ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1891

Precipitation Prediction in Ecuador Using Machine Learning Models

Jaren Acosta[1]; Sebastián Aguirre[1]; George García[1]; Esteban Pérez[1]; Andrea Pilco[1]; Angélica Quito[1]
[1]Universidad Internacional del Ecuador UIDE, Quito 170411, Ecuador,
daacostaol@uide.edu.ec, seaguirreba@uide.edu.ec, geogarciara@uide.edu.ec, esperezpe@uide.edu.ec, anpilcoat@uide.edu.ec, anquitoca@uide.edu.ec

Abstract– ***Precipitation prediction plays a critical role in water resource management, agriculture, and climate risk mitigation, particularly in regions characterized by strong climatic variability, such as Ecuador. This study investigates the application of machine learning techniques to precipitation prediction using a long-term climatic dataset spanning 1950 to 2023. Three regression models were evaluated: Decision Tree Regressor, Random Forest Regressor, and Gradient Boosting Regressor. The dataset was preprocessed through temporal decomposition, logarithmic transformation of precipitation, and cyclical encoding of seasonal effects to capture long-term trends and annual variability. Model performance was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Results show the Gradient Boosting Regressor achieving the best performance (MAE = 27.63 mm, RMSE = 37.29 mm, R^2 = 0.84). Scenario-based predictions and sensitivity analysis further demonstrate the model's physical consistency and practical applicability. The findings confirm the potential of machine learning models, particularly gradient boosting, as reliable tools for precipitation prediction and exploratory climate analysis in Ecuador.***

Keywords-- Precipitation prediction, machine learning, gradient boosting, climate variability, Ecuador.

I. INTRODUCTION

Precipitation is a crucial component of the hydrological cycle and the most important and active variable associated with atmospheric circulation in meteorological and climate studies. Precipitation forecasting is a field of study in water resource management, agriculture, and disaster preparedness, especially where climate variability and natural weather phenomena such as El Niño are present, as is the case in Ecuador. Data are crucial for precipitation prediction; this includes global precipitation data, products based on ground stations, satellite imagery from geostationary and low Earth orbit satellites, sensor data, and reanalysis data [1].

The techniques initially used for precipitation forecasting rely on statistical methods that analyze historical time series of climatic variables, such as precipitation, temperature, and atmospheric pressure. These models seek to identify patterns, trends, cycles, and mathematical relationships in the observed data to extrapolate future behavior. On the other hand, physical models, also known as numerical weather prediction (NWP) models, approach precipitation forecasting from a deterministic perspective, explicitly solving the fundamental equations of atmospheric dynamics [2].

Study [3] used the ARIMA (Autoregressive Integrated Moving Average) model to predict regional temperatures and precipitation over 2 to 20 years, specifically in the United States, using historical temperature and precipitation data from 93 cities. In [4], the Equitable Threat Score (ETS) was used to evaluate the performance of quantitative precipitation forecast (QPF) models, particularly in small areas such as the island of Taiwan. Daily precipitation data were collected from weather stations on the island, and the observed rainfall distribution for August 7 and 8, 2009, is shown on the maps. The study demonstrated that the ETS may not be suitable for evaluating the performance of prediction models when the precipitation area is large relative to the verification area. Another widely used model was the Holt-Winters exponential smoothing, a time-series forecasting method that captures both trends and seasonal factors. It was used for precipitation forecasting in the Abiansemal district of Bali, Indonesia. The main objective was to determine the best model for forecasting precipitation and to facilitate agricultural planning, especially during the most favorable planting seasons [5]. The study used monthly, in [3], proposed a model based on Functional Data Analysis (FDA) to predict precipitation and to calculate irrigation requirements for maize cultivation in Ecuador. The model uses temperature and wind speed as functional covariates to estimate the amount of monthly precipitation, obtaining an R^2 of 0.97 and a prediction error of 0.14. Study [6] used Bayesian networks to analyze climate teleconnections and their impact on precipitation in Ecuador, thereby improving understanding of climate behavior and supporting decision-making regarding the country's water resources.

Statistical and physical models have been widely used. However, the inherent complexity of precipitation formation processes, along with the high spatial and temporal variability of the phenomenon, has motivated the exploration of alternative approaches, such as machine learning models, which seek to complement or overcome the limitations of traditional methods [2]. The current state of machine learning (ML) techniques in precipitation prediction is presented below.

A. *Literature Review*

Precipitation prediction has progressively evolved towards machine learning-based models, driven by the increasing availability of climate datasets and the need to capture the marked non-linearity and spatiotemporal variability inherent in precipitation processes. Study [7] demonstrated that ML, such as Random Forests (RF), can be used for precipitation prediction in complex regions, such as the tropical Andes, especially when combined with global climate indices, regional atmospheric variables, and precipitation temporal memory. The work [8] proposed a Particle Swarm Optimization (PSO) and Support Vector Machine (SVM) model for precipitation prediction, addressing one of the main problems of classical SVM models: the optimal selection of hyperparameters. The data came from a ground-based weather station in Nanjing. The input variables were atmospheric pressure, sea level pressure, wind direction and speed, air temperature, and relative humidity. The target variable was classified as rain/no rain, achieving an accuracy close to 95%.

Research has shifted towards artificial neural networks (ANNs) and deep learning architectures. Study [9] focused on how neural networks can be an alternative for long-term precipitation prediction in California, overcoming the limitations of traditional statistical models in many cases and providing better predictive capacity for intense precipitation events. Similarly, [10] addressed the use of deep neural networks for short-term precipitation prediction in the continental United States. The proposed model, called FPP (Rapid Precipitation Prediction), used a 3D convolutional neural network to predict the spatial distribution of daily precipitation, based on meteorological fields such as temperature, geopotential height, relative humidity, and horizontal winds, and is trained using daily precipitation data from ERA-Interim (meteorological reanalysis) and CPC (Climate Prediction Center) between 1980 and 2018. The work [2] proposed MetNet-2, a deep neural network model capable of predicting precipitation up to 12 hours in advance without resorting to explicit physical simulations. The model was trained using weather radar data, satellite imagery, and atmospheric variables from the United States. Study [11] presented a UTrans-Net neural network model for short-term precipitation prediction. This model integrated a U-Net neural network architecture with a transformer mechanism to enhance feature extraction from multiscale meteorological data, aiming to predict future precipitation. The model predicted future precipitation from historical data using a sequential autoregressive approach. Experiments were conducted using a dataset from the China Meteorological Administration. The prediction accuracy results show that, at thresholds of 0.1 mm, 3 mm, and 100 mm, the UTrans-Net model achieved average accuracy of 59%, 78%, and 82%, respectively. In Ecuador, a study [12] predicted the precipitation phase (not the amount, but whether it falls as rain, snow, or a mixture) and sought to understand which meteorological variables control this process—a critical aspect for water resource management, glacial balance, and risk prevention in regions dependent on glacial melt, such as Quito. The work addresses the limitations of classical logistic models, which typically use only air temperature (and, in some cases, humidity), and proposes using machine learning models to capture the nonlinear complexity of the atmospheric processes that determine precipitation phase. The study employed high-resolution observational data collected at the PRAA-Morrena station, located approximately 4,725 meters above sea level on the Antisana Glacier. The artificial intelligence models (RF and ANN) clearly outperformed traditional logistic regression models across 8 of 10 evaluation metrics. Random Forest was the model with the best overall performance, especially in accurately detecting rain and snow. Recent advances in deep learning for weather forecasting have shifted towards more architectures capable of modeling complex spatiotemporal dependencies. Transformer-based models have emerged as an alternative to traditional convolutional and recurrent approaches. Recent studies have explored the application of Transformer-based deep learning architectures to precipitation forecasting [13]. For instance, study [14] conducted a systematic comparison of two Transformer-based models Earthformer and LLMDiff for short-term precipitation nowcasting using the SEVIR radar-echo dataset, which includes over 20,000 storm event samples at 1 km spatial resolution across the contiguous United States.

Although significant progress has been made in precipitation forecasting using statistical, physical, and machine learning approaches, important gaps remain in the context of long-term prediction under high climatic variability. Many existing models either rely on complex architectures and high computational costs or focus primarily on predictive accuracy without explicitly addressing seasonal representation and interpretability. In this context, this work addresses these limitations by proposing a machine learning framework that integrates long-term climatic data with explicit seasonal encoding and scenario-based analysis, aiming to improve both predictive performance and practical applicability. The main contributions of this study are summarized below.

B. *Contributions*

The main contributions of this study are summarized as follows:

1) A comprehensive machine learning framework for monthly precipitation prediction in Ecuador is proposed, integrating long-term climatic data spanning more than seven decades (1950–2023).

2) A preprocessing strategy combining temporal decomposition, logarithmic transformation of precipitation, and cyclical encoding of seasonal effects is implemented to effectively capture both long-term trends and annual variability.

3) A comparative evaluation of Decision Tree, Random Forest, and Gradient Boosting regressors is conducted.

4) Scenario-based predictions and sensitivity analysis are introduced to assess the interpretability, physical consistency, and practical applicability.

This work is structured as follows: Section II describes the proposed methodology, Section III presents the experimental results and evaluation metrics, Section IV discusses the findings and model applicability, and Section V summarizes the main conclusions of the study.

II. METHODOLOGY

This section presents the machine learning framework developed for monthly precipitation prediction. The methodology is illustrated in Fig. 1 and described in detail below.

A. Dataset

The study utilized open data provided by [15], consists of monthly climatic records for Ecuador covering the period from 1950 to 2023, resulting in a total of 888 observations. The data were downloaded on November 04, 2025. The dataset includes variables shown in Table I, including air temperature (tasmax, °C), relative humidity (h, %), and accumulated monthly precipitation (pr, mm). In this study, precipitation (pr) was considered the target variable.

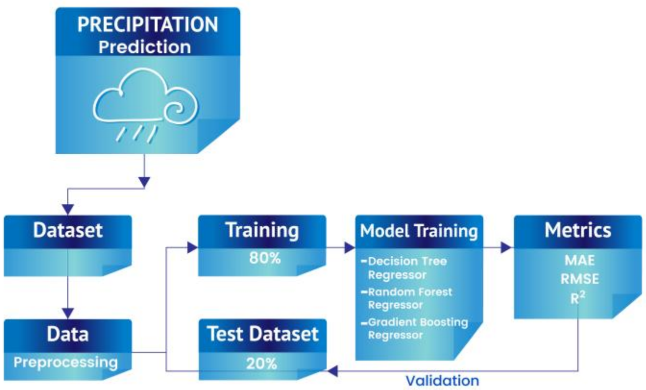

Fig. 1 Flowchart of the Monthly Precipitation Prediction Framework.

TABLE I
VARIABLES DESCRIPTION

Variable	Description	Unit	Data Type
Date	Observation date (year–month)	-	Temporal
tasmax	Monthly mean maximum air temperature	°C	Continuous
hurs	Monthly mean relative humidity	%	Continuous
pr	Monthly accumulated precipitation	mm	Continuous

B. Data preprocessing

The dataset presents a uniform temporal resolution and does not contain missing values. several preprocessing steps were applied to prepare the dataset and improve the learning capability.

1) Temporal Decomposition: Let t denote the original temporal variable expressed as year–month. The date variable was decomposed into its year and month components:

$$Year_t = year(t) \tag{1}$$

$$Month_t = month(t), Month_t \in \{1,2,....,12\} \tag{2}$$

This decomposition allows the models to separately capture long-term temporal trends (year) and seasonal variability (month), which are fundamental characteristics of precipitation time series.

2) Logarithmic Transformation of Precipitation: Precipitation data typically exhibit a positively skewed distribution with occasional extreme values. To reduce skewness and stabilize variance, a logarithmic transformation was applied to the precipitation variable P_t:

$$P_t^{\log} = \log(P_t + 1) \tag{3}$$

The constant term $+1$ avoids numerical issues when precipitation values are equal to zero. This transformation improves model convergence and reduces the influence of extreme rainfall events during training.

3) Cyclical Encoding of Seasonal Effects: Since the month variable is inherently cyclical, a direct numerical encoding may introduce artificial discontinuities (e.g., between December and January). To preserve the cyclical nature of seasonality, the month variable was transformed using sine and cosine functions:

$$Month_{\sin,t} = \sin\left(\frac{2\pi \times Month_t}{12}\right) \tag{4}$$

$$Month_{\cos,t} = \cos\left(\frac{2\pi \times Month_t}{12}\right) \tag{5}$$

This encoding ensures continuity between consecutive months and allows the models to effectively learn seasonal precipitation patterns.

4) Feature Set Construction: After preprocessing, the input feature vector for each observation t is defined as:

$$X_t = \left[Year_t, tas\max_t, hurs_t, Month_{\sin,t}, Month_{\cos,t} \right] \tag{6}$$

The target variable is defined as either the original precipitation P_t or its logarithmic transformation $P_t^{\log}$, depending on the modeling strategy adopted.

C. Machine learning modeling

This research focuses on applying machine learning techniques to predict monthly accumulated precipitation (pr) in millimeters (mm) using a long-term climate dataset for Ecuador covering the period from 1950 to 2023. The variable pr is defined as the target variable, while meteorological and temporal characteristics derived from the dataset are used as predictors. The dataset was divided into two subsets: a training set, used to fit the model to the data, and a test set, used to evaluate the model's ability to generalize to unseen data. Specifically, 80% of the dataset, comprising 710 data points, was allocated for training, while the remaining 20%, consisting of 178 data points, was reserved for testing. In this study, three predictive models were implemented: Random Forest Regressor (RFR), Decision Tree Regressor (DTR) and a Gradient Boosting Regressor (GBR). All models were implemented and executed using the scikit-learn (sklearn) library.

1) Decision Tree Regressor: A Decision Tree Regressor is composed of a hierarchical structure that includes a root node, internal decision nodes, branches, and leaf nodes. Each internal node represents a decision rule based on a threshold value of a predictor variable, while each leaf node stores a numerical prediction corresponding to the average value of the target variable within that region. This structure allows the model to approximate nonlinear relationships through a sequence of simple decision rules [16]. The construction of a Decision Tree Regressor follows a divide-and-conquer strategy. At each node, the algorithm selects the optimal feature and threshold that best split the data into two subsets. For regression tasks, this selection is commonly based on minimizing an impurity measure such as the Mean Squared Error (MSE). Formally, for a candidate split, the algorithm minimizes the weighted error of the left and right child nodes [17].

2) Random Forest Regressor: This is a supervised, ensemble-based machine learning algorithm widely used in regression tasks involving complex, nonlinear relationships between predictors and a continuous target variable [18]. The method relies on aggregating multiple decision trees, each trained on different subsets of data and features.

Formally, given an input feature vector x, the RFR prediction is calculated as the average of the predictions from all trees in the forest:

$$\hat{y}(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{7}$$

Where $T_i(x)$ denotes the prediction of the $i-th$ regression tree and N is the total number of trees in the forest.

2) Gradient Boosting Regressor: This is a supervised learning ensemble method based on decision trees, designed to address regression problems by sequentially combining multiple weak learners. Unlike bagging-based approaches that train models independently, GBR constructs trees iteratively, where each new tree is trained to correct the residual errors of the previous ensemble. This process is driven by the optimization of a differentiable loss function, where the negative gradients of the loss function are used to guide each subsequent model toward minimizing the overall prediction error [17].

D. *Training*

A testing environment was established after defining the datasets, and the training process was conducted using Google Colab. A detailed description of the training environment is provided in Table II.

TABLE II
TRAINING ENVIRONMENT

Environment	Component	Capacity
Google Colab	RAM	12.67 GB
	GPU	T4 GPU 15 GB VRAM

The hyperparameters of the Decision Tree Regressor, Random Forest Regressor and Gradient Boosting Regressor applied during the training phase are detailed in Table III. They were selected based on the non-linear behavior, seasonal variability, and the presence of extreme values inherent in the monthly precipitation data.

TABLE III
TRAINING PARAMETERS

Model	Parameter	Value
Decision Tree Regressor	Maximum Depth	10
	Minimum samples per leaf	3
	Minimum samples split	5
	Random state	42
Random Forest Regressor	Number of estimators	185
	Maximum Depth	None
	Random state	42
Gradient Boosting Regressor	Loss	Squared error
	Number of estimators	300
	Subsample	0.8
	Minimum samples split	5
	Minimum samples leaf	1
	Maximum Depth	3
	Random state	42

III. RESULTS

In this section, descriptive statistics of the dataset are presented to provide a comprehensive understanding of its distribution and characteristics. Additionally, the performance of the models utilized was evaluated using metrics to ensure a reliable assessment. ML modeling accuracy was assessed using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide complementary perspectives on prediction accuracy, error magnitude, and the proportion of variance explained by the model [19].

A. *Descriptive statistics*

Table IV summarizes the descriptive statistics of the variables used in the analysis. Precipitation (pr) exhibits substantial variability, reflecting the heterogeneous climatic conditions of the study region. The distribution shows a clear difference between minimum and maximum values, indicating the presence of both dry and wet periods. Temperature and relative humidity display comparatively lower variability, consistent with their more stable climatological behavior.

TABLE IV
DESCRIPTIVE STATISTICS OF THE CLIMATIC VARIABLES

Variable	Mean	Std. Dev.	Min	25%	Median
tasmax	24.45	0.99	21.31	23.77	24.53
Hurs	85.00	2.2	74.82	83.53	85.33
Pr	253.64	82.29	84.41	190.26	236.24

Fig. 2 illustrates the time series of monthly precipitation from 1950 to 2023. The horizontal axis represents the observation year, while the vertical axis corresponds to monthly precipitation values. The time series shows pronounced temporal variability, characterized by alternating wet and dry periods and occasional peaks associated with extreme precipitation events. This behavior highlights the non-linear and highly variable nature.

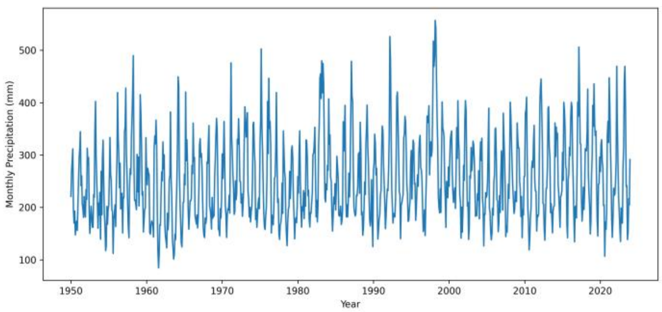

Fig. 2 Monthly Precipitation Time Series.

The histogram of monthly precipitation values is shown in Fig. 3. The distribution is positively skewed, with a higher frequency of moderate precipitation values and a long tail corresponding to extreme rainfall events. This skewed behavior justifies applying a logarithmic transformation during preprocessing.

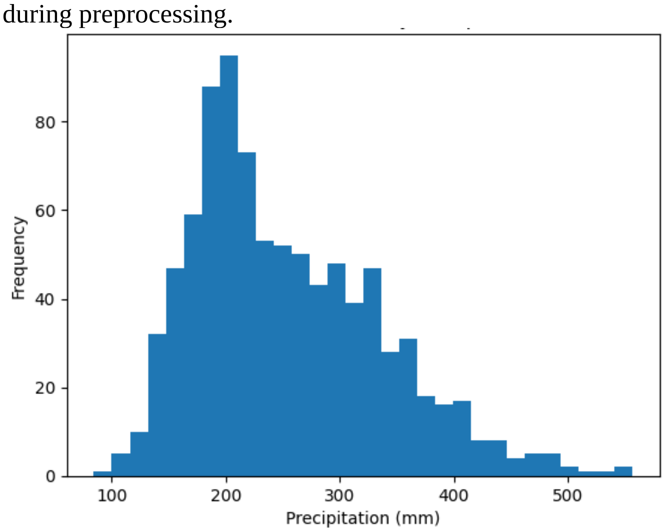

Fig. 3 Distribution of Monthly Precipitation.

Fig. 4 presents boxplots of the climatic variables. Precipitation shows a wider interquartile range and several high-end outliers, highlighting the presence of extreme rainfall events. In contrast, tasmax and hurs exhibit narrower distributions, indicating relatively stable behavior across the study period.

B. Decision tree regressor performance

The Decision Tree Regressor achieved an MAE of 33.649 mm, an RMSE of 46.934 mm, and an R^2 value of 0.747, indicating that the model can explain approximately 74.7% of the variance in monthly precipitation.

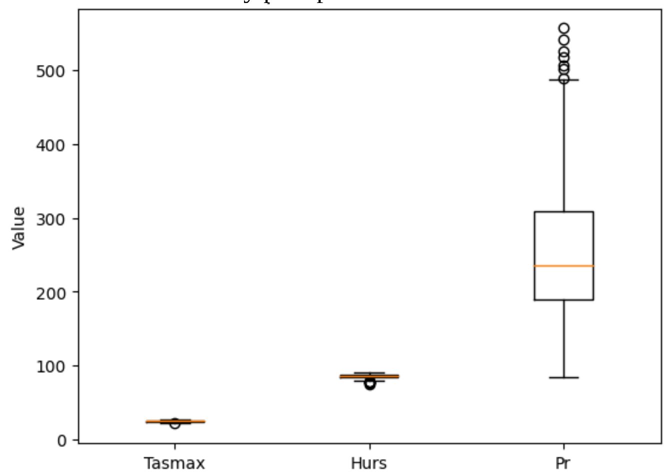

Fig. 4 Boxplot of Climatic Variables.

While the model captures the general precipitation patterns, the relatively higher RMSE suggests sensitivity to extreme precipitation events, which is a known limitation of single-tree models when dealing with highly variable climatic data.

C. Random forest regressor performance

The Random Forest model achieved an MAE of 30.8 mm, an RMSE of 42.7 mm, and an R^2 value of 0.79, demonstrating an enhanced ability to model the nonlinear relationships governing precipitation dynamics.

D. Gradient boosting performance

The Gradient Boosting Regressor (GBR) achieved a MAE of 27.633 mm, a RMSE of 37.289 mm, and an R^2 value of 0.841. These results indicate that the model explains approximately 84.1% of the variability in monthly precipitation.

E. Models comparison

Table V summarizes the performance of the evaluated machine learning models. The comparative analysis indicates a clear performance improvement when moving from a single decision tree to ensemble-based approaches. While the Random Forest Regressor outperforms the Decision Tree Regressor by reducing prediction errors and increasing explained variance, the Gradient Boosting Regressor achieves the best overall performance across all evaluation metrics. GBR exhibits the lowest MAE and RMSE values and the highest R^2, demonstrating superior accuracy and robustness in modeling monthly precipitation under the high climatic variability observed in Ecuador.

IV. Discussion

To complement the regression analysis and performance metrics, scenario-based precipitation predictions were conducted to evaluate the practical applicability of the proposed Gradient Boosting Regressor. Table VI summarizes the climatic and temporal input variables used for each prediction scenario, along with the corresponding expected precipitation ranges derived from climatological behavior and the model-predicted precipitation values.

TABLE V
Performance Comparisons of Machine Learning Models

Model	MAE (mm)	RMSE (mm)	R^2
Decision Tree Regressor	33.649	46.934	0.747
Random Forest Regressor	30.8	42.7	0.79
Gradient boosting	27.63	37.28	0.84

TABLE VI
Performance Comparisons of Machine Learning Models

Test	Tasmax (°C)	Hurs (%)	Year	Month	Expected pr (mm)	Predicted pr (mm)
1	25.0	80.0	2024	February	180–220	200.71
2	23.5	85.0	2024	March	240–300	274.70
3	26.0	78.0	2025	October	100–150	131.26

The input variables in Table VI represent realistic atmospheric and seasonal conditions within the observed data distribution. The expected precipitation values serve as climatological benchmarks based on known seasonal behavior in Ecuador, while the predicted values are quantitative outputs generated by the trained Gradient Boosting Regressor.

The close correspondence between expected ranges and predicted values indicates that the model produces physically meaningful and climatologically coherent precipitation estimates. Higher humidity and wet-season months (e.g., March) lead to increased predicted precipitation, whereas drier seasonal conditions (e.g., October) result in lower precipitation estimates. While the presence of extreme values emphasizes the importance of using machine learning models capable of capturing complex patterns and variability in precipitation dynamics.

To further illustrate the predictive behavior and physical consistency of the proposed model, a sensitivity analysis was conducted by varying relative humidity while keeping the remaining input variables fixed at representative values. As shown in Fig. 5, the predicted precipitation increases monotonically with relative humidity, which is consistent with established atmospheric processes governing rainfall formation. This behavior indicates that the Gradient Boosting Regressor captures meaningful relationships between key meteorological drivers and precipitation, rather than relying solely on statistical fitting. The smooth and coherent response of the model reinforces its interpretability and supports its applicability for exploratory climate analysis and scenario-based decision support.

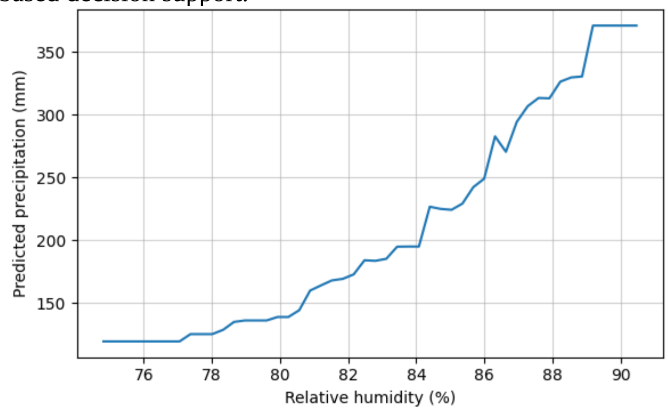

Fig. 5 Model Sensitivity to Relative Humidity using GBR Model.

From a comparative perspective, the proposed models outperform simple baselines, including climatological averages, demonstrating their ability to capture intra-annual variability beyond static seasonal expectations. However, when compared with previously reported models in Ecuador, such as the Functional Data Analysis (FDA) approach in [3], which achieved an R^2 of 0.97, it is important to consider differences in temporal resolution. The FDA model operated at a monthly aggregated level with smoother variability, whereas the present study addresses a more complex prediction task with higher variability.

Despite these results, several limitations must be acknowledged. The dataset is limited in temporal extent and corresponds to a specific high-altitude location, which may restrict generalization. Consequently, the climatological baseline was approximated from the available data rather than derived from a standard 30-year reference period.

V. Conclusions

This study presented a machine learning-based approach for precipitation prediction in Ecuador using long-term climatic data and seasonal feature encoding. Among the evaluated models, the Gradient Boosting Regressor achieved the best overall performance, explaining approximately 84% of the variability in monthly precipitation while reducing prediction errors compared to the Decision Tree and Random Forest models. Regression diagnostics, scenario-based predictions, and sensitivity analysis confirmed that the model not only achieves high predictive accuracy but also exhibits physically consistent behavior with respect to key meteorological drivers such as relative humidity and seasonal cycles.

The proposed approach provides an effective balance among accuracy, interpretability, and computational efficiency, making it suitable for exploratory forecasting and decision-support applications in water resource management, agriculture, and climate variability assessment. Future work may extend this framework by incorporating additional atmospheric variables, spatial information, or climate indices to further enhance predictive capability under changing climatic conditions.

References

[1] Q. Sun, C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, "A review of global precipitation data sets: Data sources, estimation, and intercomparisons," *Reviews of geophysics*, vol. 56, no. 1, pp. 79-107, December 2018.

[2] L. Espeholt, S. Agrawal, C. Sønderby, et al., "Deep learning for twelve hour precipitation forecasts," *Nature Communications*, vol. 13, no. 5145, pp. 1-10, September 2022.

[3] Y. Lai and D. A. Dzombak, "Use of the Autoregressive Integrated Moving Average (ARIMA) model to forecast near-term regional temperature and precipitation," *Wea. Forecasting*, vol. 35, pp. 959-976, April 2020.

[4] C. Wang, "On the calculation and correction of equitable threat score for model quantitative precipitation forecasts for small verification areas: The example of Taiwan," *Wea. Forecasting*, vol. 29, pp. 788-798, August 2014.

[5] E. Hendri and S. Fadhlia, "Time series data analysis: The Holt–Winters model for rainfall prediction in West Java," *International Journal of Applied Mathematics, Sciences, and Technology for National Defense*, vol. 2, pp. 1–8, March 2024.

[6] R. Ávila and D. Ballari, "A Bayesian network approach to identify climate teleconnections within homogeneous precipitation regions in Ecuador," in Information and Communication Technologies of Ecuador (TIC.EC), vol. 1099, Eds. Cham, Switzerland: Springer, 2020, pp. 21-35.

[7] M. Montenegro, R. Célleri, J. Orellana-Alvear, *et al.*, "Precipitation forecasting using random forest over an Ecuadorian Andes basin," *Meteorol. Atmos. Phys.*, vol. 137, no. 5, December 2024.

[8] J. Du, Y. Liu, Y. Yu, and W. Yan, "A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO–SVM) algorithms," *Algorithms 2017*, vol. 10, no. 57, pp. 1-15, April 2017.

[9] D. Silverman and J. A. Dracup, "Artificial neural networks and long-range precipitation prediction in California," *Journal of Applied Meteorology and Climatology*, vol. 39, pp. 57–66, January 2000.

[10] G. Chen and W.-C. Wang, "Short-term precipitation prediction for the contiguous United States using deep learning," *Geophysical Research Letters*, vol. 49, April 2022.

[11] H. Cao, Y. Wu, Y. Bao, X. Feng, S. Wan, and C. Qian, "UTrans-Net: A model for short-term precipitation prediction*," Artificial Intelligence and Applications*, vol. 1, April 2023.

[12] L. Campozano, L. Robaina, L. F. Gualco, L. Maisincho, M. Villacís, T. Condom, D. Ballari, and C. Páez, "Parsimonious models of precipitation phase derived from random forest knowledge: Intercomparing logistic models, neural networks, and random forest models," *Water 2021*, vol. 13, no. 21, pp. 1-23, October 2021.

[13] L. Su, X. Zuo, R. Li, X. Wang, H. Zhao, and B. Huang, "A systematic review for transformer-based long-term series forecasting," *Artificial Intelligence Review*, vol. 58, no. 3, pp. 1-29, January 2025.

[14] C. Lu and Q. Pan, "Application and comparison of two transformer-based deep learning models in short-term precipitation nowcasting," *Water*, vol. 18, p. 757, March 2026.

[15] "Climate Knowledge Portal – Download Data," World Bank. [Online]. Available: https://climateknowledgeportal.worldbank.org/download-data (Accessed: Nov. 04, 2025).

[16] J. Kushwah, A. Kumar, S. Patel, R. Soni, A. Gawande, and S. Gupta, "Comparative study of regressor and classifier with decision tree using modern tools," *Materials Today: Proceedings*, vol. 56, pp. 3571–3576, April 2022.

[17] A. Gupta, A. Bansal, and K. Roy, "Solar energy prediction using decision tree regressor," In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 489–495, May 2021.

[18] Z. El Mrabet, N. Sugunaraj, P. Ranganathan, and S. Abhyankar, "Random forest regressor-based approach for detecting fault location and duration in power systems," *Sensors*, vol. 22, no. 458, pp. 1-19, January 2022.

[19] X. Li, W. Li, and Y. Xu, "Human age prediction based on DNA methylation using a gradient boosting regressor," *Genes (Basel)*, vol. 9, no. 9, pp. 1-15, August 2018.