International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Recent Advances in Computer Applications and Information Technology
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Performance Analysis of Recurrent Neural Network Models for Rainfall Prediction in India

Vajrang Khadatkar, Sarvadnya Yawalkar

G H Raisoni University, Amravati, Maharashtra, India

Abstract

Rainfall plays a crucial role in shaping agriculture, water resource planning, and disaster management, especially in countries like India where millions of livelihoods depend on the seasonal monsoon. Accurate rainfall forecasting is therefore of great importance, yet traditional statistical models often struggle to capture the highly non-linear, uncertain, and time-dependent nature of rainfall patterns. With the advancement of artificial intelligence, deep learning techniques such as Recurrent Neural Networks (RNNs) have emerged as promising alternatives for handling sequential climate data. In this project, five RNN-based models—Simple RNN, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM, and Stacked LSTM—were implemented and evaluated on India's historical rainfall dataset covering the years 1901 to 2015. The dataset was preprocessed through normalization and transformed into time-series sequences for effective learning. Each model was trained and compared using regression metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), along with confusion matrices to classify rainfall levels into categories. The experimental findings reveal that advanced architectures like Bidirectional LSTM and Stacked LSTM achieved superior accuracy compared to the baseline Simple RNN. These results highlight the strong potential of deep learning methods in improving rainfall forecasting and support their application in climate modeling and decision-making for agriculture and disaster preparedness.

***KEYWORDS**: Rainfall Forecasting, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Time-Series Prediction, Deep Learning, Climate Modeling, Predictive Analysis, Monsoon Predictive.*

1. Introduction

Rainfall variability significantly influences agricultural production, water resource allocation, and disaster risk management in India. The Indian monsoon system exhibits complex spatial and temporal patterns, often leading to irregular rainfall distribution across regions. These fluctuations directly affect crop yields, reservoir storage, groundwater recharge, and flood management systems. Consequently, accurate and reliable rainfall prediction is essential for ensuring food security, economic stability, and climate resilience. Historically, rainfall forecasting has relied on statistical and empirical approaches. Methods such as regression modeling, time-series decomposition, and non-parametric trend detection have been widely applied to analyze long-term precipitation behavior. Research studies including [1] and [11] examined rainfall trends using statistical techniques and identified significant changes in monsoon variability across India. Although these approaches provide valuable insights into general rainfall tendencies, they often lack the capability to effectively model nonlinear relationships and long-term temporal dependencies present in meteorological data. With the advancement of computational intelligence, machine learning techniques have been introduced to improve forecasting accuracy. Ensemble learning and hybrid statistical–machine learning frameworks have demonstrated improved performance compared to conventional models [5][7]. Furthermore, the integration of spatial and temporal characteristics has enhanced predictive reliability in rainfall modeling [8][9]. Deep learning methods, particularly Recurrent Neural Networks (RNNs), have emerged as powerful tools for sequential data modeling. Architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are specifically designed to retain historical information over extended time horizons, making them well suited for rainfall prediction tasks [3][6]. Recent investigations also emphasize spatiotemporal deep learning frameworks and climate-aware modeling approaches to further enhance forecasting performance [10][14].

Despite these developments, limited studies provide a detailed comparative evaluation of multiple RNN architectures using long-term historical rainfall data of India. Therefore, this study aims to analyze and compare different recurrent neural network models to determine the most effective architecture for rainfall forecasting

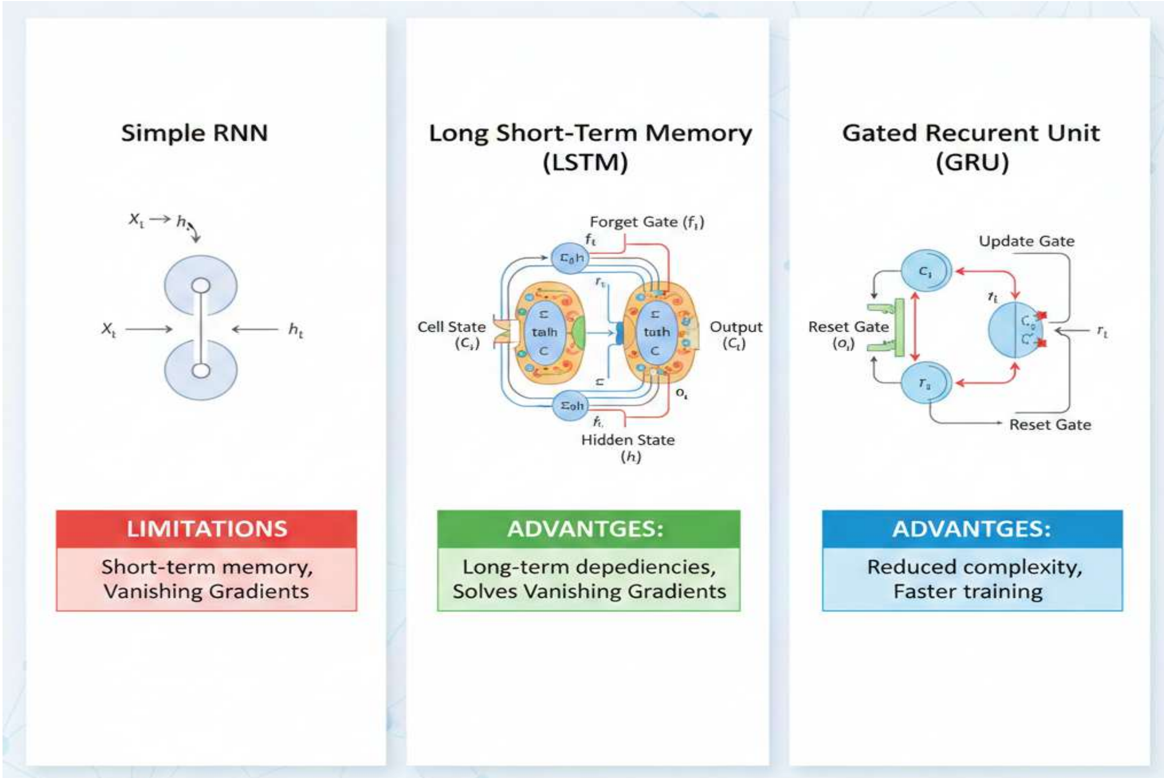

Fig 1: Comparative Architecture for Rainhall Forecasing

2. Literature Review

Rainfall trend assessment and prediction have evolved from classical statistical approaches to advanced machine learning and deep learning frameworks.

Initial studies primarily focused on statistical trend analysis techniques. Praveen *et al.* [1] applied non-parametric methods alongside machine learning models to evaluate rainfall changes across India, highlighting considerable spatial and temporal variability. Similarly, Guhathakurta and Rajeevan [11] investigated long-term rainfall patterns and reported noticeable shifts in monsoon behavior over decades. Mondal and Jain [4] conducted district-level rainfall trend analysis in the Gangetic region, revealing heterogeneous distribution patterns across districts.

Regional investigations have also emphasized the localized nature of rainfall variability. Choudhari *et al.* [2] compared rainfall trends in Konkan-Goa and Marathwada meteorological subdivisions, identifying substantial regional differences that necessitate tailored predictive frameworks. Singh *et al.* [12] analyzed seasonal rainfall patterns in central India, reinforcing the complexity of monsoon-driven precipitation systems. The introduction of machine learning methods marked a significant advancement in rainfall forecasting. Bhardwaj *et al.* [5] assessed multiple machine learning and ensemble techniques, demonstrating that ensemble-based approaches often achieve superior predictive performance. Rahman *et al.* [7] incorporated spatiotemporal modeling with machine learning algorithms, resulting in improved long-term rainfall predictions.

Deep learning approaches have further enhanced forecasting capability. Kumar *et al.* [3] compared machine learning, deep learning, and time-series techniques across an altitudinal gradient, concluding that LSTM-based models effectively capture complex temporal dependencies. Ahmad *et al.* [6] performed a comparative evaluation of rainfall prediction models and confirmed the robustness of deep learning architectures in modeling nonlinear rainfall dynamics. Recent research has also explored spatiotemporal deep learning frameworks. Hewage *et al.* [8] proposed a deep learning-based environmental prediction framework that integrates spatial and temporal components. Guo *et al.* [9] developed a spatial-temporal ensemble model for rainfall intensity forecasting, demonstrating enhanced prediction accuracy. Kumar *et al.* [10] utilized Earth Observation datasets to investigate long-term rainfall intensities, emphasizing the value of incorporating spatial datasets into predictive modeling.

Climate change impacts on rainfall patterns have been widely examined. Rajeevan *et al.* [13] analyzed changes in rainfall extremes across India and reported increasing variability in intensity and frequency. Das and Deka [14] studied the influence of climate change on rainfall distribution in northeastern India, highlighting potential implications for water resources and agriculture.

Overall, the literature reflects a clear transition from traditional statistical analysis toward intelligent predictive systems. However, comprehensive comparative analysis of different RNN architectures using long-duration Indian rainfall datasets remains relatively limited. This study contributes by systematically evaluating multiple recurrent neural network models to identify the most effective approach for rainfall forecasting in India.

3. Methodology

3.1. Dataset Information

The experimental study is based on the *Rainfall in India (1901–2015)* dataset obtained from Kaggle. The dataset contains more than 36,000 monthly rainfall records collected from various states and districts across India over a period of more than one century. Each record includes attributes such as State, District, Year, Month, and Rainfall (in millimeters), providing both temporal and spatial information necessary for accurate forecasting.

For this research, the primary features used were Year, Month, and Rainfall, as the objective was to model temporal rainfall trends and seasonal dependencies. The long historical coverage enables the learning models to capture seasonal cycles, inter-annual variability, and long-term climatic patterns. Additionally, district-level granularity allows region-specific analysis, which is useful for agricultural planning and water resource management. The public availability of the dataset ensures transparency and reproducibility of the research.

Dataset Link: https://www.kaggle.com/datasets/rajanand/rainfall-in-india`

3.2. Algorithms Used

To model the sequential characteristics of rainfall data, multiple Recurrent Neural Network (RNN)–based architectures were implemented and compared. The Simple RNN was used as a baseline model, as it maintains a hidden state to process sequential inputs. However, due to the vanishing gradient problem, it has limited capability in capturing long-term dependencies.To address this limitation, Long Short-Term Memory (LSTM) networks were employed. LSTM models introduce memory cells and gating mechanisms that regulate information flow, enabling them to retain long-term temporal information. Similarly, the Gated Recurrent Unit (GRU) simplifies the LSTM structure by combining certain gates, reducing computational complexity while maintaining effective performance.

Bidirectional LSTM was also implemented to enhance contextual learning by processing input sequences in both forward and backward directions. This structure allows the model to utilize both past and future contextual information within a defined window. Furthermore, a Stacked LSTM architecture consisting of multiple LSTM layers was developed to learn deeper and more complex hierarchical rainfall patterns.

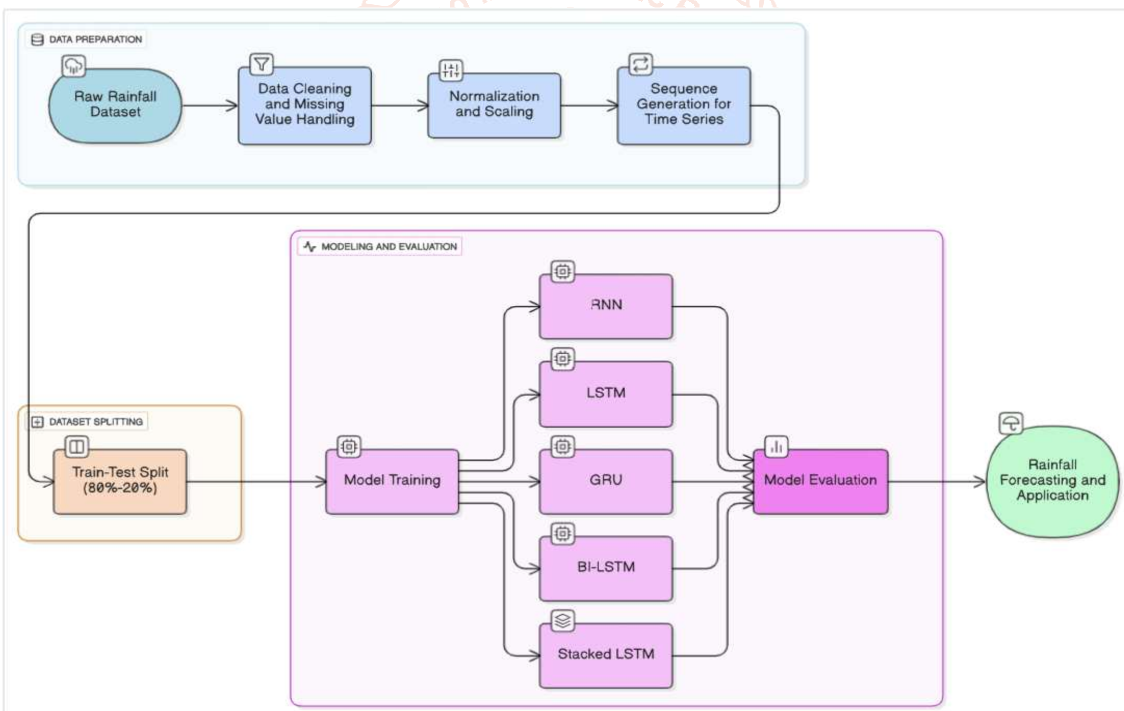

Fig 2: System Architecture

3.3. Data Preprocessing

The raw rainfall dataset was first examined for missing values and inconsistencies. Missing entries were handled using interpolation techniques to maintain the continuity of the time series. Minor anomalies were smoothed to reduce noise while preserving the overall rainfall trend.

To improve training stability and convergence speed, Min–Max normalization was applied to scale rainfall values within a uniform range. This step ensures that large magnitude values do not dominate the learning process and allows the neural network models to train efficiently.

3.4. Sequence Generation

Since rainfall forecasting is a time-series prediction problem, the dataset was transformed into supervised learning format using a sliding-window approach. For each district, rainfall values from the previous twelve months were used as input features to predict the rainfall of the subsequent month. This method preserves temporal continuity and enables the models to learn seasonal and sequential dependencies effectively.

3.5. Model Training

The prepared dataset was divided into training and testing sets using an 80:20 ratio. The training data were used to optimize model parameters, while the testing data were reserved for performance evaluation. All RNN-based models were trained using Backpropagation Through Time (BPTT) with the Adam optimizer. Mean Squared Error was used as the loss function to minimize prediction error during training.

3.6. Model Evaluation

The predictive performance of the models was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Prediction Accuracy. RMSE measures the square root of the average squared difference between actual and predicted values, thereby penalizing larger errors more heavily. MAE calculates the average absolute difference between predicted and observed rainfall values, providing an easily interpretable measure of error. Prediction Accuracy represents the percentage closeness of predicted values to actual rainfall measurements and allows comparison among different architectures.

3.7. Forecasting and Application

After model validation, each trained architecture was used to generate rainfall forecasts on unseen test data. The best-performing model was selected based on minimum error values and overall predictive consistency. The proposed framework can support practical applications such as agricultural planning, reservoir management, and disaster preparedness, particularly in regions vulnerable to floods and droughts.

4. Result

Among the five implemented architectures—Simple RNN, LSTM, GRU, Bidirectional LSTM, and Stacked LSTM—the Simple RNN model demonstrated the best overall performance on the rainfall dataset. The comparison was conducted using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).The Simple RNN achieved the lowest RMSE of 369.76 mm, MAE of 246.44 mm, and the highest R^2 score of 0.8757, indicating superior predictive accuracy compared to the other architectures. The plot of actual versus predicted rainfall values shows that the predicted curve closely follows the general trend of observed rainfall data. Although minor smoothing is observed during sharp fluctuations, the model successfully captures the primary seasonal and temporal patterns present in the dataset.In contrast, LSTM, GRU, Bidirectional LSTM, and Stacked LSTM exhibited slightly higher error values. While these advanced architectures are designed to capture long-term dependencies, their performance improvement over the Simple RNN was marginal for this particular dataset. This suggests that the rainfall time-series characteristics in this study were sufficiently modeled by the simpler recurrent structure without requiring deeper or bidirectional architectures.

Overall, the results indicate that increasing model complexity did not necessarily lead to improved forecasting accuracy for the given dataset.

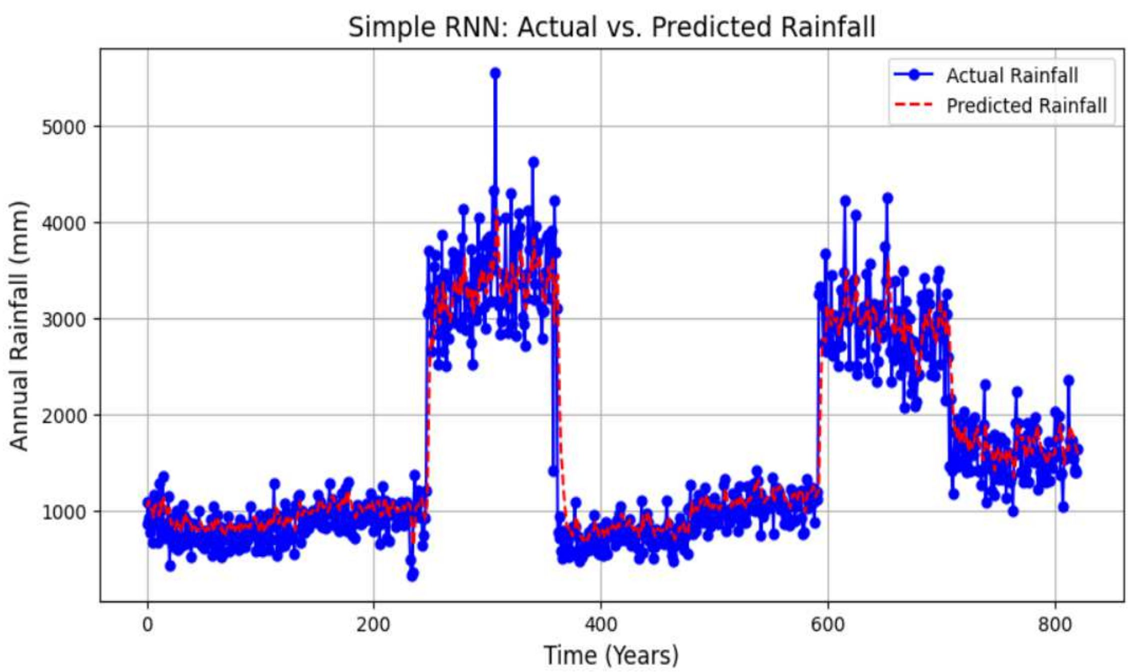

Fig 3: Actual vs Predicted value In Simple RNN

Model	RMSE	MAE	R^2 Score
Simple RNN	369.76 mm	246.44 mm	0.8757
LSTM	375.33 mm	247.67 mm	0.8719
GRU	377.61 mm	249.73 mm	0.8703
Bidirectional LSTM	374.53mm	249.16 mm	0.8724
Stacked LSTM	377.82 mm	249.38 mm	0.8702

Table I: Comparative Performance Metrics Across RNN Models

Table I compares various RNN models using RMSE, MAE, and R^2 Score, showing that the Simple RNN achieved the best overall accuracy, while Bidirectional and Stacked LSTM models also performed competitively.

5. Conclusion

This study evaluated five Recurrent Neural Network–based architectures for rainfall prediction using India's historical rainfall dataset (1901–2015). Based on comparative performance analysis, the Simple RNN model achieved the best results, with the lowest RMSE and highest R^2 score among all evaluated models. The findings demonstrate that even a basic recurrent neural network can effectively capture temporal dependencies and seasonal rainfall patterns when appropriate preprocessing and sequence generation techniques are applied. The study highlights that model selection should be guided by empirical evaluation rather than architectural complexity alone. The proposed rainfall forecasting framework can support applications in agricultural planning, water resource management, and climate risk assessment. Future work may focus on incorporating additional meteorological variables, exploring hybrid deep learning architectures, and implementing real-time prediction systems to further enhance forecasting capability.

Future Work

While our models showed strong performance, the path to a fully optimized and practical forecasting system involves several critical next steps:

Higher Data Resolution: We must shift from annual data to monthly or weekly rainfall figures. This finer granularity is essential for capturing short-term variability and crucial seasonal patterns, allowing us to generate more practical, detailed forecasts.

Embrace Multivariate Complexity: Rainfall is influenced by many factors. Future research will enhance accuracy by transitioning to a multivariate approach, incorporating critical variables like temperature, humidity, and atmospheric pressure. This gives the model a richer context for prediction.

Develop Hybrid Architectures: We can explore novel hybrid models, such as combining a Convolutional Neural Network (CNN) for extracting spatial features (e.g., from satellite imagery) with an LSTM for temporal processing. This strategic blending of architectures aims to achieve superior predictive power.

Precision Tuning: To ensure maximum performance, we will conduct an exhaustive hyperparameter optimization using advanced techniques like Bayesian Optimization. This rigorous tuning process will refine model parameters and significantly boost forecast accuracy.

Real-Time System Deployment: The final phase involves taking our best model and deploying it as part of a real-time weather forecasting system. This requires building a continuous data pipeline and creating an accessible API or user interface to deliver instant, practical predictions.

References

[1] S. Praveen, S. Gontia, and P. K. Bhunya, "Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches," Scientific Reports, vol. 10, no. 1, pp. 1–21, June 2020, doi: 10.1038/s41598-020-67228-7.

[2] B. S. Choudhari, A. K. Lohani, and N. K. Goel, "District Wise Comparative Analysis of Rainfall Trends in Konkan-Goa and Marathwada Meteorological Subdivision of India," Advances in Environmental and Engineering Science, vol. 9, no. 10, pp. 1–12, Nov. 2021.

[3] A. Kumar, R. Kumar, and A. Garg, "Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas," Scientific Reports, vol. 14, art. no. 26614, Nov. 2024, doi: 10.1038/s41598-024-77687-x.

[4] K. Mondal and S. Jain, "District-wise trend analysis of rainfall pattern in last century (1901-2000) over Gangetic region in West Bengal, India," Journal of Climatology & Weather Forecasting, vol. 3, no. 4, Dec. 2015.

[5] A. Bhardwaj, P. Jain, and M. Garg, "Rainfall Prediction: Analysis of Machine Learning Algorithms and Ensemble Techniques," in Proc. 2021 Int. Conf. on Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1–6, doi: 10.1109/CONIT51480.2021.9673275.

[6] M. Ahmad, A. Khan, and S. Khan, "Comparative analysis of different rainfall prediction models: A case study of Aligarh City, India," Results in Engineering, vol. 22, art. 102175, Apr. 2024, doi: 10.1016/j.rineng.2024.102175.

[7] M. H. Rahman, S. Ahmed, and M. R. Islam, "Spatiotemporal analysis and predicting rainfall trends in a tropical monsoon-dominated country using MAKESENS and machine learning techniques," Scientific Reports, vol. 13, art. no. 13786, Aug. 2023, doi: 10.1038/s41598-023-41132-2.

[8] B. D. Hewage, A. Perera, and M. Brilakis, "A novel framework for spatio-temporal prediction of environmental data using deep learning," Scientific Reports, vol. 10, art. no. 22243, Dec. 2020, doi: 10.1038/s41598-020-79148-7.

[9] Y. Guo, S. Li, and J. Wang, "Rainfall intensity prediction by a spatial-temporal ensemble," in Proc. IEEE Int. Joint Conf. on Neural Networks (IJCNN), Hong Kong, 2008, pp. 2336–2342, doi: 10.1109/IJCNN.2008.4634030.

[10] S. Kumar, A. Singh, and R. Patel, "Long-Term Spatiotemporal Investigation of Various Rainfall Intensities over Central India Using EO Datasets," Climate, vol. 11, no. 2, art. 27, Feb. 2024, doi: 10.3390/cli11020027.

[11] V. Guhathakurta and E. Rajeevan, "Analysis of long-term rainfall trends in India," Hydrological Sciences Journal, vol. 55, no. 4, pp. 484–496, 2010, doi: 10.1080/02626667.2010.481373.

[12] R. K. Singh, P. Kumar, and A. Verma, "Assessing seasonal variation and trends in rainfall patterns of Madhya Pradesh, Central India," Journal of Water and Climate Change, vol. 14, no. 10, pp. 3692–3716, Oct. 2023, doi: 10.2166/wcc.2023.441.

[13] M. Rajeevan, J. Bhate, and A. K. Jaswal, "Changing Frequency and Intensity of Rainfall Extremes over India from 1951 to 2003," Journal of Climate, vol. 22,

no. 18, pp. 4737–4746, Sept. 2009, doi: 10.1175/2009JCLI2896.1.

[14] A. Das and B. Deka, "Assessing Climate Change Impact on Rainfall Patterns in Northeastern India and Its Consequences on Water Resources and Rainfed Agriculture," Climate, vol. 6, no. 1, art. 2, Jan. 2025, Doi: 10.3390/cli6010002.

[15] M. Ahmad, A. Khan, and S. Khan, "Comparative analysis of different rainfall prediction models: A case study of Aligarh City, India," Results in Engineering, vol. 22, art. 102175, Apr. 2024, Doi: 10.1016/j.rineng.2024.102175.