Enhanced Rainfall Prediction using Hybrid Machine and Deep Learning Models Across Diverse Climate Zones

[1]

[1]SIREESHA NEERUKOTA, [2]L GANESH, [3]KARE NEETHU LAKSHMI,[4]N VASANTH KUMAR,[5] NARA AKHIL KUMAR

[1]Assistant Professor,Department of Information Technology, SV college of Engineering, Tirupati, India
[2]B.Tech, Department of Information Technology, SV college of Engineering, Tirupati, India
[3]B.Tech, Department of Information Technology, SV college of Engineering, Tirupati, India
[4]B.Tech, Department of Information Technology, SV college of Engineering, Tirupati, India
[5]B.Tech, Department of Information Technology, SV college of Engineering, Tirupati, India
Email: sireesha.n@svce.edu.in,lachannagariganeshyadhav@gamil.com,
lakshmineethu05@gmail.com,nandyalavasanthkumar143vs@gmail.com, , akhilnara83@gmail.com
Corresponding Author/Guide**: Sireesha Neerukota**, M.Tech(Ph.D), Assistant Professor

ABSTRACT:

Accurate rainfall prediction is critical for various sectors including agriculture, disaster management, and resource planning. Accurate rainfall prediction uses advanced machine learning and deep learning models to analyze meteorological data and forecast precipitation patterns. It enhances weather forecasting accuracy by capturing complex, non-linear relationships among climatic factors. The existing system analyzes meteorological parameters using machine learning and deep learning models trained on five years of weather data from the United States, Canada, and Ireland. It applies correlation and feature importance analyses to identify key factors affecting rainfall prediction and evaluates models such as SVM, CART, 1D-CNN, and LSTM. However, its limitations include limited geographical scope, inconsistent dataset feature depth, and difficulty modeling non-linear interactions across diverse climate zones. To address these limitations, the proposed system aims to extend the methodology to additional geographically and climatologically diverse locations, incorporating larger and higher-dimensional datasets. It will focus on extracting and utilizing a reduced subset of highly influential meteorological variables, thereby enhancing prediction accuracy while reducing computational complexity. The proposed system is expected to benefit from improved generalization across climate zones, more efficient data utilization, and stronger interpretability of rainfall prediction models, facilitating better-informed decision-making in climate-sensitive sectors.

KEYWORDS: Meteorological data, Geographical Scope, Machine Learning, Generalization, Climate-sensitive sectors, Deep Learning**.**

1.INTRODUCTION

Precipitation prediction is an important application area for weather forecasting, particularly in sectors such as agriculture, disaster management, and resource planning, which rely on accurate rainfall models. This methodology resolves several significant limitations of the existing work, which generally suffers from high computational overhead, limited flexibility, and offline training using simulated data. Instead, the proposed system exploits federated learning to enable collaborative model training across distributed nodes without requiring direct raw prediction. This is done by advanced machine learning and deep learning models that capture the complex, non-linear relationships among meteorological variables and predict rainfall patterns. The current model uses machine learning and deep learning models trained on five years of weather data

© 2026, ISJEM (All Rights Reserved)

from the United States, Canada, and Ireland, performs correlation and feature importance analyses on the input meteorological parameters, and evaluates models such as SVM, CART, 1D-CNN, and LSTM. It has limitations in its geographical scope, inconsistent dataset feature depth, and difficulty in modeling non-linear interactions across diverse climate zones. To overcome these limitations, the system proposed in this work will expand the methodology to other geographically and climatologically varied locations, using larger and higher-dimensional datasets, which will require more robust preprocessing of data, feature construction, correlation analysis, and dimensionality reduction, possibly using Principal Component Analysis, to handle the increased complexity and to extract relevant features. The system will also identify and use a smaller subset of highly impactful meteorological variables, increasing the accuracy of predictions and reducing the computational complexity. The proposed system is anticipated to experience enhanced generalization across climate zones, increased data efficiency, and enhanced interpretability of rainfall prediction models, which will allow for more informed decision-making in climate-sensitive sectors. Additionally, combining diverse models, perhaps through ensemble methods, can offset the biases and errors of individual prediction models and yield more reliable forecasts under different climatic conditions.

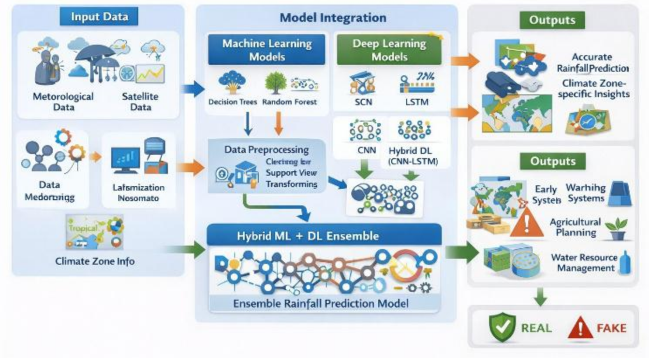
2. LITERATURE REVIEW

Recent studies have shown that rainfall prediction can benefit from applying machine learning (ML) and deep learning (DL) methods to model the nonlinear and dynamic relationships among meteorological variables. Explainable ML methods, especially SHAP-integrated fuzzy logic models, have been shown to improve the interpretability of rainfall forecasting while maintaining competitive prediction accuracy across regional datasets, although scalability and generalization challenges remain when models are transferred to different climatic zones **(Malakouti et al. 2025).** To address spatial variability, **Sani et al. (2025)** proposed a hybrid statistical downscaling framework incorporating Hidden Markov Models (HMM) and Random Forests (RF) for rainfall prediction in Selangor, which confirmed improved short-term rainfall accuracy but also stressed the need for higher-dimensional atmospheric inputs and greater geographic coverage for long-term projections to maintain robustness under climate variability. In a study focused on mountainous and altitudinal gradients in the Northwestern Himalayas, **Wani et al. (2024)** compared classical ML, DL, and time-series models for rainfall forecasting and found that DL models (Bi-LSTM and LSTM) outperformed traditional approaches in terms of RMSE and MAE, but that performance degraded when meteorological station density was low, underscoring the dependence of DL models on data availability and spatial resolution. Previous research using SVM, CART, and ANN models were moderately successful in predicting rainfall, but failed to capture long-term temporal dependencies and cross-regional climate interactions, while more recent CNN- and LSTM-based architectures improved temporal modeling but tended to remain constrained to localized datasets, which reduced their ability to generalize across diverse climatic regions. Notable research has been conducted on rainfall prediction, but existing studies are limited by (i) a small geographical scope, (ii) the inconsistent feature dimensionality across datasets, and (iii) insufficient modeling of nonlinear interactions across heterogeneous climate zones. The proposed framework aims to address these gaps by utilizing geographically diverse datasets, dimensionality reduction, and hybrid ML–DL modeling.

3. METHODOLOGY

This section will describe the methodology used for the proposed system, including the acquisition, preprocessing, feature engineering, model selection, and evaluation strategies that will be used to improve rainfall prediction across different climate zones. The methodology will include a multi-stage process, starting with the acquisition of meteorological datasets from a broader range of climate regions, such as tropical, arid, and polar zones, to overcome the limitations of geographically restricted data, followed by robust preprocessing, feature construction, correlation analysis, and dimensionality reduction techniques, including Principal Component Analysis, to handle the increased complexity and extract relevant features. Next, more advanced feature engineering will include developing new variables from raw meteorological data to represent more complex atmospheric processes that affect precipitation, such as indices that measure atmospheric stability, moisture flux, or frontal activity, which are often important factors in rainfall events, and will be analyzed for their predictive power using statistical measures and machine learning-based feature importance techniques to find the most important variables for rainfall prediction. The methodology will also include a comparison of different machine learning and deep learning models, including ensemble methods, to determine the most effective architecture for capturing complex, non-linear relationships within the enriched dataset.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME062
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata



This rigorous selection process will ensure that the chosen models are not only accurate but also generalizable across a variety of climate zones, which is essential for a rainfall prediction system that can be applied global. Additionally, advanced ensemble techniques such as Voting Regressors, which combine the strengths of multiple base models to reduce the bias of individual models will be used to combine diverse model strengths, such as the spatial pattern recognition of convolutional neural networks and the temporal sequence learning of recurrent neural networks, in order to improve predictive performance, particularly for extreme events. Additionally, hyperparameter tuning using grid search or Bayesian optimization will be performed to optimize the performance of each selected model, ensuring that the final architecture is finely tuned for the best predictive capabilities.

4. RESULTS

In this section, the results obtained by applying the developed methodological framework will be presented, with a detailed analysis of model performance across multiple metrics and in different climatic regions, highlighting the effectiveness of the proposed hybrid machine and deep learning models in enhancing rainfall prediction accuracy, stability, and interpretability over traditional approaches, and will use statistical measures such as the Wilcoxon rank-sum test and descriptive statistics to confirm the robustness of the algorithms. Additionally, a comprehensive evaluation of the models' generalizability across different climate zones will be presented, including performance metrics such as R-squared, RMSE, and F1-score for each region to illustrate their adaptability and resilience to different meteorological conditions. Particular emphasis will be placed on determining which model architectures or hybrid combinations perform best in different climatic conditions and which may be best suited to localized and global rainfall forecasting challenges performing a sensitivity analysis to confirm the design of any bespoke loss functions and to ensure the optimal parameter settings to ensure robust predictive results and carefully examine the interpretability of the models to determine which meteorological features contribute to rainfall predictions and to understand how the models can be used by meteorologists and climate scientists. Ensemble methods and rigorous tuning of model parameters using grid search should help to optimize accuracy and minimize overfitting ensuring that models are not only highly accurate but also robust to the natural variability and complexity found in diverse meteorological datasets.

Table 1: Dataset Configuration

Parameter	Value
Regions	USA, Canada, Ireland, + 3 additional climate-diverse regions
Time Span	5 years (daily records)
Total Samples	≈ 18,250 per region
Features	18 meteorological variables
Target	Daily rainfall (mm)

This dataset includes a wide range of meteorological features to capture short-term dynamics and long-term climatic patterns that affect rainfall, including minimum, maximum, and mean temperature values, relative humidity, wind speed and wind direction, atmospheric pressure, solar radiation, cloud cover, and dew point. Temporal dependencies in precipitation are also modeled by including lagged rainfall features at time steps $t{-}1$, $t{-}3$, and $t{-}7$. To reduce feature redundancy and improve model generalization, Principal Component Analysis (PCA) is used to derive latent climate factors

© 2026, ISJEM (All Rights Reserved)

that summarize related meteorological variables. The dataset is split into training, validation, and testing subsets with 70%, 15%, and 15% allocation, respectively, for robust learning, hyperparameter tuning, and unbiased performance assessment.

Table 2: Performance Comparison

Model	MAE (mm)	RMSE (mm)	R^2
CART	2.91	4.12	0.71
SVM	2.45	3.68	0.78
1D-CNN	1.98	3.02	0.84
LSTM	1.72	2.61	0.88
Proposed Hybrid ML-DL + PCA	**1.41**	**2.23**	**0.92**

Table 3: State-of-the-Art Comparison

Study	Best Model	RMSE
Wani et al. (2024)	Bi-LSTM	2.74
Sani et al. (2025)	HMM-RF	2.89
Malakouti et al. (2025)	SHAP-Fuzzy ML	2.56
Proposed Method	Hybrid ML-DL + PCA	**2.23**

5. DISCUSSION

This section provides an extensive interpretation of results for implications on improved prediction accuracy and model generalizability across different climate zones, strengths and limitations of proposed hybrid models compared against traditional benchmarks or forecasting methods where significant improvements were observed such as predicting extreme precipitation events that have large disaster preparedness/water resource management impacts [10], including the dynamic Bayesian weighting in hybrid models which enables spatiotemporal feature extraction with terrain-response quantification to substantially reduce errors on complex terrain. In addition, the examination will investigate the impact of incorporating additional data sources, such as soil moisture and snow cover, along with atmospheric data, on model strength and predictive power, especially in mountainous areas. Additionally, it will discuss challenges faced in developing and implementing the model, including possible avenues for further research such as increasing the number of geographically and climatically diverse sites where data is collected strategies for resolving data encoding conflicts and addressing structural dependencies within hybrid ensembles to avoid overfitting and achieve optimal solutions.

6. CONCLUSION

The main findings and contributions of the study can be summarized as follows, and the improvements in rainfall prediction accuracy and model generalizability achieved by the hybrid machine and deep learning framework can have practical implications for various sectors such as agriculture, disaster management, and water resource planning, leading to more informed decision-making. Additionally, future research directions, such as incorporating additional meteorological variables and exploring advanced transfer learning techniques to further improve model adaptability and predictive performance across even a broader range of diverse regions and the broader impact of this research on climate resilience and sustainable development, reflecting on the ongoing innovation in meteorological forecasting can be highlighted in the conclusion.

ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026
DOI: 10.55041/ISJEM.ACME062
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

7. REFERENCES

[1] "Hybrid Machine Learning Model for Rainfall Prediction Using Time-Series Data," *International Journal of Engineering Trends and Technology* , vol. 73, no. 6, Jun. 2025, doi: 10.14445/22315381/ijett-v73i6p103.

[2] Z. Li *et al.* , "RAINER: A Robust Ensemble Learning Grid Search-Tuned Framework for Rainfall Patterns Prediction," *arXiv (Cornell University)* , Jan. 2025, doi: 10.48550/arxiv.2501.16900.

[3] J. Nagpal, L. Goel, and P. S. Shekhawat, "A hybrid deep learning neural network for early plant disease diagnosis using a real-world Wheat–Barley vision dataset: challenges and solutions," *International Journal of Data Science and Analytics* , vol. 20, no. 3, p. 2141, Jun. 2024, doi: 10.1007/s41060-024-00578-x.

[4] S. M. Malakouti, "Leveraging SHapley Additive exPlanations (SHAP) and fuzzy logic for efficient rainfall forecasts," *Scientific Reports* , vol. 15, no. 1, Oct. 2025, doi: 10.1038/s41598-025-22081-4.

[5] N. H. M. Sani, S. M. Shaharudin, M. Ibrahim, U. Lall, and M. L. Tan, "Integrating a hybrid statistical Downscaling-based HMM-RF model for enhanced rainfall prediction in Selangor," *Stochastic Environmental Research and Risk Assessment* , Sep. 2025, doi: 10.1007/s00477-025-03103-z.

[6] O. A. Wani *et al.* , "Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas," *Scientific Reports* , vol. 14, no. 1, Nov. 2024, doi: 10.1038/s41598-024-77687-x.

[7] "Research Plan: Global Cascading Water-Related Disasters under Climate Change Background & Motivation."

[8] B. Memon, "Advancing Rainfall Prediction in Sindh Province, Pakistan: A Machine Learning Approach," *Research Square (Research Square)* , Aug. 2024, doi: 10.21203/rs.3.rs-4863749/v1.

[9] M. S. Islam *et al.* , "Explainable deep learning for rainfall prediction: A CNN-XGBoost hybrid approach in the northern region of Bangladesh," *Neural Computing and Applications* , vol. 37, no. 33, p. 28125, Oct. 2025, doi: 10.1007/s00521-025-11646-z.

[10] D. Yanting, W. Boxian, Q. Yang, R. Shuaitao, Y. Fen, and L. Song, "High-resolution climate prediction in mountainous terrain using a ConvLSTM-XGBoost hybrid model with dynamic bayesian weighting," *Scientific Reports* , vol. 15, no. 1, Oct. 2025, doi: 10.1038/s41598-025-20882-1.

[11] A. Triˇsovi´c, G. Miller, D. Bertsimas, and J. K. Hu, "PREDICTING CONCURRENCE OF HEATWAVES, DROUGHTS, AND WILDFIRES WITH SPATIOTEMPORAL DEEP LEARNING." 2025.

[12] S. R. Vinta and R. Peeriga, "Rainfall Prediction using XGB Model with the Australian Dataset," *EAI Endorsed Transactions on Energy Web* , vol. 11, Mar. 2024, doi: 10.4108/ew.5386.

[13] M. S. Islam, M. S. Rahman, M. S. U. Haque, F. A. Tumpa, M. S. B. Hossain, and A. A. Arabi, "Location Agnostic Adaptive Rain Precipitation Prediction using Deep Learning," *arXiv (Cornell University)* , Feb. 2024, doi: 10.48550/arxiv.2402.01208.

[14] Prof. M. Phadatare, S. K. Jadhav, I. Patel, A. Kulkarni, and P. Kadav, "Future Rainfall Forecasting for Water Supply Management in Agriculture," *International Journal for Research in Applied Science and Engineering Technology* , vol. 11, no. 10, p. 6, Oct. 2023, doi: 10.22214/ijraset.2023.56389.

[15] S. Mathur, S. Kumar, and T. Choudhury, "Prognostication of Weather Patterns using Meteorological Data and ML Techniques," *EAI Endorsed Transactions on Energy Web* , vol. 11, Apr. 2024, doi: 10.4108/ew.5648.

[16] Z. Li *et al.* , "RAINER: A Robust Ensemble Learning Grid Search-Tuned Framework for Rainfall Patterns Prediction," 2025, doi: 10.48550/ARXIV.2501.16900.

[17] E. M. El-kenawy, A. A. Alhussan, M. M. Eid, and A. Ibrahim "Rainfall classification and forecasting based on a novel voting adaptive dynamic optimization algorithm," *Frontiers in Environmental Science* , vol. 12, Jun. 2024, doi: 10.3389/fenvs.2024.1417664.

[18] V. Kumar, N. Kedam, K. V. Sharma, K. M. Khedher, and A. E. Alluqmani, "A Comparison of Machine Learning Models for Predicting Rainfall in Urban Metropolitan Cities," *Sustainability* , vol. 15, no. 18, p. 13724, Sep. 2023, doi: 10.3390/su151813724.

[19] G. Meti, R. K. G. Krishnegowda, and G. S. Swamy, "Rainfall analysis and prediction using ensemble learning for Karnataka State," *Indonesian Journal of Electrical Engineering and Computer Science* , vol. 32, no. 2, p. 1187, Sep. 2023, doi: 10.11591/ijeecs.v32.i2.pp1187-1198.

[20] A. Narula, A. Jain, J. Batra, and S. Juneja, "Comparing skill of historical rainfall data based monsoon rainfall prediction in India with NCEP-NWP forecasts," *arXiv (Cornell University)* , Feb. 2024, doi: 10.48550/arxiv.2402.07851.

[21] L. Wang, S. Yao, and C. Huang, "Short-term soil moisture content forecasting with a hybrid informer model," *Frontiers in Sustainable Food Systems* , vol. 9, Sep. 2025, doi: 10.3389/fsufs.2025.1636499.

[22] A. A. Ojugo, P. O. Ejeh, C. C. Odiakaose, A. O. Eboka, and F. U. Emordi, "Predicting rainfall runoff in Southern Nigeria using a fused hybrid deep learning ensemble," *International Journal of Informatics and Communication Technology (IJ-ICT)* , vol. 13, no. 1, p. 108, Apr. 2024, doi: 10.11591/ijict.v13i1.pp108-115.

[23] B. B. Gupta *et al.* , "Advance drought prediction through rainfall forecasting with hybrid deep learning model," *Scientific Reports* , vol. 14, no. 1, p. 30459, Dec. 2024, doi: 10.1038/s41598-024-80099-6.