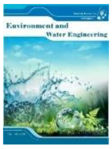
Environ. Water Eng., 2026, 12(1), 95-107

EWE

Environment and Water Engineering

Homepage: www.jewe.ir



ISSN: 2476-3683

Impact of features in rainfall prediction with explainable artificial intelligence - SHAP

Mehul Patel[1*] and Ankit Shah[2]

[1]Ph.D. Scholar, Department of Information Technology, Sankalchand Patel University, Visnagar, 384315, Gujarat, India.
[2]Professor (Dr.), Department of Big Data Analytics, Adani Institute of Digital Technology Management, Gandhinagar, 382423, Gujarat, India.

ARTICLE INFO

Paper Type: Research Paper

Received: 09 January 2026
Revised: 24 February 2026
Accepted: 10 April 2026
Published: 18 May 2026

Keywords
Rainfall,
Explainable Artificial Intelligence,
XAI,
Artificial Neural Network,
Remote Sensing,
SHAP

***Corresponding author:**
Mehul Patel
mspatelit_spce@spu.ac.in

ABSTRACT

Predicting when rain will fall is essential for managing water resources and mitigating hydrological disasters. This work aimed to identify the impact of feature variables by using an explainable deep learning strategy to analyze causes of heavy rainfall. The research developed an Artificial Neural Network (ANN) and SHAP-based model to measure how meteorological features contribute to heavy rainfall. This model improves interpretability and system transparency for identifying dependent variables. Unlike black-box models, the proposed framework shows how each meteorological feature affects model output, enabling transparent identification of the causes of heavy rainfall. Model training used hourly rainfall from eight densely populated Indian cities, making the model robust across different climates. Findings show that meteorological features strongly linked to SHAP value trends typically explain the model's predictions. This framework enhances rainfall forecasting and provides a generalizable, explainable method for hydrological modeling and decision support. The study's originality lies in creating an integrated ANN-SHAP framework that produces hourly rainfall estimates and systematically quantifies the influence of meteorological variables across urban regions.

How to ctie this paper: Patel. M., & Shah. A. (2026). Impact of features in rainfall prediction with explainable artificial intelligence - SHAP. *Environ. Water Eng*., *12*(1), 95-107.

1. Introduction

a. Background

A challenge in metropolitan areas is predicting short-term precipitation events accurately. Heavy precipitation in densely settled areas can lead to rapid flooding and structural failure, resulting in significant economic losses. Rainfall in India and other parts of the world with tropical/monsoon climates typically exhibits significant nonlinearity and relatively high spatiotemporal variability. Both of those characteristics are caused by multiple factors, including interactions among thermodynamic, dynamic, and surface-level meteorological processes, which vary in intensity and/or interactions depending on local and regional processes.

Numerical weather models based on physical laws may reliably forecast precipitation for all event sizes. However, at urban and sub-regional levels, their accuracy drops due to coarse spatial resolution and sensitivity to initial conditions. Therefore, data-driven approaches are valuable for predicting rainfall in these areas, mainly because they use high-resolution observational data.

The application of recent developments in machine learning and deep learning has yielded improved predictive ability for rainfall and other hydro-meteorological attributes by uncovering non-linear relationships from previous datasets (Schultz et al., 2021; Abdellaoui & Mehrkanoon, 2020; He et al., 2024). Rainfall prediction has been applied using artificial neural networks, recurrent neural networks, Long Short-Term Memory (LSTM) networks, and convolutional networks across multiple time scales, including hourly and seasonal (Schultz et al. 2021; Abdellaoui & Mehrkanoon 2020; Arcodia et al. 2023; He et al. 2024). In general, these models outperform traditional statistical methods, as evidenced by their ability to capture multivariate dependencies among temperature, humidity, barometric pressure, wind speed and direction, and cloud variables (Arcodia et al., 2023).

However, their growing complexity has also imposed a fundamental limitation, as nearly all deep-learning models operate as black boxes, providing little or no indication of how each independent variable influences the dependent variable(s) (Dikshit & Pradhan, 2021). The lack of interpretability poses a major barrier to the operational use of machine-learning-based rainfall prediction models, especially for high-risk applications such as flood prediction and urban water management. Without a transparent explanation of how algorithms arrived at their predictions, it is nearly impossible to judge whether the outcomes they produce are physically plausible, valid across geographical boundaries, or consistent under changing climate regimes (Dikshit & Pradhan, 2021). Developing explainable Artificial Intelligence (XAI) to understand complex machine-learning model decision-making has become a fast-growing field in response to this challenge (Başağaoğlu et al., 2022; Sahakyan et al., 2021; Chandra et al., 2023; Zounemat-Kermani & Kheimi, 2026; Robles et al., 2026).

In hydro-meteorological and rainfall-related studies, deep learning-based predictions are increasingly being interpreted using XAI techniques and compared with well-established physical processes to validate model outputs. The most utilized post-hoc explanation methods include LIME, Integrated Gradients, and Shapley Additive exPlanations (SHAP). Of these methods, SHAP has gained widespread use because of its solid theoretical underpinnings relating to cooperative game theory, as well as its ability to identify both local and global explanations for a model's output while satisfying consistency and additivity, respectively (Jesus et al. 2021; Lundberg & Lee, 2017; Retsch et al. 2022). Recent research has demonstrated the applicability of SHAP in predicting precipitation, predicting runoff, modeling flooding, and downscaling climate (Senocak et al. 2023; Wu et al. 2023; Rampal et al. 2022; Lamane et al., 2025; Mosavi et al., 2018), and has provided a means to identify the most impactful meteorological factors driving model output and the influence of each factor on the output directionality.

b. Limitations

Despite recent advances, the current literature has several limitations.

i) Limited geographic scope: Many XAI-based rainfall studies focus on one region or basin, so they lack spatial generality (Dikshit & Pradhan, 2021).

ii) Insufficient analysis of attribution stability: Most studies report only feature importance. They rarely examine stability and directionality across cities with similar climates (Başağaoğlu et al., 2022).

iii) Model-specific gaps: Few models use explainability for ANN rainfall forecasts, especially with high-resolution hourly urban data (Bialek et al. 2022; Schultz et al., 2021; Dikshit & Pradhan, 2021).

These deficiencies limit the real-world applicability of XAI for comparative urban rainfall assessment and decision-making support. There is also a significant need for interpretable, data-driven rainfall forecasts to help manage the impacts of growing urbanization and increasing rainfall extremes in India (Samanta et al., 2026; Shukla et al., 2026; Kemmannu et al., 2025). For instance, studies using deep learning (Schultz et al., 2021; Abdellaoui & Mehrkanoon, 2020) have produced usable rainfall databases—both hourly and seasonal. However, these are primarily black-box systems. Their predictive ability cannot be confidently assessed relative to the overall prediction. There is also no established understanding of which meteorological variables contribute positively or negatively to outcomes across different urban locations. While machine learning techniques are suitable for predicting rainfall and hydro-meteorological variables in India, most previous studies have focused on accuracy rather than interpretability. Very few provide details on how meteorological data contribute to rainfall outcomes in different urban contexts (Schultz et al., 2021; Abdellaoui & Mehrkanoon, 2020; Dikshit & Pradhan, 2021; Mohapatra et al., 2021; Priya Ashok & Pekkat, 2024; Materia et al., 2024). Additionally, this study establishes a valid baseline model for data-driven rainfall forecasts in India using meteorological variables. It captures multivariate dependencies among these predictors (Chamundeswari et al., 2026; Basha & Begum, 2025; Chittaragi et al., 2025).

c. Contribution

This research integrates an ANN weather forecasting system with a SHAP interpretability model. It provides a systematic method for quantifying the importance of different meteorological features for hourly rainfall forecasts in major Indian cities. Interpretability is a primary output, as it provides a global feature attribution method to evaluate the importance, direction, and consistency of meteorological parameters across cities. The goal is to quantify and understand the impact of meteorological features on ANN-based rainfall prediction using SHAP, and to analyze how these influences vary across cities, creating a more transparent, interpretable, and usable forecasting system.

2. Materials and Methods

a) Research methodology

The methodology utilized in this study combines data-driven and explainability-based techniques to develop a robust nonlinear rainfall prediction model for several urban sites and to quantitatively evaluate the contribution of each meteorological variable to the model predictions using a theoretically sound XAI framework (Başağaoğlu et al., 2022; Zounemat-Kermani & Kheimi, 2026). This approach also provides a transparent, reproducible, and methodologically rigorous basis for analysis rather than only focusing on predictive performance (Dikshit & Pradhan, 2021; Robles et al., 2026).

b) Data collection and preprocessing

The data set consists of hourly observations of meteorological parameters and rainfall across eight major Indian cities (i.e., Bengaluru, Mumbai, Delhi, Hyderabad, Jaipur, Kanpur, Nagpur, and Pune) from January 2009 to January 2020 (Soneji 2020). These cities were selected because they represent a wide range of climatic and geographic conditions and exhibit similar levels of urban density and rainfall data availability.

The dataset contains one target variable (i.e., rainfall) and twenty-two predictor variables associated with the meteorology (including atmospheric, thermodynamic, and surface variables), which include measures related to temperature, humidity, atmospheric pressure, wind characteristics, cloud cover, visibility, and astronomy-related variables (Mohapatra et al., 2021; Priya Ashok & Pekkat, 2024).

The data preprocessing step was performed to ensure the data was consistent and ready for machine learning analysis. Initially, the number of records with missing values for each variable was identified, and records with excessive missing values (based on a predefined threshold) were excluded to avoid contaminating the model training data. For all remaining missing values, numerical meteorological variables were imputed using time-aware interpolation methods when appropriate (exampl e.g., linear interpolation for short gaps). For isolated or non-sequential gaps, missing values were replaced with the city's mean. This ensured temporal continuity and minimized the artificial variance introduced during this step. Outliers and anomalous values are filtered using the Interquartile Range (IQR) and domain-specific thresholds, such as no negative rainfall values and pressure values that fall outside a reasonable range for typical atmospheric pressure. Anomalies found were either corrected if they could be identified as clear recording errors or eliminated if they could not be verified. All continuous predictor variables were normalized using min-max scaling to the range [0, 1]. The purpose of normalizing continuous predictor variables is to improve numerical stability during Artificial Neural Network (ANN) training and to prevent scale dominance among a variety of meteorological variables (He et al., 2024; Abdellaoui & Mehrkanoon, 2020). Normalization parameters were calculated solely from the training dataset and subsequently applied to the test dataset to prevent data leakage. To maintain short-term rainfall dynamics, hourly-resolution temporal features were retained. In the correlation-based feature filtering method, no predictors were pre-removed, allowing attribute analysis to quantify both strong and weaker contributors without biasing the attribution assessment.

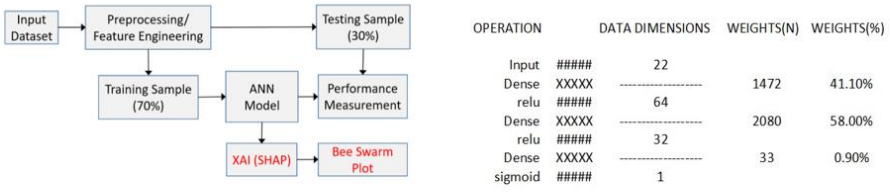
To use conventional methodology for rain/flood surveys and the best approach for robustification of generalization assessment for model development; 70% of the data were assigned to train; 30% of the data were assigned to verify results. Five-fold cross-validation was applied to the training set to improve generalization volatility test during model construction (Schultz et al., 2021; Chamundeswari et al., 2026).

c) Rainfall prediction model

A particular neural network will be used to predict rainfall because it can provide a good approximation of complex and non-linear relationships among multiple variables (e.g., temperature, humidity, pressure, etc.) related to the occurrence or amount of rainfall (Schultz et al., 2021; Abdellaoui & Mehrkanoon, 2020). While there are certain recurrent architectures, such as Long Short-Term Memory (LSTM) networks, which have been specifically designed to capture long-term time dependencies, and there are Convolutional Neural Networks (CNN) which are very good at extracting spatial features (Abdellaoui & Mehrkanoon, 2020); the current research study is using hourly tabular format meteorological variables instead of structured spatial fields and/or long sequence (i.e., temporal) memory modelling. The input variables in question contain current atmospheric conditions at the same time as the rainfall event; therefore, as seen from the preliminary analyses, there is little to no benefit in using longer than hourly resolution time series data to model extended temporal sequences of meteorological variables that might have an impact on rainfall development/occurrence. Compared with the use of LSTM-based models, feed-forward ANN's represent a much less complex computation and have a far lower likelihood of being over-fitted when the data set size and temporal depth are at a moderate level (Bialek et al., 2022; Chittaragi et al., 2025). Furthermore, CNN architectures tend to perform better on gridded data (e.g., radar/satellite imagery), whereas the data used in this research will be meteorological data collected from different geographical locations via a network of monitoring stations (He et al., 2024). Although the primary goal of this research is to use SHAP to create explainable predictions, ANN architectures also provide an easily interpretable baseline for features and their contributions to rainfall prediction and thus do not introduce additional structural dependencies that could destabilize attributions of rainfall-producing features from ANNs (Tahmasbi et al., 2025). The study used an Artificial Neural Network (ANN) as the primary nonlinear modeling approach to isolate and compare the effects of meteorological variables on precipitation across multiple locations, while providing computational efficiency and cross-city comparability of results (Bialek et al., 2022; Basha & Begum, 2025).

While LSTM can capture long-term temporal dependencies, the selected ANN will likely achieve performance on an hourly tabular dataset equivalent to that of LSTM models, but with lower model complexity and more stable attribution of features to the associated output variables. While CNN architectures perform well on spatial data such as radar and satellite imagery, CNNs' advantage over ANNs for structured meteorological inputs is not significant, whereas ANNs have a greater ability to be used with SHAP for feature contribution analysis. Thus, an ANN was selected to achieve a deliberate trade-off between its capacity for nonlinear modeling, architectural simplicity, interpretability (via SHAP), and applicability to structured hourly meteorological data.

Figure 1 shows a visual representation of the ANN architecture, consisting of three layers: an input layer with 22 nodes representing predictor variables; two fully connected hidden layers with 64 and 32 nodes, respectively; and an output layer with one node (the predicted hourly rainfall). The hidden layer used the rectified linear unit (ReLU) activation function to improve the representation of nonlinear relationships while reducing the potential for vanishing gradients. The model was trained with the Adam optimizer (learning rate = 0.001) and the mean squared error as the loss function. The hyperparameters (number of hidden layers, number of nodes, batch size, and number of training epochs) were determined empirically based on cross-validation results.



(a) XAI with ANN (b) ANN Structure

Fig 1. Model Architecture

In the final configuration of the ANN, the number of training epochs was 100, and the batch size was 32, providing adequate performance across stability of convergence, computational efficiency, and overfitting prevention. Model performance was evaluated on an independent test set using a range of complementary evaluation metrics (e.g., mean squared error, root mean squared error, mean absolute error, coefficient of determination).

d) Explainability framework with SHAP

The ANN model produced predictions that were black box. To provide clarity, the trained ANN model was integrated with the SHAP (Shapley Additive exPlanations) framework, an alternative to traditional model-interpretation methods. SHAP was chosen for its robust theoretical foundation in cooperative game theory and its ability to compute feature attributions that are both consistent and additive for complicated non-linear models (Lundberg & Lee, 2017; Jesus et al., 2021; Retsch et al., 2022).

In this study, SHAP values were computed globally to quantify the individual contribution of each meteorological variable to rainfall prediction across all observations in the training data set. It was more useful for the authors to understand, globally, the long-term (physically meaningful) causes of rainfall than to focus on many individual cause-and-effect relationships for any one time period or location. The authors used representative sampling to compute SHAP values with the large amount of collected data, thereby conserving computing resources while preserving true feature attributions (Senocak et al., 2023; Wu et al., 2023).

SHAP values were evaluated using summary and bee swarm graphics to assess feature importance, directionality (positive or negative impact), and variability across the samples in the SHAP graphics. By investigating both the size and directionality of SHAP attributions, the approach provides insight into how specific meteorological variables may contribute to changes in predicted rainfall by evaluating their effects on increases and decreases (Rampal et al., 2022; Lamane et al., 2025; Mosavi et al., 2018).

e) Cross-city comparison of findings

A significant contribution of this research is the provision of explanations that allow meaningful comparison across multiple urban areas. This study does not aggregate all cities into a single combined model when interpreting SHAP values, but instead examines results from each city to determine how consistently or inconsistently the effects of each feature differ across jurisdictions with different climates. Thus, one can assess whether the dominant predictors are universal or location-dependent, thereby improving the operational relevance of the findings (Dikshit & Pradhan, 2021; Başağaoğlu et al., 2022; Samanta et al., 2026).

f) Novelty and ability to reproduce the methodology

While the individual use of ANN and SHAP are not new discoveries, the combined use of ANN and SHAP in a systematic way to analyze rainfall within urban areas on an hourly basis while focusing on global feature-attributions and consistency among cities is novel (Zounemat-Kermani & Kheimi, 2026; Materia et al., 2024). In addition, the framework supports validating data-driven models by enabling interpretation aligned with physical knowledge, provided that there are no overly restrictive feature selection constraints in the data set.

All modeling steps, evaluation metrics, and explainability methods were carried out using reproducible workflows, with standardized performance metrics allowing comparison with previous rainfall prediction work (Robles et al., 2026; Schultz et al., 2021).

3. Results

In this section of the report, the predictive performance of the proposed ANN model and the SHAP-based explainability results are discussed. The presentation of results focuses on comparative validation and analytical interpretation rather than descriptive reporting. Model performance is evaluated in terms of generalization ability and the consistency of feature influence across eight major Indian cities.

a) Predictive performance between cities

Predictive performance was consistent across all eight cities when the ANN model was assessed using several error and goodness-of-fit metrics on an independent dataset (MSE, RMSE, MAE, and R^2). Across all cities, RMSE values were generally low to moderate, suggesting reasonable control of error in estimating hourly rainfall amounts. MAE values were, on average, quite small, suggesting robustness against extreme outliers (see Table 1). The R^2 values showed moderate variability, reflecting the intrinsic temporal variability of hourly rainfall and the lack of large-scale atmospheric predictors; such variable R^2 values had previously been observed in fine temporal-resolution rainfall modeling work using surface meteorological data.

Table 1. Performance metrics

City	ACC.	MSE	RMSE	R2	MAE
Bengaluru	0.8652	0.1286	0.3586	0.2405	0.0765
Bombay	0.79	0.2716	0.5211	0.2296	0.1715
Delhi	0.944	0.0686	0.2619	0.2221	0.0541
Hyderabad	0.9023	0.0722	0.2686	0.3094	0.0619
Jaipur	0.9493	0.0392	0.1979	0.3563	0.0272
Kanpur	0.9361	0.0638	0.2525	0.233	0.0455
Nagpur	0.8658	0.2122	0.4606	0.1518	0.1095
Pune	0.8753	0.1069	0.327	0.2353	0.0776

Different R^2 values were recorded across the metropolitan cities. Jaipur and Hyderabad demonstrated moderate explanatory power, with lower R^2 values than those recorded for Nagpur and Mumbai. One factor that may explain this disparity in R^2 values is that Mumbai's coastal monsoon climate is characterized largely by sudden, short-term rainfall events driven by moisture surges from the Arabian Sea and mesoscale convective systems. Sudden bursts of precipitation result in significant hourly-scale variance and intermittency, making it difficult to accurately model precipitation. In addition, substantial skewness in rainfall distribution at its peak during the monsoon can distort error metrics such as RMSE, which are sensitive to large deviations. Nagpur, an inland city with heavy convective summer rainfall driven by land-atmosphere interaction and thermal instability, may also show lower R^2 values because surface-level meteorological variables alone cannot fully explain the localized nature of thunderstorm rainfall. Conversely, because upper air circulation variables were likely absent in this analysis, model R^2 values were lower than MAE values. Rainfall data often have a high proportion of zero or near-zero values due to the discontinuous nature of hourly precipitation data, leading to extreme class imbalance in regression.

In cities with more erratic precipitation patterns, predictable performance tends to be lower than in municipalities with less erratic precipitation and more consistent available moisture over time. Performance variability indicates that model limitations are not the only contributing factor; complexities in the climate and the characteristics of the underlying data distribution also influence variability. The variability underscores the inherent difficulty of conducting fine-scale rainfall forecasts and suggests that when assessing model evaluation metrics across heterogeneous urban centers, contextual climate interpretations are important. The variability also reflects the challenges of modeling highly intermittent precipitation events at hourly time scales, consistent with conclusions from recent rainfall and hydrological/climatic studies using deep learning methods (Senocak et al. 2023; He et al. 2024). Importantly, the performance levels achieved by the proposed modeling frameworks are comparable to or better than those of ANN and LSTM rainfall models at comparable temporal resolutions, providing evidence for their validity. Messy models aren't new and have been well-described in the literature. Modeling results from linear regression, ARIMA, or basic decision tree models yield higher errors and lower generalizability for nonlinear rainfall processes, especially at hourly intervals. Overall, the ANN is well-suited as a nonlinear predictive model and is therefore a good choice for future explainability analyses, given its superior performance compared to traditional approaches.

a) Global feature attribution using SHAP

Feature attribution distributions for the test samples are presented in bee-swarm plots (Fig 2A–H), illustrating the total contribution and overall direction of contribution for each feature across the cities sampled. The bee-swarm plots show the distribution of feature attributions across all test samples, as well as the overall magnitude and direction of each feature attribution for each of the 8 cities. In these bee-swarm plots, each feature is ranked vertically based on the mean absolute SHAP value to illustrate each feature's overall contribution to the model's predictions. The mean absolute SHAP value is presented on the horizontal axis such that positive SHAP values (i.e., higher amount of predicted rainfall) is represented as an increase in the value of the predicted amount of rainfall, while the negative SHAP values (i.e., lower predicted rainfall) are represented as a decrease in predicted rainfall.

In addition, a color gradient from blue to red is used for each feature, representing feature intensity and allowing the reader to simultaneously visualize both the feature's intensity and the direction of its contribution to the model's predicted rainfall amounts (Mansoor et al., 2025). The SHAP summary values show that only a sparser number of dominant variables drive the prediction of total rainfall, while the remaining number of predictors contribute relatively small contributions to the predicted rainfall amounts. This finding aligns with other XAI-based hydroclimatic studies and supports the physical plausibility of these derived relationships (Başağaoğlu et al., 2022).

All cities supported the idea that variables related to humidity, cloud cover, atmospheric pressure, and apparent temperature are consistently the most important positive predictors of rainfall. Meanwhile, variables like wind speed, ambient temperature, visibility, and wind chill were the least important negative contributors. These results are consistent with established meteorological understanding and support the physical understanding of ANN models, despite their inherent nonlinearity.

c) City level comparative explainability

City-level comparative explainability (SHAP) was also evaluated by comparing SHAP attributions for cities throughout the study, as illustrated in Figure 2 (a-h).

Figure 2(a) shows the SHAP summary plot of the global impact of each feature on the prediction of hourly rainfall in Bengaluru. Among the features used to predict rainfall, humidity has the greatest impact, followed closely by wind speed, wind direction, and wind gust speed. The strong predictive power of humidity suggests that moisture is a crucial component of rainfall and that higher humidity is associated with higher rainfall predictions. Wind-related variables also strongly influence rainfall predictions, suggesting that airflow patterns and atmospheric moisture transport are key drivers of precipitation. Wind gust speed has both positive and negative effects on rainfall predictions, suggesting that rain events are highly variable and that conditions in the surrounding atmosphere fluctuate (e.g., from one moment to the next) within a small area.

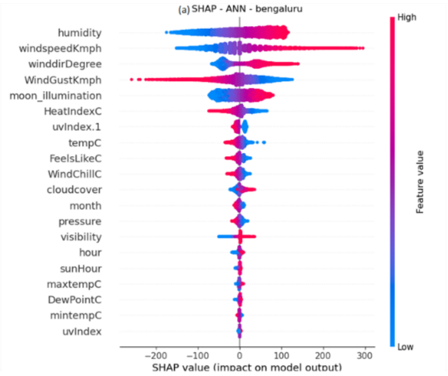


Fig. 2 (a) Feature importance shown by Bee Swarm Plot for Bengaluru

Of the next two important variables, moon illumination and heat index both have a moderate impact on rainfall predictions. This is most likely due to their ability to influence the atmospheric energy balance and create favorable conditions for precipitation. The remaining variables, such as temperature, cloud cover, and pressure, have the smallest influence on predicted rainfall, while the UV index, dew point, and minimum temperature have the least influence. Therefore, the two primary components of rainfall in Bengaluru are moisture availability and wind dynamics.

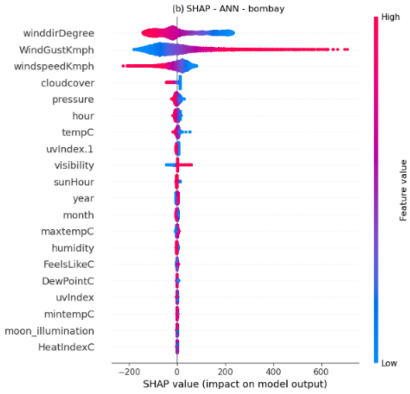


Fig 2. (b) Feature importance shown by Bee Swarm Plot for Bombay

Figure 2(b) shows how each input contributed to the expected hourly rainfall in Bombay; the most important predictor is wind direction, followed by wind gust speed, and then wind speed. The heavy reliance on these wind-related features reflects the major factors that influence rainfall in Bombay: coastal flow patterns, moisture transport, and monsoon wind dynamics. In general, high wind gust speeds exhibit a relatively strong positive relationship with rainfall intensity. This suggests a strong correlation between high excess wind speeds and precipitation. The large variability in wind direction also reflects how it shapes where and how moisture-rich air masses move. Cloud cover and pressure also contribute moderately to rainfall, indicating that they support the atmosphere's fluidity, which ultimately allows atmospheric instability to persist, a condition necessary for rain to occur. In contrast, variables such as temperature, visibility, and humidity make relatively small contributions to the total rainfall. All of this indicates that the widespread SHAP values for each input variable suggest that the weather-related processes that produce rain in Bombay are non-linear; the dynamic conditions of the winds produce far more of the expected rainfall than the other meteorological inputs.

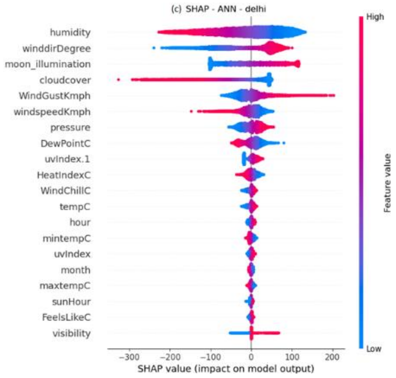


Fig. 2 (c) Feature importance shown by Bee Swarm Plot for Delhi

Figure 2(c) displays the global importance of features used for estimating hourly rainfall in Delhi (SHAP summary plot). Humidity had the greatest impact on rainfall prediction, followed by wind direction (1), moonlight (2), cloud cover (3), and wind speed (4). Generally, higher humidity values were associated with positive SHAP contributions to predicted rainfall. This finding supports the previous conclusion of a very strong relationship between moisture and rainfall. Wind-related features (such as wind direction, wind speed, and wind gust) were also shown to significantly influence rainfall modeling, underscoring the important role atmospheric winds play in both atmospheric circulation and seasonal changes in airflow that affect rainfall over Delhi. Atmospheric pressure was negatively correlated with predicted rainfall, such that higher atmospheric pressure was associated with a lower likelihood of rainfall. The broad range of SHAP values across observations suggests the nonlinear, highly variable nature of rainfall in Delhi, particularly due to urban development and convective weather.

The global feature importance for hourly rainfall prediction in Hyderabad is shown by the SHAP summary plot for Figure 2(d). The results show that the most influential variable in predicting rainfall occurrence in Hyderabad was the hour of the day. The second most influential variable was wind gust speed, followed by wind direction, wind speed, and moon illumination, in that order. The strong association with the hour indicates that short-term temporal patterns and diurnal atmospheric conditions strongly influence rainfall in Hyderabad.

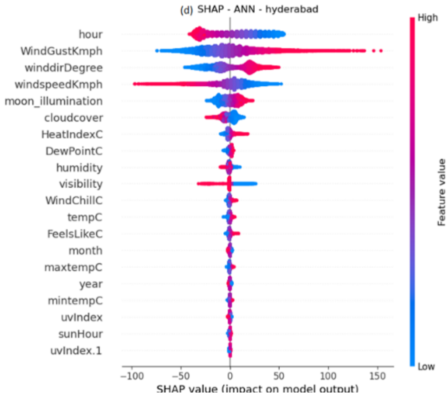


Fig. 2 (d) Feature importance shown by Bee Swarm Plot for Hyderabad

In addition, local airflow dynamics and moisture transport are likely the primary reasons why wind-related variables, such as wind gust speed, wind direction, and wind speed, significantly influence when and where rain occurs. Cloud cover, heat index, and dew point temperature all display moderate influence, supporting the idea that they contribute to atmospheric instability and precipitation development. As expected, humidity and visibility both contribute less than in coastal locations, providing stronger links with convective and wind-driven processes than relying solely on long-term moisture accumulation. The variability of SHAP values across observations illustrates the non-linear, location-specific nature of rainfall processes in Hyderabad.

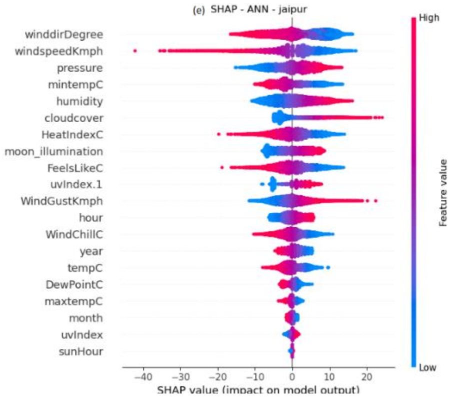


Fig. 2 (e) Feature importance shown by Bee Swarm Plot for Jaipur

Figure 2(e) shows a SHAP summary plot that represents the total contribution from each feature to predicting the hourly amount of rainfall for the city of Jaipur, India. The results show that wind direction is the strongest variable affecting rainfall, followed closely by wind speed, atmospheric pressure, minimum temperature, humidity, and cloud cover. These findings suggest that rainfall in Jaipur is primarily a result of regional air flow and the transport of moisture associated with seasonal monsoon airflow. In terms of rainfall, both humidity and cloud cover are well above average, indicating that moisture is very important for precipitation formation. In terms of their negative relationship with rainfall predictions, atmospheric pressure values are above average, indicating a lower probability of rainfall. Also, rainfall amounts, minimum temperatures, and heat indices are moderately related because they influence the development of unstable atmospheric conditions that promote convective rainfall. The range of SHAP values for all observation points indicates that rainfall is a highly variable process that occurs in non-linear ways. Rainfall is going to be extremely variable in terms of precipitation amount across all observations in a semi-arid urban climate.

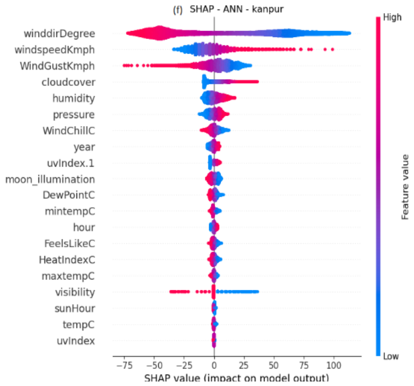


Fig. 2 (f) Feature importance shown by Bee Swarm Plot for Kanpur

For Kanpur's hourly rainfall forecast, the summary plot in Figure 2(f) shows the global comparative importance of variables, broken down by how much each contributed to predicting which hours were expected to have rainfall. Wind direction was the most significant influencing variable for hourly rainfall prediction in Kanpur, followed by wind speed, wind gust speed, cloud cover, humidity, and atmospheric pressure. The significance of wind-related variables supports the conclusion that the distribution and transport of moisture are critical influences on rainfall in Kanpur, driven by regional airflow patterns and seasonal weather systems. Higher humidity and greater cloud cover generally produced positive SHAP contributions when predicting hourly rainfall, indicating that moisture is a key component of precipitation development. Atmospheric pressure showed an inverse relationship with rainfall probability: higher atmospheric pressure was associated with lower rainfall probability. Wind chill and dew point also had moderate contributions to the prediction of hourly rainfall, showing their influence on local atmospheric stability and moisture conditions. The wide distribution of SHAP values across observations illustrates the nonlinear and variable nature of the rainfall process in Kanpur, particularly given the urban, inland climate.

The overall feature importance for hourly rainfall predicted in Nagpur is shown in the SHAP summary plot depicted in Figure 2(g). Wind direction was identified as the most influential variable in predicting rainfall; the next greatest influence factors are wind gust speed, wind speed, moon illumination, heat index, and humidity.

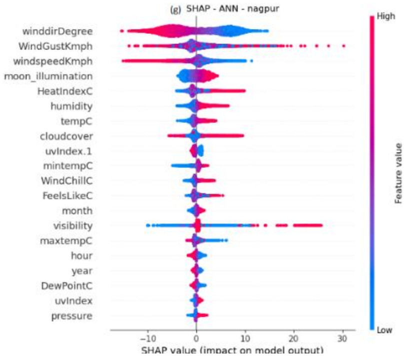


Fig. 2 (g) Feature important shown by Bee Swarm Plot for Nagpur

The significant contribution of wind-related variables indicates that rainfall in Nagpur is strongly influenced by regional airflow characteristics, seasonal wind circulation, and moisture transport. Furthermore, the positive SHAP contributions of higher humidity values to rainfall occurrence establish moisture as a critical component of precipitation development. Moderate contributions to rainfall were made by both cloud cover and temperature-related variables, such as the heat index and temperature; both reflect the significance of atmospheric instability in generating convective rainfall. The contribution of atmospheric pressure to rainfall in Nagpur was much lower than in other cities, indicating that while atmospheric pressure provides stability to the atmosphere, dynamic wind conditions are the driving force behind rainfall. The wide variability in SHAP contributions across observations further underscores the nonlinear, location-specific nature of rainfall processes in Nagpur, especially given its central, inland location.

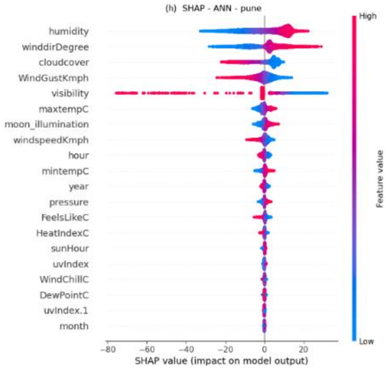


Fig. 2 (h) Feature importance shown by Bee Swarm Plot for Pune

Figure 2(h) shows the SHAP summary plot for the as obtained by utilizing the SHAP method for hourly rainfall predictors in Pune. Humidity is the most important predictor of hourly rainfall in Pune, followed by wind direction, cloud cover, wind gust speed, and visibility. Humidity is highly influential because it is a key driver of rainfall in Pune. Higher humidity increases the forecasted hourly rainfall.

Wind also plays an important role in rainfall prediction because it moves moisture in the air before it precipitates as rain. Cloud cover influences rainfall, but not as much as the other predictors, because it relates to atmospheric instability and convective processes. Visibility negatively affects hourly rainfall, suggesting that clearer skies are typically associated with less rainfall. Temperature-related predictors and pressure have smaller impacts on hourly rainfall than do the other predictors. Other predictors, such as soil type, have a negligible effect on precipitation forecasts. The variation in SHAP values indicates that the processes driving rainfall amounts in Pune are nonlinear, site-specific, and primarily driven by moisture and wind dynamics.

Although the overall ordering of feature importance remains generally stable, variations in magnitude and direction of influence occur for each feature based upon location. For instance, humidity and cloud cover are the most influential factors in improving rainfall prediction in coastal and monsoon-dominated cities; however, pressure-related features exert their greatest influence farther from the coast. Winds also exhibit marked intercity differences in their influence on rainfall prediction, with strong negative contributions from wind gust speed observed in cities such as Kanpur and Bengaluru, while positive or neutral contributions were observed in several other cities. This level of variation provides evidence that atmospheric circulation patterns, which are unique to each location, influence the degree to which winds contribute to rainfall formation, thereby highlighting the need for localized interpretation of feature comparisons rather than relying solely on aggregate feature comparison. These results build on existing XAI-based studies of rainfall, which report global feature importance but do not account for spatial/urban variation. The comparative analysis across multiple cities demonstrates that explainable model outputs are sensitive to differences in climatic context, providing additional operationally relevant data.

d) Relationship between feature values and attribution direction

The use of SHAP bee swarm analysis to identify systematic relationships between feature magnitudes and attribution signs revealed consistent patterns indicating a direct relationship between rainfall predictions and certain weather conditions. For example, higher humidity levels, cloud cover, and apparent temperatures were positively correlated with predicted rainfall, whereas higher wind speeds and temperatures were negatively correlated with it. These patterns are expected to remain consistent across the entire (or almost the entire) data set and are therefore more indicative of stable learned relationships than of spurious correlations. Additionally, the ability of multiple features to exert both positive and negative influences within the same model demonstrates that SHAP is superior at interpreting feature importance compared to traditional methods, which generally lack directional information. By illustrating how features contribute to predictions, SHAP offers us a more complete understanding of the factors driving rainfall and enables validated predictions to be developed through physical interpretation rather than data-driven methods.

e) Summary of comparative findings

Overall, the findings suggest that, while the performance of the ANN model is comparable to existing approaches for predicting rainfall, it also provides a clear, quantitative explanation for the influence of features on the predictions. The consistency of dominant predictors across cities supports the generalization of the learned relationships, whilst city-specific differences in the magnitude of attributions indicate local climatic differences not explained by the models. Together, these findings provide a compelling foundation for the following discussion of the physical interpretation, limitations, and implications of using ANN models for urban rainfall prediction.

4. Discussion

The objective of this research was to improve the understandability of machine learning rainfall predictions by using SHAP-based explainability alongside ANN models and by assessing and comparing feature contributions across several major cities in India (Başağaoğlu et al., 2022; Zounemat-Kermani & Kheimi, 2026). The findings show that explainable artificial intelligence has the potential to provide physical and meaningful explanations of rainfall drivers while also producing predictions of equal quality to those of other current machine learning techniques (Dikshit & Pradhan, 2021; Senocak et al., 2023). This section presents an interpretation of the study's findings, compares them with other studies on rainfall and explainable artificial intelligence (XAI), and explains the implications. The practical implications of the results are discussed, and the research's limitations are critically examined.

a) Correspondence with previous rainfall and XAI research

The more significant overall conclusion was that the features in both datasets were dominated by the same predictors based on physical and interpretable moisture availability from previous research (Senocak et al. 2023; Wu et al. 2023), thus confirming that the current study's findings can be extended to operational usage of this type of explainable modeling framework. This is further corroborated by monthly-aggregated comparisons, which show that moisture availability and atmospheric pressure remain the top predictors of rainfall and runoff. In the current urban study, it was confirmed that ANN and SHAP procedures can learn relationships consistent with the scientific literature and physical principles, despite their black-box properties (Senocak et al., 2023; Wu et al., 2023). The attribution patterns identified in this study are physically explainable and support the current understanding of atmospheric dynamics (Rampal et al. 2022; He et al. 2024). The clarity of the attribution will positively affect users' confidence when deploying models to forecast high-risk activities (Başağaoğlu et al., 2022). The ability to interpret the attribution provides individuals with greater confidence that the model's results are internally consistent, potentially enabling their use in decision making contexts. The current study prioritizes explainability alongside model performance, rather than the performance-based approaches emphasized by previous studies such as Schultz et al. (2021), which have prioritized performance at the cost of interpretability and usefulness. This will allow for an improved understanding of rain drivers compared to those studies. Furthermore, the LSTM-based modeling approach used by Abdellaoui & Mehrkanoon (2020) primarily focuses on capturing temporal dependencies and does not allow the user to assess how features contribute to model performance or how they may relate to the underlying processes driving rainfall. In contrast, the current study, compared with Arcodia et al. (2023), shows that a simple artificial neural network architecture enhanced with SHAP can provide both interpretability and comparable predictive performance.

b) Variability in each individual city as well as variability in local climate

The variability in forecasts across different locations clearly demonstrates that a context-driven approach is necessary for operational forecasting. Similar findings have been reported in hydrology, where spatial heterogeneity and local climate controls are critical for predicting rainfall performance in some urban areas; therefore, highlighting the importance of interpreting a specific area model accordingly (Zounemat-Kermani & Kheimi, 2026; Senocak et al., 2023). For example, if you were to look at the differences between wind variables across major US cities, you would see that there is such diverse behavior in how those wind variables behave due to different local or regional wind fields and liquid water transfer mechanisms. There is also considerable variability found in many regional rainfall studies for both India and globally; localized atmospheric circulations, coastal influences, and/or geographical influences are all determinants for how rainfall occurs (Mohapatra et al., 2021; Priya Ashok & Pekkat, 2024; Shukla et al., 2026; Liu et al., 2008; Buytaert et al., 2006)

Mohapatra et al. (2021) investigated rainfall patterns across the Indian meteorological subdivisions and found large differences in both timing and amount of precipitation between different regions. They also found that machine learning algorithms yield different results depending on the climate characteristics of each area being evaluated. Likewise, Priya Ashok & Pekkat (2024) demonstrated that rainfall prediction in Guwahati should include location-specific predictors, as the rainfall predicted by these models depends heavily on the extent to which the urban climate has influenced the model's past performance. In coastal cities such as Mumbai, rainfall is affected by moisture transport from the Arabian Sea and interactions with monsoonal flows, whereas in inland cities such as Delhi and Jaipur, rainfall is more strongly influenced by convective instability and variations in temperature and pressure, based on historical records (Shukla et al., 2026).

Studies on explainable artificial intelligence (AI) also support this variability. Dikshit and Pradhan (2021) found that feature importance in models predicting environmental outcomes is highly dependent on the model's location. Also, they found that aggregating feature attribution masks important regional differences. Başağaoğlu et al. (2022) also state that context-based interpretability in hydro-climatic modeling must consider local climate or geographical conditions, as these largely control hydrologic behavior.

These findings are important for the development of machine-learning-based precipitation predictions at the urban scale. If a prediction is made without features related to city-specific climate and without relying on aggregated or global feature importance, the variability across cities may lead to erroneous conclusions and models that do not generalize well.

Table 2. Performance metrics

Mumbai's Coastal Rainfall Dynamics and Atmospheric Drivers (Shukla et al., 2026)	Validates why Mumbai's R^2 was lower and why temperature attribution differed from inland cities like Delhi
Regional Rainfall Variability (Mohapatra et al., 2021)	Provides physical grouping that justifies treating coastal and inland cities as distinct entities
Limitations of Data-Driven Models (Materia et al., 2024)	Supports the argument that aggregate models may mask local climatic differences

Therefore, this research demonstrates, through an evaluation of the SHAP attributions across different cities, that the model's explainability is context specific; thus, the results of explainability should also be interpreted contextually. This differs from Wu et al. (2023), who typically analyze feature importance rather than examining how it varies across cities, thereby limiting the ability to understand how local climates affect feature attribution. In contrast to Rampal et al. (2022), who provide a global overview of feature importance trends, the current study demonstrates that feature attribution does not occur uniformly and depends heavily on the urban environment and the climatic conditions of the study location.

c) Interpretable urban rainfall prediction- benefits of incorporating SHAP into ANN

In addition to providing transparency of modeling processes, SHAP and ANN allow for the identification of a smaller subset of predictors that can then aid in determining which sensors should be monitored, as well as how to collect urban meteorological data for use in predicting urban rainfall. The ability to differentiate between the positive and negative effects of predictor features can lead to a better understanding of prediction uncertainty and the model's performance during extreme or anomalous weather events. From an operational standpoint, an explainable rainfall prediction model allows practitioners to build confidence in their predictions because there is evidence that the predictions are based on physically feasible variables. This is especially critical for urban flood management and early warning systems, as decision-makers need not only accurate forecasts but also a clear explanation of how the model produced them (Kemmannu et al., 2025; Shukla et al., 2026). SHAP can simultaneously analyze the direction and strength of each feature's contribution, whereas previous statistical and machine learning models only show the magnitude of importance, providing a better overall understanding of how rainfall works. This also distinguishes SHAP from previous XAI studies in hydrology (Başağaoğlu et al., 2022), which mainly focused on ranking features rather than analyzing the reliability and direction of effects across multiple locations.

d) Broadened practical implications

The practical implications for these results go beyond model interpretability. Most urban flood management early warning systems are based on threshold-based rainfall triggers with a limited understanding of the meteorological factors behind predicted rainfall. By establishing stable, positive contributing factors (e.g., moisture-related variables and cloud cover), urban authorities will have tools to interpret forecast rainfall as a function of clear, identifiable atmospheric conditions rather than a purely numerical value. Increased interpretability could foster greater confidence in rapid model implementation during critical flooding events, when emergency response decisions must be made quickly. Additionally, identifying locality-specific variability in variable influence will have direct applications for localized flood mitigation efforts. Urban drainage design, reservoir management, and stormwater infrastructure prioritization will be improved by more accurately identifying which meteorological variables have the greatest predictive impact in the specific climatic zone (Kemmannu et al., 2025; Samanta et al., 2026). For example, coastal cities experiencing moisture surges would require different monitoring priorities than residents living in an inland convective-dependent area. Therefore, establishing clear links between explainable rainfall formulas and both predictive accuracy and informing risk-based urban planning and adaptive climate resilience will be achieved.

e) Trade-off between model performance and interpretability

Although the ANN model generates performance results like recent studies for deep learning for rainfall prediction, most cities produce R^2 values indicative of the inherent difficulty in hourly rainfall prediction from hour to hour (Schultz et al., 2021; He et al., 2024). Since rainfall occurs through multiscale atmospheric processes (i.e., many of which cannot be explained by using only surface level meteorological variables), the research suggests to not interpret explainability as a replacement for predictive skill; however, use of global SHAP analysis clearly highlights stable and interpretable drivers across a wide array of data but likely will not reflect the event specific dynamics that are captured through local and/or temporal SHAP analyses. Future research could build on the present framework by considering the combined use of global and local explainability to better capture transient rainfall mechanisms. Unlike most previous studies, where interpretability is viewed as a secondary result, here interpretability is the main goal of the model, helping determine how operationally useful those rainfall models will be for real-world decisions (Zounemat-Kermani & Kheimi, 2026; Robles et al., 2026).

f) Limitations and future directions

There are numerous limitations to consider. First, the analysis is performed using a fixed set of surface-level meteorological predictors; thus, the study does not include upper-air or global climate indices, which may also improve predictive skill and the physical accuracy of predictions. Computational constraints required the use of sampling-based SHAP calculations, which may introduce slight approximation errors. The study is limited to a single model class, namely ANN, which restricts the ability to validate the generalization of the findings through a comparison of explainability using different model architectures; however, findings indicate that combining ANN-based predictions with SHAP-based explainability forms a reliable and interpretable framework for analysis of urban rainfall, expanding the application of XAI in hydro-meteorological modeling and establishing a foundation for the development of more transparent and operationally relevant forecasting systems for rainfall (Başağaoğlu et al., 2022; Materia et al., 2024). A further practical limitation is the computational cost of conducting SHAP analyses for large datasets and deep learning models, as exact SHAP value calculations grow exponentially with increasing feature counts, rendering them infeasible for high-dimensional problems without approximate methods. In this report, sampling-based methods were utilized to reduce the computational burden associated with SHAP analyses; however, the use of sampling-based methods may introduce minor variance to estimate attribution, as well as increase computing time to scale to additional cities or longer periods of time (Jesus et al., 2021; Lamane et al., 2025; Mosavi et al., 2018). When it comes to operational development, whether at the national or real-time forecast level, the important question is computational efficiency. Future studies should consider opportunities for optimizing models through taking a model-specific approach with SHAP optimizations, introducing parallelization methodologies, and/or using other types of models to explain the model outputs with an appropriate balance between how well the output can be attributed (i.e., fidelity) and the extent to which it can be carried out (i.e., scalability). In addition, benchmark comparisons of time and memory consumption across architectures, such as artificial neural networks, long short-term memory (LSTM) networks, and ensemble models, would provide very useful information for large-scale implementation.

Conclusions

This research article makes a substantive contribution to the theoretical understanding of explainability in AI, specifically regarding the prediction of rainfall using data-driven techniques such as artificial neural networks (ANNs). The study demonstrates a method for transforming black-box ANN models into interpretable ANN frameworks using SHAPley values (SHAP). In addition to understanding feature-level attribution through the quantification of the importance, directionality, and consistency of atmospheric data (i.e., meteorological variables), this will aid researchers in using their data-driven rainfall predictions more accurately within one or more urban locations. Ultimately, this research provides a structured methodology for linking data-driven results to physically interpretable ones, thereby enhancing the scientific credibility of machine learning methods in hydrology.

Although many of the aforementioned studies have made important contributions to the field, there remain gaps. There is relatively low use of explicitly spatial data (e.g., radar and/or satellite images), a lack of long-term time-series dependencies in model construction, and a lack of methods to evaluate model results for extreme rainfall events (i.e., rare events). Additionally, hybrid deep learning architectures and real-time applications for operational forecasting systems remain underexplored.

Future research should extend this framework to incorporate spatial datasets and hybrid architectures, including convolutional neural networks and long short-term memory (LSTM) models, to capture both spatial and temporal dependencies in rainfall prediction. Future research should also investigate model performance under extreme weather conditions and assess the validity of SHAP-based interpretations in real-time forecasting environments. Finally, future studies should extend the framework across additional geographies and integrate domain-specific physical properties to improve the generalizability and operational effectiveness of these predictive systems.

Declarations

Acknowledgemnet

This research could not have been completed without the PARAM SHAVAK supercomputing facility (available through Sankalchand Patel University, Visnagar) for performing experiments; the authors also wish to acknowledge the assistance of the Indian Institute of Remote Sensing, Dehradun, in providing valuable training support during this study.

Funding

No funding was provided for this research.

Conflict of interest

There are no known conflicts of interest from any Author.

Ethics approval and consent to participate

N/A, as this article does not involve either human or animal subjects.

Consent for publication

All Authors agree to publish this manuscript as submitted.

Data availability

The data are publicly available at https://www.kaggle.com/dsv/1129180

Materials availability

Not Applicable

Code availability

The code can be provided upon request.

Author contribution

The manuscript was written through the contributions of all authors.

References

Abdellaoui, I. A., & Mehrkanoon, S. (2020). Deep multi-stations weather forecasting: Explainable recurrent convolutional neural networks. *ArXiv Preprint ArXiv*:2009.11239. https://doi.org/10.48550/arXiv.2009.11239

Arcodia, M. C., Barnes, E. A., Mayer, K. J., Lee, J., Ordonez, A., & Ahn, M. S. (2023). Assessing decadal variability of subseasonal forecasts of opportunity using explainable AI. *Environmental Research: Climate*, *2*(4), 045002. https://doi.org/10.1088/2752-5295/aced60

Başağaoğlu, H., Chakraborty, D., Lago, C. D., Gutierrez, L., Şahinli, M. A., Giacomoni, M., Furl, C., Mirchi, A., Moriasi, D., & Şengör, S. S. (2022). A review on interpretable and explainable artificial intelligence in hydroclimatic applications. *Water*, *14*(8), 1230. https://doi.org/10.3390/w14081230

Basha, S. S., & Begum, B. N. (2025). Weight optimized stacked LSTM with conditional random fields using self-adaptive generalized normal distribution optimizer for crop yield forecasting. *Applied Soft Computing*, *184*, 113760. https://doi.org/10.1016/j.asoc.2025.113760

Bialek, J., Bujalski, W., Wojdan, K., Guzek, M., & Kurek, T. (2022). Dataset-level explanation of heat demand forecasting ANN with SHAP. *Energy*, 261, 125075. https://doi.org/10.1016/j.energy.2022.125075

Buytaert, W., Celleri, R., Willems, P., De Bièvre, B., & Wyseure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *Journal of Hydrology*, *329*(3-4), 413-421. https://doi.org/10.1016/j.jhydrol.2006.02.031

Chamundeswari, G., Kumar, V. S., Bairisetty, P. B., Aggala, C., Yakkala, N. K., & Venkatesh, C. V. C. (2026, January). A Next-Generation AI based Predictive Model for Indian Rainfall using Real-Time API Data. *In 2026 International Conference on AI-Driven Smart Systems and Ubiquitous Computing (ICAUC)* (1-5). IEEE. https://doi.org/10.1109/ICAUC68182.2026.11441247

Chandra, H., Pawar, P. M., Elakkiya, R., Tamizharasan, P. S., Muthalagu, R., & Panthakkan, A. (2023). Explainable artificial intelligence for soil fertility prediction. *IEEE Access*, *11*, 97866-97878. https://doi.org/10.1109/ACCESS.2023.3311827

Chittaragi, A., Patil, B., Kumar, M. P., & Devanna, P. (2025). Hybrid Random Forest–Artificial Neural Network Model Based Forecasting of Anthracnose in Bottle Gourd Across Different Transplanting Windows. *Smart Agricultural Technology*, *12*, 101477. https://doi.org/10.1016/j.atech.2025.101477

Dikshit, A., & Pradhan, B. (2021). Interpretable and explainable AI (XAI) model for spatial drought prediction. *Science of the Total Environment*, 801, 149797. https://doi.org/10.1016/j.scitotenv.2021.149797

Haykal, H., Permanasari, A., & Rabbani, K. A. (2022). The legal status of fiduciary guarantee belongs to members of save and loan cooperatives due to bankruptcy. *Journal of Advances in Humanities and Social Sciences*, *8*(1), 9-18. https://doi.org/10.20474/jahss-8.1.2

He, R., Zhang, L., & Chew, A. W. Z. (2024). Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning. *Expert Systems with Applications*, 235, 121160. https://doi.org/10.1016/j.eswa.2023.121160

Jesus, S., Belém, C., Balayan, V., Bento, J., Saleiro, P., Bizarro, P., & Gama, J. (2021). How can I choose an explainer? An application-grounded evaluation of post-hoc explanations. *In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 805–815. ACM. https://doi.org/10.1145/3442188.3445941

Kemmannu, P. K., Udayakumar, M., Tamilselvan, V., Lingineni, S. L., Kumar, A., & Paul, P. M. (2025, September). A Hybrid AI-Powered Framework for Accurate Urban Flood Prediction and Real-Time Early Warning Systems. *In 2025 6th International Conference on Smart Electronics and Communication (ICOSEC)*, (1819-1824). IEEE. https://doi.org/10.1109/ICOSEC67334.2025.11459765

Lamane, H., Mouhir, L., Moussadek, R., Baghdad, B., Kisi, O., & El Bilali, A. (2025). Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. *International Journal of Sediment Research*, *40*(1), 91-107. https://doi.org/10.1016/j.ijsrc.2024.10.002

Liu, Y., Gupta, H., Springer, E., & Wagener, T. (2008). Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. *Environmental Modelling & Software*, *23*(7), 846-858. https://doi.org/10.1016/j.envsoft.2007.10.007

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30. https://doi.org/10.48550/arXiv.1705.07874 Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Mansoor, M., Khan, T. I., Jam, F. A., & Alasmari, M. (2025). From donations to devotion: how cause-related marketing frames drive brand evangelism through cognitive and social pathways in hospitality. *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/ijchm-02-2025-0233

Materia, S., García, L. P., vanStraaten, C., O, S., Mamalakis, A., Cavicchia, L., Coumou, D., deLuca, P., Kretschmer, M., & Donat, M. (2024). Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives. *WIREs Climate Change*, *15*(6), e914. https://doi.org/10.1002/wcc.914

Mohapatra, G., Rakesh, V., Purwar, S., & Dimri, A. P. (2021). Spatio-temporal rainfall variability over different meteorological subdivisions in India: Analysis using different machine learning techniques. *Theoretical and Applied Climatology*, *145*(1), 673-686. https://doi.org/10.1007/s00704-021-03644-7

Priya Ashok, S., & Pekkat, S. (2024). Performance assessment of rainfall forecasting models for urban Guwahati City using machine learning techniques and singular spectrum analysis. *Journal of Water and Climate Change*, *15*(4), 1565-1587. https://doi.org/10.2166/wcc.2024.465

Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolino, C., Noll, B., & Meyers, T. (2022). High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand. *Weather and Climate Extremes*, 38, 100525. https://doi.org/10.1016/j.wace.2022.100525

Retsch, M., Jakob, C., & Singh, M. (2022). Identifying relations between deep convection and the large-scale atmosphere using explainable artificial intelligence. *Journal of Geophysical Research: Atmospheres*, *127*(3), e2021JD035388. https://doi.org/10.1029/2021JD035388

Robles, K. P. V., Solmerin, J. G., Pugat, G. C. E., & Monjardin, C. E. F. (2026). A Review of the Advances and Emerging Approaches in Hydrological Forecasting: From Traditional to AI-Powered Models. *Water*, *18*(1), 119. https://doi.org/10.3390/w18010119

Sahakyan, M., Aung, Z., & Rahwan, T. (2021). Explainable artificial intelligence for tabular data: A survey. *IEEE Access*, 9, 135392–135422. https://doi.org/10.1109/ACCESS.2021.3116481

Samanta, R., Singh, V., Bajpai, R., Lodh, A., Routray, A., & Srivastava, A. (2026). Advances in the monitoring and forecasting of urban extreme meteorological events: A bibliometric review. *Natural Hazards*, *122*(5), 186. https://doi.org/10.1007/s11069-025-07933-6

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., & Stadtler, S. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200097. https://doi.org/10.1098/rsta.2020.0097

Senocak, A. U. G., Yilmaz, M. T., Kalkan, S., Yucel, I., & Amjad, M. (2023). An explainable two-stage machine learning approach for precipitation forecast. *Journal of Hydrology*, 627, 130375. https://doi.org/10.1016/j.jhydrol.2023.130375

Shukla, K. K., Mishra, A. K., Kumar, G., Sahu, D. K., Bhute, S., & Adhikary, S. (2026). Atmospheric drivers of the 26 May 2025 heavy rainfall event over Mumbai: insights from observations and reanalysis. *Natural Hazards*, 122, 380. https://doi.org/10.1007/s11069-026-08118-5

Soneji, H. (2020). Historical weather data for Indian cities [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/1129180

Tahmasbi, S. K., Kouhanestani, Z. K., & Joorabian, S. (2025). Investigating the uncertainty of atmospheric general circulation models in simulating the future runoff of Abu Al-Abbas Watershed, Khuzestan, Iran. *Environment*, *11*(2), 116-124.

Wu, J., Wang, Z., Dong, J., Cui, X., Tao, S., & Chen, X. (2023). Robust runoff prediction with explainable artificial intelligence and meteorological variables from deep learning ensemble model. *Water Resources Research*, *59*(9), 1-24. https://doi.org/10.1029/2023WR035676

Zounemat-Kermani, M., & Kheimi, M. (2026). Explainable Artificial Intelligence in Hydrology: A Review. *Water Resources Management*, *40*(3), 106. https://doi.org/10.1007/s11269-025-04435-9