**Association of different wet periods during monsoon with the productivity of major kharif crops in Gujarat**

.

ABSTRACT

|  |
| --- |
| Agriculture is reliant on monsoon precipitation as it replenishes water storage systems and influences areas that depend on tanks, canals, and groundwater for irrigation purposes. During times of scant rainfall, the monsoon can significantly impact reservoir and groundwater levels, thereby influencing agricultural productivity and the Indian economy.. The study analyzed rainfall and crop yield data for the major kharif crops (paddy, pearl millet, groundnut, cotton, castor, maize, sesame, and pigeon pea) in 19 districts of Gujarat, which had more than 30 years of data available. Stepwise Multiple Linear Regression (SMLR), Artificial Neural Network (ANN), XGBoost Regression, Random Forest, and Support Vector Regression (SVR). These methods were used to determine which combinations of monthly rainfalls contributed significantly to crop yield. The research was divided into a training process, in which 70% of the data was utilized, and a validation process, in which 30% of the remaining data was utilized.For comparisons among various methods, error metrics of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) were calculated.The measures helped detailed evaluation of the accuracy and reliability of each technique in forecasting crop yield based on a monthly rain pattern. The highest R2 value was obtained with the use of XGBoost method when it produced strong correlation between rainfall and crop yield, while the RMSE value remained the lowest showing improved predictability in the estimation of the crop yield. Random Forest model followed, then Artificial Neural Network, Support Vector Regression, and Stepwise Linear Regression models. |

*Keywords: {Random Forest (RF), Artificial Neural Network (ANN), Support Vector Regression (SVR) and Stepwise Linear Regression (SLR)}*

1. INTRODUCTION

In India, rainfed agriculture is practiced on more than half of the cropland, with over 80% of the annual rainfall received from the southwest monsoon during June to September. This period, known as the Kharif season, is crucial for the cultivation of various crops, while the post-rainy period, referred to as the Rabi season (October to March), also plays a significant role in agricultural productivity. The crops grown in these seasons rarely overlap, except for a few vegetables and perennial horticultural and plantation crops, and significantly differ in their water requirements. The timing and distribution of monsoon rains not only influence the performance of *Kharif* crops but also impact the crops grown in the subsequent season, which utilize the residual moisture from the monsoon.

The significance of studying the association of rainfall with crop productivity in Gujarat cannot be overstated. Gujarat, a state in western India, has diverse agro-climatic zones and experiences varied rainfall patterns, which significantly affect agricultural output. Understanding these associations can help in better planning and management of agricultural practices, ultimately enhancing crop productivity and ensuring food security.Rainfed agriculture accounts for over 51% of the country's net planted area and approximately 40% of total food output (Singh et al., 2019). However, it is characterized by low levels of productivity and input usage, coupled with the vagaries of the monsoon emanating from climate change, resulting in wide variation and instability in crop yields (Mall et al., 2006). As the demand for food grains increases, improving the productivity of rainfed areas is essential. Properly managed rainfed areas have the potential to produce more food and grow faster than irrigated areas, which have reached a productivity plateau (Wani et al., 2009).Several studies have shown the importance of weather and climate in explaining crop yields. For instance, unfavorable weather conditions in the early growth stages of grain maize may limit leaf size and photosynthetic capacity, while adverse conditions in later stages (heatwaves and drought) can reduce the number of silks produced, resulting in poor pollination and kernel development (Lobell et al., 2014). Similar effects have been observed in other crops. For example, a study on Robusta coffee yield in Vietnam and Indonesia suggested a 14% yield reduction for every 1°C increase in mean minimum/maximum temperatures above 16.2/24.1°C during the growing season (Pham et al., 2019). Extreme weather events, such as frosts, heat waves, or prolonged drought, have severe detrimental effects on plant growth and development, and thus crop production (Wheeler & Von Braun, 2013).

Machine learning has become a powerful tool in various fields, including agriculture, where it is increasingly used to forecast crop yields. Various machine learning algorithms, such as Random Forest (RF), Support Vector Regression (SVR), and Artificial Neural Networks (ANN), have been employed to model the relationship between weather and crop yield. For example, RF has been used to monitor winter wheat and investigate the effects of various feature combinations on classification accuracy (Belgiu & Drăguţ, 2016), while XGBoost has outperformed RF and SVM in crop mapping based on spectral properties of distinct crops (Ma et al., 2019). Other studies have used supervised neural networks to determine the appropriate combination of resources, such as plant nutrients and micronutrients, that affect plant growth (Montesinos-López et al., 2021).

Despite the advancements in machine learning, this study focuses on using RF, ANN, SVR, and Stepwise Linear Regression (SLR) due to their robustness and interpretability in handling large datasets and complex relationships between variables. These models have been proven effective in various agricultural applications, providing reliable results with a balance of accuracy and computational efficiency (Crane-Droesch, 2018).

However, recent literature suggests the potential of hybrid models, such as LSTM-ALO (Long Short-Term Memory with Ant Lion Optimization), ELM-PSOGWO (Extreme Learning Machine with Particle Swarm Optimization and Grey Wolf Optimization), SVR-SAMOA (Support Vector Regression with Self-Adaptive Multi-objective Optimization Algorithm), ANFIS-GBO (Adaptive Neuro-Fuzzy Inference System with Gorilla Baboon Optimization), SVR-WCAMFO (Support Vector Regression with Weighted Crowding Avoidance Multi-Objective Optimization), OP-ELM (Optimally Pruned Extreme Learning Machine), and LSSVM-GSA (Least Squares Support Vector Machine with Gravitational Search Algorithm), for hydrological variable modeling (Feng et al., 2020; Ahmadi et al., 2020; Dehvari et al., 2021; Ghosh & Chatterjee, 2021). While these advanced models offer promising results, they often require extensive computational resources and expertise, which may not always be practical for broader agricultural applications. Additionally, the primary aim of this study is to establish a foundational understanding of the rainfall-yield relationship using well-established machine learning techniques before exploring more complex hybrid models.Furthermore, the study suggests comparing the chosen models with additional models like MARS (Multivariate Adaptive Regression Splines), M5Tree (a decision tree model), and LSTM (Long Short-Term Memory networks) to validate the results. These models are known for their flexibility and capability to handle non-linear relationships and time-series data, making them suitable for agricultural yield prediction (Shin et al., 2019; Quinlan, 1992; Hochreiter & Schmidhuber, 1997).

The primary objective of this study is to explore how different wet periods during the monsoon season influence the productivity of major *Kharif* crops in Gujarat. This investigation is guided by two hypotheses. First, we hypothesize a significant correlation between specific monthly combinations of rainfall and the yield of these crops. By analyzing historical data spanning several decades across various districts of Gujarat, we aim to identify which rainfall patterns most strongly correlate with favorable or adverse crop outcomes. Second, we hypothesize that machine learning techniques can effectively predict crop yield based on these identified rainfall patterns. By leveraging advanced algorithms such as Random Forest (RF), Support Vector Regression (SVR), Artificial Neural Networks (ANN), and Stepwise Linear Regression (SLR), we seek to develop predictive models capable of accurately forecasting crop yields under different monsoon scenarios. These hypotheses form the foundation of our research, aiming to enhance our understanding of the complex relationship between monsoon rainfall and agricultural productivity in Gujarat.In the wake of climate change, population growth, and increasing food demand, timely, accurate, and reliable crop yield estimation is essential. This study aims to find the best significant association of monthly rainfall combinations with yield in each district of Gujarat using RF, ANN, SVR, and SLR techniques, providing a basis for future research that could incorporate more advanced hybrid models.

2. material and methods

**Study Area**

Gujarat is a western state in India, bordered by Rajasthan, Madhya Pradesh, Maharashtra, and the Arabian Sea (Fig .1). It has the longest coastline in the country. It is located between latitudes 20°06 –24°42 N and 68°10 –74°28 E, at an altitude of 137 m above MSL. Due to the unique shape of Gujarat, the state has many boundaries and a large variety of ecosystems in its territory. The most important one is Kutch, which is located in the northwest of the state and shares a northern boundary with the provinces of Sindh in Pakistan. Kutch is bordered on the south by the Bay of Kutch, on the west by the Arabian Sea, and on the east by the Rann (Kutch Lake) in the interior. It has diverse ecosystems and receives monsoon rainfall from June to October. The southern part of Gujarat gets more rainfall than the northern part. The Saurashtra and Gulf of Cambay regions receive over 600mm of rainfall. The semi-desert region of Kachchh gets the lowest annual rainfall of about 250mm. Gujarat also experiences intense rainfall events occasionally.

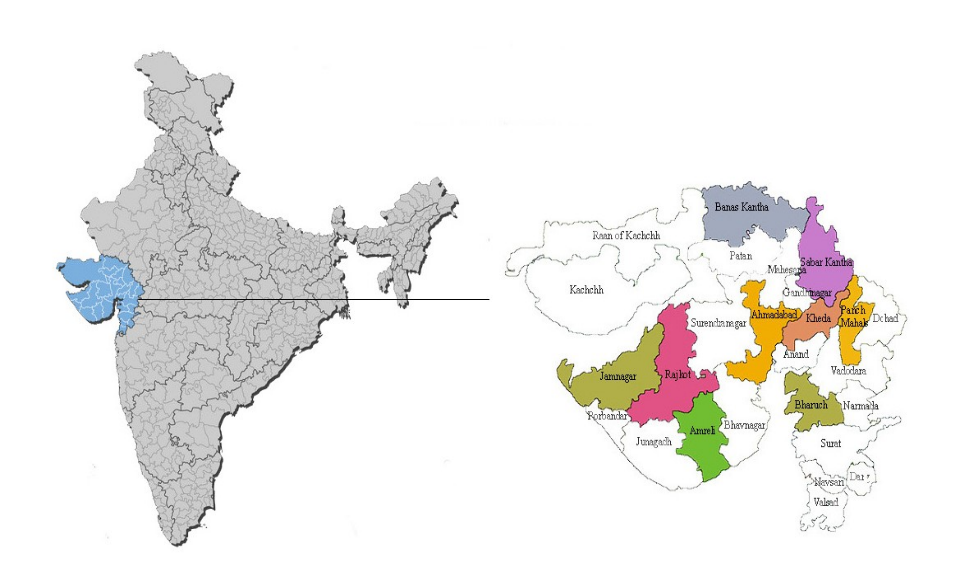


Fig. 1 — Study area of Highlighted districts of Gujarat



Fig.2 —Annual Rainfall of 19 districts of Gujarat

**Data**

Daily time series rainfall records from 1970 to 2020 year of 19 districts(Fig. 2) ,Gujarat were collected from rain gauge stations/observatories of State Agricultural Universities (SAUs), the India Meteorological Department(IMD) and the Gujarat State Disaster Management Authority(GSDMA). District wise productivity data for major crops was collected from the Directorate of Agriculture, Government of Gujarat. In this study, monthly rainfall and yield data of the major 8 kharif crops (paddy, pearl millet, groundnut, cotton, castor, maize, sesame and pigeon pea) for 19 districts of Gujarat with 30 years of data were collected. We considered the districts with major crop growing areas.

**Analysis**

A time series of historical yield data for eight major kharif crops (paddy, pearl millet, groundnut, cotton, castor, maize, sesame, and pigeon pea) was collected across 19 major crop-growing districts of Gujarat, covering a 30-year period. To ensure consistency and comparability of crop yield over time, we applied a detrending technique, which is widely used in agricultural risk analysis to remove long-term trends caused by technological improvements, behavioral changes, and climate variability (Ray et al., 2012). This process made the yield data stationary and more suitable for statistical modeling.Both crop yield (response variable) and rainfall (predictor variables) were detrended to reduce bias and improve the reliability of the models. Only those rainfall–yield relationships that showed statistically significant positive correlations (p < 0.01 and p < 0.05) were selected for modeling. Once significant associations were identified, we applied supervised regression techniques to predict yield based on rainfall.We compared traditional and machine learning models: Stepwise Regression, Random Forest (RF), Artificial Neural Network (ANN), and XGBoost Regression. The entire modeling process was carried out in MATLAB and Python. (Fig 3)The dataset was split into 70% training and 30% testing sets before feature design. Supervised learning algorithms were trained using this split, where the training set was used for model development and the testing set for final performance evaluation.Instead of a manual trial-and-error method, we employed Grid Search to systematically tune the hyperparameters of each model. Grid Search explored multiple combinations of parameters (e.g., the number of estimators, learning rate, tree depth for XGBoost) to identify the best-performing configuration for each algorithm. This method ensured better generalization and reduced the chances of overfitting.Since the input variables had different scales and units, the dataset was normalized using the Min-Max normalization technique to reduce noise and improve convergence. Other normalization methods (e.g., standard scaler, power transformer, and absolute scaler) were tested but did not outperform Min-Max normalization. After model training, the predictions were back-transformed to their original scale for interpretation.To validate the models and prevent overfitting, we adopted 5-fold cross-validation, which balances computational efficiency and model stability. We tested other k-values (3 and 10), but 5-fold provided consistent results with lower computational cost. Repeated cross-validation was also used to assess robustness.

**Detrending**

**Reject**

**Machine learning techniques (ANN,SVR,SLR,XGBoost and RF)**

**Data**

**Five-fold cross validation**

**Training set**

**Model building**

**Prediction accuracy**

**Testing set**

**Significant with rainfall combinations (June,July,Aug,Sept, June\_July,July\_Aug,July\_Aug\_Sept,June\_July\_Aug\_Sept)**

**No**

**yes**

**Pearson correlation coefficient**

(**Paddy, pearl millet, groundnut, cotton, castor, maize, sesame, pigeon pea)**

**Feature selection**

**(R2,RMSE and MAE**)

**Crop**

Fig. 3 —The experimental workflow of technique selection and validation

For model evaluation, we used the Coefficient of Determination (R²), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These metrics helped measure the accuracy and reliability of each model in predicting crop yields based on rainfall inputs.

|  |  |
| --- | --- |
| **Table. 1—Parameter settings for the machine learning models used for association between rainfall and crops** | |
| Model | Best parameters |
| ANN | Number of fully connected layers=1  First layer size=100  Activation= ReLU  Iteration limit =1000  Regularization strength (Lambda)= 0  Standardize data= Yes |
| SVR | Kernel function=Gaussian  Kernel scale= 0.25  Box constraint=Automatic  Epsilon= Auto  Standardize data=Yes |
| SLR | Initial terms= Linear  Upper bound on terms= Interactions  Maximum number of steps=1000 |
| RF | bootstrap=True  n\_estimators = 10  criterion = "squared\_error”  min\_samples\_split = 2  min\_samples\_leaf= 1  min\_weight\_fraction\_leaf= 0.0  max\_features= 1.0 |
| XGBoost | booster='gbtree’  n\_estimators=10  max\_depth=6  learning\_rate=0.3  colsample\_bytree=1  subsample=1 |

3. results and discussion

Correlation of monthly rainfall and its combination with crop yield

In this study, the correlation analysis method was utilized to assess the relation between rainfall and crop yield. The null hypothesis (H0) state that there exists a correlation between climate variations (specifically erratic rainfall) and the productivity of crops (paddy, pearl millet, groundnut, cotton, castor, maize, sesamum and pigeon pea), indicating that crop yield relies on the amount of rainfall. Conversely, the alternative hypothesis (H1) suggests that the amount of rainfall does not determine crop yield. 35 demonstrated that the yearly amount of rainfall has no impact on crop yield, but rather it is the seasonal or monthly rainfall that influences it. To determine the strength of the linear relationship between the variables of seasonal rainfall and annual crop yield, the Pearson product moment correlation coefficient was employed. The resulting value can range from +1 to -1. A value of 0 signifies no association between the variables. A value greater than 0 indicates a positive association, meaning that as rainfall increases, so does the crop yield, thereby supporting H0 and rejecting H1. Conversely, a value less than 0 reveals a negative association, implying that as rainfall increases, crop yield decreases. Hence, H1 is accepted and H0 is rejected.



Fig. 4 —Major crops with best associated districts of Gujarat

**Paddy**

Correlation analysis was carried between dependent variable (annual paddy yield) and independent variables (average rainfall values for July-August, August-September, June-July-August, July-Aug-Sept, and June-July-Aug-Sept), for Ahmedabad and Panchmahal as displayed in Table 2.It indicated that there was correlation between monsoon months and rice yield.From Table 2, it was observed that July-August and June-July-August were common monsoon months in Ahmedabad and Panchmahal.In Ahmedabad , by using the mean score method .36-37 In July-August, the XGBoost model scored the highest R2(0.99) and lowest RMSE (0.02), followed by Random Forest (0.79 and 94.28), ANN (0.56 and 159.74), SVR (0.43 and 181.04), and SLR (0.21 and 213.06), respectively, in accuracy. In Panchmahal, by using the mean score method, we found that June-July-August were the best among other months. In the June-July-August months, XGBoost has the highest R2 (0.99) and lowest RMSE (15.9), followed by Random Forest (0.90 and 114.19), ANN (0.58 and 157.36), SVR (0.41 and 187.75) and SLR (0.22 and 215.62) in accuracy of prediction.

**Groundnut**

According to Fig 4, there is a significant correlation between the groundnut yield in Rajkot and Jamnagar districts. This correlation is influenced by various monsoon months, such as September, June-July, July-August, August-September, June-July-August, July-August-September and June-July-August-September ,June-July, July-August, and June-July-August are the monsoon months that are common to both Rajkot and Jamnagar(Table 2).When utilizing the mean score method in Rajkot, it was found that June-July performed the best compared to other months. During this period, XGBoost achieved the highest R2(0.99) and the lowest RMSE (13.07). Random Forest had an R2(0.81) and RMSE (78.86) followed by ANN with R2 (0.72) and RMSE (115.25). SVR had R2(0.45) and RMSE (160.48), while SLR had an R2 (0.16) and RMSE (199.92). Similarly, in Jamnagar, following the mean score method, June-July was also determined to be the best month compared to the others. XGBoost outperformed during this month with highest R2 (0.99) and RMSE (15.9). Random Forest achieved R2 (0.8) and an RMSE (163.73), while ANN obtained an R2 value of 0.57 and an RMSE of 179.07. SVR had an R2 (0.47) and RMSE (200.41), whereas SLR had an R2(0.34) and RMSE (222.70).

**Castor**

The relationship between castor yield and the monsoon months of June-July and June-July-Aug-Sept in Banaskantha and Kheda Districts shows a strong positive correlation. June-July is regarded as the most favorable month for castor yield in Banaskantha according to the mean score method. XGBoost had the highest prediction accuracy R2(0.99) and the lowest RMSE (15.9) followed by Random Forest (0.84 and 158.32), ANN (0.53 and 169.42), SVR (0.30 and 206.47) and SLR (0.14 and 228.69) respectively.Similarly, in Kheda, June-July were the best months according to mean score method. The XGBoost model had the highest R2(0.99) and the lowest RMSE (15.9) followed by Random Forest (0.84 and 158.32), ANN (0.53 and 169.42), SVR (0.30 and 206.47) and SLR (0.14 and 228.69) respectively.

**Cotton**

The correlation coefficient of cotton yield in Rajkot and Amreli District (Fig 4) was significantly correlated with the monsoon months of June, June-July, July-Aug, Aug-Sept, June-July-Aug, July-Aug-Sept, and June-July-Aug-Sept (Table 2). The castor yield in Rajkot and Amreli District (Fig 4) was significantly correlated with the monsoon months of June, June-July, July-August, August-September, June-July-August, July-August-September, and June-July-August-September (Table 2). In Rajkot and Amreli, the main monsoon months were June, June-July, July-August, and June-July-August. Through mean scoring, we found that July was the best month in Rajkot. XGBoost performs the best with R2 (0.99) and the lowest RMSE (8.97) this month, followed by Random Forest (0.62 and 118.71), ANN (0.53 and 131.18), SVR (0.22 and 168.59) and SLR (0.08 and 182) respectively.Amreli performed the best in July compared to other months, based on the mean score. In terms of model performance, the XGBoost model has the greatest R2 (0.99) and the lowest RMSE (15.9), followed by Random Forest has (0.87 and 118.71), ANN (0.86 and 131.18), SVR (0.50 and 168.59) and SLR (0.19, 182.50) respectively.

**Sesame**

The association between sesame yield and the monsoon months of June-July, June-July-August, July-August-September, and June-July-August-September in Jamnagar and Amreli District was found to be significant (Fig 4) (Table 2). It was observed that June-July and June-July-August were the typical monsoon months in Jamnagar and Amreli, respectively. When using the mean scoring approach, it was determined that June-July was the most favorable period in Jamnagar. Among the models used, XGBoost showed the highest R2 value (0.99) and the lowest RMSE value (15.91) in June-July, followed by Random Forest (0.84 and 147.58), ANN (0.86 and 44.47), SVR (0.25 and 101.37), and SLR (0.24 and 102.19). In Amreli, the best month among others was found to be June-July-August using the mean score method. XGBoost achieved the highest R2 (0.99) and the lowest RMSE (10.68) during this period, followed by Random Forest (0.72 and 100.88), ANN (0.54 and 151.60), SVR (0.24 and 194.67), and SLR (0.13 and 208.44).

**Pearl Millet**

The yield of pearl millet in Rajkot and Banaskantha Districts exhibited a significant positive correlation coefficient with various monsoon months, including June, June-July, July-August, August-September, June-July-August, July-August-September, and June-July-August-September (Table 2). Utilizing the mean score method, it was determined that July was the most favorable month for pearl millet production in Rajkot. This conclusion was supported by the XGBoost model, which achieved the highest R2 (0.99) and lowest RMSE (4.62), followed by Random Forest (0.80 and 175.06). ANN (0.39 and 148.56) followed by SVR (0.15 and 175.06) and SLR (0.11 and 179.76) respectively. In Banaskantha, by using the mean score method, we found June-July was the best among other months. This month, XGBoost has the highest R2 (0.99) and lowest RMSE (4.51), followed by Random Forest (0.83 and 49.89), ANN (0.77 and 130.84), SVR (0.22 and 242.33) and SLR (0.12 and 258.45) respectively. We found that in Rajkot and Banaskantha, XGBoost has the lowest MAE, followed by Random Forest, ANN, SVR and SLR.

**Maize**

The positive correlation coefficient of maize yield in Sabarkantha and Bharuch districts indicates a significant association with the September monsoon month ( Table 2). Upon analyzing, it becomes evident that September was a common monsoon month in both Sabarkantha and Bharuch. Specifically, in the region of Sabarkantha, September proved to be particularly effective for maize yield. During this month, the algorithm known as XGBoost demonstrated the highest R2 (0.99) and RMSE (7.15), followed by Random Forest (0.92 and 75.88), ANN (0.39, 68.21), SVR (0.77, 135.63) and SLR (0.11, 136.51) respectively. Similarly, in the region of Bharuch, September also demonstrated effectiveness for maize yield. During this month, XGBoost exhibited highest R2 (0.99) and lowest RMSE (15.99) followed by Random Forest (0.89 and 131.81), ANN (0.39 124.53), SVR (0.07 and 155.17) and SLR (0.10 and 152.59), respectively.

Pigeon Pea(Tur)

In Panchamahal District, there was a positive and significant correlation between pigeon pea yield and the monsoon months of September, July-August-September, and June-July-Aug-Sept(Table 2). September was found to be a common monsoon month in Panchamahal, and it was particularly effective for pigeon pea yield. Among the different prediction models, XGBoost performed the best, with the highest R2 (1.00) and RMSE (0.00), followed by Random Forest (0.85 and 63.65), ANN (0.84 and 54.37), SVR (0.62 and 84.44) and SLR (0.42 and 103.98) respectively.

These findings highlight the efficacy of ensemble models incorporating gradient descent boosting and bagging techniques, as they consistently deliver more accurate predictions with diminished

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table .2—Pearson correlation coefficient of 8 major *kharif* crop yield with rainfall** | | | | | | | | | | | |
| Monsoon months | PADDY | | GROUNDNUT | | | | CASTOR | | PEARL MILLET | | |
| Ahmedabad  (45 yr) | Panchmahal  (48 yr) | | Rajkot  (50 yr) | | Jamnagar  (35 yr) | Banaskantha  (42 yr) | Kheda  (31 yr) | Banaskantha  (42 yr) | | Rajkot  (50 yr) |
| June | NS | NS | | | (0.38) \*\* | NS | (0.38) \*\* | NS | | NS | (0.50) \*\* |
| July | NS | (0.44) \*\* | | | NS | (0.56) \*\* | NS | (0.56) \*\* | | (0.32) \* | (0.33) \* |
| August | (0.37) \* | NS | | | (0.36) \*\* | NS | (0.36) \*\* | NS | | NS | NS |
| September | NS | (0.53) \*\* | | | (0.53) \*\* | (0.44) \*\* | (0.53) \*\* | NS | | NS | NS |
| June -July | NS | (0.46) \*\* | | | (0.40) \*\* | (0.44) \*\* | (0.40) \*\* | (0.38) \* | | NS | (0.53) \*\* |
| July-August | (0.38) \*\* | (0.41) \*\* | | | (0.42) \*\* | (0.58) \*\* | (0.42) \*\* | NS | | (0.34) \* | (0.40) \*\* |
| August-September | (0.4) \*\* | (0.40) \*\* | | | (0.55) \*\* | (0.36) \* | (0.55) \*\* | (0.43) \* | | NS | NS |
| June-July-August | (0.34) \* | (0.48) \*\* | | | (0.53) \* | (0.61) \*\* | (0.53) \* | (0.44) \* | | NS | (0.55) \*\* |
| July-August-September | (0.40) \*\* | (0.57) \*\* | | | (0.60) \*\* | (0.68) \*\* | (0.60) \*\* | (0.39) \* | | (0.31) \* | (0.49) \*\* |
| June\_July\_August\_Sept | (0.37) \* | (0.62) \*\* | | | (0.66) \*\* | (0.69) \*\* | (0.66) \*\* | (0.48) \*\* | | NS | (0.50) \*\* |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Monsoon months | COTTON | | SESAME | | MAIZE | | PIGEON PEA (TUR) |
| Rajkot  (50 yr) | Amrelli  (22 yr) | Jamnagar  (38 yr) | Amrelli  (53 yr) | Sabarkantha  (45yr) | Bharuch  (43yr) | Panchamahal  (25 yr) |
| June | NS | NS | NS | NS | (0.31)\* | NS | NS |
| July | (0.28)\* | (0.44)\* | NS | (0.29)\* | NS | NS | NS |
| August | NS | NS | NS | NS | NS | NS | NS |
| September | (0.34)\* | NS | (0.49)\*\* | NS | (0.31)\* | (0.31)\* | (0.65)\*\* |
| June -July | (0.36)\*\* | (0.47)\* | (0.48)\*\* | (0.34)\* | NS | NS | NS |
| July-August | (0.39)\*\* | (0.55)\*\* | NS | NS | NS | NS | NS |
| August-September | (0.38)\*\* | (0.48)\* | NS | NS | NS | (0.37)\* | NS |
| June-July-August | (0.44)\*\* | (0.61)\*\* | (0.38)\* | (0.37)\*\* | NS | NS | NS |
| July-August-September | (0.48)\*\* | (0.63)\*\* | (0.59)\*\* | (0.31)\* | NS | NS | (0.51)\*\* |
| June\_July\_August\_Sept | (0.51\*\*) | (0.66)\*\* | (0.49)\*\* | (0.35)\*\* | NS | NS | (0.41)\* |

\*\* = highly significant at 1% level, \* = Significant at 5% level.

TABLE 3. **Pearson correlation coefficient of 8 major *kharif* crop yield with rainfall (Contd.)**

errors. The low MAE, RMSE, and high R2 values show that the developed model is very accurate compared to the experimental values. The R2 value of each response shows how well the model fits the data. A higher determination coefficient and lower data error indicate a higher accuracy of the data. The nonlinear behavior of the dataset could explain why the model is better at fitting the data. According to our evaluation metrics, the XGBoost model performed the best. XGBoost has highest R2 among the other Machine learning techniques. The RMSE and MAE are also the lowest. We found major crops of 19 districts were best associated with rainfall periods by XGBoost technique.

4. Conclusion

The results of this study showed that, among other machine learning techniques, the XGBoost machine learning approach was best to identify the most effective monsoon seasons for growing crucial *kharif* crops in various regions of Gujarat, as per the findings of this study. The model was optimized with the most effective configurations and provided the most precise outcomes, achieving the greatest R² and the least RMSE, in comparison to other methods such as Random Forest, ANN, SVR, and Linear Regression. The results showed that each crop had a specific time when rainfall was most helpful—for example, paddy grew best in July–August in Ahmedabad, groundnut in June–July in Rajkot and Jamnagar, and maize in September in Sabarkantha and Bharuch. These findings helped create a district-wise sowing calendar and showed how important rainfall timing is for crop growth. Overall, XGBoost proved to be a useful tool for improving crop planning based on climate.

References

1. Joshi P, Gulati A & Cummings R W, Agricultural diversification and smallholders in South Asia. Academic Foundation(2007).
2. Jat R K, Sapkota, T B, Singh R G, Jat, M L, Kumar M, & Gupta R K,Seven years of conservation agriculture in a rice–wheat rotation of eastern gangetic plains of south asia: yield trends and economic profitability,Field Crops Research, 164 (2014) 199–210.
3. <https://agricoop.gov.in/en/RainfedDiv>.
4. Darby H & Lauer J, Plant physiology: Critical stages in the life of a corn plant,UW Crop Scouting Manual, 15(20) (2000).
5. Ritchie S & Hanway J, How a corn plant develops. Special Report, Iowa State University of Science and Technology, (48) (1982).
6. Kath, J., Byrareddy, V. M., Craparo, A., Nguyen-Huy, T., Mushtaq, S., Cao, L., &Bossolasco, L,Not so robust: Robusta coffee production is highly sensitive to temperature,Global Change Biology, 26(6) (2020) 3677–3688.
7. Hatfield, J L & Prueger, J H, Temperature extremes: Effect on plant growth and development. Weather and Climate Extremes, USDA Research and Programs on Extreme Events,(10) (2015) 4–10.
8. Mathieu J A & Aires F,Statistical weather-impact models: An application of neural networks and mixed effects for corn production over the United States, Journal of Applied Meteorology and Climatology, 55(11) (2016) 2509–2527.
9. Vogel J, Rivoire P, Deidda C, Rahimi L, Sauter C, Tschumi E, Van Der Wiel K, Zhang T, & Zscheischler J, Identifying meteorological drivers of extreme impacts: An application to simulated crop yields, Earth System Dynamics, 12 (2021) 151–172.
10. Ray D K, Gerber J S, MacDonald G K, & West P C, Climate variation explains a third of global crop yield variability. Nature Communications, 6 (2015). 1–9.
11. van Oijen M, Dauzat J, Harmand J M, Lawson G, &Vaast P, Coffee agroforestry systems in Central America: II. Development of a simple processbased model and preliminary results, Agroforestry Systems, 80(3)(2010). 361–378.
12. Dormann C F, Schymanski S J, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Römermann C, Schröder B, & Singer A, Correlation and process in species distribution models: Bridging a dichotomy. Journal of Biogeography, 39(12) (2012) 2119–2131.
13. Rahn, E, Vaast, P, Läderach, P, van Asten, P, Jassogne, L, & Ghazoul, JExploring adaptation strategies of coffee production to climate change using a process-based model, Ecological Modelling, 371 (2018) 76–89.
14. Vezy R, le Maire G, Christina M, Georgiou S, Imbach P, Hidalgo H G, Alfaro, E J, Blitz-Frayret, C, Charbonnier, F, Lehner, P, Loustau, D & RoupS O ,Dynacof: A process-based model to study growth, yield and ecosystem services of coffee agroforestry systems, Environmental Modelling & Software, 124 (2020). 104609.
15. Lobell D B & Burke M B, On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology, 150(11) (2010) 1443–1452.
16. Ray D K, Gerber J S, MacDonald, G K, & West, P. C, Climate variation explains a third of global crop yield variability. Nature Communications, 6 (2015) 1–9.
17. Gornott C & Wechsung F, Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany, Agricultural and Forest Meteorology, 217 (2016) 89–100.
18. Hoogenboom G, White J W & Messina C D ,From genome to crop: integration through simulation modeling,Field Crops Research, 90(1) (2004). 145–163. https://doi.org/ 10.1016/j.scitotenv.2020.142844
19. Van Oijen M, Dauzat J, Harmand J M, Lawson G, &Vaast P, Coffee agroforestry systems in Central America: II. Development of a simple processbased model and preliminary results, Agroforestry Systems, 80(3)(2010). 361–378.
20. Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLOS ONE, 8*(6), e66428. https://doi.org/10.1371/journal.pone.0066428
21. Roberts M J, Braun N O, Sinclair T R, Lobell D B & Schlenker, Comparing and combining process-based crop models and statistical models with some implications for climate change, Environmental Research Letters, 12(9) (2017) 095010.
22. Lobell D B & Burke M B, On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology, 150(11) (2010) 1443–1452.
23. Mathieu J A & Aires F ,Statistical weather-impact models: An application of neural networks and mixed effects for corn production over the United States, Journal of Applied Meteorology and Climatology, 55(11) (2016) 2509–2527.
24. Kern A, Barcza Z, Marjanovi´c H & Árendás, T, Fodor, N., Bónis, P., Bognár, P., and Lichtenberger, J. Statistical modelling of crop yield in central europe using climate data and remote sensing vegetation indices. Agricultural and Forest Meteorology, (February) (2018) 300–320.
25. Belgiu M & Draguţ L,Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing. 114 (2016) 24-31. DOI-doi.org/10.1016/j.isprsjprs.2016.01.011
26. Xu F, Li Z , Zhang S,Huang, N, Quan Z,Zhang W, Liu X, Jiang, X Pan J & Prishchepov A,Mapping Winter Wheat with Combinations of Temporally Aggregated Sentinel-2 and Landsat-8 Data in Shandong Province, China, Remote Sens. 12 (2020) 2065.
27. Liu J, Feng Q, Gong J, Zhou J, Liang, J & Li Y,Winter wheat mapping using a random forest classifier combined with multi-temporal and multi-sensor data, Int. J. Digit. Earth, 11 (2017) 783–802.
28. Saini R & Ghosh S K, Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery,Geocarto Int,  36 (2019) 2141–2159.
29. Prins A J & Van Niekerk A, Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms,Geo-Spatial Inf. Sci,24 (2020) 215–227.
30. Shivnath G & Santanu K ,Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks, International Journal on Recent and Innovation Trends in Computing and Communication, 2(2) (2014) 292–297. <https://doi.org/10.17762/ijritcc.v2i2.2959>.
31. Vinciya P & Valarmathi A, Agriculture Analysis for Next Generation High Tech Farming in Data Mining. IJARCSSE. 6(5) (2016).
32. Zhihao H, Z. Kalbarczy k & Iyer R K , A Data-Driven Approach to Soil Moisture Collection and Prediction,IEEE Xplore, 2 (2016) 292-297.
33. Joshua V, Priyadharson, S M & Kannadasa R ,Exploration of Machine Learning Approaches for Paddy Yield Prediction in Eastern Part of Tamilnadu, Agronomy, 11 (2021) 2068. DOI-doi.org /10.3390 /agronomy 11102068.
34. Cao J, Zhang Z, Tao F, Zhang L, Luo Y,Han J & Li Z, Identifying the contributions of multi-source data for winter wheat yield prediction in China. *Remote Sens.,* *12* ( 2020)750. DOI-doi.org/10.3390/rs12050750
35. Rodriguez J D,Perez A,& Lozano J A,Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 569–575.
36. Adejuwon J D Vulnerability in Nigeria: A national-level assessment In Climate Change and Vulnerability and Adaptation (Routledge), (2013) 214-233.
37. Pepe MS, Reilly M & Fleming TR,Auxiliary outcome data and the mean score method,Journal of Statistical Planning and Inference, 42 (1994) 137–160.
38. Reilly M & Pepe MS. A mean score method for missing and auxiliary covariate data in regression models, Biometrika,82(2) (1995 ) 299–314.
39. Elhishi S , Elashry A M & El-Metwally S, Unboxing machine learning models for concrete strength prediction using XAI. Sci Rep 13, (2023) 19892. https://doi.org/10.1038/s41598-023-47169-7