Journal of Artificial Intelligence and Metaheuristic

♦ ISSN: 2833-5597 Vol: 09, No. 02, 2025 ♦ PP. 24–30

Article DOI: https://doi.org/10.54216/JAIM.090203

FULL LENGTH ARTICLE

Interpretable Rainfall Forecasting Using SHAP-Enhanced Machine Learning: A Case Study on U.S. Urban Climate Data (2024–2025)

Khaled Sh. Gaber[1,*] **Mahmoud Elshabrawy Mohamed**[1,*]

[1]Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA

Emails: khsherif@jcsis.org; mshabrawy@jcsis.org

Received: January 27, 2025 Revised: March 03, 2025 Accepted: May 08, 2025 ⋆ Corresponding author

ABSTRACT

Accurate rainfall prediction supports resilient communities, sustainable agriculture, water distribution, and disaster preparedness. Because rainfall depends nonlinearly on interacting meteorological conditions, conventional forecasting methods can be inadequate for short-term, localized events. This study develops a machine-learning framework for next-day rainfall classification using daily weather observations from 20 major U.S. cities during 2024–2025. The dataset includes temperature, relative humidity, wind speed, precipitation, cloud cover, and pressure, together with a binary *Rain Tomorrow* label. Six classifiers are evaluated: Random Forest, Support Vector Machine (RBF kernel), k-Nearest Neighbors, Logistic Regression, Naive Bayes, and Linear SVM. SHapley Additive exPlanations (SHAP) are used to interpret model predictions and feature contributions. Random Forest reports the strongest results, with precision, recall, and F1-score of 1.00; RBF SVM and KNN achieve F1-scores of 0.97 and 0.94, respectively. Logistic Regression, Linear SVM, and Naive Bayes achieve F1-scores between 0.76 and 0.77. SHAP analysis identifies precipitation and humidity as the most influential variables, consistent with meteorological expectations. These results suggest that combining an ensemble classifier with interpretable explanations can support transparent rainfall forecasting for agriculture, urban planning, and emergency response.

Keywords: Rainfall prediction ▪ SHAP ▪ Machine learning ▪ Random Forest ▪ Explainable AI

1. INTRODUCTION

Precipitation is a vital hydrological process on Earth. Accurate forecasts are central to agricultural production, environmental protection, hydrological infrastructure, emergency response, and public policy. Rainfall is the primary source of freshwater and governs processes that support ecosystem health and sustainable communities [1]. Agricultural planning, reservoir management, drainage control, and disaster-warning systems all depend on reliable rainfall information. In rain-fed regions, sub-seasonal forecasts can inform agricultural planning and on-farm decisions [2]. In urban areas, rapid land-use change and urban sprawl can increase flooding, drainage failures, traffic disruption, property damage, and risks to human life when forecasts are inadequate [3].

Rainfall anomalies are among the consequential effects associated with a changing climate. Their impacts include flash floods, landslides, prolonged droughts, fires, and agricultural disruption [4]. A warmer atmosphere can hold more moisture and intensify the hydrological cycle, increasing uncertainty and the likelihood of short, intense rainfall events as well as longer dry periods [5]. Timely forecasts can therefore provide substantial benefits to vulnerable communities and environmentally sensitive regions [6].

© 2025 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

Rainfall prediction is challenging because many interacting variables influence precipitation, including temperature, humidity, barometric pressure, wind, cloud cover, and atmospheric circulation. Large-scale ocean–atmosphere systems such as ENSO, IOD, and NAO also affect rainfall patterns [7]. The resulting system has spatial and temporal dependencies and nonlinear, non-stationary behavior [8]. Traditional Numerical Weather Prediction (NWP) models use physical equations to represent atmospheric dynamics. Although theoretically powerful, they require substantial computing resources and can struggle to resolve localized phenomena such as convective storms and mountain-influenced weather. Uncertain parameterizations and boundary conditions can also cause differences between forecast and observed rainfall, particularly for short-term or localized events [9].

These limitations have encouraged the use of data-driven Machine Learning (ML) and Deep Learning (DL) methods. Rather than relying solely on explicit physical-process models, these approaches learn statistical patterns from observations and can model high-dimensional, noisy, and nonlinear data. Support Vector Machines, decision trees, Random Forests, Gradient Boosting, and XGBoost have been applied to rainfall prediction, while recurrent architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit networks can learn temporal patterns [10, 11, 12]. Their performance nevertheless depends on the quality, quantity, and representativeness of the input data.

This study develops a dataset and evaluates machine-learning approaches for next-day rainfall prediction using daily observations from major urban and semi-urban centers across the United States during 2024–2025. Each record includes temperature, humidity, wind speed, pressure, cloud cover, and precipitation, as well as a binary *Rain Tomorrow* label. The study compares classifiers under a common evaluation framework and uses SHAP values to explain model predictions. The analysis aims to connect machine-learning methods with practical hydrometeorological forecasting and to identify models that can support future operational systems [13].

2. LITERATURE REVIEW

Rainfall prediction is important for flood preparedness, agricultural planning, water-resource management, and disaster risk reduction. Because precipitation is shaped by interacting atmospheric and terrestrial factors, data-driven ML and DL approaches are increasingly used alongside traditional statistical and physical models.

Barrera-Animas et al. compared conventional ML models with temporal neural architectures for hourly rainfall prediction using data from five UK cities covering 2000–2020. Their study evaluated LSTM, Stacked-LSTM, and Bidirectional-LSTM models alongside tree ensembles and support-vector regression; Bidirectional-LSTM and Stacked-LSTM performed well in accuracy and computational efficiency [14]. For Nigeria, Ojo and Ogunjo compared Multivariate Polynomial Regression, Artificial Neural Networks, ANFIS, SVM, and autoregressive approaches using 31 years of rainfall data. ANFIS performed best across ten consecutive months, illustrating the potential of hybrid fuzzy methods for seasonal and regional prediction [15].

For Saudi Arabia, Baljon and Sharma evaluated machine-learning classifiers for rainfall prediction using historical meteorological data; their decision-tree classifier achieved 96.1% accuracy [16]. Akhtar et al. proposed a CNN optimized with Adaptive Searched Scaling factor-based Elephant Herding Optimization, automatically tuning network structure and related parameters. Their results demonstrated the potential of evolutionary optimization for rainfall prediction [17]. Allawi et al. compared ANN models optimized with Artificial Bee Colony, Particle Swarm Optimization, and Imperialist Competitive Algorithm using 54 years of Malaysian monthly rainfall data; ICA-ANN produced the strongest comparative results [18].

Chen et al. compared LSTM and Random Forest for rainfall prediction in the distinct climatic regions of Rize and Konya, Turkey. Their 41-year time-series evaluation found that LSTM performed well across several hydrological metrics [19]. Tuysuzoglu et al. introduced an ensemble K-star classifier with probability-based aggregation and feature selection for next-day categorical rainfall prediction [20]. Ahmadi et al. studied ensemble variants of SMO regression in Iran and identified minimum relative humidity as an influential variable [21]. Satheesh et al. examined the predictability of synoptic-scale rainfall during the West African summer monsoon using a data-driven statistical forecast based on spatiotemporal patterns in satellite rainfall. Their forecast outperformed climatology and operational ensemble benchmarks along the fringes of the rainbelt, while skill remained poor within the main rainbelt [22].

For China's Taihu Basin, Gu et al. combined KNN, XGBoost, SVR, or ANN base learners in a stacking ensemble and incorporated local meteorological factors and climate indices [23]. Alqahtani et al. developed a CNN–Bidirectional-LSTM hybrid for rainfall prediction in the environmentally sensitive wetlands of southern Iraq [24]. Mohamadi et al. studied rainfall prediction using multiple inclusive models and large climate indices [25]. Related work in the Sefidrood Basin combined RBFNN and MLP models with NMR, PSO, FFA, and GA optimization, reducing RMSE by up to 72% compared with standalone models [25]. Johny et al. combined Multivariate Empirical Mode Decomposition, Time Dependent Intrinsic Cross Correlation, and LSTM with ENSO and IOD indices for all-India rainfall analysis, reporting skill 0.98, NSE 0.95, and IA 0.91 [26].

Other studies have addressed related hydrological prediction tasks. Zhao et al. evaluated 17 ML algorithms for rainfall-driven debris-flow prediction in central China; Extra Trees performed well on their event dataset [27]. Roy et al. compared metaheuristic ANFIS models for runoff prediction in UK river basins and reported strong validation performance for biogeography-based optimization [28]. Collectively, these studies illustrate the value of multivariate inputs, nonlinear learners, ensembles, and hybrid optimization. The present work focuses on next-day rainfall classification and complements prediction metrics with SHAP-based interpretation.

3. DATASET AND PREPROCESSING

3.1 Dataset Description

The study uses the USA Rainfall Prediction Dataset, which co-

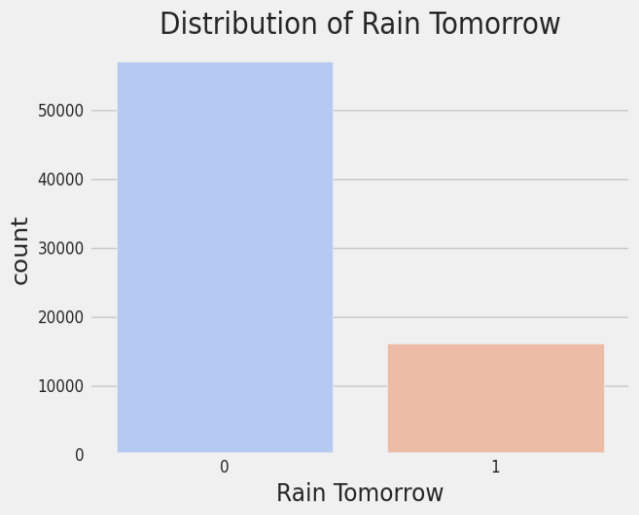


Figure 1. Distribution of the target variable, *Rain Tomorrow* (0 = no rain, 1 = rain).

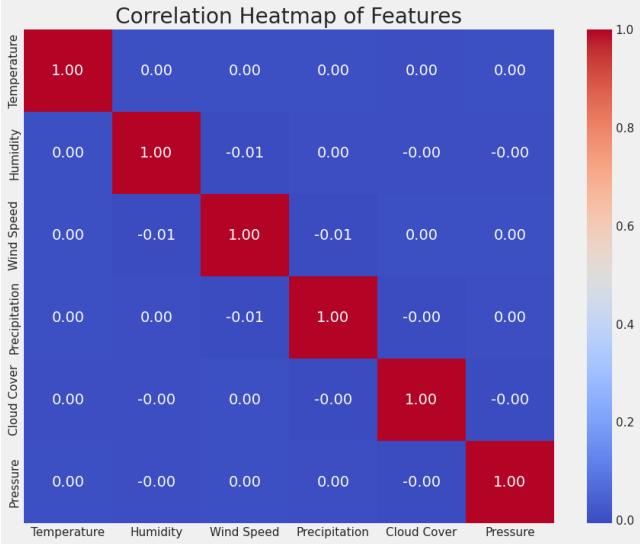


Figure 2. Pearson correlation heatmap of the meteorological features.

ntains daily weather observations from 20 major U.S. cities for 2024 and 2025 [29]. The data are intended to support temporal analysis, seasonal-pattern recognition, and rainfall-event prediction. The target variable, *Rain Tomorrow*, is binary: 1 indicates rainfall the following day, and 0 indicates no rainfall. The six meteorological input features are:

- Temperature: daily average surface-air temperature in degrees Fahrenheit.
- Humidity: relative atmospheric humidity, expressed as a percentage.
- Wind speed: surface wind movement, measured in miles per hour.
- Precipitation: rainfall observed during the measurement period, in inches.
- Cloud cover: percentage of the sky covered by clouds.
- Pressure: barometric pressure in millibars.

Figure 1 shows the distribution of the target classes. No-rain observations outnumber rain observations, so the dataset has class imbalance. Figure 2 shows the Pearson correlations among the numerical meteorological features. Most pairwise linear relationships are small, suggesting limited multicollinearity among the predictors; rainfall relationships may nevertheless be nonlinear.

3.2 Data Preprocessing

The original dataset contained no missing values. To make the workflow reproducible and extensible, the preprocessing strategy nevertheless included conditional missing-value handling. Numerical variables such as temperature, humidity, wind speed, and pressure were to be imputed with their column means using the `SimpleImputer` class from scikit-learn. Mode imputation or forward filling could be used for categorical or datetime variables if required [30].

The data types were checked for compatibility with the machine-learning pipeline. Continuous variables were represented numerically, while the categorical target *Rain Tomorrow* was encoded as an integer. Conversion steps ensured that inputs met the data requirements of each classifier.

3.3 Feature Engineering

The selected input variables were temperature, humidity, wind speed, precipitation, cloud cover, and pressure. These features represent both immediate weather conditions and factors associated with rainfall formation. Feature-importance scores from tree-based models, including Random Forest and XGBoost, indicated that precipitation and humidity were the most influential variables. This ranking is consistent with the role of atmospheric moisture and observed precipitation in rainfall patterns. Feature-importance analysis can also inform later variable reduction, interaction construction, or parameter optimization.

4. RESULTS AND DISCUSSION

This section evaluates six classifiers on the USA Rainfall Prediction Dataset (2024–2025). The binary classification task predicts whether rainfall will occur the following day. Precision, recall, and F1-score are used to compare predictive performance, and confusion matrices, ROC curves, Precision–Recall curves, and SHAP visualizations provide additional diagnostic and interpretability analyses.

4.1 Evaluation Metrics

Precision measures the fraction of predicted rain events that are correct:

$$\text{Precision} = \frac{TP}{TP+FP}, \tag{1}$$

where TP denotes true positives and FP false positives. High precision corresponds to fewer false rain alerts.

Recall measures the fraction of actual rain events identified correctly:

$$\text{Recall} = \frac{TP}{TP+FN}, \tag{2}$$

where FN denotes false negatives. High recall means that fewer rainy days are missed.

The F1-score is the harmonic mean of precision and recall:

$$F_1 = 2\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

It provides a single summary of the balance between false-positive and false-negative behavior, which is useful when class frequencies are unequal.

4.2 Applied Machine-Learning Models

Six classifiers were evaluated:

1. **Random Forest:** An ensemble of decision trees trained using bootstrap samples and randomized feature selection. Voting across trees supports generalization and helps limit overfitting. In this evaluation, Random Forest achieved precision, recall, and F1-score values of 1.00.
2. **Logistic Regression:** A linear baseline that estimates class probabilities with the logistic (sigmoid) function. Its F1-score was 0.77.
3. **RBF Support Vector Machine:** A kernel-based classifier that can represent nonlinear decision boundaries. It achieved precision of 0.98, recall of 0.97, and an F1-score of 0.97.
4. **k-Nearest Neighbors (KNN):** A non-parametric method that assigns a class by voting among nearby observations. It achieved an F1-score of 0.94.
5. **Naive Bayes:** A probabilistic classifier based on Bayes' theorem and a conditional-independence assumption. It achieved precision of 0.96, recall of 0.63, and an F1-score of 0.76.
6. **Linear SVM:** A maximum-margin linear classifier. It achieved precision of 0.84, recall of 0.71, and an F1-score of 0.77.

4.3 Classification Results

Table 1 summarizes the reported precision, recall, and F1-score values. Random Forest achieved the highest scores, while RBF SVM and KNN also performed strongly. Logistic Regression and Linear SVM had identical reported results. Naive Bayes combined high precision with lower recall, indicating that it missed a larger share of actual rain events.

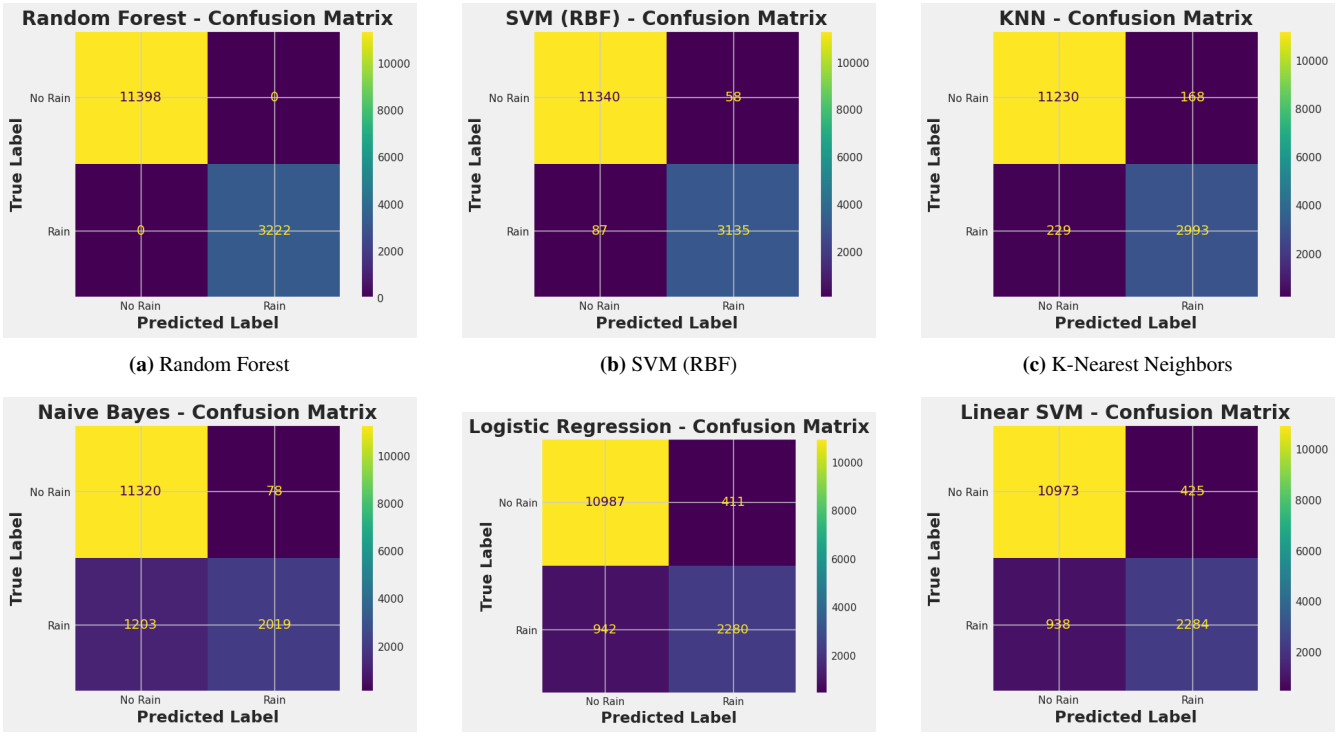
Table 1. Comparison of the classification models.

Model	Precision	Recall	F_1-score
Random Forest	1.00	1.00	1.00
Logistic Regression	0.84	0.71	0.77
SVM (RBF)	0.98	0.97	0.97
KNN	0.95	0.93	0.94
Naive Bayes	0.96	0.63	0.76
Linear SVM	0.84	0.71	0.77

Figure 3 compares the true and predicted class labels. Random Forest shows no errors in the displayed matrix. The RBF SVM and KNN matrices contain relatively few errors, whereas Logistic Regression, Linear SVM, and Naive Bayes miss more rain events. These patterns complement the aggregate precision and recall results.

The Receiver Operating Characteristic (ROC) curve describes the relationship between the true-positive rate and false-positive rate as the decision threshold varies. The curves in Figure 4 report AUC values of 1.00 for Random Forest, RBF SVM, and KNN; 0.98 for Naive Bayes; and 0.95 for Logistic Regression and Linear SVM. ROC–AUC complements fixed-threshold measures by summarizing discrimination across thresholds.

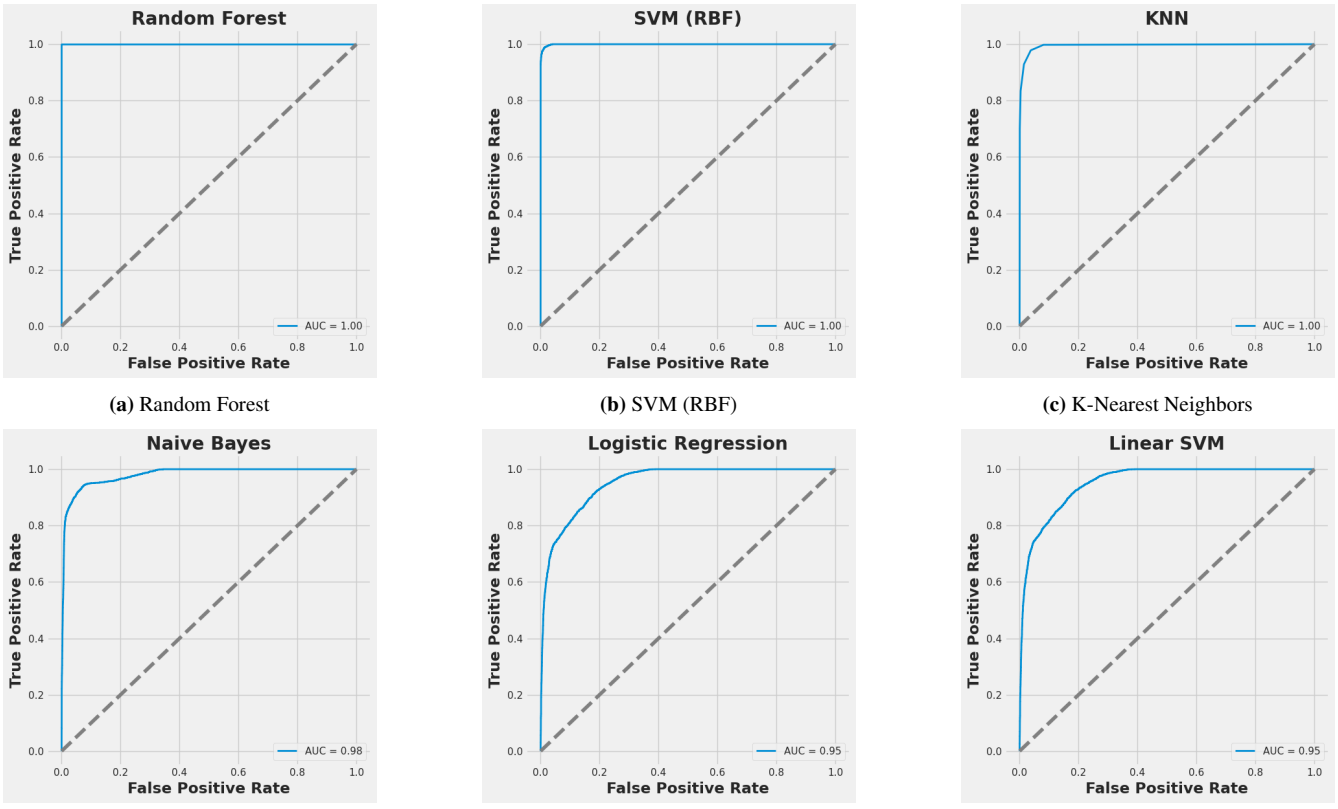
Precision–Recall analysis is especially informative when positive and negative classes are imbalanced. Figure 5 shows the trade-off between precision and recall over classification thresholds. The reported average-precision (AP) values are 1.00 for Random Forest and RBF SVM, 0.98 for KNN, 0.94 for Naive Bayes, and 0.86 and 0.85 for Logistic Regression and Linear SVM, respectively. Naive Bayes retains



(a) Random Forest **(b)** SVM (RBF) **(c)** K-Nearest Neighbors

(d) Naive Bayes **(e)** Logistic Regression **(f)** Linear SVM

Figure 3. Confusion matrices for the six rainfall-prediction classifiers, comparing actual and predicted *Rain* and *No Rain* labels.



(a) Random Forest **(b)** SVM (RBF) **(c)** K-Nearest Neighbors

(d) Naive Bayes **(e)** Logistic Regression **(f)** Linear SVM

Figure 4. ROC curves for the six classification models. The curves compare sensitivity and specificity over decision thresholds; AUC values are reported in each panel.

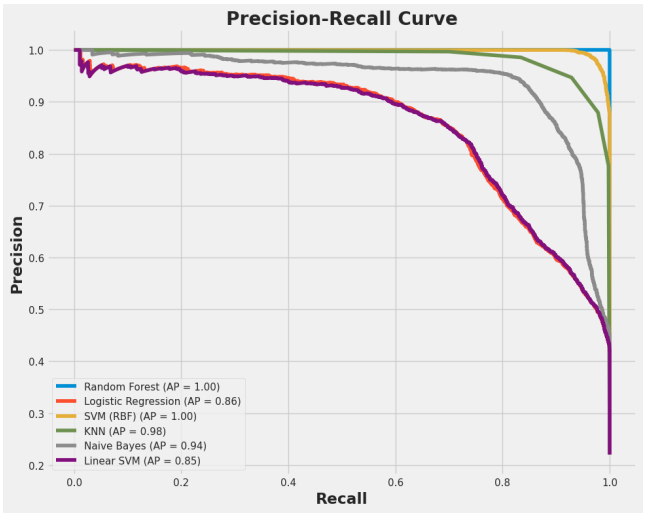


Figure 5. Precision–Recall curves for the classification models. Average precision (AP) summarizes each model's curve.

high precision over part of the recall range but declines as recall increases, consistent with its lower recall at the selected threshold.

4.4 SHAP Interpretation

SHAP interaction values describe both individual feature contributions and pairwise effects on a model's output. Figure 6 shows the interaction between temperature and humidity. The plotted interaction values vary around zero, indicating that these variables can have joint effects even when their individual contributions are small.

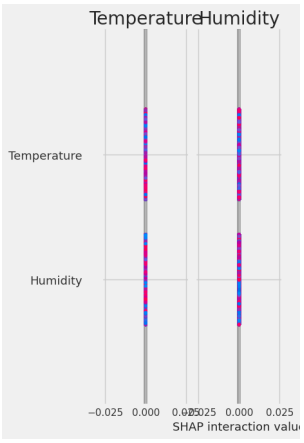


Figure 6. SHAP interaction plot for temperature and humidity. Deviations from zero indicate joint feature effects on the model output.

Figure 7 presents the SHAP summary plot, ranking features by their contributions across observations.

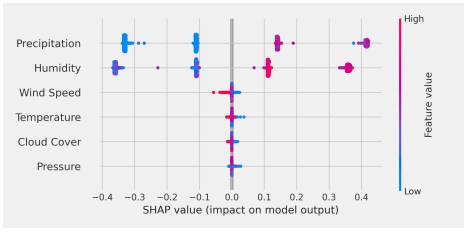


Figure 7. SHAP summary plot showing the direction and magnitude of feature effects across model predictions.

Precipitation and humidity have the greatest influence on the predictions, followed by wind speed and temperature; cloud cover and pressure have smaller contributions overall. The color gradient represents feature values, while the horizontal position shows the direction and magnitude of SHAP impact. This pattern is consistent with the expected meteorological influence of precipitation and atmospheric moisture.

5. CONCLUSION AND FUTURE WORK

This study presented an interpretable machine-learning framework for next-day rainfall classification using daily observations from 20 U.S. cities during 2024–2025. Random Forest produced the highest reported precision, recall, and F1-score, each equal to 1.00. SHAP explanations identified precipitation and humidity as the leading contributors to predictions, making the model's outputs more transparent. RBF SVM and KNN also achieved strong results, while Logistic Regression, Linear SVM, and Naive Bayes showed lower F1-scores.

Accurate forecasts paired with interpretable explanations could support agricultural planning, urban development, emergency response, and water-resource management. Future work may examine hybrid metaheuristic optimization methods such as Genetic Algorithms, Particle Swarm Optimization, and Biogeography-Based Optimization; incorporate streaming observations for real-time updates; and test scalability across larger geographic areas and longer observation periods.

REFERENCES

[1] K. Wudu, A. Abegaz, L. Ayele, and M. Ybabe, "The impacts of climate change on biodiversity loss and its remedial measures using nature based conservation approach: A global perspective," *Biodiversity and Conservation*, vol. 32, no. 12, pp. 3681–3701, 2023.

[2] L. Hirons, C. M. Wainwright, P. Nying'uro, D. Quaye, J. Ashong, C. Kiptum, N. K. Opoku, E. M. Thompson, and B. Lamptey, "Experiences of co-producing subseasonal forecast products for agricultural application in Kenya and Ghana," *Weather*, vol. 78, no. 5, pp. 148–153, 2023.

[3] F. Piadeh, K. Behzadian, and A. M. Alani, "A critical review of real-time modelling of flood forecasting in urban drainage systems," *Journal of Hydrology*, vol. 607, p. 127476, 2022.

[4] M. Kotz, A. Levermann, and L. Wenz, "The effect of rainfall changes on economic production," *Nature*, vol. 601, no. 7892, pp. 223–227, 2022.

[5] M. A. Mohamed, G. S. El Afandi, and M. El-Mahdy, "Impact of climate change on rainfall variability in the blue nile basin," *Alexandria Engineering Journal*, vol. 61, no. 4, pp. 3265–3275, 2022.

[6] D. Bari, T. Bergot, and R. Tardif, "Fog decision support systems: A review of the current perspectives," *Atmosphere*, vol. 14, no. 8, p. 1314, 2023.

[7] P. H. Hrudya, H. Varikoden, and R. Vishnu, "A review on the indian summer monsoon rainfall, variability and its association with enso and iod," *Meteorology and Atmospheric Physics*, vol. 133, pp. 1–14, 2021.

[8] A. Perez Bello, A. Mailhot, D. Paquin, and D. Paquin-Ricard, "Temperature-precipitation scaling rates: A rainfall event-based perspective," *Journal of Geophysical Research: Atmospheres*, vol. 127, no. 22, p. e2022JD037873, 2022.

[9] C. O. de Burgh-Day and T. Leeuwenburg, "Machine learning for numerical weather and climate modelling: A review," *Geoscientific Model Development*, vol. 16, no. 22, pp. 6433–6477, 2023.

[10] A. Yadav and F. Khan, "Machine learning techniques in rainfall forecasting," *Applied Soft Computing*, 2023.

[11] Y. O. Ouma, R. Cheruyot, and A. N. Wachera, "Rainfall and runoff time-series trend analysis using lstm recurrent neural network and wavelet neural network with satellite-based meteorological data: Case study of nzoia hydrologic basin," *Complex & Intelligent Systems*, pp. 1–24, 2021.

[12] E. C. L. de Oliveira, A. V. Nogueira Neto, A. P. Paes dos Santos, C. P. W. da Costa, J. C. G. de Freitas, P. W. M. Souza-Filho, R. de Lima Rocha, R. C. Alves, V. dos Santos Franco, E. C. de Carvalho, and R. G. Tedeschi, "Precipitation forecasting: From geophysical aspects to machine learning applications," *Frontiers in Climate*, vol. 5, p. 1250201, 2023.

[13] C. M. Liyew and H. A. Melese, "Machine learning techniques to predict daily rainfall amount," *Journal of Big Data*, vol. 8, pp. 1–11, 2021.

[14] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. Davila 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, p. 100204, 2022.

[15] O. S. Ojo and S. T. Ogunjo, "Machine learning models for prediction of rainfall over nigeria," *Scientific African*, vol. 16, p. e01246, 2022.

[16] M. Baljon and S. K. Sharma, "Rainfall prediction rate in Saudi Arabia using improved machine learning techniques," *Water*, vol. 15, no. 4, p. 826, 2023.

[17] M. Akhtar, A. S. A. Shatat, S. A. H. Ahamad, S. Dilshad, and F. Samdani, "Optimized cascaded cnn for intelligent rainfall prediction model: A research towards statistic-based machine learning," *Theoretical Issues in Ergonomics Science*, vol. 24, no. 5, pp. 564–592, 2023.

[18] M. F. Allawi, U. H. Abdulhameed, A. Adham, K. N. Sayl, S. O. Sulaiman, M. M. Ramal, M. Sherif, and A. El-Shafie, "Monthly rainfall forecasting modelling based on advanced machine learning methods: Tropical region as case study," *Engineering Applications of Computational Fluid Mechanics*, vol. 17, no. 1, p. 2243090, 2023.

[19] C. Chen, Q. Zhang, M. H. Kashani, C. Jun, S. M. Bateni, S. S. Band, S. S. Dash, and K.-W. Chau, "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.

[20] G. Tuysuzoglu, K. U. Birant, and D. Birant, "Rainfall prediction using an ensemble machine learning model based on k-stars," *Sustainability*, vol. 15, no. 7, p. 5889, 2023.

[21] H. Ahmadi, B. Aminnejad, and H. Sabatsany, "Application of machine learning ensemble models for rainfall prediction," *Acta Geophysica*, vol. 71, no. 4, pp. 1775–1786, 2023.

[22] A. R. Satheesh, P. Knippertz, A. H. Fink, E.-M. Walz, and T. Gneiting, "Sources of predictability of synoptic-scale rainfall during the West African summer monsoon," *Quarterly Journal of the Royal Meteorological Society*, vol. 149, no. 757, pp. 3721–3737, 2023.

[23] 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, p. 492, 2022.

[24] F. Alqahtani, M. Abotaleb, A. A. Subhi, E.-S. M. El-Kenawy, A. A. Abdelhamid, K. Alakkari, A. Badr, H. K. Al-Mahdawi, A. Ibrahim, and A. Kadi, "A hybrid deep learning model for rainfall in the wetlands of southern iraq," *Modeling Earth Systems and Environment*, vol. 9, no. 4, pp. 4295–4312, 2023.

[25] S. Mohamadi, Z. Sheikh Khozani, M. Ehteram, A. N. Ahmed, and A. El-Shafie, "Rainfall prediction using multiple inclusive models and large climate indices," *Environmental Science and Pollution Research*, vol. 29, no. 56, pp. 85 312–85 349, 2022.

[26] K. Johny, M. L. Pai *et al.*, "A multivariate emd-lstm model aided with time dependent intrinsic cross-correlation for monthly rainfall prediction," *Applied Soft Computing*, vol. 123, p. 108941, 2022.

[27] Y. Zhao, X. Meng, T. Qi, Y. Li, G. Chen, D. Yue, and F. Qing, "Ai-based rainfall prediction model for debris flows," *Engineering Geology*, vol. 296, p. 106456, 2022.

[28] B. Roy, M. P. Singh, and A. Singh, "A novel approach for rainfall-runoff modelling using a biogeography-based optimization technique," *International Journal of River Basin Management*, vol. 19, no. 1, pp. 67–80, 2021.

[29] W. Ali, "USA Rainfall Prediction Dataset (2024–2025)," Kaggle dataset, 2025.

[30] S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," *Computational and Mathematical Organization Theory*, vol. 25, no. 3, pp. 319–335, 2019.