Machine learning techniques for rainfall prediction: a systematic literature review

Deepa Sharma[1], **Anand Kumar Shukla**[1], **Punam Rattan**[2]
[1]School of Computer Applications, Lovely Professional University, Phagwara, India
[2]Department of Computer Applications, Manav Rachna University, Faridabad, India

Article Info

Article history:

Received Sep 23, 2024
Revised May 17, 2026
Accepted Jun 19, 2026

Keywords:

Deep learning
Ensemble learning
Hybrid model
Long short-term memory
Machine learning
Rainfall prediction

ABSTRACT

There are numerous aspects of human life in which knowing how much rain to expect might be beneficial. Heavy rainfall events such as flash floods and landslides, as well as droughts, can be predicted with effective rainfall forecasting. Because of reliable weather forecasts, the infrastructure required to capture rainwater and cultivate crops may be planned ahead of time. A variety of machine learning (ML) and deep learning (DL) algorithms enable accurate weather forecasting. This work seeks to provide a full overview of the numerous ML algorithms used for rainfall prediction by focusing on the technique, input parameters, and several performance measures. The review consists of 51 works divided into three sections. It is found that long short-term memory (LSTM), one of the DL algorithms, is mostly used by researchers for developing the model, but in recent years, ensemble learning and hybrid learning have also gained popularity among researchers as they give more accurate results. These methods need to be explored further.

Corresponding Author:

Punam Rattan
Department of Computer Applications, Manav Rachna University
Faridabad, Haryana, India
Email: punamrattan@gmail.com

1. INTRODUCTION

The amount of rainfall on the planet is the source of fresh water and thus has a major impact on the survival of plants and animals. Floods [1] and landslides [2], [3] result from excessive precipitation, whereas droughts and crop failures [4] result from insufficient precipitation due to infrequent or non-existent rainfall. Both situations are unfortunate for the residents concerned. The World Bank (2023) reports that 70% of natural disasters are caused by hydrometeorological events that lead to global economic losses of USD 300 billion annually. The unpredictable nature of monsoon patterns in India has resulted in agricultural losses of more than INR 30,000 crore annually, as per the reports of the Indian Meteorological Department (IMD) 2022. These figures underscore the critical necessity for precise and prompt rainfall forecasting, as even minor inaccuracies in predictions can significantly affect food security, disaster mitigation, and water resource management.

Conventional statistical and numerical models for rainfall prediction frequently encounter difficulties in accurately representing the nonlinear and highly dynamic characteristics of the rainfall processes. However, recent progress in machine learning (ML) and deep learning (DL) has enhanced prediction accuracy by extracting intricate spatial–temporal patterns directly from data. Numerous studies have utilized artificial neural networks (ANN), convolutional neural network (CNN), long short-term memory (LSTM), and hybrid ensemble architectures, yielding promising outcomes across various climatic regions.

Many existing reviews are either outdated or narrowly focused on specific model families or geographic areas, which results in a limited understanding of how current ML and DL models compare in terms of architecture, datasets, and evaluation metrics. This review seeks to bridge this gap by systematically investigating the recent progress in rainfall forecasting using ML and DL approaches. It presents: i) a comparative analysis of prominent models and algorithms, ii) a synthesis of performance metrics and methodological trends, and iii) identification of current challenges and research opportunities for creating robust, data-driven rainfall prediction systems.

This review is structured around the following research questions:

– Which ML and DL techniques, including ANN, CNN, recurrent neural network (RNN), LSTM, gated recurrent unit (GRU), and hybrid ensembles, have proven the most effective for predicting rainfall across different timeframes and geographical areas?
– What types of input parameters, such as meteorological, hydrological, and satellite-derived features, are typically used, and how do they influence the performance of the models?
– How do various performance metrics and comparative analyses highlight the strengths, weaknesses, and generalization abilities of different modeling strategies?

This review aims to offer a thorough synthesis of the current methods, benchmark outcomes, and new trends that can inform the creation of more precise and adaptable rainfall forecasting systems.

2. RESEARCH METHODOLOGY AND SYSTEMATIC LITERATURE REVIEW

2.1. Focus of review

For each paper the focus is on the category of ML and DL technique used for developing the model, dataset parameters and source, and nature of the dataset (numeric/image). The focus also includes the time period and data frequency (daily, monthly) of the dataset collected, as well as the performance metrics used for comparing the accuracy of the model with other models. In addition, the focus is on the time period for which the prediction is done (very short, short, medium, long), and geographical area of the research.

2.2. Nature and source of input parameters used in the dataset

The model uses a dataset that can be of either numeric or image type. RNN/LSTM models use numerical meteorological data, whereas CNN models use satellite image data. All authors used data from meteorological departments, weather stations, or open climate repositories. Researchers have examined both local (e.g., temperature, wind speed, humidity) and global factors e.g., El Nino, Indian Ocean Dipole (IOD), sea level pressure, that affect rainfall.

2.3. Search strategy and timeframe

An extensive literature review was performed to locate studies concerning ML methodologies for predicting rainfall. To ensure thorough coverage of the existing literature, four principal scientific databases were queried using predetermined Boolean search expressions. Table 1 provides a summary of the databases searched along with the respective Boolean search strings.

Table 1. Search strategy and keywords used for the systematic review

Aspect	Description
Databases searched	Google Scholar, ScienceDirect, IEEE Xplore, and Scopus
Boolean query used	(“rainfall forecasting” OR “rainfall prediction” OR” precipitation prediction”) AND (” machine learning” OR “deep learning” OR “neural networks” OR “ensemble learning” OR” hybrid model”)
Supplementary keywords	“Support vector machines”, “decision trees”, “time series analysis”, “weather forecasting”, “meteorology”, “remote sensing”, “data mining”, “big data”,” deep learning”, “climate change”
Publication time period	January 2019-April 2024
Language and type	Only peer-reviewed journals and conference papers in the English Language
Purpose	To identify original studies applying ML/DL/ensemble methods for rainfall prediction

2.4. Inclusion and exclusion criteria

Predefined inclusion and exclusion criteria were established prior to the screening process to ensure the relevance and quality of the selected studies. These criteria were consistently applied during the screening stages of titles, abstracts, and full texts. The inclusion and exclusion criteria are detailed as follows:

– Language: an English language abstract was required for the review.
– Rainfall prediction criterion: only studies on rainfall prediction models using ML, DL, and ensemble learning were included. Predictions for rainfall runoff, wind speed, power load, and temperature were eliminated.

– Outcome criterion: studies that compared different models and algorithms for their prediction accuracy were included. This helps in a thorough understanding of the workflow of the models providing an idea of the best techniques to be used for prediction.
– Original research: includes original research, skipping reviews, panel talks, and other studies.

2.5. Screening and PRISMA flow

Studies were identified, screened and included in accordance with preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 guidelines to ensure both transparency and reproducibility. Figure 1 provides a visual representation of the PRISMA-based selection process [5]. Initially, 190 records were gathered from Google Scholar, ScienceDirect, IEEE Xplore, and Scopus. After eliminating 50 duplicate entries, 140 papers remained. During the screening process, titles and abstracts were reviewed to determine their relevance to rainfall prediction using ML/DL methods; 40 records were dismissed because they were either unrelated or non-technical. The remaining 100 full-text articles were evaluated for eligibility. At the eligibility stage, 49 articles were excluded because of the absence of rainfall-specific analysis, missing performance metrics, or because they were reviews or conceptual papers. Ultimately, 51 studies met all criteria and were included in the study.

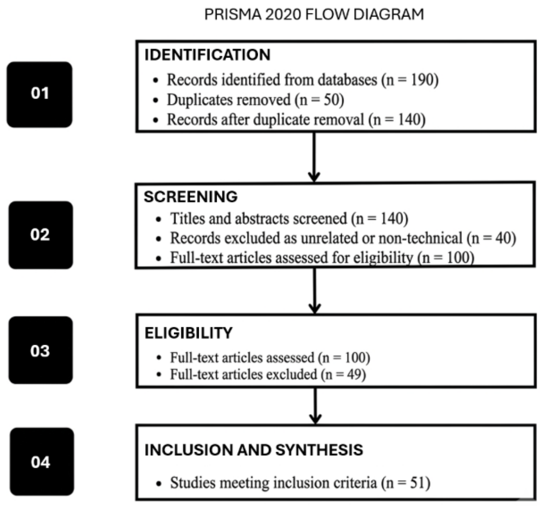

Figure 1. PRISMA framework illustrating the selection process of the papers

2.6. Technique and algorithms used for the model development

The LSTM, a form of RNN, was utilized for meteorological numeric data, and a CNN for image data (radar/satellite). Recent studies have used ensemble/hybrid models combining both image and numeric data to enhance accuracy. Big data and the internet of things (IoT) provide a wealth of data for DL models capable of capturing complex nonlinear patterns.

2.7. Determining model parameters

Regional rainfall is affected by both local and global climatic factors such as temperature, humidity, sunshine, wind, El Niño, and IOD moment. Researchers use the covariance between parameters to determine the link between rainfall and other factors to train the model with only significant parameters. For instance, Lopez *et al.* [6] found a Clausius–Clapeyron relationship between atmospheric pressure, dewpoint, and humidity that predicted precipitation, whereas Pathan *et al.* [7] found that wind speed and minimum temperature are strong indicators of rainfall.

2.8. Data extraction and synthesis

Data from each study were organized into categories such as the type of model used (ANN, CNN, RNN/LSTM, hybrid, ensemble), details about the dataset (source, duration, frequency, data type), performance indicators (root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and accuracy), and the study's geographical focus and prediction timeframe. Narrative and thematic synthesis categorized the results based on the type of algorithm and the nature of the data. The quantitative findings are detailed in Tables 2 to 4, emphasizing comparative accuracy and trends.

Table 2. Summary of models developed for rainfall prediction using different neural networks

Ref.	NN used	Type	Other models for comparison	Time period (dataset)/ dataset type/sources	Performance metrics (value)	Input parameters*
[8]	Temporal CNN	Hybrid	ARIMA, MLP, JMA ECMWF, BPNN, SVMLSTM	2015-2017 (numeric) 11 metrological stations	TR (78.64, 76.12), MAE (1.81), MSE (4.94)	T, RH, WS, SLP
[9]	RNN, LSTM	NA	Holt-winters extreme learning	1980-2013 (34 years) Numeric	RMSE, accuracy (88%), Epochs, loss, LR	T, RH, WS, SR
[10]	RNN, LSTM	NA	NA	August 2020 (numeric) BMD[1]	Accuracy (76%)	T, RH, WS, SLP
[11]	ANN (FFNN), MLP (PSO)	Hybrid	MLP	2009 (numeric) automatic weather station	$RMSE_{PSO}$ (0.14) RMSELM (0.18)	T, RH, WS
[12]	RNN BLSTM-GRU	Hybrid	MLP, LSTM, BLSTM	1997-2017 (numeric) NCHMDB[2]	MSE (0.0075), MSE (0.87), R^2 (0.87), Corr (0.938)	T, RH, WS, SR, RF
[13]	ANN, EEMD-ANN	Hybrid	Hindcast forecast model	1871–2016 (numeric) IITM Pune and IMD	R, MAE, NRMSE IA (Index of agreement)	RF
[14]	CNN, Conv1D-MLP	Hybrid	Deep MLR SVR	1941-2005(Numeric) IMD	RMSE, r, NSE	T, RH, WS, SR, SLP, RF
[15]	RNN, LSTM	NA	Comparison (different parameters)	Dec 2014-Aug 2019 (numeric) meteorological station	MAAPE (0.9644)	C(I), RF(II)
[16]	DNN, Conv3D-GRU	Hybrid	Conv2D Conv2D-GRU	Rainy days (radar images) meteorological station	MSE(2596), MAE(6790), B-MSE, B-MAE	IMG
[17]	RNN, LSTM	NA	FFNN RNN	1901-2017 (numeric) www.data.gov.in	MAPE(79) RMSE(135.4)	RF
[18]	RNN, LSTM	NA	ANN	1901-2017 (numeric) IMD Pune	MSE, RMSE(126.006), MAD, R, CS	RF
[19]	RNN, IRF-LSTM	Hybrid	ANN, SVM, LSTM, CNN-GRU	1981-2020 (numeric) state development planning	NSE(23.56), RMSE(1.5), MAE(0.58), r(0.71)	T, RH, WS, SR, RF
[20]	DNN, FMCGEP-DNN	Hybrid	MLP, BPNN SVM, Random forest	1957-2018 (numeric) ISI3, RNMI4 and AWS[5]	MSE, RMSE, MAE, R^2	RF
[21]	RNN, GA-OLSTM	Hybrid	LSTM	1901-2017 (numeric) IMD Pune	MSE(0.004), RMSE(0.078), CS, R	RF
[22]	CEEMD-CMSE-stacking	Hybrid	CEEMD-LSTM stacking	1960-2019(numeric) NA	RMSE, MAE, R^2	RF
[23]	RNN, RNN-GRU	NA	LSTM	1969-2021 (numeric) IMD	RMSE(12.71), MAE(13.72)	RF, T, WS
[24]	RNN, LSTM	NA	ARIMA, MLP, SVM	1985-2017 (numeric) NMSA[6]	RMSE, MSE, NSE, MAE, MAPE, R^2 (99.72)	T, RH, SR, WS, RF
[25]	LSTM, SVM-BiLSTM	Hybrid	LSTM, BiLSTM SVM-LSTM	1981-2020 (numeric) meteorological Bureau	MSE, NSE, and MAE	RF
[26]	KNN, XGB, SVR, ANN Stacking Model	Ensemble	Baseline models	1961-2019 (numeric) china meteorological data service centre	RMSE, MAE, R^2	C, SLP, WS, T, RH, SR
[27]	RNN-LSTM, Bi-LSTM, CNN	Ensemble	Baseline models	2015-2021 (numeric) IMD	RMSE, MAE, MSE	RF, RH, T
[28]	RNN, LSTM	NA	Random forest	1980-2020 (monthly) numeric meteorological stations	RMSE, RSR, LMI, R^2, NSE, Taylor and Violin diagrams	RF
[29]	DNN, CNN-LSTM	Hybrid	MIM TrajGRU ConvLSTM	June-Oct 2018-19 (numeric) weather stations	CSI, FAR, POD and HSS	RF
[30]	RNN, LSTM, PCA	NA	KNN, LogisticSVM, random forest, NB, NN	2012-18 (numeric) BMD[1]	Accuracy (97.14)	T, RH, WS, SLP
[31]	DNN, ConvLSTM	Hybrid	NA	May 2021 and Dec 2021 Himawari real-time images site, JMA[7]	SSIM, MSE Accuracy (0.9907_{10min}, $.9717_{30min}$ 0.9201_{60min})	IMG
[32]	RNN, LSTM	NA	NA	1901-2015 (numeric) IITM, Pune	RMSE(58.46) for test data	RF
[33]	RNN, CEEMDAN with LSTM	Ensemble	LSTM (EEMD, CEEMD, CEEMDAN)	1951-2020 (numeric) China Meteorological Data Network	MAPE, RMSE, R^2	RF
[34]	DNN, DeepEns-REM	Ensemble	SVR, DT boosting Random forest	July-Dec 2016 (numeric, image) ITCP[8]	CSI, FAR, POD, and MSE	RF

[1]Bangladesh Meteorological Department, [2]National Center of Hydrology and Meteorology Department Bhutan, [3]Indian Statistical Institute, [4]Royal Netherlands Meteorological Institute, [5]Australian Weather Station, [6]National Meteorological Service Agency, [7]Japan Meteorological Agency, [8]Italian Department of Civil Protection. Parameter codes: T = temperature variables (e.g., Tmax, Tmin, mean temperature), RH = relative humidity/dew-point temperature, WS = Wind speed and/or wind direction, SR = Solar radiation/sunshine/longwave radiation/evapotranspiration, SLP = atmospheric or sea level pressure/geopotential height, RF = rainfall/precipitation/rainfall intensity, C = climate oscillation indices (e.g., El Niño, IOD, SOI, SHAM, WP patterns), IMG = radar or satellite image-based spatial inputs, PCA = principal component analysis used for dimensionality reduction.

2.9. Bias and limitations

Some potential biases include the limitation to English-language sources and four primary databases, inconsistent metric reporting across studies, and exclusion of unpublished or preprint work. Some relevant studies might have been left out due to these limitations, which could impact how broadly the review findings can be applied. Furthermore, the wide range of datasets, evaluation methods, and experimental setups used in the studies made it difficult to conduct direct quantitative comparisons. Nevertheless, this review highlights the most significant advancements in rainfall prediction from 2019 to 2024.

3. RESULTS AND DISCUSSION

This section presents the findings of the 51 studies reviewed. The models were assessed based on their architectures, input data, and performance metrics. A clear trend emerged, showing a preference for DL techniques, particularly LSTM, GRU, and hybrid/ensemble models, which generally outperform traditional ML and statistical methods owing to their ability to capture nonlinear and temporal patterns in rainfall data. A detailed comparison of the models used in the selected studies is provided in subsection 3.1.

3.1. Overall comparisons of machine learning/deep learning models

Table 2 offers a comprehensive comparison of the ML and DL models utilized in the reviewed studies, detailing the model types, dataset features, and performance evaluation metrics. Table 2 illustrates that recurrent models, such as LSTM and GRU, excelled in handling time-series rainfall data, whereas CNN and convolutional long short-term memory (ConvLSTM) models were effective when working with spatial radar or satellite images. Hybrid and ensemble models typically achieve the highest accuracy by combining spatial and temporal feature representations.

3.2. Comparison of algorithms on same dataset

Beyond the general model comparisons, numerous studies have specifically assessed various algorithms on identical datasets using the same performance metrics to identify the most effective model. Table 3 deals with comparisons done by researchers between various learning algorithms based on the same dataset and the same performance metrics in order to find the best one. As shown in Table 3, classifiers based on ensemble and boosting techniques, such as random forest, XGBoost, and CatBoost, typically surpass individual classifiers such as naïve Bayes, k-nearest neighbors (KNN), support vector machine (SVM), and decision trees (DT). This is because they can combine multiple decision boundaries and minimize variance. Nonetheless, when datasets display pronounced nonlinear temporal patterns, DL models like LSTM often deliver superior predictive accuracy compared with traditional ML methods.

Table 3. Comparison of multiple learning algorithms evaluated on the same dataset

Ref.	Algorithm used	Dataset used	Performance metrics and conclusion
[35]	C5.0 (with and without SMOTE)	Rainfall, SR, temp, humidity, evaporation, wind speed (2005-2017)	Confusion matrix, improved accuracy when SMOTE is used
[36]	SVR-firefly model	Rainfall (1990-2014)	RMSE and NSE (better accuracy than SVR)
[37]	EMLRM, WA-SVM, ANN, and non-linear regression	47 climate variables (mean values)	RMSE, MAE, and R^2, better performance with the MapReduce algorithm
[38]	DT model produced by J48	Min, max, and avg temp, avg humidity, rainfall, sun exposure time, max and avg wind speed (2013-2019)	Accuracy: 77.8% (training data); 86% (testing data)
[39]	Random forest and logistic regression	2015-18	Confusion matrix, accuracy logistic regression
[40]	Naïve Bayes	Humidity, rainfall, wind speed, and precipitation	Confusion matrix, accuracy, error rate NBC gives an accuracy of 95.91%
[41]	XGBoost	Min, max, and avg temp, avg humidity, rainfall, sun exposure time max, and avg wind speed (2013-2019)	RMSE, MAE, XGBoost increases accuracy and reduces overfitting, and Avg humidity and min temp influence rainfall
[42]	MLR, random forest, and XGBoost	Date, evaporation, sunshine, max and min temp, humidity, wind speed and rainfall (1999-2018)	RMSE, MAE, and XGBoost
[43]	Cat boost classifier, perceptron classification algorithms	23 parameters Nov 2007 - June 2017	Accuracy, confusion matrix cat boost classifier accuracy= 81.36
[44]	Logistic reg linear DA quadratic DA, KNN, DT, gradient boosted trees, random forest bernoulli NB, and deepNN	25 parameters 2007-17	precision, F1-score, and DL model (98.26%,88.61%)
[45]	Linear and lasso reg, ridge model, KNN, and random forest	Rainfall 1901-2015	MAE, MSE, RMSE KNN, and random forest perform better
[46]	DT, KNN, and logistic regression	max and min temp, wind speed, humidity, pressure, and cloud	MAE, MSE, RMSE, KNN (84.183%), and DT (83.762%)

3.3. Comparison of model architectures

In addition to comparing algorithms, numerous studies have assessed various model architectures to identify the most effective method for rainfall prediction. Table 4 presents a comparison of different rainfall prediction models to determine the best model based on the performance metrics. Table 4 shows that the effectiveness of the rainfall prediction model depends on the network architecture. Recurrent DL models like LSTM, Stacked LSTM, and bidirectional long short-term memory (BiLSTM) outperform traditional ANN and RNN models by capturing long-term temporal dependencies in rainfall data. Models with memory gating mechanisms excel at handling sequential rainfall events during multi-step forecasting. Hybrid models combining LSTM with CNN or optimization components show improvements by integrating temporal and spatial learning. These findings show that successful rainfall prediction requires modeling complex temporal interactions rather than using static representations.

Table 4. Comparison of different rainfall prediction model architectures

Ref.	Comparison done between	Best one	Performance metrics used
[47]	MLNPP (50 hidden and Tangent) and MLNPP (100 hidden and Sigmoid)	MLNPP with 50 hidden neurons SCG-tangent	MAE and RMSE
[48]	Echo state network (ESN) and deep ESN	Deep ESN	RMSE, NRMSE, and r
[49]	PSOANFIS, ANN, and SVM	SVM	R, MAE, POD, CSI, FAR, and robustness
[50]	ANN and LSTM	LSTM	MSE, RMSE, and MAE
[51]	BDTR, DFR, NNR, and BLR	BDTR	MAE, RMSE, RAE, RSE, and R^2
[52]	NN, SVM, NB, random forest, and GA	NN	Accuracy= 96.44
[53]	Brazilian Atmospheric model, NN-TensorFlow, and NN-MPCA	NN (other) and NN-MPCA (Spring)	RMSE COV-covariance mean error (ME)
[54]	MLR, SVM, KNN, and random forest	Random forest	Accuracy (89.16), precision, and recall
[55]	MLR, KNN, SVM, ANN, DT, and random forest	Random forest	Accuracy (96.1), precision, recall, RMSE, Error, MAE, and F-measures
[56]	MLR, KNN, SVM, ANN, DT, and random forest	Random forest	Accuracy (96.1), confusion matrix, and ROC
[57]	KNN and DT	DT	Accuracy (93.13)
[58]	LSTM, stacked LSTM, BiLSTM, XGB, and ensemble	Stacked LSTM and BiLSTM	Loss, RMSE, and RMSEL

3.4. Time period for which the dataset is collected

Dataset lengths in the reviewed studies ranged from months to over a century, with most spanning 5-20 years to provide temporal diversity while minimizing historical data issues. Temporal resolution selection depends on forecasting goals: hourly data/daily data/weekly or monthly data for very short-term/ short/ medium-range rainfall intensity. Figure 2 shows the dataset duration distribution across studies.

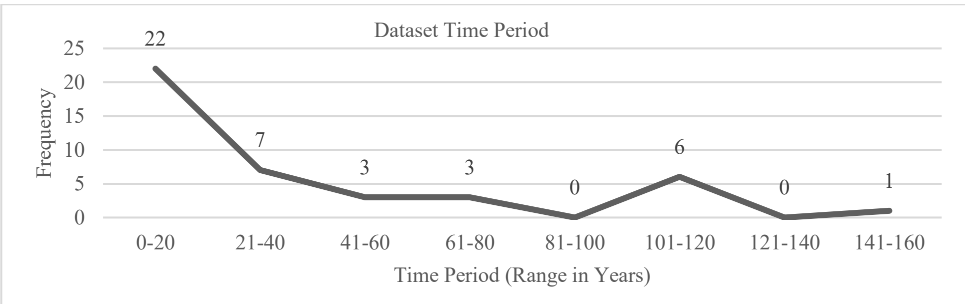

Figure 2. Distribution of the dataset time period

3.5. Frequency of input parameters used

Choosing input parameters is crucial for effective rainfall prediction models, as these directly impact the learning process. In reviewed studies, key meteorological parameters were frequently used due to their strong correlation with rainfall. As shown in Figure 3, the most common inputs are rainfall, temperature, humidity, and wind speed, followed by solar radiation, pressure, cloud cover, and climate indices like ENSO and IOD. These indicate that both local and global climate conditions influence rainfall variability.

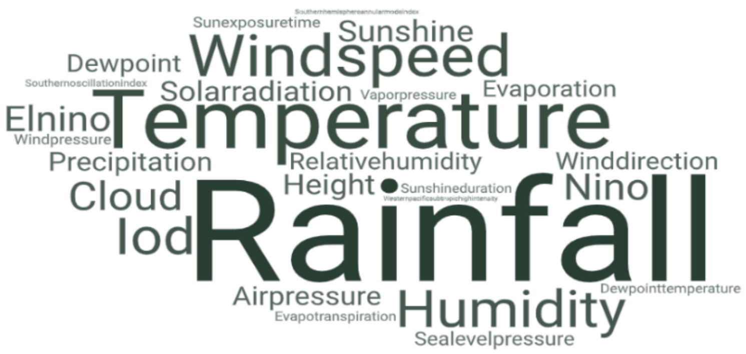

Figure 3. Word cloud representation of the most frequently used input parameters

3.6. Performance metrics used

Performance metrics are essential for evaluating rainfall prediction models' accuracy and stability. Figure 4 shows the frequency of common performance metrics. The RMSE is most widely used as it penalizes larger errors, while MAE is valued for interpretability. Correlation-based metrics like R^2 and Pearson correlation coefficient (r) evaluate prediction alignment with observed patterns. Classification-based studies use confusion matrix metrics, including accuracy, precision, recall, and F-measure. Hybrid forecasting used Nash–Sutcliffe efficiency (NSE) and mean absolute percentage error (MAPE). Figure 4 shows the distribution, with RMSE and MAE as dominant indicators.

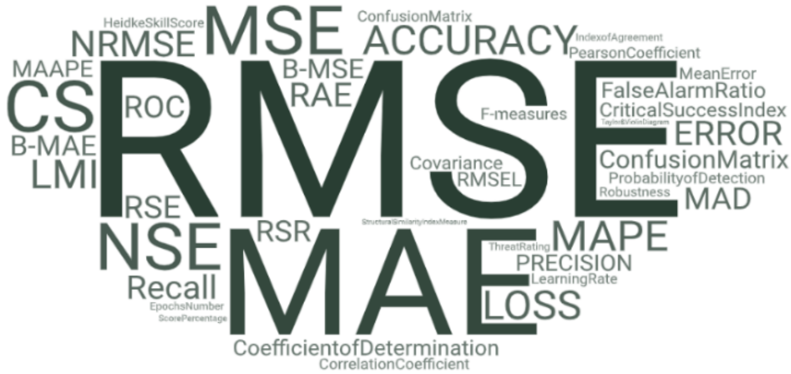

Figure 4. Word cloud representation of different performance metrics used

3.7. Increasing use of ensemble learning for making prediction models

Recent trends in rainfall prediction research have shifted towards hybrid and ensemble learning approaches. While individual models such as LSTM, CNN, or random forest capture specific patterns, hybrid frameworks combine the strengths of multiple algorithms to improve the prediction accuracy. Ensemble models such as stacking, boosting, and bagging reduce model variance by aggregating outputs, while hybrid DL architectures (e.g., CNN-LSTM, ConvLSTM, and GRU-LSTM combinations) integrate spatial and temporal learning. From 1990 to 2019, the number of papers on "ensemble learning" that were released in the core set of Web of Science has grown at a rapid pace [59]. Ganaie *et al.* [60] addressed the effectiveness of ensemble learning and the many strategies used to enhance it, including bagging, boosting, and heterogeneous ensembles. Figure 5 shows an increasing trend in using the hybrid model or ensemble model for prediction compared to earlier times. Recent studies have further reinforced this trend. Kanani *et al.* [61] introduced an ensemble method based on LSTM that greatly enhanced the accuracy of rainfall classification and regression. In a similar vein, Putra *et al.* [62] utilized multisensor data along with ensemble learning to improve high-resolution rainfall estimation, highlighting the increasing move towards hybrid and ensemble-based strategies.

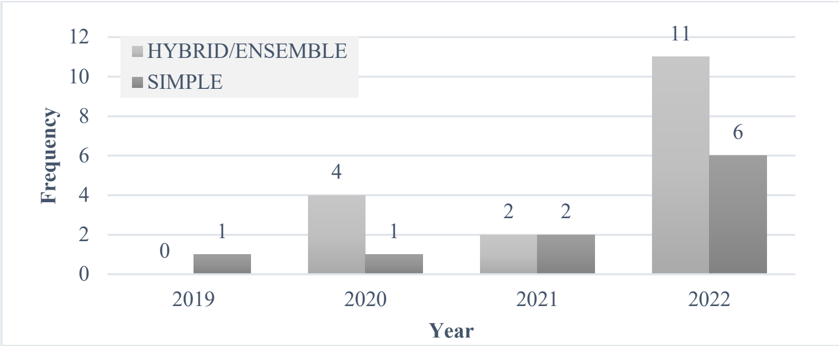

Figure 5. Year-wise frequency of hybrid/ensemble or simple models used

3.8. Finding the best model

The comparative analysis showed that recurrent DL architectures outperformed traditional ML methods in rainfall prediction. LSTM outperformed other recurrent network algorithms and could tackle complicated problems with substantial time lags [63]. Hybrid models, such as CNN-LSTM and ConvLSTM, enhance prediction by integrating temporal and spatial feature extraction, leading to robust predictions under varying conditions. Ensemble-based deep-learning frameworks that combine multiple architectures show improved generalization and reduced forecast uncertainty. These findings indicate that effective rainfall prediction models must capture temporal rainfall evolution and spatial climatic interactions. LSTM-based hybrid and ensemble models have emerged as strong candidates for operational forecasting owing to their stability and accuracy across meteorological contexts.

3.9. Finding the best algorithm

Table 3 shows algorithm choice significantly impacts rainfall prediction accuracy. Classical algorithms like KNN, SVM, DT, and logistic regression perform adequately with limited data but struggle with complex atmospheric relationships. Random forest outperforms individual classifiers like naïve Bayes, multilayer perceptron (MLP), sequential minimal optimization (SMO), DT, KNN, and SVM by modeling nonlinear interactions [64]. Logistic regression surpassed random forest with linear relationships [39], while XGBoost showed higher accuracy, demonstrating boosted ensemble advantages [42]. For time-dependent rainfall, LSTM and GRU outperform classical models by retaining long-term temporal information through memory mechanisms. No algorithm consistently outperforms others across conditions. Selection should be based on dataset characteristics: ensemble methods suit nonlinear data, while LSTM models excel in time-series forecasting.

4. CONCLUSION AND FUTURE WORK

This review highlights significant algorithmic deficiencies and proposes future research paths in multi-modal, hybrid, and real-time rainfall forecasting. Accurate rainfall prediction is critical for mitigating extreme weather impacts such as floods, landslides, and droughts. Studies have shown that model performance improves through the careful selection of meteorological parameters, network architecture, and learning strategies. While DL models, particularly LSTM-based and hybrid frameworks, show superior capability in capturing rainfall patterns, refinement is required for reliability under variable conditions. Future research should focus on hybrid and fusion-based ensemble techniques, multi-source datasets, real-time predictions, and integration of physical climate models with DL for robust rainfall forecasting systems.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Deepa Sharma	✓	✓		✓			✓		✓	✓	✓		✓	
Anand Kumar Shukla		✓		✓						✓		✓		
Punam Rattan	✓	✓		✓						✓	✓	✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] M. Babar, M. Rani, and I. Ali, "A deep learning-based rainfall prediction for flood management," in *2022 17th International Conference on Emerging Technologies, ICET 2022*, 2022, pp. 196–199, doi: 10.1109/ICET56601.2022.10004663.

[2] R. Tichavský, J. A. B.-Cánovas, K. Šilhán, R. Tolasz, and M. Stoffel, "Dry spells and extreme precipitation are the main trigger of landslides in Central Europe," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-019-51148-2.

[3] A. C. Mondini, F. Guzzetti, and M. Melillo, "Deep learning forecast of rainfall-induced shallow landslides," *Nature Communications*, vol. 14, no. 1, pp. 1–10, 2023, doi: 10.1038/s41467-023-38135-y.

[4] S. Poornima, M. Pushpalatha, R. B. Jana, and L. A. Patti, "Rainfall forecast and drought analysis for recent and forthcoming years in India," *Water*, vol. 15, no. 3, 2023, doi: 10.3390/w15030592.

[5] D. Moher, D. G. Altman, A. Liberati, and J. Tetzlaff, "PRISMA statement," *Epidemiology*, vol. 22, no. 1, 2011, doi: 10.1097/EDE.0b013e3181fe7825.

[6] A. G.-Lopez, I. C.-Paz, and M. M. Mandujano, "Algorithm to predict the rainfall starting point as a function of atmospheric pressure, humidity, and dewpoint," *Climate*, vol. 7, no. 11, 2019, doi: 10.3390/cli7110131.

[7] M. S. Pathan, J. Wu, Y. H. Lee, J. Yan, and S. Dev, "Analyzing the impact of meteorological parameters on rainfall prediction," in *2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), USNC-URSI 2021*, 2021, pp. 100–101, doi: 10.23919/USNC-URSI51813.2021.9703664.

[8] P. Zhang, W. Cao, and W. Li, "Surface and high-altitude combined rainfall forecasting using a convolutional neural network," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1765–1777, 2021, doi: 10.1007/s12083-020-00938-x.

[9] S. Poornima and M. Pushpalatha, "Prediction of rainfall using intensified LSTM-based recurrent neural network with weighted linear units," *Atmosphere*, vol. 10, no. 11, 2019, doi: 10.3390/atmos10110668.

[10] I. Salehin, I. M. Talha, M. M. Hasan, S. T. Dip, M. Saifuzzaman, and N. N. Moon, "An artificial intelligence based rainfall prediction using LSTM and neural network," in *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering, WIECON-ECE 2020*, 2020, pp. 5–8, doi: 10.1109/WIECON-ECE52138.2020.9398022.

[11] H. A.-Kader, M. A.-Elsalam, and M. Mohamed, "Hybrid machine learning model for rainfall forecasting," *Journal of Intelligent Systems and Internet of Things*, vol. 1, no. 1, pp. 5–12, 2020, doi: 10.54216/jisiot.010101.

[12] M. Chhetri, S. Kumar, P. P. Roy, and B. G. Kim, "Deep BLSTM-GRU model for monthly rainfall prediction: a case study of Simtokha, Bhutan," *Remote Sensing*, vol. 12, no. 19, pp. 1–13, 2020, doi: 10.3390/rs12193174.

[13] K. Johny, M. L. Pai, and S. Adarsh, "Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting," *Theoretical and Applied Climatology*, vol. 141, no. 1–2, pp. 1–17, 2020, doi: 10.1007/s00704-020-03177-5.

[14] M. I. Khan and R. Maity, "Hybrid deep learning approach for multi-step-ahead daily rainfall prediction using GCM simulations," *IEEE Access*, vol. 8, pp. 52774–52784, 2020, doi: 10.1109/ACCESS.2020.2980977.

[15] D. Z. Haq *et al.*, "Long short-term memory algorithm for rainfall prediction based on El-Nino and IOD data," *Procedia Computer Science*, vol. 179, pp. 829–837, 2021, doi: 10.1016/j.procs.2021.01.071.

[16] D. Sun, J. Wu, H. Huang, R. Wang, F. Liang, and H. Xinhua, "Prediction of short-time rainfall based on deep learning," *Mathematical Problems in Engineering*, vol. 2021, 2021, doi: 10.1155/2021/6664413.

[17] P. Kanchan, "Rainfall analysis and forecasting using deep learning technique," *Journal of Informatics Electrical and Electronics Engineering*, vol. 2, no. 2, pp. 1–11, 2021, doi: 10.54060/jieee/002.02.015.

[18] N. Thakur and S. Karmakar, "Deep learning approach using long short term memory technique for monthly rainfall prediction in Chhattisgarh, India," *International Journal of Scientific Research in Computer Science and Engineering*, vol. 9, no. 1, pp. 8–13, 2021, doi: 10.26438/ijsrcse/v9i1.813.

[19] U. Bhimavarapu, "IRF-LSTM: enhanced regularization function in LSTM to predict the rainfall," *Neural Computing and Applications*, vol. 34, no. 22, pp. 20165–20177, 2022, doi: 10.1007/s00521-022-07577-8.

[20] Y. Peng, D. Gong, C. Deng, H. Li, H. Cai, and H. Zhang, "An automatic hyperparameter optimization DNN model for precipitation prediction," *Applied Intelligence*, vol. 52, no. 3, pp. 2703–2719, 2022, doi: 10.1007/s10489-021-02507-y.

[21] N. Thakur, S. Karmakar, and S. Soni, "Time series forecasting for uni-variant data using hybrid GA-OLSTM model and performance evaluations," *International Journal of Information Technology*, vol. 14, no. 4, pp. 1961–1966, 2022, doi: 10.1007/s41870-022-00914-z.

[22] X. Zhang, K. Wang, and Z. Zheng, "A novel integrated learning model for rainfall prediction CEEMD-FCMSE -stacking," *Earth Science Informatics*, vol. 15, no. 3, pp. 1995–2005, 2022, doi: 10.1007/s12145-022-00819-2.

[23] G. S. Jeba, P. Chitra, and U. M. Rajasekaran, "Time-series analysis and flood prediction using a deep learning approach," in *2022 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2022*, 2022, pp. 139–142, doi: 10.1109/WiSPNET54241.2022.9767102.

[24] D. Endalie, G. Haile, and W. Taye, "Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia," *Water Supply*, vol. 22, no. 3, pp. 3448–3461, 2022, doi: 10.2166/WS.2021.391.

[25] X. Liu, X. Sang, J. Chang, Y. Zheng, and Y. Han, "Rainfall prediction optimization model in ten-day time step based on sliding window mechanism and zero-sum game," *Aqua Water Infrastructure, Ecosystems and Society*, vol. 71, no. 1, pp. 1–18, 2022, doi: 10.2166/aqua.2021.086.

[26] J. Gu, S. Liu, Z. Zhou, S. R. Chalov, and Q. Zhuang, "A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China," *Water*, vol. 14, no. 3, 2022, doi: 10.3390/w14030492.

[27] V. C. Mouli, P. Chitra, M. H. Subramanian, and S. Abirami, "A deep learning ensemble model for short-term rainfall prediction," in *2022 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2022*, 2022, pp. 135–138, doi: 10.1109/WiSPNET54241.2022.9767163.

[28] C. Chen *et al.*, "Forecast of rainfall distribution based on fixed sliding window long short-term memory," *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 248–261, 2022, doi: 10.1080/19942060.2021.2009374.

[29] S. Chen, X. Xu, Y. Zhang, D. Shao, S. Zhang, and M. Zeng, "Two-stream convolutional LSTM for precipitation nowcasting," *Neural Computing and Applications*, vol. 34, no. 16, pp. 13281–13290, 2022, doi: 10.1007/s00521-021-06877-9.

[30] M. Billah, M. N. Adnan, M. R. Akhond, R. R. Ema, M. A. Hossain, and S. Md. Galib, "Rainfall prediction system for Bangladesh using long short-term memory," *Open Computer Science*, vol. 12, no. 1, pp. 323–331, 2022, doi: 10.1515/comp-2022-0254.

[31] B. C. Majang, N. Zaini, and L. Mazalan, "Rainfall nowcasting based on satellite images using convolutional long-short term memory," in *12th International Conference on System Engineering and Technology, ICSET 2022*, 2022, pp. 67–71, doi: 10.1109/ICSET57543.2022.10010806.

[32] Shweta, H. Sehrawat, and V. Siwach, "Monsoonal rainfall forecasting using LSTM neural network," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2022*, 2022, pp. 1–5, doi: 10.1109/ICRITO56286.2022.9964816.

[33] S. Guo, Y. Wen, X. Zhang, G. Zhu, and J. Huang, "Research on precipitation prediction based on a complete ensemble empirical mode decomposition with adaptive noise–long short-term memory coupled model," *Water Supply*, vol. 22, no. 12, pp. 9061–9062, 2022, doi: 10.2166/ws.2022.412.

[34] G. Folino, M. Guarascio, and F. Chiaravalloti, "Learning ensembles of deep neural networks for extreme rainfall event detection," *Neural Computing and Applications*, vol. 35, no. 14, pp. 10347–10360, 2023, doi: 10.1007/s00521-023-08238-0.

[35] E. Kurniawan, F. Nhita, A. Aditsania, and D. Saepudin, "C5.0 algorithm and synthetic minority oversampling technique (SMOTE) for rainfall forecasting in bandung regency," *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*. 2019, doi: 10.1109/ICoICT.2019.8835324.

[36] A. D. Mehr, V. Nourani, V. K. Khosrowshahi, and M. A. Ghorbani, "A hybrid support vector regression–firefly model for monthly rainfall forecasting," *International Journal of Environmental Science and Technology*, vol. 16, no. 1, pp. 335–346, 2019, doi: 10.1007/s13762-018-1674-2.

[37] P. C. S. Reddy and A. Sureshbabu, "An adaptive model for forecasting seasonal rainfall using predictive analytics," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 5, pp. 22–32, 2019, doi: 10.22266/ijies2019.1031.03.

[38] M. T. Anwar, S. Nugrohadi, V. Tantriyati, and V. A. Windarni, "Rain prediction using rule-based machine learning approach," *Advance Sustainable Science, Engineering and Technology*, vol. 2, no. 1, pp. 1–6, 2020, doi: 10.26877/asset.v2i1.6019.

[39] S. M. Gowtham, Y. S. Ganesh, and M. M. Ali, "Efficient rainfall prediction and analysis using machine learning techniques," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 6, pp. 3467–3474, 2021.

[40] A. U. Azmi, A. F. Hadi, D. Anggraeni, and A. Riski, "Naive Bayes methods for rainfall prediction classification in Banyuwangi," in *Journal of Physics: Conference Series*, 2021, vol. 1872, no. 1, doi: 10.1088/1742-6596/1872/1/012028.

[41] M. T. Anwar, E. Winarno, W. Hadikurniawati, and M. Novita, "Rainfall prediction using extreme gradient boosting," in *Journal of Physics: Conference Series*, 2021, vol. 1869, no. 1, doi: 10.1088/1742-6596/1869/1/012078.

[42] C. M. Liyew and H. A. Melese, "Machine learning techniques to predict daily rainfall amount," *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00545-4.

[43] A. Mahadware, A. Saigiridhari, A. Mishra, A. Tupe, and N. Marathe, "Rainfall prediction using different machine learning and deep learning algorithms," in *2022 2nd Asian Conference on Innovation in Technology, ASIANCON 2022*, 2022, pp. 1–8. doi: 10.1109/ASIANCON55314.2022.9908857.

[44] M. Raval, P. Sivashanmugam, V. Pham, H. Gohel, A. Kaushik, and Y. Wan, "Automated predictive analytics tool for rainfall forecasting," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-021-95735-8.

[45] K. Raj and T. D. Amin, "Rainfall prediction using deep learning and machine learning techniques," *Research Square*, May 11, 2023, doi: 10.21203/rs.3.rs-2790967/v1.

[46] A. M. S. Vigil, R. Shwetha, S. Singh, and R. Shruthi, "Evaluating machine learning models to enhance rainfall prediction," *Revista Electronica de Veterinaria*, vol. 25, no. 1, pp. 646–652, 2024, doi: 10.69980/redvet.v25i1s.805.

[47] L. C. P. Velasco, R. P. Serquiña, M. S. A. A. Zamad, B. F. Juanico, and J. C. Lomocso, "Week-ahead rainfall forecasting using multilayer perceptron neural network," *Procedia Computer Science*, vol. 161, pp. 386–397, 2019, doi: 10.1016/j.procs.2019.11.137.

[48] M. H. Yen, D. W. Liu, Y. C. Hsin, C. E. Lin, and C. C. Chen, "Application of the deep learning for the prediction of rainfall in Southern Taiwan," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019, doi: 10.1038/s41598-019-49242-6.

[49] B. T. Pham *et al.*, "Development of advanced artificial intelligence models for daily rainfall prediction," *Atmospheric Research*, vol. 237, 2020, doi: 10.1016/j.atmosres.2020.104845.

[50] A. Samad, Bhagyanidhi, V. Gautam, P. Jain, Sangeeta, and K. Sarkar, "An approach for rainfall prediction using long short term memory neural network," in *2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020*, 2020, pp. 190–195, doi: 10.1109/ICCCA49541.2020.9250809.

[51] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, "Rainfall forecasting model using machine learning methods: case study Terengganu, Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1651–1663, 2021, doi: 10.1016/j.asej.2020.09.011.

[52] M. Lakshmitha, G. Anisha, S. Rajesh, and B. Venkatesan, "Rainfall prediction using deep learning technique," *International Research Journal of Engineering and Technology*, vol. 8, no. 10, pp. 1179–1182, 2021.

[53] J. A. Anochi, V. A. D. Almeida, and H. F. D. C. Velho, "Machine learning for climate precipitation prediction modeling over South America," *Remote Sensing*, vol. 13, no. 13, pp. 1–18, 2021, doi: 10.3390/rs13132468.

[54] J. S. A. N. W. Premachandra and P. P. N. V. Kumara, “A novel approach for weather prediction for agriculture in Sri Lanka using machine learning techniques,” in *International Research Conference on Smart Computing and Systems Engineering, SCSE 2021*, 2021, pp. 182–189, doi: 10.1109/SCSE53661.2021.9568319.

[55] K. Kaliyaperumal, A. Rahim, D. K. Sharma, R. Regin, S. Vashisht, and K. Phasinam, “Rainfall prediction using deep mining strategy for detection,” in *2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, 2021, pp. 1623–1628, doi: 10.1109/ICOSEC51865.2021.9591879.

[56] D. Hemavathi, V. R. Kumar, R. Regin, S. S. Rajest, K. Phasinam, and S. Singh, “Technical support for detection and prediction of rainfall,” in *2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, 2021, pp. 1629–1634, doi: 10.1109/ICOSEC51865.2021.9591762.

[57] S. Biruntha, B. S. Sowmiya, R. Subashri, and M. Vasanth, “Rainfall prediction using KNN and decision tree,” in *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, 2022, pp. 1757–1763, doi: 10.1109/ICEARS53579.2022.9752220.

[58] A. Y. B.-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, “Rainfall prediction: a comparative analysis of modern machine learning algorithms for time-series forecasting,” *Machine Learning with Applications*, vol. 7, 2022, doi: 10.1016/j.mlwa.2021.100204.

[59] Y. Yang, H. Lv, and N. Chen, “A survey on ensemble learning under the era of deep learning,” *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5545–5589, 2023, doi: 10.1007/s10462-022-10283-5.

[60] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: a review,” *Engineering Applications of Artificial Intelligence*, vol. 115, 2022, doi: 10.1016/j.engappai.2022.105151.

[61] S. Kanani, S. Patel, R. K. Gupta, A. Jain, and J. C. W. Lin, “An AI-enabled ensemble method for rainfall forecasting using long-short term memory,” *Mathematical Biosciences and Engineering*, vol. 20, no. 5, pp. 8975–9002, 2023, doi: 10.3934/mbe.2023394.

[62] M. Putra, M. S. Rosid, and D. Handoko, “High-resolution rainfall estimation using ensemble learning techniques and multisensor data integration,” *Sensors*, vol. 24, no. 15, 2024, doi: 10.3390/s24155030.

[63] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[64] U. Shah, S. Garg, N. Sisodiya, N. Dube, and S. Sharma, “Rainfall prediction: accuracy enhancement using machine learning and forecasting techniques,” in *PDGC 2018 - 2018 5th International Conference on Parallel, Distributed and Grid Computing*, 2018, pp. 776–782, doi: 10.1109/PDGC.2018.8745763.

BIOGRAPHIES OF AUTHORS

Deepa Sharma is a Ph.D. scholar at the School of Computer Applications, Lovely Professional University, Phagwara, India. She has a master's degree in Computer Applications (MCA) from the University of Jammu. She has cleared UGC NET (2012) and GATE (2010) exams in computer applications. Her research areas are machine learning and deep learning, especially focusing on prediction models. She can be contacted at email: apdeepasharma@gmail.com.

Prof. Dr. Anand Kumar Shukla has 22+ years of experience working as a professor and dean at the School of Computer Applications, Lovely Professional University, Phagwara, India. He has a Ph.D. in Computer Science and Engineering, along with M.Tech. (Computer Science and Engineering) and MCA. He is a member of IEEE, ACM, and CSTA (New York). He is an author of more than 100 international publications and 7 books. He can be contacted at email: anand.shukla@lpu.co.in or anandshukla0203@gmail.com.

Dr. Punam Rattan is a professor at Department of Computer Applications, Manav Rachna University, Faridabad, India, where she specialises in computer applications. She has a Ph.D. in Computer Applications from IKG Punjab Technical University, as well as dual master's degrees in Business Administration (MBA) and Computer Applications (MCA). She is a seasoned academician with 20 years of experience. She is a prolific author, having published more than 30 research papers and three books in prestigious journals and conferences, including IEEE Xplore, Scopus, and SCI-indexed publications. Her research is concentrated on the disciplines of data mining and machine learning, and she has made substantial contributions to these fields. She can be contacted at email: punamrattan@gmail.com.