Enhanced rainfall prediction in Gujarat, India using advanced machine learning models

Shaikh A.F.[1]*, Kharat P.V.[2], Pujari A.B.[3], Gunaware P.D.[4] and Darade M.M.[5]

1. TSSM's Padmabhushan Vasantdada Patil Institute of Technology, Bavdhan, Pune, Maharashtra, 411021, INDIA
2. Ajeenkya DY Patil School of Engineering Lohegaon, Pune, Maharashtra, 412105, INDIA
3. Kjei's KJ College of Engineering and Research, Pune, Maharashtra, 411048, INDIA
4. HSBPVT's GOI Faculty of Engineering Kashti, Shrigonda, Kashti, Maharashtra, 414701, INDIA
5. Anantrao Pawar College of Engineering and Research, Pune, Maharashtra, 411009, INDIA

*ayub24686@gmail.com

Abstract

Precise prediction of precipitation is essential for efficient management of water resources, planning of agriculture and readiness for disasters, particularly in areas like Gujarat, India, where climate fluctuations are common. This study uses cutting-edge machine learning methods such as XGBoost and CatBoost, to improve rainfall forecasts made from historical rainfall data. Important metrics including R-squared (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used to compare and to assess the performance of these models.

In training, testing and validation datasets, CatBoost consistently outperforms XGBoost in terms of prediction accuracy, as evidenced by greater R^2 values and lower RMSE and MAE values. These results imply that CatBoost is a better option for rainfall prediction jobs as it is more adept at identifying patterns and trends in the rainfall data. The study's outcomes have significant implications for Gujarat's ability to predict rainfall accurately. Improved predictions can aid in better planning for water storage and distribution, optimize agricultural schedules and enhance flood management strategies.

Keywords: Rainfall prediction, CatBoost, XGBoost, Water resource management.

Introduction

Rainfall prediction is an essential part of meteorology that affects daily living, agriculture, water resource management and readiness for disasters[11]. Planning and reducing the negative consequences of droughts or excessive rain require accurate forecasts[6]. For example, urban planners need exact forecasts to control flood hazards, while farmers depend on accurate rainfall projections to arrange their planting and harvesting schedules.

However, predicting rainfall is notoriously difficult due to the chaotic nature of weather systems[5]. The complexity arises from numerous interacting variables including temperature, humidity, wind patterns and topographical features. Traditional methods often fall short in capturing this complexity, leading to less accurate predictions[1]. In the past, rainfall forecasting was done in conjunction with statistical techniques and numerical weather prediction (NWP) models. Regression analysis is a statistical technique that uses past data to find correlations between different meteorological variables[13]. These methods are relatively straightforward but can oversimplify the complex interactions within weather systems. NWP models, on the other hand, use physical equations to simulate the atmosphere[14]. These models divide the atmosphere into a grid and solve equations of motion, thermodynamics and other relevant physical processes to predict future weather conditions[12,17]. Despite advances in computational power and model sophistication, NWP models still face significant challenges. They require extensive computational resources, are sensitive to initial conditions and often struggle with fine-scale weather features such as localized heavy rainfall.

Machine learning (ML) offers a novel approach to rainfall prediction by using large datasets to identify trends and anticipate results[9]. In contrast to conventional statistical methods, machine learning algorithms are capable of handling non-linear correlations and interactions between variables. This skill is very useful in meteorology as complicated and non-linear dynamics are frequently involved in weather occurrences. As ML models are exposed to new data, they may continually increase their accuracy and learn to make better predictions over time[15]. Furthermore, ML techniques may use a range of data sources including radar data, satellite images and ground-based observations, to generate more comprehensive and accurate models[18].

Recently, there has been a lot of interest in using machine learning to anticipate rainfall. Big data processing allows machine learning (ML) models to find tiny correlations and patterns that may go unnoticed by traditional approaches[4,16]. For instance, deep learning techniques, a subset of ML, can analyze satellite images to identify cloud formations indicative of rainfall[8]. These models can also incorporate temporal data to understand how weather patterns evolve over time. Several studies have demonstrated how machine learning (ML) might improve rainfall forecasting accuracy[2].

For instance, different studies have demonstrated differing degrees of success when using neural networks and support vector machines to forecast rainfall[3,10].

Despite the conspicuous absence of comprehensive studies comparing different machine learning techniques, there is growing interest in using ML to precipitation forecasting[7]. Much of the existing research focuses on individual models, often lacking a systematic evaluation of their comparative strengths and weaknesses. The variety and complexity of weather systems may also not be adequately captured by many studies because they employ small or restricted datasets. By thoroughly contrasting CatBoost with XGBoost for rainfall prediction, this work seeks to close this disparity. Using a big and diverse data set, the aim of this research is to identify the optimal approach for various scenarios and provide insights into how these models work in various conditions.

CatBoost and XGBoost are members of the gradient boosting family of algorithms. Gradient boosting is an ensemble approach that builds a powerful predictive model by integrating the predictions of several weak learners, generally decision trees. Numerous fields, including finance, healthcare and meteorology, have effectively used this approach. It performs particularly well with structured data. CatBoost is made to effectively handle categorical data. To prevent overfitting, ordered boosting is used and a variety of strategies are used to increase model speed and accuracy. CatBoost's primary benefit is its capacity to handle categorical features without requiring a lot of preprocessing which makes it a reliable option for datasets including both numerical and categorical variables.

XGBoost, is a potent technique that is frequently utilized in both practical applications and competitive predictive modeling. Performance, scalability and adaptability are well-known attributes of XGBoost. It includes several advanced features such as regularization, which prevents overfitting and a sparsity-aware algorithm for handling missing values efficiently. XGBoost can handle big datasets with many dimensions, it is a good fit for difficult jobs like rainfall forecasting.

Objective of the study

This study's main goal is to improve Gujarat's rainfall forecast accuracy by utilizing cutting-edge machine learning methods. Specifically, the study uses important performance indicators including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R^2) to assess and compares the effectiveness of CatBoost and XGBoost models in rainfall prediction based on historical rainfall data.

To create a strong predictive model that local government agencies, agricultural planners and disaster management teams may use, the study aims to determine which of the two models, CatBoost or XGBoost, produces more accurate and dependable forecasts. The region's planning for agriculture, disaster preparedness and water resource management are all intended to be enhanced by this approach.

Study area and data collection

Gujarat is the sixth-largest State in terms of area in India; it is situated on the western coast and spans over 196,024 square kilometers. Bound by the Arabian Sea to the west, Madhya Pradesh to the east, Rajasthan to the north and Maharashtra to the south, Gujarat features a diverse range of landscapes, including coastal regions, arid deserts, fertile plains and hilly terrains. Its strategic location along the Arabian Sea makes it a crucial hub for trade and commerce, with significant ports like Kandla and Mundra. Gujarat's diverse geography has an impact on the State's variable climate. There are three primary seasons in the climate which are summer, monsoon and winter. Summers, which run from March to June, are usually hot and dry, with some locations seeing temperatures as high as 40°C (104°F).

The southwest monsoon winds, which bring significant rainfall, are what define the monsoon season, which runs from June to September. The moderate and dry winters that span November through February have temperatures between 10°C and 25°C (50°F and 77°F). The rainfall distribution in Gujarat is highly uneven, with the southern and eastern regions receiving more rainfall compared to the arid northwestern parts. The Saurashtra and Kutch regions are particularly prone to droughts, while the areas around Surat and Valsad in the south receive substantial rainfall. Agriculture, emergency preparedness and the management of water resources are all severely hampered by this rainfall fluctuation. The research region is shown in fig. 1.

Importance of rainfall prediction in Gujarat: Rainfall prediction is of paramount importance in Gujarat due to the State's economic dependence on agriculture, which employs a significant portion of its population. Major crops such as cotton, ground nuts, millet and sorghum are heavily reliant on the monsoon rains. By assisting farmers in making well-informed decisions regarding planting, irrigation and harvesting, accurate rainfall forecasts can boost agricultural output and lower the likelihood of crop failure. In addition to agriculture, rainfall prediction is critical for managing Gujarat's water resources. The State has several major rivers including the Narmada, Tapi and Sabarmati, which are vital for irrigation, drinking water and industrial use.

Effective management of these water resources requires precise information about rainfall patterns and volumes to ensure adequate water supply and prevent water scarcity. Gujarat is also susceptible to severe weather phenomena like floods and cyclones. The State's long coastline makes it susceptible to tropical cyclones, which can bring heavy rainfall and cause widespread damage. Rainfall predictions that are precise and timely are crucial for disaster planning and response because they allow authorities to take early action against flooding, evacuate populations that are at risk and give early warnings.

Challenges in rainfall prediction for Gujarat: Rainfall prediction in Gujarat presents several challenges due to the State's diverse geography and climatic variability. It is challenging to create a forecast model that works for all scenarios due to the incredibly unequal rainfall distribution.

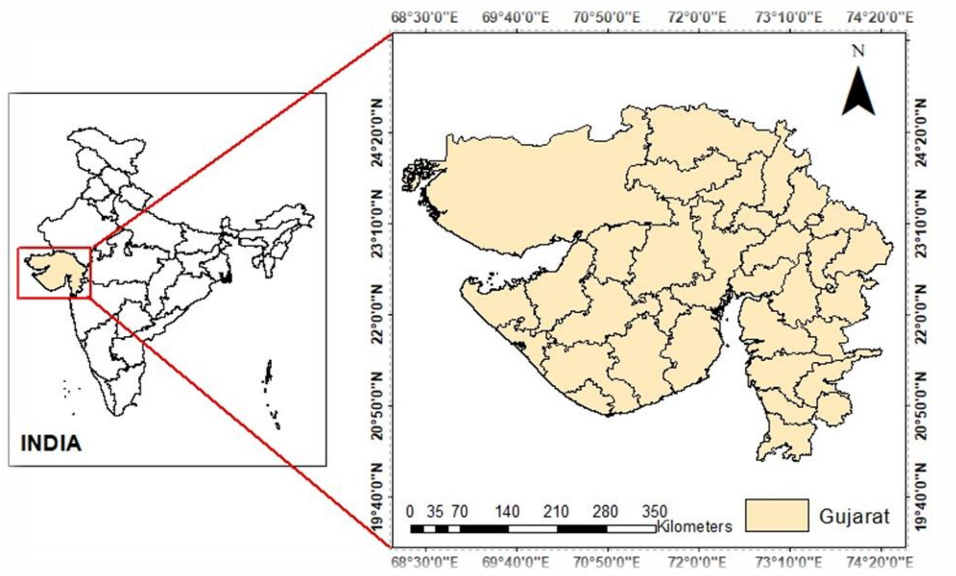

Figure 1: Study area

The coastal areas may receive heavy rainfall from cyclones while the interior regions might experience drought conditions simultaneously. Complex models that can take into consideration regional variances and produce precise projections for various parts of the state, are needed due to this geographical heterogeneity.

A further obstacle is the scarcity of high-resolution meteorological data. Many parts of Gujarat, particularly the remote and rural areas, lack adequate weather monitoring infrastructure. The lack of data might make it difficult to create and validate reliable prediction models. Predicting rainfall with great precision is further complicated by the intricate interactions of several meteorological elements including temperature, humidity, wind patterns and terrain.

Data collection and preprocessing: The Ministry of Jal Shakti of the Indian Government oversees a large database called WRIS (India's Water Resources Information System), from which this data was gathered. The dataset will include daily rainfall records from multiple locations within the State. The collected daily rainfall data from WRIS covers an extensive time span of four decades, precisely ranging from 1980 to 2021. The first step in ensuring the quality and appropriateness of the data for machine learning models is preprocessing, which includes handling missing values, eliminating outliers and normalizing the data to maintain consistency. The data will then be divided into training and validation sets in order to assess how well the CatBoost and XGBoost models perform.

Material and Methods

The technique for this work is meant to construct reliable rainfall forecast models utilizing historical data. The method begins with data collecting to capture the State's geographical rainfall variability. Next, the dataset is divided into testing, validation and training sets, with 15% going toward validation, 70% going toward training and 15% set aside for testing. Two advanced machine learning algorithms, CatBoost and XGBoost, are selected for their efficiency in handling complex datasets. CatBoost excels in managing categorical data and reducing overfitting with ordered boosting while XGBoost offers scalability and regularization to prevent overfitting.

Feature engineering is used in the training process to generate more predictive features and model training is used to identify patterns in rainfall. Model validation is conducted using evaluation metrics. After training, a comparative analysis of CatBoost and XGBoost models is performed to evaluate their effectiveness across different regions in Gujarat. This study analyzes each model's capabilities and determines the best suited algorithm for given circumstances.

CatBoost: CatBoost is a gradient boosting algorithm that minimizes overfitting and effectively manages categorical data. The steps and essential formulas for CatBoost are as follows:

Step 1: Set the model's initial value to a constant, typically the average of the desired values.

$$F_0(x) = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (1)$$

where $F_0(x)$ is the initial model prediction (usually the mean of target values). N represents the number of observations y_i is the true target value for the ith observation.

Step 2: CatBoost uses ordered boosting, which trains models in a way that reduces prediction shift and overfitting by training on ordered subsets of data.

Step 3: Converts categorical features into numerical values using one-hot encoding which converts categories into binary columns. Target statistics replaces categorical values with the mean target value for that category.

Step 4: Determine the loss function's gradient in relation to the current model.

$$g_i^{(m)} = \frac{\partial \mathrm{L}(y_i,\ F_{m-1}(x_i))}{\partial \mathrm{F}_{m-1}(x_i)} \quad (2)$$

where $g_i^{(m)}$ is the gradient for the ith observation in the mth iteration. $\mathrm{L}(y_i,\ F_{m-1}(x_i))$ is the loss function and $\mathrm{F}_{m-1}(x_i)$ is the prediction from the previous iteration.

Step 5: Fit a decision tree to the negative gradient values.

$$h_m(x) = Decision\,Tree(x, -g^{(m)}) \quad (3)$$

where $h_m(x)$ is the new tree fitted to the negative gradient.

Step 6: Update the model with the new tree.

$$F_x(x) = F_{m-1}(x) + \eta * h_m(x) \quad (4)$$

where η is the learning rate and $F_x(x)$ is the updated model after the mth iteration.

Step 7: Repeat the previous six stages until convergence is reached, or for a predetermined number of iterations.

XGBoost: XGBoost is a scalable and flexible gradient boosting algorithm that includes regularization to prevent overfitting. Here are the steps and key formulae involved in XGBoost.

Step 1: Set the model's initial value to a constant, often the average of the desired values.

$$F_0(x) = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (5)$$

where $F_0(x)$ is the initial model prediction (usually the mean of target values).

Step 2: Determine the loss function's gradient and Hessian, or second-order derivative, with relation to the present model.

$$g_i^{(m)} = \frac{\partial \mathrm{L}(y_i,\ F_{m-1}(x_i))}{\partial \mathrm{F}_{m-1}(x_i)} \quad (6)$$

$$h_i^{(m)} = \frac{\partial^2 \mathrm{L}(y_i,\ F_{m-1}(x_i))}{\partial \mathrm{F}_{m-1}(x_i)^2} \quad (7)$$

where $g_i^{(m)}$ is the gradient for the ith observation in the mth iteration. $h_i^{(m)}$ is the Hessian (second-order derivative) for the ith observation in the mth iteration. $\mathrm{L}(y_i,\ F_{m-1}(x_i))$ is the loss function and $\mathrm{F}_{m-1}(x_i)$ is the prediction from the previous iteration.

Step 3: Construct a decision tree using the gradient and Hessian values. The objective is to minimize the following regularized loss:

$$L^{(m)} = \sum_{i=1}^{N}[g_i^{(m)}\ h_m(x_i) + \frac{1}{2}h_i^{(m)} h_m(x_i)^2] + \Omega(h_m) \quad (8)$$

where $L^{(m)}$ is the regularized loss function for the mth iteration. The regularization term for the tree's complexity, $\Omega(h_m)$, is commonly defined as follows:

$$\Omega(h_m) = \gamma\mathrm{T} + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \quad (9)$$

where γ is the parameter controlling the number of leaves, T is the number of leaves in the tree, λ is the regularization parameter and w_2 is the weights of the leaves.

Step 4: Prune the tree to avoid overfitting, using metrics such as the gain to decide whether to split a node or not.

Step 5: Update the model with the new tree.

$$F_m(x) = F_{m-1}(x) + \eta\ h_m\ (x) \quad (10)$$

where $F_m(x)$ is the updated model after the mth iteration and η is the learning rate.

Step 6: Continue steps 2 through 5 until convergence is reached, or for a predetermined number of iterations.

Model evaluation

Three important assessment metrics are utilized to evaluate the effectiveness of the CatBoost and XGBoost models in rainfall prediction:

Mean absolute error (MAE): The MAE quantifies the average size of the errors between the expected and actual values. It provides a sense of the average deviation between the forecasts and the actual numbers.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (11)$$

where N is the total number of observations, y_i is the actual value for the ith observation, $\hat{y}_i$ is the predicted value for the ith observation and the absolute difference $|y_i - \hat{y}_i|$is averaged over all observations to give the MAE.

Root mean squared error (RMSE): The square root of the average squared discrepancies between expected and actual values is measured by RMSE. The squaring of differences prior to averaging highlights greater mistakes more than MAE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (12)$$

where the squared differences $(y_i - \hat{y}_i)^2$ are averaged and then the square root is taken to compute the RMSE.

R-squared (R^2): The percentage of the dependent variable's variation that the independent variables in the model account for, is expressed statistically as R-squared. It gives a sense of how well the data match the model.

$$RMSE = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \quad (13)$$

where $\bar{y}_i$ is the mean of the actual values, $\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ represents the sum of squared residuals (prediction errors) and $\sum_{i=1}^{N}(y_i - \bar{y}_i)^2$ represents the total variance in the actual values.

Results

Using training, testing and validation datasets, the performance of the CatBoost and XGBoost models in forecasting rainfall is shown. MAE, RMSE and R2 are the evaluation metrics that are employed in the comparison. These measurements offer a thorough grasp of the precision and dependability of the models.

Training results: Table 1 summarizes the CatBoost and XGBoost models' performance on the training dataset. The evaluation metrics indicate how well the models fit the training data. From table 1, it is evident that CatBoost outperforms XGBoost in all the metrics on the training dataset. With a lower MAE of 0.12 than XGBoost's 0.15, CatBoost's predictions are more accurate than XGBoost's. Similarly, the RMSE for CatBoost is 0.16, lower than the 0.18 for XGBoost, which suggests that CatBoost makes fewer large prediction errors. Furthermore, compared to XGBoost, which has an R^2 of 0.90, CatBoost has a better R^2 of 0.92, meaning it explains more variation in the target variable.

Validation results: Usually, the validation dataset is used to choose the model. Table 2 displays the CatBoost and XGBoost models' performance on the validation dataset. Table 2 indicates that CatBoost outperforms XGBoost on the validation dataset as well. The MAE for CatBoost is 0.13, lower than XGBoost's 0.16, suggesting more accurate predictions. The RMSE for CatBoost is 0.17, compared to 0.20 for XGBoost, indicating fewer large errors. The R^2 value for CatBoost is 0.90, higher than XGBoost's 0.87, showing that CatBoost explains more variance in the validation data.

Testing results: The models' ability to generalize to previously encountered data is assessed using the testing dataset. The performance metrics for the CatBoost and XGBoost models on the testing dataset are presented in table 3. CatBoost continues to outperform XGBoost on the testing dataset. The MAE for CatBoost is 0.14, compared to 0.17 for XGBoost, indicating better prediction accuracy. The RMSE values further support this, with CatBoost at 0.19 and XGBoost at 0.21, indicating that CatBoost has fewer large errors in its predictions. The R^2 value for CatBoost is 0.88, higher than the 0.85 for XGBoost, demonstrating that CatBoost better captures the variance in the testing data.

Discussion

The results across the training, testing and validation datasets consistently show that CatBoost outperforms XGBoost in predicting rainfall as shown in figures 2 to 4. CatBoost achieves lower MAE and RMSE values, indicating higher accuracy and fewer large errors.

Additionally, CatBoost has higher R^2 values, demonstrating a better fit to the data and capturing more variance in the target variable. Several factors contribute to the superior performance of CatBoost:

a) CatBoost is especially made to handle categorical features in an efficient manner, which is useful when dealing with datasets that contain categorical variables.
b) CatBoost's ordered boosting method improves its generalization capacity by minimizing overfitting and prediction shift.
c) CatBoost's regularization strategies aid in preventing overfitting, which improves performance on unobserved data.

While XGBoost is a powerful and widely used gradient boosting algorithm, CatBoost's tailored features for categorical data and its ordered boosting method give it an edge in this rainfall prediction study.

Table 1
Training Results

Metric	CatBoost	XGBoost
MAE	0.12	0.15
RMSE	0.16	0.18
R^2	0.92	0.90

Table 2
Validation Results

Metric	CatBoost	XGBoost
MAE	0.13	0.16
RMSE	0.17	0.20
R^2	0.90	0.87

Conclusion

This work offers a thorough method for predicting Gujarat's rainfall using cutting-edge machine learning models including XGBoost and CatBoost. The goal was to assess and contrast these models' performances using the important metrics of MAE, RMSE and R^2. The results from the training, testing and validation datasets consistently demonstrate that CatBoost outperforms XGBoost in predicting rainfall. CatBoost achieved lower MAE and RMSE values, indicating higher prediction accuracy and fewer large errors. Additionally, the higher R^2 values for CatBoost suggest that it explains a greater proportion of variance in the rainfall data compared to XGBoost. The findings suggest that CatBoost is a more effective and reliable model for rainfall prediction than XGBoost, making it a preferred choice for similar predictive modeling tasks.

Table 3
Testing Results

Metric	CatBoost	XGBoost
MAE	0.14	0.17
RMSE	0.19	0.21
R^2	0.88	0.85

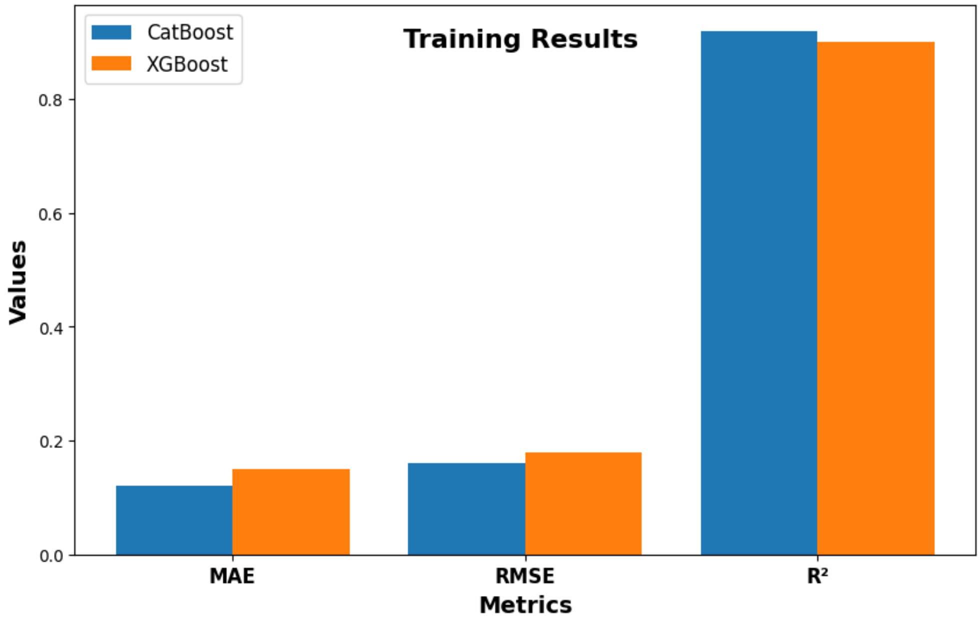

Figure 2: Comparison of CatBoost and XGBoost performance on training data

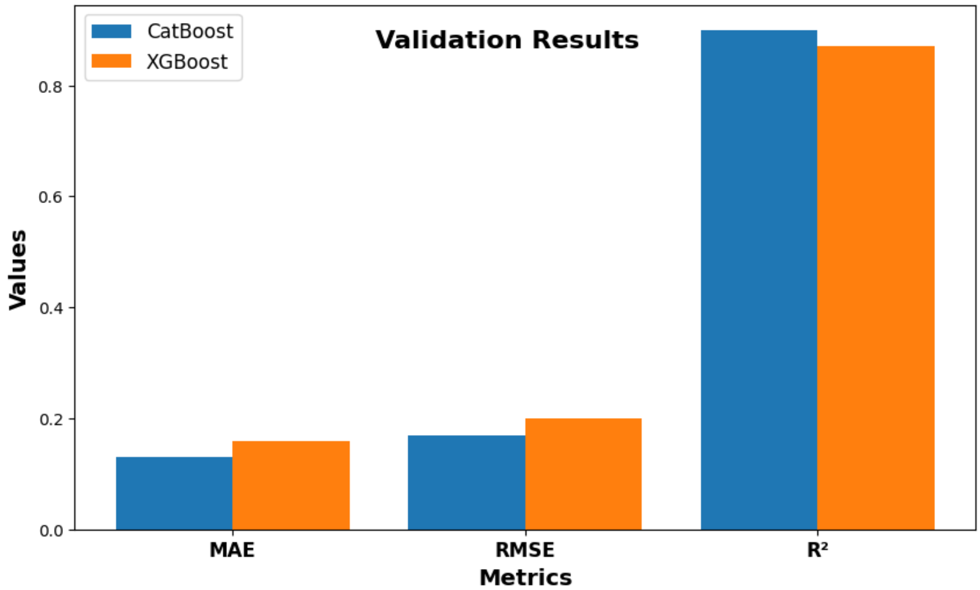

Figure 3: Comparison of CatBoost and XGBoost performance on validation data

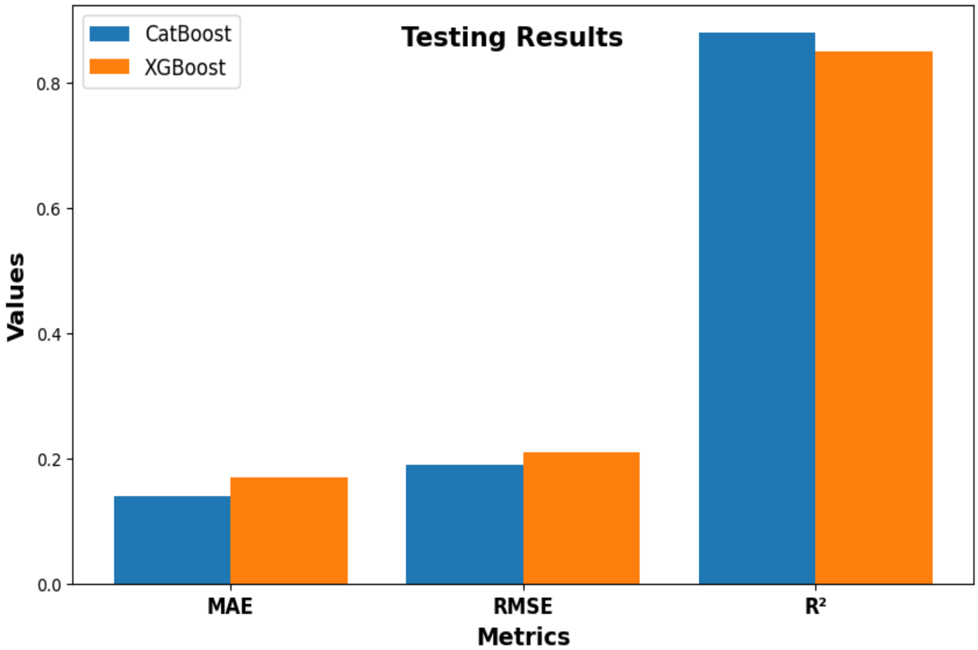

Figure 4: Comparison of CatBoost and XGBoost performance on testing data

This study emphasizes how crucial it is to use cutting-edge machine learning methods to accurately anticipate rainfall. The findings of this study have important ramifications for Gujarat's capacity to forecast rainfall with precision. Improved rainfall prediction can aid in better water resource management, helping to plan for water storage and distribution more efficiently.

Accurate forecasts can also support agricultural planning, allowing farmers to optimize planting schedules and irrigation practices, ultimately enhancing crop yields and reducing the risk of crop failure due to unexpected weather changes. Reliable rainfall forecast is also essential for managing and preparing for disasters. Authorities can reduce the potential damage to infrastructure and loss of life from floods by taking preventive steps in advance of severe rainfall events. The study's cutting-edge machine learning models, especially CatBoost, offer a strong tool for improving Gujarat's rainfall predictions' precision and dependability, strengthening the area's overall sustainability and resilience to climatic shocks.

References

1. Alsulamy S., Kumar V., Kisi O., Kedam N., Rathnayake N., Enhancing water level prediction using ensemble machine learning models: A comparative analysis, *Water Resources Management*, Vol., 1-20 **(2025)**

2. Baudhanwala D., Mehta D. and Kumar V., Machine learning approaches for improving precipitation forecasting in the Ambica River basin of Navsari District, Gujarat, *Water Practice & Technology*, **19(4)**, 1315–1329 **(2024)**

3. Behzad M., Asghari K., Eazi M. and Palhang M., Generalization performance of support vector machines and neural networks in runoff modelling, *Expert Systems with Applications*, **36(4)**, 7624–7629 **(2009)**

4. Cramer S., Kampouridis M., Freitas A.A. and Alexandridis A.K., An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives, *Expert Systems with Applications*, **85**, 169–181 **(2017)**

5. Dhanya C.T. and Nagesh Kumar D., Nonlinear ensemble prediction of chaotic daily rainfall, *Advances in Water Resources*, **33(3)**, 327–347 **(2010)**

6. Fung K.F., Huang Y.F., Koo C.H. and Soh Y.W., Drought forecasting: A review of modelling approaches 2007–2017, *Journal of Water and Climate Change*, **11(3)**, 771–799 **(2020)**

7. Gholami Rostam M., Sadatinejad S.J. and Malekian A., Precipitation forecasting by large-scale climate indices and machine learning techniques, *Journal of Arid Land*, **12(5)**, 854–864 **(2020)**

8. Gupta R. and Nanda S.J., Cloud detection in satellite images with classical and deep neural network approach: A review, *Multimedia Tools and Applications*, **81(22)**, 31847–31880 **(2022)**

9. He R., Zhang L. and Chew A.W.Z., Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning, *Knowledge-Based Systems*, **251**, 109125 **(2022)**

10. He Z., Wen X., Liu H. and Du J., A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region, *Journal of Hydrology*, **509**, 379–386 **(2014)**

11. Lebel L., Local knowledge and adaptation to climate change in natural resource-based societies of the Asia-Pacific, *Mitigation and Adaptation Strategies for Global Change*, **18(7)**, 1057–1076 **(2013)**

12. Lovejoy S., The future of climate modelling: Weather details, macroweather stochastics—or both?, *Meteorology*, **1(4)**, 414–449 **(2022)**

13. Mehta D., Dhabuwala J., Yadav S.M., Kumar V. and Azamathulla H.M., Improving flood forecasting in Narmada River basin using hierarchical clustering and hydrological modelling, *Results in Engineering*, **20**, 101571 **(2023)**

14. Mohandas S., Weather and climate modeling. In: Social and Economic Impact of Earth Sciences, Springer Nature Singapore, Singapore, 121–141 **(2023)**

15. Mosavi A., Ozturk P. and Chau K., Flood prediction using machine learning models: Literature review, *Water*, **10(11)**, 1536 **(2018)**

16. Pirone D., Cimorelli L., Del Giudice G. and Pianese D., Short-term rainfall forecasting using cumulative precipitation fields from station data: A probabilistic machine learning approach, *Journal of Hydrology*, **617**, 128949 **(2023)**

17. Salahi A., Ashrafzadeh A. and Vazifedoust M., Assessing the forecasting accuracy of intense precipitation events in Iran using the WRF model, *Earth Science Informatics*, **17(3)**, 2199–2211 **(2024)**

18. Tiwari D.K., Kumar V., Goyal A., Khedher K.M. and Salem M.A., Comparative analysis of data driven rainfall-runoff models in the Kolar River basin, *Results in Engineering*, **23**, 102682 **(2024)**.

(Received 13th June 2025, accepted 19th July 2025)