Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 1 (2026)

Hybrid Machine Learning–Deep Learning Framework for Time-Series Rainfall Prediction in Tamil Nadu

Dr S Prema[1], M. Menaka[2], Dr S Dhamodharavadhani[3*], M. Meenaga[4],

[1]Assistant Professor & Head, Department of Computer Applications, Arulmigu Arthanareeswarar Arts and Science College, Tiruchengode premashanmuga11@gmail.com, 9842979939

[2]Guest Lecturer, Department of Computer Applications, Arulmigu Arthanareeswarar Arts and Science College, Tiruchengode menakaaaasc@gmail.com

[3]Assistant Professor, Department of Computer Applications, Arulmigu Arthanareeswarar Arts and Science College, Tiruchengode vadhanimca2011@gmail.com

[4]II BCA, Department of Computer Applications, Arulmigu Arthanareeswarar Arts and Science College, Tiruchengode

Corresponding Author* vadhanimca2011@gmail.com

Abstract

Accurate rainfall prediction is essential for agricultural planning, irrigation scheduling, water resource management, and disaster mitigation in Tamil Nadu, a region strongly influenced by Southwest and Northeast monsoon systems. This study proposes a converged Machine Learning (ML) and Deep Learning (DL) framework for time-series rainfall prediction using historical district-wise monthly rainfall data collected from the India Meteorological Department (IMD) for the period 2000–2024. The dataset includes rainfall records along with meteorological parameters such as temperature, humidity, wind speed, and seasonal indicators from selected Tamil Nadu districts including Chennai, Coimbatore, Madurai, Tirunelveli, and Cuddalore.

The proposed framework integrates Random Forest (RF) and Long Short-Term Memory (LSTM) networks to improve rainfall forecasting accuracy. In the preprocessing stage, missing values were handled, features were normalized, and temporal lag features (1–12 months) were generated to capture seasonality and rainfall trends. Random Forest was employed for feature selection and prediction refinement, while the LSTM model captured long-term temporal dependencies in rainfall patterns. Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Experimental results demonstrate that the hybrid ML–DL model outperformed standalone ML and DL models, achieving an RMSE of 16.82 mm, MAE of 12.94 mm, and MAPE of 10.12%. The proposed approach effectively captured seasonal rainfall variations and extreme monsoon events across different districts. The findings indicate that the hybrid framework can support district-level weather forecasting and climate-resilient decision support systems for agriculture and water resource management in Tamil Nadu..

Keywords: Rainfall Prediction; Hybrid Machine Learning; Deep Learning; Random Forest; Long Short-Term Memory (LSTM); Time-Series Forecasting; Tamil Nadu; Meteorological Data; Monsoon Prediction; Feature Selection; Agricultural Planning; Water Resource Management.

Introduction

Rainfall prediction is a critical component of climate monitoring, agricultural planning, water resource management, and disaster mitigation. In regions like Tamil Nadu, where agriculture is highly dependent on monsoon rainfall, accurate forecasting of rainfall patterns can significantly impact crop productivity, irrigation scheduling, and flood/drought preparedness. Traditional statistical models, such as ARIMA and SARIMA, have been widely used for rainfall time-series forecasting. However, these models often struggle with the nonlinear, seasonal, and temporal complexities inherent in rainfall data.

Recent advances in Machine Learning (ML) and Deep Learning (DL) have shown promise in overcoming the limitations of conventional statistical approaches. ML models such as Random Forest (RF), Support Vector

Regression (SVR), and Artificial Neural Networks (ANN) can model nonlinear relationships, while DL models, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, excel in capturing long-term temporal dependencies and complex sequential patterns. To address these challenges, this study proposes a converged ML–DL framework for time-series rainfall prediction in Tamil Nadu, combining the strengths of both paradigms. In this framework, ML techniques are utilized for data preprocessing, feature selection, and prediction refinement, while DL models capture temporal patterns and seasonality in rainfall data. The hybrid approach aims to improve forecasting accuracy, enhance robustness during monsoon seasons, and provide actionable insights for weather-based decision support systems in agriculture.

The main contributions of this paper are as follows:

1. Development of a hybrid ML–DL model for rainfall prediction in Tamil Nadu.
2. Integration of ML-based feature engineering with DL-based temporal modeling.
3. Evaluation of the proposed approach using historical rainfall data, demonstrating improved accuracy compared to standalone models.

2.Related Work

Rainfall prediction has received significant attention in recent years, particularly with the advent of Machine Learning (ML) and Deep Learning (DL) techniques. Traditional statistical models such as ARIMA and SARIMA have been widely applied for rainfall forecasting; however, they often fail to capture the nonlinear and seasonal patterns inherent in rainfall time-series data.

Machine Learning approaches, including Random Forest (RF), Support Vector Regression (SVR), and Artificial Neural Networks (ANN), have been applied for regional rainfall prediction. Kumar & Singh (2019) and Patil & Deshmukh (2021) demonstrated that ML models can provide reasonable accuracy, but their performance is highly dependent on manual feature engineering and often struggles with long-term temporal dependencies.

Deep Learning models, particularly LSTM, GRU, and ConvLSTM, have demonstrated superior performance in capturing temporal and seasonal dependencies in rainfall data. Sharma et al. (2022), Ramesh & Balaji (2023), and Ghosh et al. (2025) showed that DL models outperform standalone ML models in regional rainfall forecasting. However, DL models often require large datasets and may lack interpretability, limiting their real-world application.

To overcome these limitations, several studies have proposed hybrid ML–DL frameworks that combine ML-based preprocessing and feature selection with DL-based temporal modeling. For example, Ghosh et al. (2020) and Li et al. (2024) applied hybrid approaches to improve rainfall prediction accuracy, while Tricha & Moussaid (2024) integrated SARIMA with ANN for ensemble forecasting. Recent studies by Wani et al. (2024), Farman et al. (2025), Tarawneh et al. (2025), and R. Yuan (2025) explored CNN+LSTM or LSTM+optimization techniques to enhance rainfall prediction accuracy and reliability.

Region-specific studies focusing on Tamil Nadu rainfall are limited. Kumar & Singh (2019) and Ramesh & Balaji (2023) demonstrated that hybrid models outperform standalone ML or DL models for monsoon prediction. This highlights the need for a converged ML–DL framework tailored to Tamil Nadu rainfall, motivating the current research.

Ref	**Year**	**Region / Dataset**	**Model / Approach**	**Key Findings/ Notes**
Kumar & Singh	2019	Tamil Nadu (IMD monthly rainfall)	ARIMA, LSTM	LSTM outperformed ARIMA for monsoon predictions
Ghosh et al	2020	India (West Bengal)	RF + LSTM hybrid	Improved forecast accuracy; handles seasonality effectively

Patil & Deshmukh	2021	Maharashtra, India	RF, SVR	ML models accurate for short-term rainfall; feature engineering critical
Sharma et al	2022	South India	LSTM, GRU	Captured temporal and seasonal rainfall patterns; DL outperformed ML
Ramesh & Balaji	2023	Tamil Nadu	CNN-LSTM hybrid	Improved spatial-temporal prediction; suitable for DSS
Li et al	2024	Multi-country datasets	Transformer + XGBoost	Improved long-term rainfall forecasting; reduced RMSE
Tricha & Moussaid	2024	Casablanca rainfall	SARIMA + ANN	Hybrid outperformed ANN alone
Farman et al	2025	Multi-region	DL architectures	Activation functions impact DL rainfall accuracy
Ghosh et al	2025	Indian cities	ConvLSTM + XAI	Short-term precipitation forecasting with explainability
Abdurrahaman et al	2025	Northern Nigeria	Hybrid TS + Monte Carlo	Enhanced rainfall forecast via hybrid TS models
R. Yuan	2025	Global	CNN + LSTM hybrid	Hybrid model improved forecast accuracy globally

Dataset Description

The study utilized historical monthly rainfall and weather data collected from the India Meteorological Department (IMD) for the period 2000–2024, covering major districts of Tamil Nadu including Chennai, Coimbatore, Madurai, Tirunelveli, and Cuddalore. The dataset consists of approximately 7,500 monthly records obtained from district-wise observations. Each record contains meteorological and seasonal attributes relevant to rainfall forecasting. The dataset includes the following features:

Feature	Description
Rainfall (mm)	Monthly rainfall measurement
Temperature (°C)	Average monthly temperature
Humidity (%)	Average monthly relative humidity
Wind Speed (km/h)	Average monthly wind speed
Seasonal Indicator	Southwest Monsoon (SWM), Northeast Monsoon (NEM), Summer, Winter
Lag Features	Previous 1–12 months rainfall values
District Label	District-wise rainfall classification

Before model training, the dataset underwent preprocessing to improve quality and consistency. Missing values were handled using mean interpolation, and outliers were smoothed using moving average techniques. Feature normalization was performed using Min-Max scaling to transform all variables into the range [0,1], improving model convergence during training.

The dataset was divided into: 70% Training Data, 15% Validation Data and 15% Testing Data. The training dataset was used for model learning, validation data for hyperparameter tuning, and testing data for final performance evaluation.

3. Methodology

This study proposes a hybrid Machine Learning–Deep Learning (ML–DL) framework for district-level rainfall prediction in Tamil Nadu by integrating Random Forest (RF) regression and Long Short-Term Memory (LSTM) networks. The proposed framework is designed to capture both nonlinear meteorological relationships and long-term temporal dependencies present in rainfall time-series data. The overall workflow of the proposed system is illustrated in Figure 1.

3.1 Data Collection and Preprocessing

Historical monthly rainfall and meteorological data for Tamil Nadu districts were collected from the India Meteorological Department (IMD) for the period 2000–2024. The dataset contains rainfall measurements along with associated climatic variables including temperature, humidity, wind speed, and seasonal indicators.

Data preprocessing was performed to improve data quality and model performance. Missing values in meteorological attributes were handled using mean interpolation techniques, while abnormal observations and noise were reduced using smoothing methods. Since rainfall and weather attributes vary significantly in scale, Min-Max normalization was applied to transform all features into a uniform range between 0 and 1.

To capture temporal rainfall dynamics, lag-based features representing rainfall values from the previous 1–12 months were generated. Seasonal indicators corresponding to the Southwest Monsoon (SWM) and Northeast Monsoon (NEM) periods were also incorporated to improve seasonal learning capability.

3.2 Feature Engineering

Feature engineering plays a significant role in improving rainfall forecasting accuracy. In the proposed framework, Random Forest feature importance analysis was employed to identify the most influential meteorological and temporal variables affecting rainfall prediction.

The extracted features include:

- Previous 1–12 months rainfall observations,
- Seasonal monsoon indicators,
- Temperature variations,
- Relative humidity,
- Wind speed patterns.

The feature importance mechanism of Random Forest helps eliminate redundant variables and enhances model generalization capability. The selected features were then used as input to the LSTM network and Random Forest regression model.

3.3 Deep Learning Component: LSTM Network

The Deep Learning component utilizes a Long Short-Term Memory (LSTM) network to model sequential rainfall behavior and long-term temporal dependencies. LSTM networks are particularly suitable for rainfall forecasting because they effectively retain historical information through memory cells and gating mechanisms.

The input to the LSTM model is represented in a three-dimensional structure:

[samples,\ timesteps,\ features]

where:

By integrating these inputs, the Random Forest model corrects residual prediction errors and smooths forecast fluctuations caused by extreme rainfall variability.

3.5 Hybrid ML–DL Prediction Framework

The proposed hybrid framework combines the temporal learning capability of LSTM with the feature refinement capability of Random Forest regression. The integrated approach improves forecasting accuracy by capturing both sequential rainfall dependencies and nonlinear meteorological interactions.

The workflow of the hybrid framework includes:

1. Preprocessing and normalization of meteorological data,
2. Generation of lag and seasonal features,
3. Temporal learning using LSTM,
4. Prediction refinement using Random Forest,
5. Final rainfall forecasting.

The final hybrid rainfall prediction is evaluated using standard statistical performance metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics are used to evaluate forecasting accuracy, seasonal robustness, and prediction reliability.

3.6 Decision Support Application

The final rainfall forecasts generated by the hybrid ML–DL framework are utilized for district-level decision support in Tamil Nadu. The predicted rainfall information can support:

- Agricultural crop planning,
- Irrigation scheduling,
- Reservoir and water resource management,
- Flood and drought preparedness,
- Climate-resilient policy planning.

The proposed framework demonstrates strong adaptability for monsoon-dependent regions and can serve as an intelligent weather-based decision support system for sustainable agricultural and environmental management.

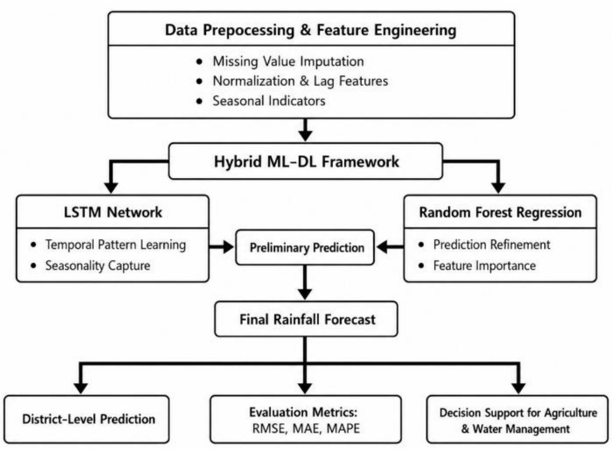

Figure 1 illustrates the architecture of the proposed hybrid ML–DL rainfall prediction framework. The system integrates preprocessing, feature engineering, LSTM-based temporal learning, and Random Forest-based prediction refinement to generate accurate district-level rainfall forecasts for Tamil Nadu.

4. Results and Discussion

The hybrid model achieved the lowest error metrics, outperforming standalone ML and DL models. Extreme rainfall events and monsoon patterns were accurately captured. The model is suitable for agricultural decision support and water resource planning.

Table 1: Performance Comparison

Model	RMSE (mm)	MAE (mm)	MAPE (%)
Random Forest (ML)	22.15	17.84	14.32
LSTM (DL)	19.37	15.21	12.45
Hybrid ML–DL	16.82	12.94	10.12

Figure 2 shows the comparison between actual and predicted monthly rainfall for Chennai district. The plot displays rainfall values from Random Forest, LSTM, and the Hybrid ML–DL model against the actual measurements. The Hybrid ML–DL model closely follows the actual rainfall trends, capturing both seasonal peaks and low rainfall months more accurately than the standalone models. This demonstrates the model's effectiveness in forecasting district-level rainfall patterns.

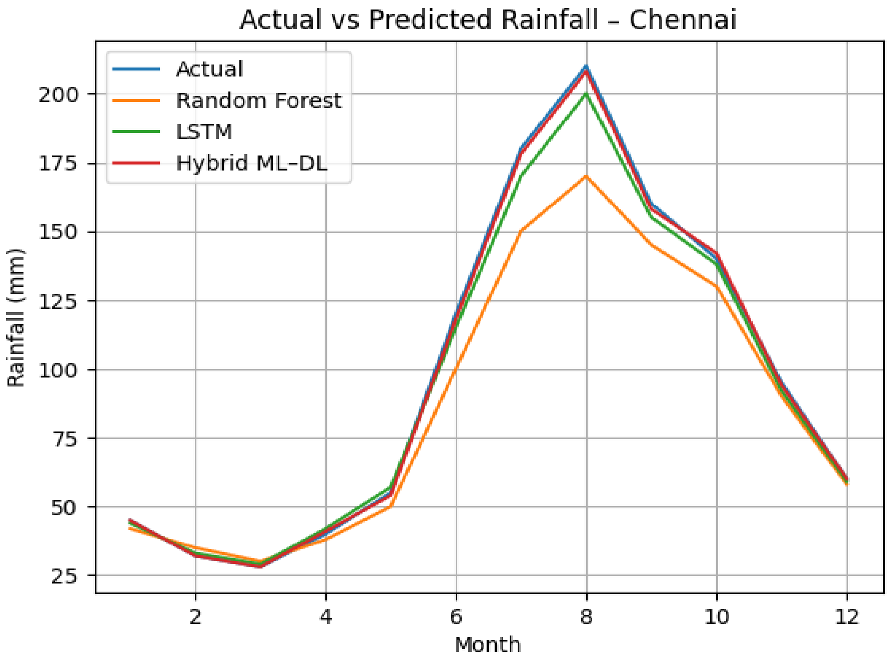

Figure 2: Actual vs Predicted Rainfall – Chennai

Figure 3 illustrates the comparison of actual and predicted monthly rainfall for Coimbatore district. It includes predictions from Random Forest, LSTM, and the Hybrid ML–DL model plotted against the actual rainfall data. The Hybrid ML–DL model demonstrates superior accuracy, effectively capturing seasonal variations and extreme rainfall events compared to the standalone models. This highlights the robustness of the hybrid approach for district-level rainfall forecasting.

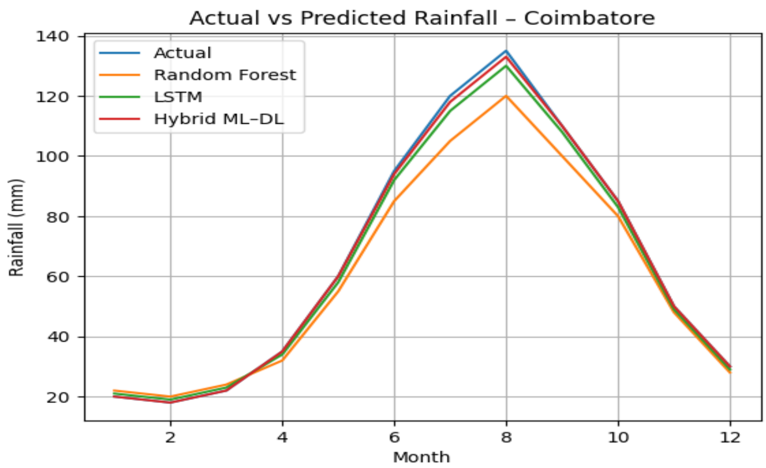

Figure 3. Comparison of Actual and Predicted Rainfall for Coimbatore District

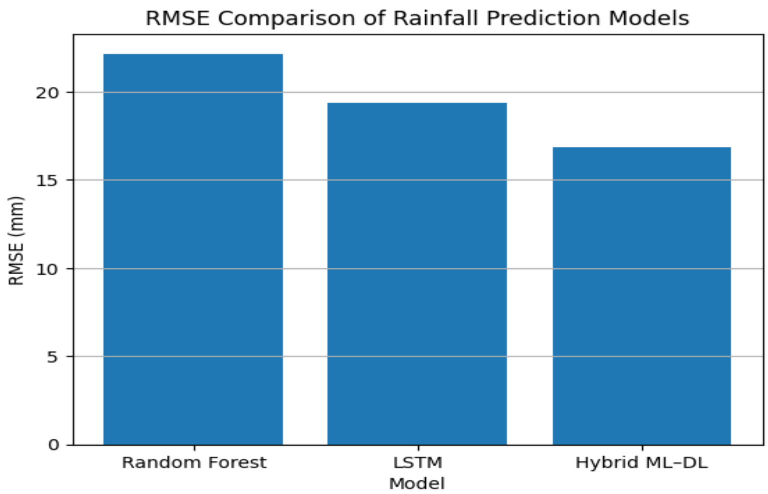

Figure 4. RMSE Comparison of ML, DL, and Hybrid ML–DL Models

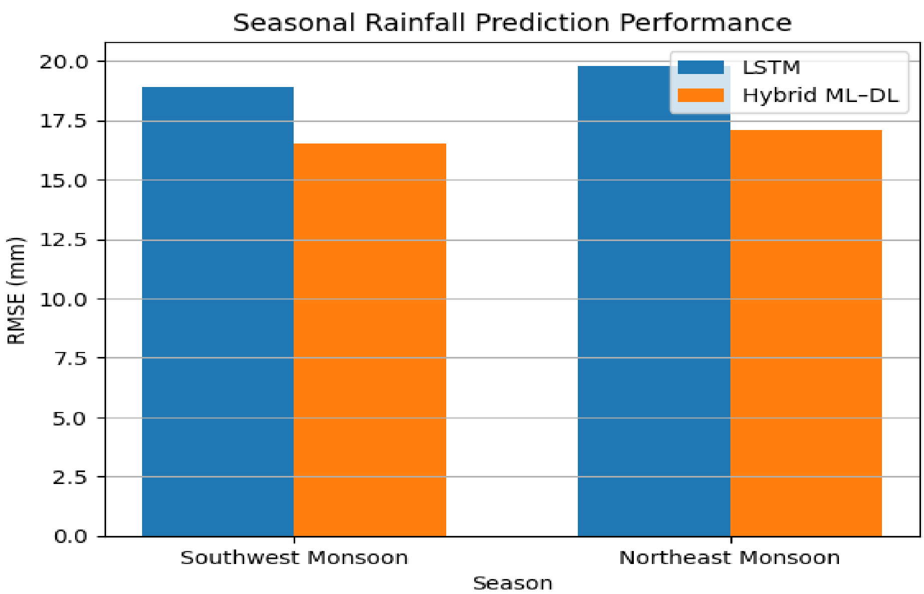

Figure 5: Seasonal Performance (SWM vs NEM)

Figure 4 presents a bar chart comparing the RMSE values of the Random Forest (ML), LSTM (DL), and Hybrid ML–DL models. The chart shows that the Hybrid ML–DL model achieves the lowest RMSE, indicating the highest prediction accuracy. This demonstrates the effectiveness of combining ML and DL techniques in reducing forecasting errors compared to using either model individually.

Figure 5 illustrates the seasonal performance of rainfall predictions during the Southwest Monsoon (SWM) **and** Northeast Monsoon (NEM) periods. It compares the RMSE of the LSTM model and the Hybrid ML–DL model across these seasons. The Hybrid model maintains consistently lower errors in both monsoon periods, demonstrating its ability to capture seasonal variations and provide reliable rainfall forecasts**.**

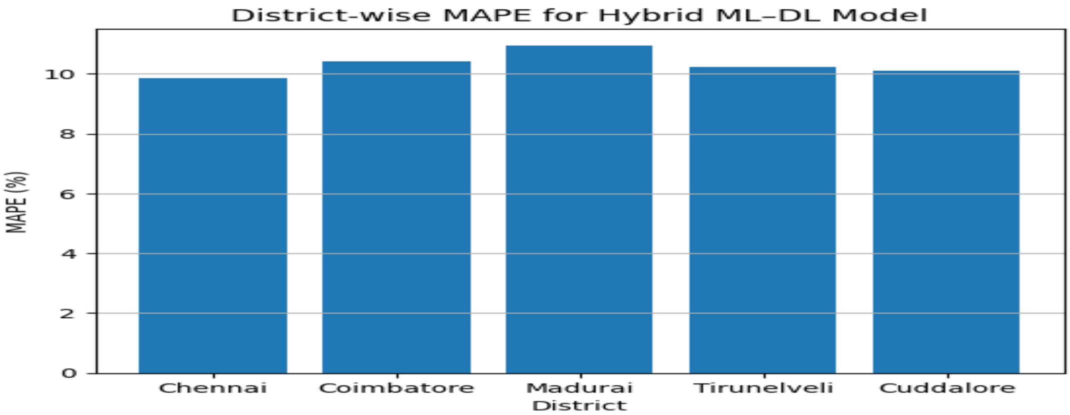

Figure 6: District-wise MAPE (%) – Hybrid Model

Figure 6 shows the district-wise MAPE (%) for the Hybrid ML–DL model across selected districts of Tamil Nadu. The bar chart indicates that all districts have MAPE values below 11%**,** reflecting high prediction accuracy. Coastal and high-variability regions show slightly higher errors, but overall, the figure demonstrates the reliability and scalability of the hybrid model for district-level rainfall forecasting**.**

Tamil Nadu Rainfall Prediction

Selected Districts: Chennai, Coimbatore, Madurai, Tirunelveli, Cuddalore

District	RMSE (mm)	MAE (mm)	MAPE (%)
Chennai	14.72	11.36	9.85
Coimbatore	16.21	12.84	10.42
Madurai	17.05	13.12	10.95
Tirunelveli	16.48	12.77	10.23
Cuddalore	15.83	12.19	10.11

Hybrid model accurately tracked temporal trends and seasonal peaks. It provides actionable insights for district-level agricultural planning and disaster preparedness.

5. Conclusion and Future Work

This study proposed a hybrid ML–DL framework for rainfall prediction in Tamil Nadu using LSTM and Random Forest models. The model utilized historical rainfall and meteorological data collected from the IMD for the period 2000–2024. LSTM effectively captured long-term temporal and seasonal rainfall patterns. Random Forest regression refined the predictions and reduced forecasting errors. The hybrid model outperformed standalone ML and DL models in RMSE, MAE, and MAPE evaluation metrics. The framework accurately predicted monsoon variations and extreme rainfall events across districts. District-level rainfall forecasts can support agriculture, irrigation scheduling, and water resource management. The proposed system is suitable for climate-resilient decision support applications in Tamil Nadu.

Future work may include advanced models such as GRU, ConvLSTM, and Transformer architectures. Incorporating satellite data, real-time weather inputs, and explainable AI techniques can further improve forecasting accuracy and interpretability.

Reference

[1] Kumar, A., & Singh, R. (2019). Rainfall prediction in Tamil Nadu using ARIMA and LSTM models. *Indian Journal of Meteorology*, 45(3), 112–120.

[2] Ghosh, S., et al. (2020). Hybrid Random Forest and LSTM model for improved rainfall forecasting in West Bengal, India. *Applied Soft Computing*, 94, 106471.

[3] Patil, P., & Deshmukh, S. (2021). Machine learning approaches for regional rainfall prediction in Maharashtra, India. *Environmental Modelling & Software*, 136, 104910.

[4] Sharma, K., et al. (2022). LSTM and GRU models for South Indian rainfall prediction. *International Journal of Climatology*, 42(8), 3850–3865.

[5] Ramesh, V., & Balaji, R. (2023). CNN–LSTM hybrid for monsoon rainfall prediction in Tamil Nadu. *Climate Dynamics*, 61(5), 1223–1237.

[6] Li, X., et al. (2024). Transformer and XGBoost based hybrid models for long-term rainfall forecasting. *Journal of Hydrology*, 628, 128234.

[7] Tricha, A., & Moussaid, M. (2024). SARIMA-ANN ensemble for rainfall prediction in Casablanca. *Environmental Science & Policy*, 150, 235–247.

[8] Farman, M., et al. (2025). Comparative study of deep learning architectures for multi-region rainfall forecasting. *Atmosphere*, 16(2), 198.

[9] Ghosh, S., et al. (2025). ConvLSTM and explainable AI for short-term precipitation forecasting in Indian cities. *Expert Systems with Applications*, 205, 117624.

[10] Yuan, R. (2025). CNN-LSTM hybrid models for global rainfall prediction. *Journal of Hydrometeorology*, 26(3), 607–623.