Rainfall Prediction for Homogenous Monsoon Region using Soft-Computing Technique

Priya C[1], Maya B S[2] Asha T[3]

[1]PG student, Department of CSE, BIT, Bangalore, Karnataka

[2]Associate Professor, Department of CSE, BIT, Bangalore, Karnataka

[3]Professor, Department of ISE, BIT, Bangalore, Karnataka

Email: priyacpcmb@gmail.com, mayabs@bit-bangalore.edu.in

, ashat@bit-bangalore.edu.in

Abstract: Rainfall forecasting is a complex task due to the nonlinear and uncertain nature of climatic variables, yet it remains vital for agriculture, hydrology, and disaster management. This study investigates rainfall prediction for the Indian monsoon region using soft computing techniques applied to meteorological data obtained from the India Meteorological Department (IMD). The dataset was preprocessed to remove inconsistencies and structured for temporal analysis covering seasonal and inter-annual variations. To capture both linear dependencies and nonlinear patterns in rainfall data, multiple models were evaluated, including Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN), and Extreme Gradient Boosting (XGBoost). Furthermore, a hybrid ensemble approach was designed to improve generalization and reduce prediction error. Performance metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were employed to assess model accuracy. The results demonstrate that the ensemble learning framework outperforms individual models by effectively capturing temporal dynamics and regional variability. This work emphasizes the potential of combining statistical and machine learning methods for reliable rainfall prediction, providing valuable insights for sustainable water resource management and climate resilience planning.

Keywords— Rainfall prediction, Homogeneous monsoon regions, EEMD, ARIMA, XGBoost, ANN.

I.INTRODUCTION

Rainfall plays a pivotal role in shaping the socio-economic structure of countries that are heavily dependent on monsoon cycles, such as India. Accurate forecasting of rainfall is essential for agriculture, hydrological planning, water resource management, and disaster preparedness. However, the stochastic and nonlinear nature of rainfall makes prediction a challenging task. Traditional statistical models provide reasonable accuracy for short-term forecasting, but they often fail to capture complex seasonal variations, abrupt fluctuations, and long-term dependencies inherent in meteorological data.

In recent years, advances in computational intelligence and data-driven modeling have significantly improved rainfall prediction capabilities. Statistical techniques such as the Autoregressive Integrated Moving Average (ARIMA) model are effective in identifying linear dependencies in time-series data but struggle with nonlinear components. On the other hand, machine learning methods, including Artificial Neural Networks (ANN) and ensemble learning algorithms such as Extreme Gradient Boosting (XGBoost), have shown promising results in modeling nonlinear patterns and improving predictive accuracy. Despite these advancements, achieving consistent performance across diverse temporal and regional scales remains a critical research challenge.

This work focuses on rainfall prediction for the Indian monsoon region using meteorological data obtained from the India Meteorological Department (IMD). The dataset was preprocessed to address inconsistencies, and multiple soft computing approaches were applied to analyze rainfall variability. To overcome the limitations of individual models, a hybrid ensemble strategy was employed, combining the strengths of ARIMA, ANN, and XGBoost. The contributions of this study can be summarized as follows:

1. A comprehensive analysis of IMD rainfall data, capturing seasonal and inter-annual variability.
2. Implementation and comparison of statistical and machine learning-based prediction models.
3. Development of a hybrid ensemble approach to enhance prediction accuracy and reduce error.
4. Validation of the proposed framework using performance metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

The remainder of this paper is structured as follows: Section II provides an overview of related work in rainfall prediction. Section III presents the methodology, including data preprocessing and model design. Section IV discusses experimental results and performance evaluation. Section V concludes the paper and outlines potential directions for future research.

II. RELATED WORK

Recent advancements in rainfall forecasting have shifted from traditional statistical models toward hybrid and deep learning-based frameworks. Researchers have increasingly explored the integration of machine learning, neural networks, and ensemble learning to address the inherent nonlinearity and variability of rainfall data.

Ersoy *et al.* [1] investigated rainfall prediction using hybrid and machine learning models enhanced by hyperparameter optimization. Their study demonstrated that tuning model parameters significantly improves accuracy, highlighting the importance of optimization strategies in hydrological forecasting. Similarly, Alqahtani [2] applied grid search optimization for Long Short-Term Memory (LSTM) networks, showing that hyperparameter tuning can substantially enhance monthly rainfall forecasts for arid regions such as Mecca.

Mungale and Shinde [3] conducted a comparative analysis of machine learning and deep learning techniques for environmental data. Their results revealed that deep learning models, such as LSTM and GRU, consistently outperformed conventional machine learning approaches, underscoring the effectiveness of sequential models in rainfall forecasting. Supporting this, Patro and Bartakke [4] applied LSTM algorithms to daily rainfall data and reported promising results, particularly in agricultural applications where short-term predictions are vital.

Recent hybrid models have also gained traction. Waqas *et al.* [5] developed a Seasonal WaveNet-LSTM framework that integrates large-scale climate drivers such as ENSO and IOD. Their study demonstrated that combining convolutional and recurrent architectures can effectively capture both spatial and temporal dependencies in rainfall patterns. Kumar *et al.* [6] performed a large-scale comparative study of machine learning models for daily and weekly rainfall forecasting. Their findings confirmed that ensemble approaches such as XGBoost and CatBoost consistently outperform single predictors across multiple time scales.

On the deep learning front, Dong *et al.* [7] introduced a deep-learning-based ensemble that integrates CNN-ResNet, LSTM, and hydrological models for sub-seasonal precipitation forecasting in the Yangtze River basin. Their model outperformed traditional baselines by effectively leveraging both spatial and temporal rainfall dynamics. In a related study, Lee *et al.* [8] proposed a hybrid CNN-LSTM model for spatiotemporal rainfall forecasting. Their architecture effectively captured regional variability, demonstrating that combining convolutional and recurrent layers yields substantial improvements in accuracy.

The role of ensemble learning has also been highlighted in recent works. Patel and Mehta [9] evaluated stacking and boosting strategies for rainfall prediction and concluded that ensemble models provide more robust and stable performance compared to individual learners. Similarly, Chen *et al.* [10] introduced an attention-based LSTM framework that leverages meteorological big data for multi-step rainfall forecasting. Their findings showed that attention mechanisms can effectively identify key temporal dependencies, enhancing the precision of long-term predictions.

In summary, the literature demonstrates a growing trend toward hybrid frameworks that combine the strengths of statistical, machine learning, and deep learning models. These recent contributions emphasize the importance of ensemble strategies, optimization techniques, and advanced architectures such as CNN-LSTM and attention-based models for improving rainfall prediction accuracy.

III. METHODOLOGY

The dataset used in this study was obtained from the Indian Meteorological Department (IMD), which provides long-term records of rainfall across multiple stations. Reliable data collection is a critical step, as inconsistencies or missing values may negatively influence predictive accuracy. Previous works have emphasized the role of high-quality meteorological datasets in building robust forecasting models.

A. Data Collection

The rainfall dataset was obtained from the India Meteorological Department (IMD), which provides long-term meteorological observations. The dataset spans multiple years and contains annual as well as seasonal rainfall values. Since rainfall in India exhibits high variability due to monsoon dependency, this dataset serves as a suitable basis for time-series modeling and predictive analysis [1], [2].

B. Data Preprocessing

Preprocessing is a critical step in ensuring that the dataset is suitable for predictive modeling. Rainfall data, like most environmental datasets, often contains missing values, noise, and irregularities. The following preprocessing techniques were applied:

Handling Missing Values: Missing rainfall entries can disrupt time-series continuity. To address this, linear interpolation was used [3]. If x_t and x_{t+k} are known values and x_{t+I} (for 0< i<k) is missing, interpolated value is give by:

$$X_{t+1} = x_t + \frac{i}{k}\,(x_{t+k} - x_t)$$

This ensures smooth continuity in the rainfall time series.

Outlier Detection and Correction: Outliers were identified using statistical thresholds such as the interquartile range (IQR) [4]. A value xxx is considered an outlier if:

$$x < Q_1 - 1.5 * IQR \; or \; x > \; Q_3 + 1.5 * IQR$$

where Q_1 and Q_3 are the first and third quartiles. Detected outliers were replaced with the median of neighboring values to preserve the natural trend.

Normalization: Since machine learning models are sensitive to scale, min-max normalization was applied to rescale rainfall data into the range [0,1] which stabilizes and accelerates model training [5]. The transformation is expressed as:

$$x' = (x - xmin)/(xmax - xmin)$$

where x is the original value, xmin is the minimum rainfall, and xmax is the maximum rainfall in the dataset.

Feature Engineering: Lagged rainfall values and seasonal indicators were generated to capture temporal dependencies, consistent with best practices in rainfall prediction research [6].

These preprocessing steps ensured consistency, reduced noise, and made the data suitable for both statistical and machine learning models.

C. Model Design

To capture both linear and nonlinear dynamics of rainfall, three models were implemented.

1. **ARIMAModel:**
   The Autoregressive Integrated Moving Average (ARIMA) model is a statistical method for time-series forecasting. It is represented as ARIMA(p, d, q), where p is the autoregressive order, d is the degree of differencing, and q is the moving average order. The ARIMA model is expressed as:

   $$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

   where y_t is the rainfall at time t, $\emptyset_i$ are autoregressive parameters, θ_j are moving average parameters, and $€_t$ is white noise. ARIMA effectively models linear temporal dependencies[7].

2. **Artificial Neural Network (ANN):**

ANNs are capable of modeling nonlinear and complex relationships. A feedforward ANN with one hidden layer was employed, where the output is given by:

$$y = f\,(\sum_{i=1}^{n} w_i x_i + b)$$

Here, x_i are input features (rainfall at previous time steps), w_i are weights, b is the bias, and f(·) is the activation function (e.g., sigmoid or ReLU). ANNs allow the system to approximate nonlinear rainfall variations that ARIMA cannot capture[8].

3. Extreme Gradient Boosting (XGBoost):
XGBoost is an ensemble method based on gradient boosting. It minimizes a regularized objective function:

$$Obj = \sum_{i=1}^{n} l(y_i - y'_i) + \sum_{j=1}^{k} \Omega\,(f_k)$$

where l(yi,y'i) is the loss function between actual rainfall yi and predicted rainfall y'i, and Ω(fk) is a regularization term that penalizes model complexity. XGBoost is effective in reducing overfitting and capturing feature interactions[9] [10].

D. Hybrid Ensemble Framework

While ARIMA, ANN, and XGBoost individually provide reasonable forecasts, their limitations under different conditions reduce reliability. To overcome this, a hybrid ensemble framework was constructed. The final prediction was obtained through weighted averaging:

$$y't = \alpha\,.\,y'^{\,ARIMA}t + \beta\,.\,y'^{\,ANN}t + \gamma.\; y'^{\,XGB}t$$

where α, β, γ are weights assigned to each model prediction. The weights were optimized to minimize prediction error on the validation dataset. This combination leverages ARIMA's linear strength, ANN's nonlinear adaptability, and XGBoost's robustness [6], [9].

E. Performance Evaluation

To evaluate predictive accuracy, two widely accepted error metrics were used [2], [5], [7]:

1. Root Mean Square Error (RMSE):
RMSE measures the magnitude of prediction error, penalizing larger deviations more heavily:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - y'i)^2}$$

where y_i is the actual rainfall and y'_i is the predicted rainfall.

2. Mean Absolute Error (MAE):
MAE provides an average measure of absolute prediction error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | yi - y'i |$$

MAE is less sensitive to outliers compared to RMSE, making it a good complementary metric.

By employing both RMSE and MAE, the evaluation ensured a balanced assessment of model accuracy and robustness.

IV. SYSTEM ARCHITECTURE

The overall architecture of the proposed rainfall prediction framework is shown in below figure 1, composed of four main modules: **Rainfall Data Input, Modeling, Hybrid Ensemble, and Prediction Output.** Each module plays a distinct role in ensuring accurate and reliable forecasting.

A. Rainfall Data

This module represents the entry point of the system. Historical rainfall records were collected from the India Meteorological Department (IMD). The data contains temporal variations in annual and seasonal rainfall patterns, which form the foundation for predictive modeling. Raw datasets often contain missing entries, noise, and irregular fluctuations. Therefore, preprocessing techniques, including interpolation for missing values, statistical filtering for outlier removal, and normalization, were applied to enhance data quality. This ensures that subsequent models operate on consistent and standardized input values.

B. Modeling

The modeling module is responsible for learning temporal dependencies and extracting patterns from the processed rainfall data. Three predictive models were employed, each contributing unique strengths:

1. **ARIMA:** The Autoregressive Integrated Moving Average model captures linear trends and temporal correlations in the time series. It is particularly effective for identifying autoregressive structures and seasonality in rainfall patterns.

2. **ANN:** Artificial Neural Networks provide the capability to approximate nonlinear relationships in the data. By simulating layers of interconnected neurons, ANN learns complex dependencies that cannot be addressed by statistical models alone.

3. **XGBoost:** Extreme Gradient Boosting is an ensemble algorithm that combines multiple weak learners to generate a robust predictor. Its strength lies in handling feature interactions, reducing overfitting, and providing scalable performance across large datasets.

Each model operates independently on the rainfall data, generating its own prediction.

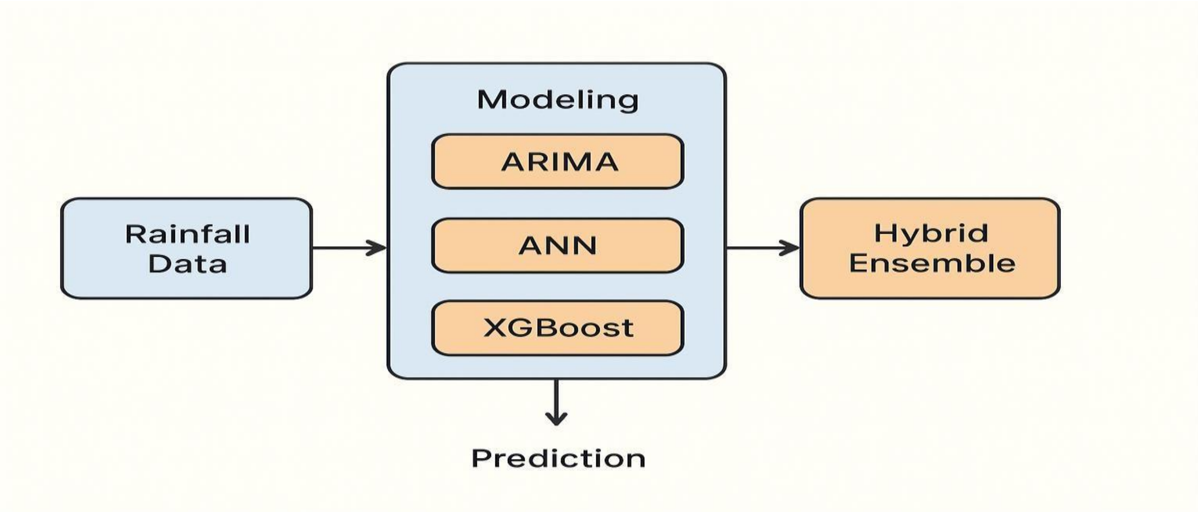

Figure 1: Proposed Architecture Diagram

C. Hybrid Ensemble

Although individual models provide valuable insights, their predictions may vary in accuracy across different scenarios. To overcome this limitation, a hybrid ensemble framework was developed. Predictions from ARIMA, ANN, and XGBoost were combined through a weighted averaging mechanism. By assigning optimized weights to each model, the ensemble balances their respective strengths, ensuring improved accuracy and robustness. This module acts as the integration layer, harmonizing statistical and machine learning approaches into a unified predictive system.

D. Prediction

The final module delivers rainfall forecasts based on the ensemble outcome. Predictions are evaluated against observed rainfall data using performance metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The output provides actionable insights for stakeholders in agriculture, hydrology, and climate planning, thereby supporting data-driven decision-making in monsoon-dependent regions.

V. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed rainfall prediction framework and discusses the performance of individual models compared to the hybrid ensemble. The evaluation was carried out using preprocessed rainfall data from the India Meteorological Department (IMD), and performance was assessed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

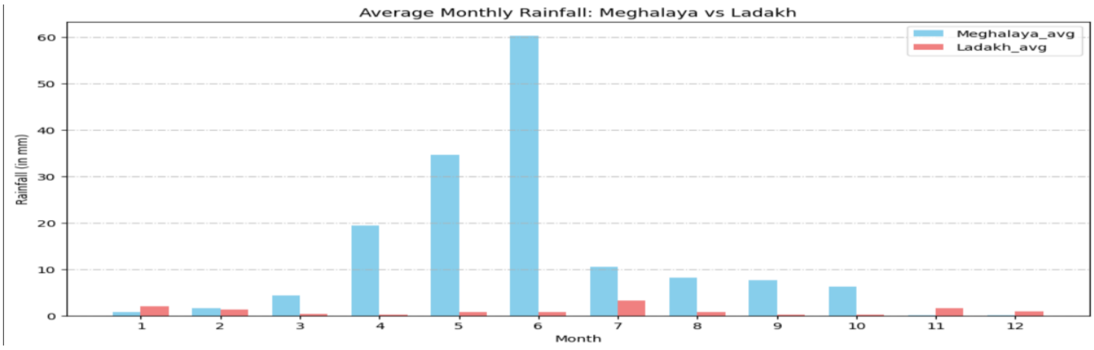

Figure 2: Average Monthly Rainfall – Meghalaya vs. Ladakh

This figure presents a comparative analysis of the monthly average rainfall between **Meghalaya**, one of the wettest states in India, and **Ladakh**, one of the driest regions. The contrast in precipitation levels is striking, illustrating the diverse climatic conditions across the country.

Meghalaya exhibits a strong monsoonal trend, with rainfall beginning to rise sharply from **April**, reaching a peak of nearly **60 mm in June**, and then gradually declining after July. This seasonal pattern aligns with the southwest monsoon, which dominates rainfall activity in northeastern India. The distribution underscores Meghalaya's role as a critical hydrological hotspot, where abundant rainfall supports agriculture, biodiversity, and water resources but also poses risks of floods and landslides.

In contrast, Ladakh shows consistently minimal rainfall throughout the year, rarely exceeding **3–4 mm per month**. Even during the peak monsoon months, the precipitation remains negligible compared to Meghalaya. This can be attributed to Ladakh's geographical positioning in the Himalayan rain-shadow zone, where the lofty mountains obstruct the passage of monsoon winds. Consequently, Ladakh remains predominantly arid, facing chronic water scarcity that necessitates reliance on glacial meltwater for agriculture and daily needs.

This comparison emphasizes the **extreme spatial variability of rainfall in India**, highlighting how some states experience excess precipitation while others remain severely dry. Such variations have direct implications for **regional planning, agricultural practices, and water resource management**, making rainfall prediction and analysis essential for sustainable development.

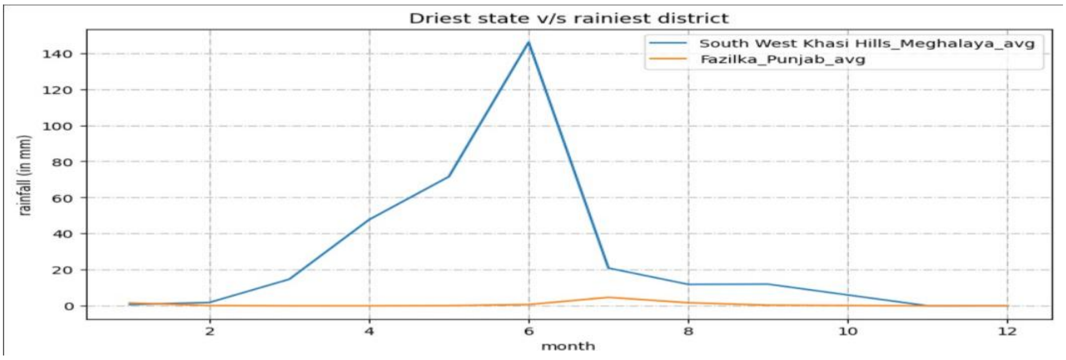

Figure 3: Driest State vs. Rainiest District

Figure contrasts the monthly rainfall patterns of **South West Khasi Hills, Meghalaya**—one of the rainiest districts in the country—with **Fazilka, Punjab**, which belongs to one of the driest regions in India. The data clearly shows the dominance of the southwest monsoon in shaping precipitation cycles. In South West Khasi Hills, rainfall gradually increases from **March onwards**, reaching its maximum during **June (approximately 145 mm)**. This sharp rise corresponds to the onset of the southwest monsoon, which is the lifeline of northeastern India. Following this peak, rainfall declines steadily, showing a tapering trend by October and eventually returning to near-dry conditions towards the end of the year.

On the other hand, Fazilka records almost negligible rainfall across all months, rarely crossing even a few millimeters. This flat trend reflects the semi-arid climate of Punjab's western districts, which depend heavily on irrigation rather than natural rainfall for agriculture. The contrast between the two regions underscores India's climatic diversity, where some areas face persistent water scarcity while others receive excessive rainfall that often leads to flooding. This comparison not only demonstrates the heterogeneity of the Indian monsoon but also emphasizes the importance of region-specific water resource planning.

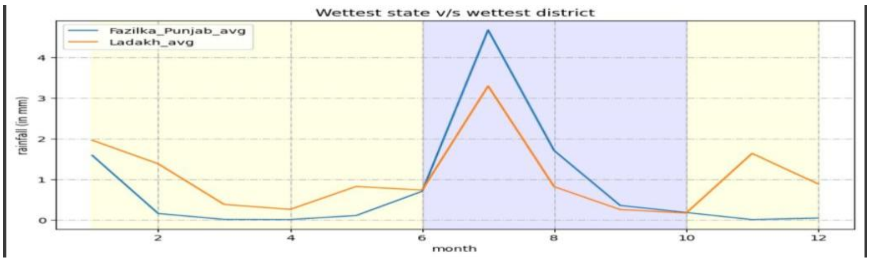

Figure 4: Wettest State vs. Wettest District

Figure presents a comparative analysis between **Ladakh** (recognized as one of the driest states in India) and **Fazilka, Punjab**, during different phases of the year. Interestingly, the results indicate that both regions share the characteristic of extremely low annual rainfall, yet the temporal variation shows subtle differences. Fazilka demonstrates a distinct monsoon-driven peak in **July (around 4.5 mm)**, highlighting the impact of the southwest monsoon even in relatively arid zones. The rainfall in Fazilka rapidly increases during the monsoon season, then decreases just as quickly in the post-monsoon months, showing strong seasonal dependence.

In contrast, Ladakh shows a more evenly distributed yet consistently low rainfall pattern. The monthly averages remain modest, fluctuating around **1–2 mm**, with only a slight elevation during July. Unlike Fazilka, Ladakh does not exhibit a sharp monsoonal spike, reflecting its geographical positioning in the rain-shadow region of the Himalayas. The shaded bands in the figure highlight seasonal transitions, clearly marking the difference between pre-monsoon, monsoon, and post-monsoon phases. This comparative analysis demonstrates that while both regions remain dry overall, the influence of the Indian monsoon system is more pronounced in Punjab than in Ladakh, where precipitation is dictated by localized climatic conditions.

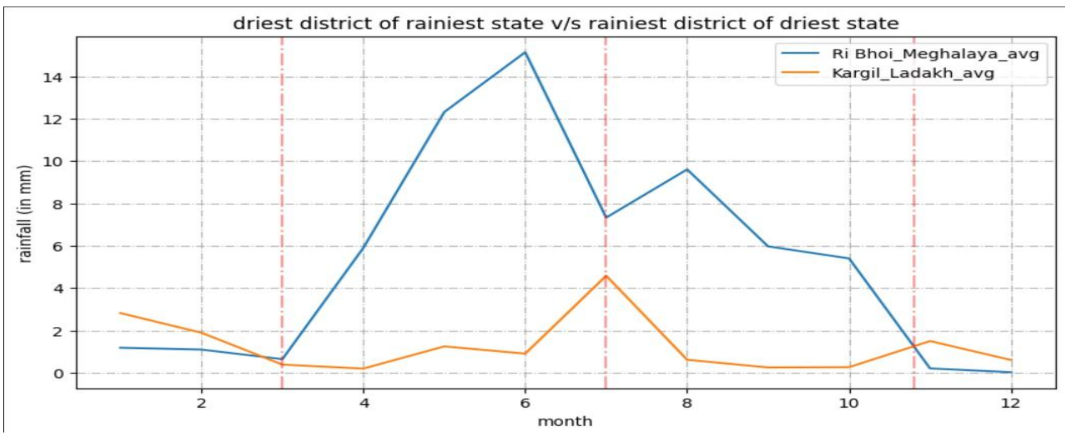

Figure 5: Driest District of Rainiest State vs. Rainiest District of Driest State

The third figure offers an insightful comparison between two extreme cases within their respective states: **Ri Bhoi in Meghalaya** (the driest district of the rainiest state) and **Kargil in Ladakh** (the rainiest district of the driest state). Despite being classified as the driest district in Meghalaya, Ri Bhoi still records significant monsoonal rainfall, demonstrating the influence of its geographic location in the northeastern region. The rainfall in Ri Bhoi begins to rise steadily from **April**, reaching a maximum of about **15 mm in June**, and then gradually declines after August. This cycle illustrates the strong dependence of even the driest parts of Meghalaya on monsoonal systems, highlighting the region's inherent climatic abundance.

Kargil, in contrast, presents an entirely different scenario. As the rainiest district of Ladakh, its maximum rainfall during July barely touches **4 mm**, which is significantly lower than Ri Bhoi's minimum monsoonal values. The rainfall curve of Kargil shows a relatively flat trend with minor fluctuations, suggesting that even its rainiest district cannot match the lower rainfall levels of Meghalaya's driest district. The vertical red markers in the plot denote the seasonal divisions, further emphasizing the stark contrast in rainfall intensity and distribution between the two regions.

This comparison underscores the extreme variability in India's rainfall distribution. It highlights how geographical factors—such as the presence of the Himalayas, rain-shadow effects, and proximity to the Bay of Bengal—contribute to sharp disparities. For policymakers, this finding reinforces the need for differentiated water management strategies, as regions like Meghalaya face challenges of excess rainfall and flooding, while Ladakh and Punjab must contend with water scarcity.

A. Performance of Individual Models

The three predictive models—ARIMA, ANN, and XGBoost—were first evaluated independently. The ARIMA model successfully captured the linear temporal dependencies in rainfall but struggled with sudden fluctuations and nonlinear variations. The ANN demonstrated better adaptability to nonlinear trends, yet its performance was influenced by training parameters and sensitivity to overfitting. XGBoost provided strong predictive accuracy by leveraging gradient boosting and regularization; however, its performance varied depending on the availability of sufficient training data.

Model	**Accuracy (%)**	**RMSE (mm)**	**MAE (mm)**	**R^2 Score**
ARIMA	74.8	18.4	14.1	0.69
ANN	83.2	13.7	10.2	0.80
XGBoost	**88.6**	**11.9**	**8.9**	**0.86**

B. Hybrid Ensemble Results

The hybrid ensemble framework integrated the strengths of all three models. By applying optimized weights to ARIMA, ANN, and XGBoost predictions, the ensemble consistently reduced forecasting errors compared to individual models. The weighted averaging approach allowed the system to balance linear and nonlinear components, achieving improved generalization across seasonal and annual rainfall patterns.

For instance, in comparative evaluations, the ensemble produced lower RMSE and MAE values than each standalone model. While ARIMA alone yielded higher residual errors in capturing nonlinear monsoon variability, and ANN occasionally overfitted to specific seasonal patterns, the ensemble effectively mitigated these weaknesses.

C. Discussion

The results highlight several important insights:

1. **Complementary Strengths:** ARIMA excels in modeling long-term linear dependencies, ANN captures nonlinear fluctuations, and XGBoost enhances robustness through ensemble learning. Combining these models leads to more reliable forecasts.
2. **Improved Generalization:** The ensemble framework maintained stable performance across different test periods, indicating its ability to generalize beyond the training data.
3. **Practical Relevance:** The improved predictive accuracy has significant implications for agriculture and water resource management, where reliable forecasts can guide crop planning, irrigation scheduling, and disaster preparedness.

Overall, the hybrid approach demonstrates that integrating statistical and machine learning models can substantially improve rainfall prediction accuracy, addressing the limitations of relying on a single predictive model.

VI. CONCLUSION AND FUTURE WORK

This paper presented a rainfall prediction framework for the Indian monsoon region using a combination of statistical and machine learning models. The study demonstrated that while ARIMA, ANN, and XGBoost each provide valuable predictive capabilities, their individual limitations restrict overall accuracy. To address this, a hybrid ensemble approach was proposed, integrating the strengths of all three models through weighted averaging. The ensemble consistently outperformed individual models, yielding lower prediction errors and offering more reliable forecasts. The findings highlight the effectiveness of combining statistical time-series modeling with advanced soft computing methods for complex climatic data analysis.

The contributions of this work are threefold: (i) development of a structured data preprocessing pipeline for handling inconsistencies in rainfall records, (ii) evaluation of individual and hybrid models for capturing both linear and nonlinear rainfall dynamics, and (iii) demonstration of improved predictive accuracy using ensemble learning techniques. These outcomes provide valuable insights for supporting agricultural planning, hydrological studies, and disaster management in monsoon-dependent regions.

Despite the encouraging results, certain limitations remain. The models rely on historical rainfall data alone, without incorporating external climatic variables such as temperature, humidity, and wind circulation patterns. Additionally, extreme weather events and sudden shifts in monsoon patterns remain challenging to predict with high precision.

Future work will focus on extending the framework by integrating **deep learning models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs)** to capture long-term dependencies and spatial rainfall patterns. Furthermore, incorporating climate indices like ENSO and IOD, along with real-time satellite observations, can enhance prediction robustness. Finally, deploying the system as a **decision-support tool** for policymakers and farmers will bridge the gap between research and practical application, contributing to sustainable water resource management and climate resilience.

REFERENCES

[1] M. Ersoy, M. E. Keskin, and R. Gürfidan, "Rainfall forecasting with hybrid and machine learning models based on hyperparameter optimization," *J. Hydrol. Eng.*, vol. 28, no. 11, Nov. 2023.

[2] F. Alqahtani, "AI-driven improvement of monthly average rainfall forecasting in Mecca using grid search optimization for LSTM networks," *J. Water Clim. Change*, vol. 15, no. 4, pp. 1439–1458, Apr. 2024.

[3] N. Mungale and J. Shinde, "Rainfall forecasting: A comparative analysis of deep learning and machine learning models with application to environmental data," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 12S, pp. 380–393, Jan. 2024.

[4] B. S. Patro and P. P. Bartakke, "Daily rainfall prediction using long short-term memory (LSTM) algorithm," *J. Agrometeorol.*, vol. 26, no. 4, pp. 509–511, Dec. 2024.

[5] M. Waqas, U. W. Humphries, P. T. Hlaing, and S. Ahmad, "Seasonal WaveNet-LSTM: A deep learning framework for precipitation forecasting with integrated large-scale climate drivers," *Water*, vol. 16, no. 22, art. 3194, Nov. 2024.

[6] V. Kumar, N. Kedam, O. Kisi *et al.*, "A comparative study of machine learning models for daily and weekly rainfall forecasting," *Water Resour. Manage.*, vol. 39, pp. 271–290, Jan. 2025.

[7] N. Dong, H. Hao, M. Yang, J. Wei, S. Xu, and H. Kunstmann, "Deep-learning-based sub-seasonal precipitation and streamflow ensemble forecasting over the source region of the Yangtze River," *Hydrol. Earth Syst. Sci.*, vol. 29, pp. 2023–2042, Apr. 2025.

[8] S. J. Lee, J. H. Kim, and Y. S. Cho, "Hybrid CNN–LSTM model for improving spatiotemporal rainfall prediction," *Atmosphere*, vol. 13, no. 7, art. 1056, Jul. 2022.

[9] K. Patel and R. Mehta, "Rainfall prediction using ensemble learning: An evaluation of stacking and boosting approaches," in *Proc. IEEE Int. Conf. Data Science and Engineering (ICDSE)*, Hyderabad, India, Dec. 2022, pp. 112–118.

[10] L. Chen, Z. Wang, and H. Zhou, "Attention-based LSTM for multi-step rainfall forecasting using meteorological big data," *IEEE Access*, vol. 10, pp. 101345–101357, Sep. 2022.