iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: VII Month of publication: July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84383

Machine Learning Approach for Rainfall Prediction Using Multiple Regression Techniques

Paina Suvarna Latha[1], Doma Gangadhar[2]
[1]MCA Student, Department of MCA, AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam, Andhra Pradesh-531173
[2]Associate Professor, Department of MCA, AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam, Andhra Pradesh-531173

Abstract: Rainfall prediction is one of the most significant research areas in meteorology because of its direct influence on agriculture, water resource management, disaster preparedness, and environmental sustainability. Accurate rainfall forecasting assists governments, farmers, and disaster management agencies in making informed decisions regarding crop planning, irrigation scheduling, flood prevention, and drought management. Conventional statistical forecasting methods often fail to capture the complex nonlinear relationships existing among climatic variables, leading to inaccurate predictions. Machine Learning (ML) techniques have emerged as efficient alternatives by automatically learning hidden patterns from historical weather datasets. This paper presents a machine learning-based rainfall prediction system using Multiple Linear Regression (MLR), Random Forest Regressor, and K-Nearest Neighbor (KNN) Regression algorithms. The proposed model utilizes historical meteorological parameters such as temperature, humidity, atmospheric pressure, wind speed, and previous rainfall records to predict future rainfall with improved accuracy. Initially, the collected dataset undergoes preprocessing techniques including missing value handling, normalization, feature selection, and data transformation. The processed dataset is divided into training and testing subsets for model development and validation. The performance of each regression algorithm is evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2 Score, and prediction accuracy. Experimental results demonstrate that the Random Forest Regressor achieves superior prediction performance compared with Multiple Linear Regression and KNN Regression by effectively handling nonlinear relationships and reducing prediction errors. The proposed system provides an efficient and reliable rainfall forecasting solution that can support agricultural planning and environmental monitoring.

Keywords: Machine Learning, Rainfall Prediction, Multiple Linear Regression, Random Forest Regressor, KNN Regression, Weather Forecasting, Regression Analysis

I. INTRODUCTION

Rainfall is one of the most influential climatic factors affecting agriculture, water resources, transportation, power generation, and environmental sustainability. The occurrence and intensity of rainfall significantly influence food production, groundwater availability, and natural disaster management. In developing countries such as India, where agriculture forms the backbone of the economy, accurate rainfall prediction plays a crucial role in improving crop productivity and minimizing losses caused by floods and droughts. Traditional rainfall prediction methods mainly depend on statistical models and meteorological observations. Although these approaches provide useful estimates, they often fail to model the complex nonlinear interactions among atmospheric variables. Weather conditions continuously change due to seasonal variations, climate change, geographical diversity, and environmental factors, making rainfall prediction a challenging task. Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have significantly improved the capability of weather forecasting systems. Machine learning algorithms can analyze massive historical datasets, discover hidden relationships among weather parameters, and generate highly accurate prediction models without requiring explicit programming. These techniques automatically learn from previous observations and continuously improve prediction performance. In this research, a machine learning framework is proposed to predict rainfall using Multiple Linear Regression (MLR), Random Forest Regressor, and K-Nearest Neighbor (KNN) Regression algorithms. Historical weather datasets containing multiple meteorological attributes are utilized for training and testing the prediction models. Before model construction, the collected data undergoes preprocessing operations such as handling missing values, normalization, feature engineering, and data cleaning to improve prediction accuracy.

The developed system compares the performance of all three regression models using various statistical evaluation metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score. Based on experimental analysis, the model providing the minimum prediction error and maximum prediction accuracy is selected as the optimal rainfall prediction model.

The proposed system contributes toward developing an intelligent rainfall forecasting solution capable of supporting farmers, government agencies, environmental researchers, and disaster management authorities. Accurate rainfall prediction enables better agricultural planning, efficient water resource utilization, and timely disaster preparedness, thereby reducing economic losses and improving sustainable development.

II. LITERATURE REVIEW

Rainfall prediction has attracted considerable attention from researchers due to its significant role in agriculture, water resource management, and disaster mitigation. Various statistical and machine learning techniques have been proposed to improve prediction accuracy by analyzing historical meteorological data.

Li et al. (2018) proposed a machine learning-based rainfall prediction framework using Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Networks (ANN). Their study demonstrated that Random Forest outperformed traditional statistical approaches because of its capability to model nonlinear relationships among weather parameters. The authors concluded that feature selection plays a vital role in improving forecasting accuracy.

Aswin et al. (2019) developed deep learning models based on Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) for rainfall forecasting. The proposed models effectively captured temporal weather patterns and achieved lower prediction errors compared with conventional regression techniques. However, deep learning models require large datasets and high computational resources.

Sankhadeep Chatterjee et al. (2019) introduced a hybrid neural network approach for rainfall prediction using K-Means clustering followed by Artificial Neural Networks. Their proposed model improved forecasting accuracy by identifying similar climatic patterns before prediction. Although the prediction accuracy increased, the computational complexity was relatively higher.

Minghui Qiu et al. (2019) proposed a Multi-task Convolutional Neural Network for short-term rainfall prediction using data collected from multiple weather observation stations. Their model effectively utilized spatial relationships among weather stations and significantly improved forecasting performance. However, the system required continuous real-time sensor data for optimal performance.

Sandeep Kumar Mohapatra et al. (2020) conducted a comparative analysis of various machine learning algorithms including Decision Tree, Random Forest, Linear Regression, and Artificial Neural Networks for rainfall prediction using historical meteorological datasets. The study concluded that ensemble learning methods consistently produced higher prediction accuracy than individual regression models.

Although several machine learning techniques have demonstrated promising results, many existing systems suffer from limitations such as insufficient feature engineering, inadequate preprocessing, overfitting, and reduced generalization capability. Therefore, selecting suitable regression algorithms along with effective preprocessing techniques is essential for achieving reliable rainfall prediction.

III. EXISTING SYSTEM

Traditional rainfall prediction systems primarily depend on numerical weather forecasting models and statistical analysis of historical climate data. These approaches estimate future rainfall based on predefined mathematical equations and meteorological observations. Although statistical models have been extensively used for decades, their prediction capability is often limited when handling highly nonlinear climatic relationships. Recent studies have introduced Artificial Intelligence and Machine Learning techniques for rainfall prediction. Algorithms such as Support Vector Machines, Artificial Neural Networks, Decision Trees, and Deep Learning models have improved forecasting accuracy to some extent. However, several existing systems still face multiple challenges due to the complexity and uncertainty associated with atmospheric conditions. Most conventional rainfall prediction models require extensive manual feature engineering and are highly sensitive to noisy or missing data. Moreover, many deep learning approaches demand large computational resources and longer training time, making them unsuitable for real-time prediction in resource-constrained environments.

A. Limitations of Existing System

1) Low prediction accuracy for highly nonlinear weather conditions.
2) Difficulty in handling missing and noisy datasets.
3) High computational complexity for deep learning models.
4) Poor scalability with continuously increasing weather data.
5) Limited capability in identifying hidden relationships among meteorological variables.
6) Increased prediction error due to inadequate preprocessing techniques.

These limitations motivate the development of an efficient machine learning framework capable of producing more reliable rainfall predictions.

IV. PROPOSED SYSTEM

The proposed rainfall prediction system utilizes supervised machine learning techniques to forecast rainfall using historical meteorological data. The system incorporates Multiple Linear Regression (MLR), Random Forest Regressor, and K-Nearest Neighbor (KNN) Regression algorithms to analyze weather parameters and predict rainfall with improved accuracy.Initially, the historical weather dataset is collected from publicly available meteorological repositories. The dataset contains multiple climatic attributes including temperature, humidity, atmospheric pressure, wind speed, evaporation, sunshine duration, cloud cover, and previous rainfall observations.

The collected data undergoes preprocessing to eliminate missing values, remove duplicate records, normalize numerical features, and transform categorical attributes into numerical representations. Feature selection techniques are then applied to identify the most influential weather parameters contributing to rainfall prediction. The processed dataset is divided into training and testing subsets using a 70:30 ratio. Each regression algorithm is trained independently using the training dataset. After model training, prediction performance is evaluated using unseen testing data. Finally, the prediction accuracy of all three regression models is compared using standard evaluation metrics.

The proposed framework aims to reduce prediction errors while improving forecasting accuracy by effectively modeling nonlinear relationships present in meteorological data.

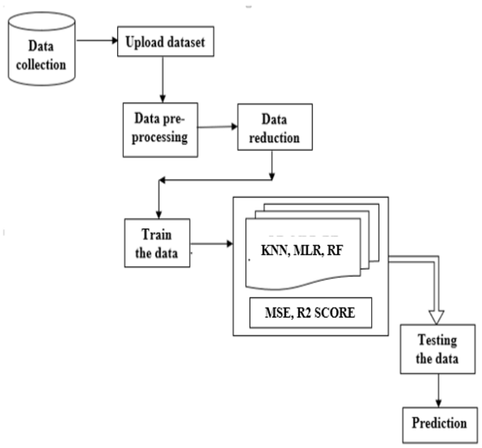

Figure 1: Proposed System Architecture

The proposed workflow consists of the following phases:

1) Dataset Collection
2) Data Pre-processing
3) Feature Selection
4) Data Splitting
5) Model Training
6) Rainfall Prediction
7) Performance Evaluation
8) Comparative Analysis

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

V. METHODOLOGY

The proposed rainfall prediction model employs supervised machine learning techniques to estimate rainfall using historical meteorological data. The complete workflow consists of dataset acquisition, preprocessing, feature engineering, model training, prediction, and performance evaluation. The objective is to develop an efficient prediction model capable of accurately forecasting rainfall using multiple climatic parameters. Initially, the historical weather dataset is collected from publicly available meteorological repositories. The dataset contains important weather attributes such as temperature, humidity, atmospheric pressure, wind speed, cloud cover, evaporation, sunshine duration, and rainfall measurements. These parameters significantly influence rainfall occurrence and therefore serve as the primary input variables for prediction. Since real-world datasets often contain incomplete, inconsistent, and noisy information, the collected data undergoes several preprocessing operations. Missing values are handled using appropriate statistical techniques, duplicate records are removed, categorical attributes are encoded, and numerical features are normalized to ensure uniformity among variables. These preprocessing techniques improve the learning capability of machine learning algorithms and reduce prediction errors.

After preprocessing, the dataset is divided into training and testing datasets using a 70:30 ratio. The training dataset is utilized to construct prediction models, whereas the testing dataset evaluates the generalization capability of each algorithm. Three supervised machine learning algorithms, namely Multiple Linear Regression (MLR), Random Forest Regressor (RFR), and K-Nearest Neighbor (KNN) Regression, are implemented and trained independently.

The trained models generate rainfall predictions using unseen weather observations. Finally, the performance of each model is compared using statistical evaluation metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2 Score). The model exhibiting the highest prediction accuracy and lowest prediction error is selected as the optimal rainfall prediction model.

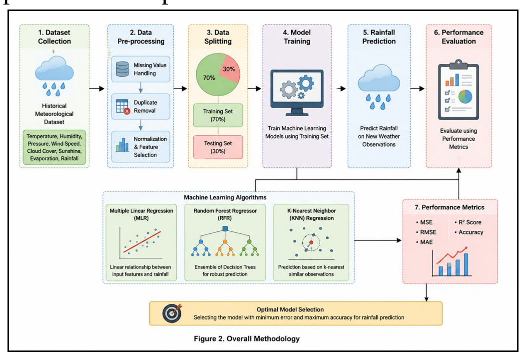

Figure 1: Overall Methodology

A. Dataset

The proposed system utilizes a historical rainfall dataset containing meteorological observations collected over several years. The dataset consists of multiple weather parameters that directly or indirectly influences rainfall patterns. Each record represents the atmospheric conditions observed during a specific period.The dataset contains approximately several thousand instances with multiple numerical attributes representing environmental conditions. Before model development, each attribute is carefully analyzed to understand its contribution toward rainfall prediction.

The major input attributes considered in this study are listed below.

Table 1: Dataset Attributes

S.No	Attribute	Description
1	Temperature	Average atmospheric temperature (°C)
2	Humidity	Relative humidity (%)
3	Pressure	Atmospheric pressure (hPa)
4	Wind Speed	Wind velocity (km/h)
5	Cloud Cover	Percentage of cloud coverage
6	Sunshine	Daily sunshine duration (hours)
7	Evaporation	Water evaporation level
8	Rainfall	Historical rainfall measurement (mm)

The rainfall attribute serves as the target variable, whereas all remaining attributes function as predictor variables.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

VI. RESULTS

The proposed rainfall prediction system was implemented using Python and evaluated on a historical rainfall dataset containing multiple meteorological parameters. The experiments were conducted using Multiple Linear Regression (MLR), Random Forest Regressor (RFR), and K-Nearest Neighbour (KNN) Regression algorithms. Before model training, the dataset was preprocessed by handling missing values, removing duplicate records, normalizing numerical features, and selecting significant attributes. The processed dataset was divided into 70% training data and 30% testing data for model development and evaluation.

The prediction performance of each algorithm was analyzed using standard regression evaluation metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2 Score). Lower values of MSE, RMSE, and MAE indicate better prediction capability, whereas a higher R^2 Score represents improved goodness of fit.

The experimental analysis demonstrates that the Random Forest Regressor consistently achieved better prediction accuracy than Multiple Linear Regression and KNN Regression. The ensemble learning strategy employed by Random Forest effectively captures nonlinear relationships among meteorological variables and reduces overfitting, leading to superior forecasting performance.

A. Home Page

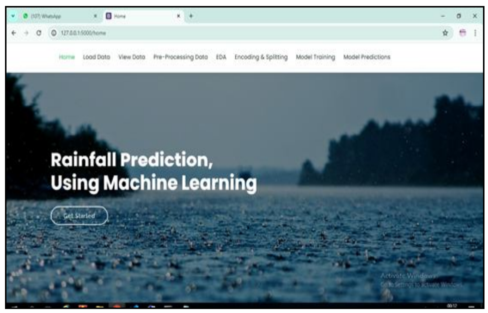

Figure 3: Rainfall Prediction home page

B. Upload Dataset

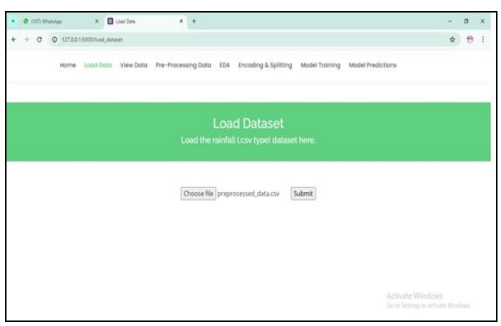

Figure 4: Upload Dataset

Figure 5.Annual Rainfall Prediction

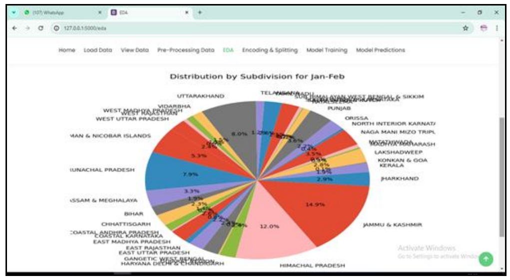

Figure 6: Distribution of Annual Rainfall

VII. CONCLUSION

Rainfall prediction is an important application of machine learning that supports agriculture, water resource management, disaster preparedness, and environmental planning. In this work, a machine learning-based rainfall prediction system was developed using Multiple Linear Regression (MLR), Random Forest Regressor (RFR), and K-Nearest Neighbor (KNN) Regression algorithms. The proposed framework includes data preprocessing, feature selection, model training, and performance evaluation to improve the accuracy of rainfall prediction.

Experimental analysis demonstrated that the Random Forest Regressor produced superior performance compared with Multiple Linear Regression and KNN Regression. The model achieved lower prediction errors and a higher coefficient of determination due to its ensemble learning capability and robustness against overfitting. Proper preprocessing of meteorological data further enhanced prediction accuracy by eliminating inconsistencies and improving data quality. The proposed system provides an efficient, reliable, and scalable rainfall prediction solution that can assist farmers, researchers, meteorological departments, and disaster management authorities in making informed decisions. The integration of machine learning techniques into weather forecasting has significant potential to improve prediction accuracy and support sustainable resource management.

REFERENCES

[1] A. S. Kumar, P. Geetha, and R. Vinaya Kumar, “Deep Learning Models for Rainfall Prediction,” International Conference on Communication and Signal Processing, pp. 125–130, 2019.
[2] S. Chatterjee, B. Datta, S. Sen, and N. Dey, “Rainfall Prediction Using Hybrid Neural Network Approach,” International Conference on Signal Processing, Telecommunications and Computing, pp. 89–95, 2019.
[3] M. Qiu, P. Zhao, K. Zhang, J. Huang, and W. Chu, “A Short-Term Rainfall Prediction Model Using Multi-Task Convolutional Neural Networks,” IEEE International Conference on Data Mining, pp. 395–404, 2019.
[4] S. K. Mohapatra, A. Upadhyay, and C. Gola, “Rainfall Prediction Based on 100 Years of Meteorological Data,” International Conference on Computing and Communication Technologies, pp. 201–207, 2020.
[5] M. Chattopadhyay and S. Chattopadhyay, “Predictive Models for Indian Summer Monsoon Rainfall Using Support Vector Machine,” Theoretical and Applied Climatology, vol. 121, no. 3, pp. 621–634, 2018.
[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
[7] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[8] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[9] N. S. Altman, “An Introduction to Kernel and Nearest Neighbor Nonparametric Regression,” The American Statistician, vol. 46, no. 3, pp. 175–185, 1992.
[10] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, New York, NY, USA: Springer, 2013.
[11] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann, 2012.
[12] F. Chollet, Deep Learning with Python, 2nd ed. Manning Publications, 2021.
[13] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[14] S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Packt Publishing, 2019.
[15] T. Mitchell, Machine Learning. McGraw-Hill, 1997.