International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05969
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Rainfall Intensity Prediction System using Machine Learning Techniques

M.Chandan Kumar[1], N.Srujana[2], P.Aswanth[3], T.Dharma[4]

Supervisor**: Ms.M.Sowjanya**,M.Tech(PhD), Assistant Professor, Dept. of CSE, VIET

[1]*Department of CSE (AIML), Visakha Institute of Engineering and Technology, Andhra Pradesh, India*
[2]*Department of CSE (AIML), Visakha Institute of Engineering and Technology, Andhra Pradesh, India*
[3]*Department of CSE (AIML), Visakha Institute of Engineering and Technology, Andhra Pradesh, India*
[4]*Department of CSE (AIML), Visakha Institute of Engineering and Technology, Andhra Pradesh, India*

Abstract - This study presents the development of a machine learning-based rainfall prediction system using historical meteorological data. Rainfall prediction plays a crucial role in agriculture, water resource management, and disaster prevention. Traditional forecasting methods often fail to capture the complex and non-linear relationships between atmospheric parameters, leading to inaccurate predictions. In this project, a dataset containing weather parameters such as temperature, dew point, humidity, sea level pressure, visibility, wind speed, month, and day was used to train machine learning models. Data preprocessing techniques including handling missing values using median imputation, outlier removal using the Interquartile Range (IQR) method, and log transformation for skewness reduction were applied to improve data quality. Among various machine learning algorithms, Extreme Gradient Boosting (XGBoost) was selected due to its superior performance in handling non-linear data and providing high prediction accuracy. Hyperparameter tuning was performed using GridSearchCV to optimize model performance. The model was evaluated using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared score. Furthermore, the trained model was deployed using Streamlit to create an interactive web application that allows users to input real-time weather parameters and obtain rainfall predictions along with graphical visualizations.

***Key Words*:** Machine Learning, Rainfall Prediction, XGBoost, Data Pre-processing, Stream lit, Regression

1.INTRODUCTION

Rainfall prediction plays a crucial role in agriculture, disaster management, and water resource planning. Traditional methods often fail to accurately predict rainfall due to the complex and nonlinear nature of atmospheric conditions. Machine learning (ML) techniques provide an advanced approach to analysing meteorological data, improving the accuracy of forecasts. This project implements ML algorithms such as Multivariate Linear Regression (MLR), Random Forest.. (RF), and Extreme Gradient Boosting (XGBoost) to predict rainfall intensity based on historical meteorological data.
Accurate prediction of rainfall helps mitigate risks associated with droughts and floods, optimise irrigation planning, and improve water resource management. However, traditional forecasting techniques, such as numerical weather prediction models and statistical approaches, have limitations in capturing the dynamic nature of meteorological parameters. With advancements in computational power and data availability, machine learning has emerged as a promising solution for weather forecasting.

ML models can analyse large datasets, identify hidden patterns, and generate predictions based on past trends. These models can incorporate multiple meteorological features such as temperature, humidity, wind speed, and atmospheric pressure to enhance prediction accuracy. This project explores various ML algorithms and evaluates their effectiveness in predicting rainfall intensity.

The Rainfall Intensity Prediction System is a smart, data-driven application developed using machine learning algorithms to forecast daily rainfall with greater precision. Built on models like Multivariate Linear Regression, Random Forest, and XGBoost, it analyzes key meteorological inputs such as temperature, humidity, wind speed, and air pressure to deliver timely and accurate predictions. To ensure accessibility and user engagement, the system includes a web interface developed using Streamlit, enabling real-time user inputs and instant visual feedback.

2.LITERATURE REVIEW

Machine learning has gained significant attention in meteorology due to its ability to analyse complex patterns and relationships in weather data. Traditional weather forecasting methods, including numerical weather prediction (NWP) models, rely on mathematical simulations of atmospheric processes. While these models provide valuable insights, they often struggle with high computational costs and limitations in handling non-linear dependencies among meteorological variables.

Gnanasankaran and Ramaraj (2020) applied multiple linear regression models to predict rainfall using Indian meteorological data. Their findings highlighted that while linear regression can provide baseline predictions, it fails to capture nonlinear relationships effectively. Srinivas et al. (2020) further explored machine learning strategies based on weather radar data, demonstrating that advanced techniques like decision trees and ensemble methods outperform traditional regression models in terms of accuracy.

Deep learning techniques have also been investigated for rainfall forecasting. Zeelan et al. (2020) employed artificial neural networks (ANNs) and deep learning approaches, concluding that ANN-based models performed significantly better in capturing rainfall variability. Aswin et al. (2018) implemented CNNs and RNNs to predict rainfall intensity, achieving high precision through the integration of spatiotemporal meteorological data.

Thirumalai et al. (2017) focused on heuristic prediction models using machine learning techniques, showing that Random Forest and XGBoost significantly enhanced prediction accuracy compared to conventional approaches. Kusiak et al. (2013) demonstrated that integrating radar reflectivity data with machine learning algorithms can enhance the precision of rainfall predictions, emphasizing that feature selection and hyperparameter tuning play a critical role in optimising prediction models.

Overall, the literature underscores the potential of machine learning in rainfall forecasting. However, challenges remain, including the need for high-quality datasets, real-time data integration, and the selection of optimal models for specific climatic conditions

3.METHODOLOGY

3.1 System Overview

The proposed system leverage

es machine learning models trained on historical weather data to improve prediction accuracy. It integrates diverse meteorological parameters such as temperature, humidity, wind speed, and atmospheric pressure. The system implements three ML algorithms: Multivariate Linear Regression (MLR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost).

3.2 Data Collection and Preprocessing

A dataset containing weather parameters (temperature, dew point, humidity, sea level pressure, visibility, wind speed, month, and day) spanning 2012-2022 was used. The preprocessing pipeline included: (1) Missing value imputation using median values; (2) Outlier removal through the Interquartile Range (IQR) method, capping values at Q1 - 1.5*IQR and Q3 + 1.5*IQR; (3) Log transformation of the target variable to address skewed data distributions; and (4) Feature normalization using StandardScaler.

3.3 Machine Learning Algorithms

Multivariate Linear Regression (MLR) establishes relationships between multiple independent variables and rainfall. The general equation is: $Yi = \beta1Xi1 + \beta2Xi2 + ... + \beta pXip + \varepsilon i$, where Yi represents predicted rainfall, Xi represents independent meteorological variables, β represents regression coefficients, and εi is the error term.

Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates outputs to improve accuracy. It handles non-linearity and missing data effectively, and provides feature importance to identify key meteorological parameters.

Extreme Gradient Boosting (XGBoost) is an advanced algorithm that optimises decision trees through gradient boosting. It implements regularisation techniques to prevent overfitting and achieves superior accuracy with minimal computational cost. Hyperparameter tuning using GridSearchCV explored parameters including n_estimators [100, 200, 300], learning_rate [0.01, 0.05, 0.1], and max_depth [3, 6, 9].

3.4 System Architecture

The system architecture consists of multiple layers: Data Acquisition Layer (gathers historical weather data), Preprocessing Layer (cleans and transforms raw data), Feature Selection and Engineering Layer (identifies key

parameters), Model Training Layer (implements ML algorithms), Prediction and Evaluation Layer (generates forecasts and evaluates performance), and a User Interface Layer (interactive Streamlit web application).

3.5 Algorithm - Prediction Workflow

Step 1: Load historical meteorological dataset (2012-2022)

Step 2: Preprocess data (impute missing values, remove outliers, apply log transform)

Step 3: Split data (80% train, 20% test) and scale features

Step 4: Train XGBoost model with GridSearchCV hyperparameter tuning

Step 5: Accept user inputs via Streamlit interface

Step 6: Preprocess inputs, apply trained scaler, and generate prediction

Step 7: Display predicted rainfall (mm) with MSE, MAE, R² metrics and visualizations

4.RESULTS AND DISCUSSION

The Rainfall Intensity Prediction System was evaluated using meteorological data from 2012 to 2022. The system demonstrated strong performance across multiple evaluation metrics. The XGBoost model with GridSearchCV tuning achieved the best parameters: colsample_bytree = 0.9, learning_rate = 0.05, max_depth = 3, n_estimators = 100, and subsample = 1.0.

The model was evaluated on a 20% test set using Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) score. XGBoost outperformed both Multivariate Linear Regression and Random Forest in capturing the nonlinear relationships between meteorological parameters and rainfall intensity.

The system correctly handled diverse input conditions. The interactive Streamlit web interface allowed users to input real-time weather parameters including temperature (°C), dew point (°C), humidity (%), pressure (hPa), visibility (km), wind speed (km/h), month, and day. The application returned the predicted rainfall in millimeters along with model performance metrics.

The bar chart comparison of actual versus predicted rainfall values across 20 test samples demonstrated close alignment between model predictions and ground truth values. The scatter plot of actual vs. predicted rainfall further confirmed the model's reliability across the range of rainfall intensities present in the dataset. Ensemble learning techniques, particularly XGBoost, demonstrated superior performance in handling the nonlinear and complex meteorological patterns present in the dataset.

Fig. 1 shows the rainfall prediction input parameters interface. Fig. 2 illustrates the model predictions versus actual rainfall values, demonstrating the system's accuracy.

5.CONCLUSIOSNS

This paper presented a Rainfall Intensity Prediction System using machine learning techniques that successfully addresses the limitations of traditional forecasting methods. By integrating advanced algorithms including Multivariate Linear Regression (MLR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), the system provides accurate and reliable rainfall predictions based on historical meteorological data.

The results demonstrate that ML-based approaches outperform traditional statistical models in terms of accuracy and adaptability. The integration of robust data preprocessing, hyperparameter tuning via GridSearchCV, and a Streamlit-based user interface makes the system both scientifically sound and practically accessible for farmers, disaster management officials, and policymakers.

Key findings include: (1) XGBoost exhibited superior accuracy among the algorithms tested; (2) Comprehensive data preprocessing including IQR-based outlier removal and log transformation significantly improved model reliability; (3) The interactive web interface democratizes access to ML-based weather forecasting for non-technical users.

Future enhancements will focus on incorporating deep learning architectures such as LSTM networks for capturing temporal dependencies, integrating geospatial and satellite data, implementing real-time IoT data pipelines for dynamic model updates, and developing mobile application platforms to improve accessibility in rural and remote regions.

REFERENCES

1. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD, 785-794.

3. Gnanasankaran, N., & Ramaraj, E. (2020). Rainfall prediction using multiple linear regression. International Journal of Advanced Research in Computer Science.

4. Zeelan, B., et al. (2020). Deep learning approaches for rainfall forecasting. IEEE Access.

5. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

6. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.

7. Rahman, M. M., & Lateh, H. (2016). Meteorological drought forecasting using machine learning techniques. Theoretical and Applied Climatology, 123(3-4), 613-623.

8. Pothuraju, V.V. Satyanarayana, et al. (2025). AI-Powered Recommender Systems for E-Commerce. In: IEEE Proceedings of ICRISET-2025.