IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 10, October 2025

DOI: 10.17148/IARJSET.2025.121005

A Review of Predictive Models for Agro-Meteorological Data Using Machine Learning

Danish Nawaz[1], Manisha S[2], Chandan Hegde[3]

Student, Department of MCA, Surana College, Bengaluru, India[1]

Student, Department of MCA, Surana College, Bengaluru, India [2]

Assistant Professor, Department of MCA, Surana College, Bengaluru, India [3]

Abstract: Karnataka's agricultural sector is very sensitive to climatic variability with regard to timing and distribution of monsoon rainfall. In response to the urgent need for reliable Agro-Meteorological forecasting, this study investigated the potential use of two supervised machine learning models Decision Tree and Random Forest for precipitation forecasting in Karnataka. The models were trained using historical meteorological datasets that were developed from the results of a number of climate parameters related to, for example, atmospheric circulation patterns, sea surface temperature, and specific monsoon-related indices. For both models, the combination of recursive partitioning and ensemble learning incorporated numerous distributed climate factors and their associated nonlinear relationships. Both models were then critically compared, with Random Forest being determined as most superior and a better overall estimation in terms of quality and reliability, trend association and being more robust to overfitting. In summary, these results demonstrate that data-driven approaches can dramatically improve the geographic and temporal quality of forecasts in this region. The framework proposed is a scalable, interpretable, and practical tool to aid climate resilient agricultural planning in Karnataka, and more generally Agro-climatic zones in India.

Keywords: Karnataka agriculture, climatic variability, monsoon rainfall forecasting, agro-meteorology, machine learning, Decision Tree, Random Forest, precipitation prediction, climate parameters, ensemble learning, nonlinear relationships, overfitting robustness, data-driven forecasting, climate-resilient agriculture, agro-climatic zones of India

I. INTRODUCTION

Predictive rainfall is fundamental for agriculture, water management, and hazard mitigation in places that rely heavily on monsoon systems. Rainfall—a phenomenon influenced by complex atmospheric dynamics—often occurs in nonlinear and non-stationary regimes, which is difficult to capture using traditional forecasting model techniques. Although traditional regression-based methods are capable at some scales, they often are unable to sufficiently generalize to novel physiographic and climatic zones.

The recent developments in artificial intelligence (AI) and machine learning (ML) have opened new possibilities for rainfall forecasting. Research shows that data fusion using AI which couples multiple satellite and reanalysis datasets minimizes both systematic and random errors in rainfall estimates. Hybrid applications, for example empirical mode decomposition using random forest regression, have demonstrated improved performance in capturing annual rainfall variability in Kerala, India. Comparing geostatistical techniques against ML models, machine learning (especially ensemble algorithms) continues to provide improved performance for dealing with the spatial heterogeneity of rainfalls Additionally, smart city apps have used machine learning fusion models that combine decision trees, Naïve Bayes, KNN, and SVM approaches using fuzzy logic and predict from real time urban weather data. Also, in Ethiopia, extreme gradient boosting (XGBoost) has demonstrated a greater effectiveness than traditional regression approaches, in predicting daily rainfall. As a whole, these studies emphasize the potential for hybrid, ensemble, and AI-based models to overcome the drawbacks of traditional approaches. This article provides their approaches and results, to provide a clear picture of the current developments in rainfall prediction, and to identify gaps for future research.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 10, October 2025

DOI: 10.17148/IARJSET.2025.121005

II. LITERATURE REVIEW

Advancements in weather forecasting have recently emphasized the potential for machine learning models, especially Decision Trees (DT) and Random Forests (RF), to predict rainfall patterns which are important for agricultural planning. Jayasree et al. (2023) developed a hybrid model consisting of Empirical Mode Decomposition (EMD) and Random Forest to predict annual rainfall for Kerala. Based on the EMD methodology, which decomposes rainfall time series signals into different intrinsic components, they have sought to build upon literature, which demonstrated that their hybrid EMD-RF was more accurate and robust than the ARMA and RF models alone. This study focused on predicting annual rainfall and therefore did not capture the level of regional and seasonal variability related to agricultural contexts such as in Karnataka.

Likewise, Xiang et al. (2020) employed Decision Tree and Stochastic Forest algorithms that were used to predict summer rainfall in Chongqing, China. Their work validated ensemble tree-based models had higher trend consistency and predictive accuracy than single-factor models. Their methods were thorough, but did not account for geographical aspects of Indian monsoonal trends, nor did they consider agricultural implications.

In another study, Khan and Bhuiyan (2021) used multiple satellite and reanalysis datasets and fused them through multiple machine learning approaches—including RF—to enhance rainfall estimates over the Upper Blue Nile Basin. While the method decreased errors considerably, it did not address location-based agricultural information, nor it performed an analysis on rural-urban or crop-sensitive sub-regions（2）.

In the agricultural-oriented literature, Suljug et al. (2024) analyzed a total of 19 different machine learning models for predicting agrometeorological parameters affecting maize crops in Croatia. Random Forest was arguably one of the best performing models overall, but particularly effective when the data was severed by land use in urban, rural and suburban areas. Although their results were inspiring, they cannot be accurately extrapolated to the conditions in India including the semi-arid and monsoonal lands in Karnataka.

At a more regional level, Agnihotri and Mohapatra (2011) used a statistical stepwise regression to analyze variations in monsoon rainfall across 19 stations in Karnataka. Predictive skill of POP model was reasonable, but less flexible and was unable to capture any potential nonlinear relationships inherent in the data compared to tree-based algorithms such as DT and RF..

Birant et al. (2025) proposed a more contemporary innovation: the Temporal Random Tree (TRT) model, which weighs more recent information more heavily in rainfall classification tasks. While the TRT model gained improved performance over a traditional Random Forests, it remains largely unexplored in the Indian context of assessing rainfall.

Liyew and Melese (2021) carried out systematic comparative research between the Random Forest, XGBoost, and linear models in Ethiopia and found that ensemble models are usually easier to use and generate better predictive results in daily rainfall forecast. Their study also confirmed the strength of tree-based ensembles but did not examine monsoon specific criteria.

In summary, these studies confirm that Decision Tree and Random Forest algorithms are useful in rainfall modeling. However, most do not focus on the agro climatic complexity of Karnataka or specifically lack the interpretability and understanding required for farmers' decision making. The current research aims to address that gap by modifying and applying these models in the approximation of the physiographic regions of Karnataka, and investigating their applicability for supporting climate-smart agriculture.

Title	Authors / Year	Objective	Links	Findings	Research Gap
Decision Trees & Random Forest vs Other Models in Kashmir	2022, Kaul et al.	Compare DT, RF, SVM, KNN etc. for rainfall prediction in Kashmir	https://www.mdpi.com/2073-445x/11/12/2180	RF and ANN most accurate (~81 %), DT moderate (~78 %)	Seasonal factor plateau; no feature selection based on meteorological indices
The Application of a Decision Tree and	Chinese Meteorology Researchers (2020)	Build a decision tree classifier for summer precipitation in	https://www.mdpi.com/2073-4433/11/5/508	Random Forest reduced overfitting and improved classification of	Focused on one region and only summer season; lacks temporal

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 10, October 2025

DOI: 10.17148/IARJSET.2025.121005

Stochastic Forest Model in Summer Precipitation Prediction in Chongqing		Chongqing and evaluate Random Forest ensemble.		summer precipitation events.	component modeling beyond static features.
Regression Tree Ensemble Rainfall–Runoff Forecasting Model and Its Application to Xiangxi River, China	Zhai, A., Fan, G., Ding, X., Huang, G. (2022)	Use ensemble regression trees to forecast rainfall–runoff relationships.	https://www.mdpi.com/2073-4441/14/3/463	Ensemble tree models produced accurate runoff forecasts based on rainfall inputs.	Focus on hydrological runoff rather than direct rainfall prediction; further work needed on meteorological variable inclusion.
Artificial Intelligence-Based Techniques for Rainfall Estimation Integrating Multisource Precipitation Datasets	MDPI Atmosphere Authors (2021)	Fuse multiple precipitation sources and apply Decision Tree and Random Forest models.	https://www.mdpi.com/2073-4433/12/10/1239	Tree-based models with hyperparameter tuning achieved superior error metrics; identified key variables.	Integration only at gauge-regional scale; lacks temporal dynamics and daily forecasts for agriculture.
A Comparative Study of Machine Learning Models for Predicting Meteorological Data in Agricultural Applications	MDPI Electronics (2022)	Evaluate DT, RF, SVM, and GPR for predicting multiple agrometeorological parameters.	https://www.mdpi.com/2079-9292/13/16/3284	Bagged Trees and GPR performed well; DT and RF performed decently for some targets.	Focused on parameters other than rainfall; lacks study on rainfall quantity/time and local-scale food production.
A new machine learning method for rainfall classification: temporal random tree	PeerJ Paper (2025)	Introduce Temporal Random Tree for binary rainfall classification with temporal weighting.	https://peerj.com/articles/cs-3022/	Improved binary classification performance and higher interpretability than complex models.	Only classification (rain/no rain), not regression; needs extension to quantitative rainfall prediction.
Machine learning techniques to predict daily rainfall amount	Big Data Journal (2021)	Evaluate MLR, RF, XGBoost to predict daily rainfall using station data.		XGBoost outperformed others; RF still acceptable.	DT algorithms not deeply examined; lacks multi-location validation and sensor data integration.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 10, October 2025

DOI: 10.17148/IARJSET.2025.121005

Hybrid EMD-RF Model for Predicting Annual Rainfall in Kerala, India	2023, Jayasree et al.	Enhance annual rainfall forecasting using EMD + RF decomposition	https://www.mdpi.com/2076-3417/13/7/4572	Hybrid model outperformed plain RF and ARMA models (lower MAE, higher R^2)	Annual scale only; no seasonal/monthly resolution; state-wide only
Rainfall Prediction System Using ML Fusion for Smart Cities	2022, Gupta et al.	Use hybrid ML frameworks (SVM+DT etc.) for rainfall yes/no classification	https://www.mdpi.com/1424-8220/22/9/3504	Hybrid SVM+DT reached ~80 % accuracy over 14-year data	Classified only binary rainfall occurrence; not amount or seasonal prediction
Precipitation occurrence modelling over Karnataka	2012, Agnihotri	Predict daily summer monsoon rainfall occurrence via statistical model	https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/met.246	POP model had >95 % detection along coast; skillful vs IMD methods	Used climatology-based POP; no ML approach

III. FUTURE ENHANCEMENT

This investigation can be extended to take a systematic approach to develop rainfall prediction models for Karnataka's agro-climatic zones based on Decision Tree (DT) and Random Forest (RF) algorithms. The sub-parts of this section present the methodology designed as the intersection of the technical robustness of machine learning workflows and the practical importance of agrometeorological forecasting for decision-making within farming.

1. Study Area and Climatic Context

Karnataka, located in southern India, has various rainfall patterns due to its wide-ranging topography and physiography. There are three meteorological divisions within Karnataka, Coastal Karnataka (CK), South Interior Karnataka (SIK) and North Interior Karnataka (NIK). Rainfall prediction differs in each zone because of elevation, distance from the Western Ghats, and atmospheric conditions.

2. Data Collection

Daily historical weather data from 1981 to 2020 can be collected from the India Meteorological Department (IMD) and Indian Institute of Tropical Meteorology (IITM). The dataset includes:

- Daily rainfall (mm)
- Maximum and minimum temperature (°C)
- Relative humidity (%)
- Atmospheric pressure (hPa)
- Wind speed (km/h)
- Sea surface temperature (SST)
- Monsoon onset and retreat dates
- ENSO and IOD indices (from NOAA and IITM sources).

3. Data Preprocessing

To ensure high-quality input for the models, the dataset needs to undergo multiple preprocessing steps:

- **Missing value treatment:** Linear interpolation and mean imputation **will be applied** based on temporal continuity and climatic season.
- **Outlier detection:** Z-score and IQR-based methods **will be used** to identify and handle extreme anomalies.
- **Normalization:** Min-Max scaling **will be used** to transform all variables within the 0–1 range to prevent feature dominance during training.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 10, October 2025

DOI: 10.17148/IARJSET.2025.121005

Temporal aggregation: In addition to daily data, 7-day and 15-day rolling averages **will be calculated** to capture short-term climatic trends.

IV. CONCLUSION

This review has highlighted the emerging importance of machine learning to address the challenges of agro-meteorological forecasting. Regression approaches, such as the Probability of Precipitation (POP) framework in regards to forecasting rainfall developed in Karnataka, provided useful first steps for understanding rainfall dynamics, but cannot address the nonlinear aspects of climatic dynamics. Further progress has shown the efficacy of AI, in that exploring multisource data via integration models designed with algorithms such as Random Forest, Gradient Boosting, and neural networks, has significantly decreased the errors of rainfall estimations. Hybrid models have also been introduced for example, empirical mode decomposition combined with Random Forest, which improved prediction accuracy by managing the non-stationary properties of the rainfall time series.

Numerous comparative studies have documented the comparative utility of the machine learning ensembles, which routinely outperform classical statistical and geostatistical approaches in addressing the spatial heterogeneity and temporal variability of585 precipitation.

As well, urban applications in particular have responded to the variety rainfall prediction approaches using developed ensemble approaches as well as developing urban frameworks of combining tools under intelligent cities applications tapping into smart city potential and possibilities respectively.

This summary is quite damning on the classical models so far in terms of practical solutions but as previously discussed there is good reason to suggest some confidence with machine learning, given that they offer not only the improved prediction quality but arguably improved interpretability and scalability over traditional models and methods. However, much work remains to do. Most of the studies reported, have either restricted their examples to single regions or short timescales, and thus generalisation is problematic for agro-climatic zones with significant physiographic diversity particularly regions similar to Karnataka. Moreover, several points require further inquiry across continents, but very few studies address spatio-temporal heterogeneity including a few that used temporally distributed sub-samples.

REFERENCES

[1]. Agnihotri, G., & Mohapatra, M. (2011). Prediction of occurrence of daily summer monsoon precipitation over Karnataka. *Meteorological Applications*, 19(2), 130–139

[2]. Khan, R. S., & Bhuiyan, M. A. E. (2021). Artificial intelligence-based techniques for rainfall estimation integrating multisource precipitation datasets. *Atmosphere*, 12(10), 1239

[3]. Jayasree, A., Sasidharan, S. K., Sivadas, R., & Ramakrishnan, J. A. (2023). Hybrid EMD-RF model for predicting annual rainfall in Kerala, India. *Applied Sciences*, 13(7), 4572

[4]. Farooq, I., Bangroo, S. A., Bashir, O., Shah, T. I., Malik, A. A., Iqbal, A. M., et al. (2022). Comparison of random forest and kriging models for soil organic carbon mapping in the Himalayan region of Kashmir. *Land*, 11(12), 2180

[5]. Rahman, A., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M. A., & Mosavi, A. (2022). Rainfall prediction system using machine learning fusion for smart cities. *Sensors*, 22(9), 3504

[6]. Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8, 153

[7]. Xiang, Y (2020). Decision tree and stochastic forest methods for rainfall prediction in Chongqing, China.

[8]. Suljug, A (2024). Comparative analysis of ML models for agrometeorological predictions in Croatia.

[9]. Birant, D (2025). Temporal Random Tree model for rainfall prediction.

[10]. https://mausam.imd.gov.in/responsive/rainfallinformation_state.php?msg=C