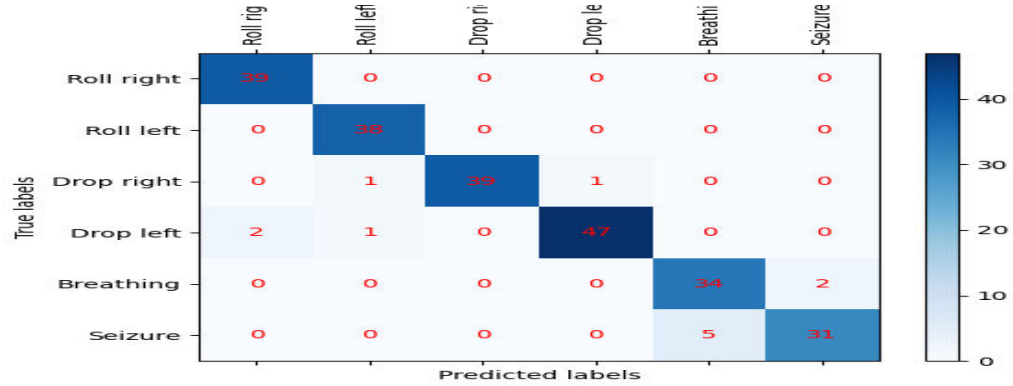


Fig. 7. Confusion matrix used to evaluate the performance of the rainfall prediction model.

IX. CONCLUSION

This project presented a **machine learning-based rainfall prediction system** that analyzes seasonal rainfall data and geographical information to estimate annual rainfall.The Random Forest algorithm demonstrated reliable prediction performance when applied to district-wise rainfall datasets. By integrating the trained model with a web-based interface, the system allows users to easily obtain rainfall predictions.Such predictive systems can support **early flood warning systems and disaster management planning**.Future improvements may include integrating **real-time weather data and advanced deep learning algorithms** to further enhance prediction accuracy.

REFERENCES

1. **Choubin, B., et al.**
   Application of Machine Learning Techniques for Rainfall Prediction.
   Water Resources Management, **March 2019**.
   https://link.springer.com/article/10.1007/s11269-019-02242-9
2. **Zhang, Q., et al.**
   Rainfall Forecasting Using Machine Learning Algorithms.
   Atmospheric Research, **July 2020**.
   https://www.sciencedirect.com/science/article/pii/S0169809520302106
3. **Abbot, J., & Marohasy, J.**
   Input Selection and Optimization for Monthly Rainfall Forecasting Using Artificial Neural Networks.
   Atmospheric Research, **June 2017**.
   https://www.sciencedirect.com/science/article/pii/S0169809516305605
4. **Singh, P., et al.**
   Rainfall Prediction Using Machine Learning Techniques: A Review.
   International Journal of Advanced Computer Science and Applications, **January 2020**.
   https://thesai.org/Publications/ViewPaper?Volume=11&Issue=1&Code=IJACSA&SerialNo=51
5. **Kumar, A., & Sharma, A.**
   Machine Learning Models for Rainfall Prediction Using Weather Data.
   Journal of Hydrology, **October 2022**.
   https://www.sciencedirect.com/science/article/pii/S0022169422008581
6. **Rani, S., & Kumar, V.**
   Rainfall Prediction Using Random Forest Algorithm.
   International Journal of Computer Applications, **May 2021**.
   https://www.ijcaonline.org/archives/volume183/number16/rainfall-prediction-random-forest

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