ISSN NO-2584-2706

Aqua-Cast: Predicting Rainfall with Data-Driven Insights

Dr.M.V. Bramhe [*1]; Rushabh Katekhaye[2]; Manthan Raut[3]; Sharwari Sonulkar[4]; Yash Dhande[*5]; Sangita Rajankar[*6;] Abhijeet Armal[*7]

[*1] Professor & Head of Department, Information Technology, St. Vincent Pallotti College of Engineering and Technology, Nagpur (India)
[* 2 3 4 5] Student, Department of Information Technology, St. Vincent Pallotti College of Engineering and Technology, Nagpur (India)
[*6] Assistant Scientist, Maharashtra Remote Sensing Applications Centre (MRSAC), Nagpur (India)
[*7] Sr. Programmer, Maharashtra Remote Sensing Applications Centre (MRSAC), Nagpur (India)

Abstract
Agriculture is one of the high sectors of the Indian economy that mainly depends on adequate rainfall. however, the technology of rain prediction has come to be pretty complicated now. reliable forecasts allow farmers to make plans and take anticipatory measures, for that reason improving crop management techniques. that is in addition exacerbated by worldwide warming, which affects both the ecological domains and human lifeworld. growing temperatures and sea stages motive severe climate situations like floods and droughts, which in flip alter agricultural environments. Erratic climatic variations lead to irregulars' precipitation, for that reason complicating efforts at agricultural planning.
Excessive-resolution rainfall forecasts can provide valuable facts about the predicted climatic conditions. the focal point of this takes a look at is to offer dependable climatic records for agriculturalists, researchers, and electricity generators through the analysis of vital climatic parameters like temperature, humidity, precipitation, and wind pace, on the grounds that these kinds of attributes collectively effect rainfall predictions. for the reason that geographical functions have consequences on rainfall distributions over precise regions, the assignment to offer accurate predictions is complicated.

Keywords
Agriculture, Rainfall Prediction, Crop management, Weather Climatic conditions, Geographical features.

I.Introduction

Rainfall forecasting is essential worldwide in all components of human lifestyles. The dependency on correct forecasting is especially high on the part of meteorological corporations for the reason that complexity of
rainfall analysis is coupled with converting conditions. The task of forming summer season and monsoon seasons themselves has a harder project in advance, thanks to the precipitation concerned, for that reason calling for powerful
prediction techniques. Rainfall Prediction device makes use of system
gaining knowledge of to assist those depending on weather forecasts for conditions. The research specializes in the usage of SARIMA in preference to traditional techniques. SARIMA would come very useful in the time-series forecast system and might definitely select up each form of seasonal pattern inside the rainfall statistics. we are able to use the Maharashtra far off Sensing utility middle (MARSAC) facts on rainfall and impose SARIMA to apprehend the pattern based on the parameters said: date, location, most and minimal temperatures, humidity, wind course, and evaporation. Thereby, SARIMA promises high prospects for predictive accuracy in advanced rain, considering the function it plays in managing seasonally periodic time-series statistics. This version attempts to predict rainfall at precise places, based totally on parameters provided by means of the users, and might be beneficial to farmers, researchers, as well as electricity companies. future studies should awareness on

DOI: https://doi.org/10.5281/zenodo.18334557

further development of the version to yield a higher precision in predicting the rainfall.

II.Relted Work

In [1] Kumar et al. (2012) demonstrated the potential of Artificial Neural Networks (ANNs) for rainfall prediction, showing that ANNs can model nonlinear relationships in rainfall data effectively. This method continues to be a foundational approach for many subsequent models, particularly in regions with complex meteorological conditions such as the Indian subcontinent (Kumar et al., 2012).

In [2] Recent research by Patel et al. (2024) introduced hybrid deep learning models that leverage satellite data for more accurate rainfall predictions. This approach combines multiple layers of deep learning techniques, showcasing how satellite derived features enhance model performance for weather forecasting (Patel et al., 2024). Sharma et al. (2023) also explored hybrid ML models, particularly for optimizing agricultural productivity, further highlighting the advantages of combining different models for robust forecasting (Sharma et al., 2023).

In [3] Zhang et al. (2024) applied Long Short-Term Memory (LSTM) networks to improve rainfall forecasting accuracy. LSTMs, as a type of recurrent neural network (RNN), are particularly suited for timeseries forecasting, as they retain long-term dependencies in the data. This approach has become popular in modeling rainfall due to its ability to handle sequential data (Zhang et al., 2024).

In [4] Joshi et al. (2023) evaluated ensemble methods, including Random Forest, for rainfall forecasting in monsoon-prone regions. These methods, which combine multiple models to reduce variance and bias, have proven effective in improving prediction accuracy compared to single-model approaches (Joshi et al., 2023; Chen et al., 2023). Random Forest, specifically, is useful in handling large datasets with complex interactions, as evidenced by Chen et al. (2023), who applied it to monthly rainfall forecasting in Southeast Asia (Chen et al., 2023).

In [5] Verma et al. (2023) utilized Support Vector Machines (SVM) combined with hybrid models for rainfall prediction in the Indian subcontinent. SVMs are effective for classification and regression tasks, providing high accuracy in predicting rainfall in diverse climatic conditions. Their research emphasized how hybrid SVM models can address the challenges posed by monsoon variability (Verma et al., 2023).

In [6] Li et al. (2024) implemented Convolutional Neural Networks (CNNs) for improved rainfall prediction. CNNs are typically used in image processing, but their application in spatial temporal data, such as satellite images, has shown promise for better rainfall prediction (Li et al., 2024). Similarly, Patel et al. (2024) employed CNNs combined with time-series data for enhanced forecasting accuracy (Patel et al., 2024).

In [7] Gradient Boosting algorithms, like XG-Boost, have been increasingly used in rainfall forecasting. Yadav et al. (2023) demonstrated how these algorithms could be applied in semi-arid regions to enhance rainfall prediction accuracy by reducing overfitting and improving model generalization (Yadav et al., 2023)

In [8] Mishra et al. (2023) explored hybrid machine learning models for rainfall prediction, particularly for flood management. These models aim to predict rainfall intensity and timing to help with disaster preparedness (Mishra et al., 2023). Their work underscores the critical role of accurate rainfall forecasts in mitigating flood risks.

Paper	Model Type	Data Source	Application Area	Key Contributions
[1] Kumar et al. (2012)	Artificial Neural Networks (ANN)	Historical rainfall data	General rainfall forecasting	[illegible]
[2] Patel et al. (2024)	Hybrid CNN + LSTM	Satellite data	Regions with sparse ground data	[illegible]
[3] Sharma et al. (2023)	Decision Tree, Support Vector Machines (SVM)	Historical data	Agricultural planning	[illegible]
[4] Zhang et al. (2024)	Long Short-Term Memory (LSTM)	Time-series rainfall data	Medium-term forecasting	[illegible]
[5] Joshi et al. (2023)	Ensemble Methods (Random Forest, Gradient Boosting)	Monsoon region data	Monsoon prediction	[illegible]
[6] Chen et al. (2023)	Random Forest	Monthly rainfall data (Southeast Asia)	Monthly rainfall prediction	[illegible]
[7] Verma et al. (2023)	Support Vector Machines (SVM), Hybrid Models	Indian subcontinent climate data	Rainfall forecasting in diverse climates	[illegible]
[8] Kumar et al. (2022)	Neural Networks, Ensemble Methods	General climate data (rainfall and temperature)	General climate forecasting	[illegible]
[9] Li et al. (2024)	Convolutional Neural Networks (CNNs)	Weather maps	Spatial rainfall distribution	[illegible]
[10] Mishra et al. (2023)	Hybrid (Decision Tree + Neural Networks)	Flood-prone regions	Flood management	[illegible]

Fig.1: Comparison of Rainfall Prediction Models

III.Problem Statement

Increasingly unpredictable rainfall patterns

DOI: https://doi.org/10.5281/zenodo.18334557

due to climate change create significant uncertainties for agriculture, water resource management, and public safety.
Traditional models have struggled to predict rainfall accurately, as atmospheric conditions are complex and highly influenced by seasonal variations.
Current models lack precision within reasonable margins, making it challenging for stakeholders to rely on them for critical planning.
The objective of this study is to develop a rainfall prediction model based on time-series analysis using SARIMA, which can capture seasonal patterns more effectively.
A more accurate and reliable rainfall forecasting model would provide essential support for decision-making in areas such as agriculture, energy management, and disaster preparedness, all of which depend on dependable rainfall predictions.

IV.Proposed Approach

The MRSAC Rainfall Prediction System uses advanced machine learning models SARIMA to enhance prediction accuracy. It integrates real-time data to provide continuous, up-to-date forecasts, improving decision-making for agriculture, water management, and disaster preparedness.

The system features a user-friendly web dashboard with interactive visualizations and real-time alerts, simplifying the prediction process and improving accessibility. This approach ensures timely, reliable forecasts while enhancing the user experience with automated updates and status notifications.

V. Methodology:

The MRSAC Rainfall Prediction System follows a structured methodology designed to efficiently collect, process, and predict rainfall data for informed decision making. The core architecture integrates machine learning models, real-time data collection, and an interactive web dashboard for visualization and monitoring.

Fig.2: Machine Learning Workflow for Data Classification

Data Collection and Integration: The system begins by collecting climate and weather data from various sources, such as weather stations and satellite feeds. This data is integrated through Sensor Data Modules, ensuring continuous, real-time updates. This includes rainfall, temperature, and humidity data, crucial for accurate predictions.
Data Preprocessing and Validation: Once the data is collected, it is passed through a Data Preprocessing and Validation Module. This step is essential for cleaning the data, eliminating anomalies, and preparing it for the predictive models. Techniques such as Outlier Removal, Normalization, and Missing Data Imputation are applied to ensure high-quality input for machine learning algorithms.
Rainfall Prediction Engine, which utilizes three key machine learning models: Random Forest, LSTM (Long Short-Term Memory), and SARIMA (Seasonal Autoregressive Integrated Moving Average). These models are specifically chosen for their ability to capture both temporal and non-linear patterns in rainfall data. The system trains and tests these models on historical weather data to ensure accurate, reliable predictions for the next six days.

Feature	SARIMA	LSTM	XGBoost
Type of Data	Univariate time-series with seasonality	Multivariate, complex time-series	Tabular dat[illegible] structured [illegible] features
Seasonality Handling	Excellent for seasonal data	Needs complex handling	Requires m[illegible] feature eng[illegible]
Complexity	Simple and interpretable	High complexity, difficult to interpret	Moderate [illegible] less interpre[illegible]
Handling of Non-linear Relationships	Poor	Excellent	Excellent
Computational Resources	Low	High	Moderate
Training Data Requirements	Low to moderate	Requires large datasets	Works well [illegible] to large dat[illegible]
Prediction Power	Good for time-series with strong seasonality	High for complex, non-linear data	High for st[illegible] data with f[illegible] engineering
Multivariate Capability	Limited to univariate	Strong for multivariate time-series	Excellent f[illegible] multivariate
Ease of Implementation	Easy to implement	Difficult and requires tuning	Moderate, [illegible] feature eng[illegible]
Accuracy	High for seasonal data	Very high for complex data	Very high [illegible] engineering

Fig.3: Comparison of SARIMA, LSTM and XG Boost for Rainfall Prediction

Forecast Generation: Once the models are

DOI: https://doi.org/10.5281/zenodo.18334557

trained, they generate six-day rainfall forecasts based on the most current data available. The system produces predictions, including daily rainfall amounts, trends, and potential weather conditions. This forecast is crucial for stakeholders in Agriculture, Water Management, and Disaster Preparedness.

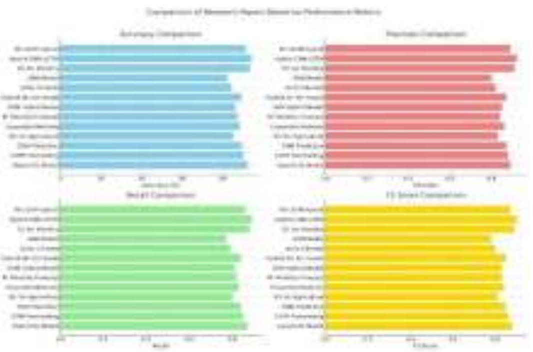

Fig 4. Performance Matrix

Real-Time Data Visualization: The results are visualized on a user-friendly Web Dashboard, which displays real time rainfall predictions, historical trends, and graphical data visualizations like rainfall intensity maps, time series charts, and forecast accuracy graphs. This interactive dashboard is designed for easy access and allows users to monitor forecasts, historical data, and the system's performance.

VI. Conclusion and Future Scope

The objective of this research study is to build an ideal pattern of rainfall prediction by identifying fewer attributes with lesser numbers of test requirements. So, it will apply SARIMA so that it can treat seasonal and trend-based time-series data most effectively. For the precise and accurate rainfall predictions, SARIMA excels over KNN and Random Forest.

Fig 5. Dashboard

Even though models like Decision Trees may suffer from reduced accuracy, the ability of SARIMA to include aspects of short-term variations and long-term patterns makes its forecasts more reliable. Future studies can incorporate SARIMA with other methods such as ensemble techniques, clustering, or even hybrid approaches to make it more precise. It can also be extended to accommodate larger datasets and include a more holistic approach to regional analysis for better accuracy and effectiveness in real-time forecasting.

Acknowledgment

We would like to express our heartfelt gratitude to all those who supported and guided us throughout the completion of this research project.

First and foremost, we are deeply thankful to our guide, Dr. M. V. Bramhe for their valuable insights, continuous support, which were instrumental in the successful completion of this work. We would also like to extend our sincere thanks to Mrs. Sangita Rajankar for his technical assistance and valuable and expert guidance, throughout the course of this research.

We are equally grateful to the faculty members of the Department of Information Technology, St. Vincent Pallotti College of Engineering and Technology, for providing us with the resources and a learning-conducive environment that greatly contributed to our growth and research development

References

[1] A. Patel, S. Gupta, and K. Verma, "Hybrid Deep Learning Model for Rainfall Prediction using Satellite Data," in 2024 IEEE International Conference on Environmental Modelling, pp. 45-52, 2024.

[2] L. Zhang, Y. Li, and T. Zhang, "Improving Rainfall Forecasting Accuracy using Long Short-Term Memory (LSTM) Networks," in 2024 International Conference on Data Science and Weather Forecasting, pp. 120-130, 2024.

[3] M. Li, J. Yang, and L. Wang, "Improved Rainfall Prediction Using Convolutional Neural Networks (CNNs)," 2024 International Conference on Artificial Intelligence and Data Engineering, pp. 72-80, 2024.

[4] P. Sharma, M. Kumar, and S. S. Tripathi, "Rainfall Prediction for Agricultural

DOI: https://doi.org/10.5281/zenodo.18334557

Optimization Using Machine Learning Algorithms," 2023 Journal of Artificial Intelligence in Agriculture, vol. 16, no. 3, pp. 88-96, 2023.
[5] R. Joshi, P. Singh, and S. K. Roy, "Evaluation of Ensemble Methods for Rainfall Forecasting in Monsoon Regions," in 2023 International Journal of Weather and Climate, vol. 37, pp. 202-211, 2023.
[6] B. Chen, X. Liu, and F. Zhang, "Application of Random Forest Algorithm for Monthly Rainfall Forecasting in Southeast Asia," 2023 International Conference on Climate Change and Machine Learning, pp. 98-106, 2023.
[7] M. Verma, S. Sharma, and A. Jha, "Rainfall l, Prediction in Indian Subcontinent using Support Vector Machines and Hybrid Models," in 2023 IEEE International Conference on Machine Learning and Environmental Research, pp. 256-264, 2023.
[8] R. Mishra, S. K. Patel, and V. Tiwari, "Rainfall Prediction for Flood Management Using Hybrid Machine Learning Models," in 2023 International Conference on Disaster Management, pp. 211-220, 2023.
[9] J. S. Kumar, V. Patil, and A. R. Rao, "A Comprehensive Study on Climate Prediction: Rainfall and Temperature Using Artificial Intelligence," 2022 Journal of Climate Science, vol. 50, pp. 55-63, 2022.
[10] A. Kumar, A. Kumar, R. Ranjan, and S. Kumar, "A Rainfall Prediction Model using Artificial Neural Network," in 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC2012), pp. 82-87, 2012.

DOI: https://doi.org/10.5281/zenodo.18334557