

Factors Influencing Agricultural Yield, Quality, and Profitability: A Data-Driven Analysis of Major Crops in Maharashtra

Deepak S. Sonawane*¹, Dr. Ashok Y. Tayade²

¹ Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

Email ID: deepaksonawanevpkbiet@gmail.com,

²Department of Statistics, Dr Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar

Email ID : aytayade@gmail.com,

***Corresponding Author**

Deepak S. Sonawane

Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

Email ID: deepaksonawanevpkbiet@gmail.com

ABSTRACT

This research provides an integrated climate-soil-crop model using big data in relation to agricultural yield, crop quality, and crop profit from Maharashtra. A composite Total Yield Index was created to evaluate the productivity of different crops across the region. Machine Learning (ML) classifiers, including Linear Regression, Random Forest models, and XGBoost classifiers, were used in conjunction with statistical analysis to identify the main variables affecting crop yield. Results show that temperature and nitrogen have a high degree of effect on crop yield while rainfall does not have much of an effect. Random Forest provided the best overall prediction accuracy, demonstrating a higher level of accuracy in capturing the non-linear relationships associated with the dataset. Utilizing Explainable AI (using Shapley values) was critical in providing interpretability for the ML methods. A data-driven integrated model or systemic approach provides improved decision-making capabilities and insight into sustainable approaches to operate in agriculture.

Keywords: - Agricultural Yield, Machine Learning (ML), Data-Driven Farming, Precision Agriculture, SHapley Additive Explanations (SHAP), Climate Factors.

INTRODUCTION

Over many years, agriculture has been transforming itself through the use of data-based technologies and new analytical methods. There is an increasing need for an integrated, systematic way of analysing agriculture in areas, such as Maharashtra, where agricultural output will be determined by climate, soil and resources. Traditional methods are important but are frequently too simple to provide insights about the multiple ways, both positive and negative, that all these factors interact and affect the crop, the quality of the crop, and the profitability of those crops.

Machine learning, in conjunction with appropriate statistical methods, can be used as a tool to help understand agricultural systems by using their physical processes and how they respond to interventions and changes. By building on this background, this research aims to provide a base of compiled agricultural research and to quantify a comprehensive, data-driven approach to evaluate key determinants affecting crop performance and promote sustainable agricultural practices.

It is no longer only advantageous to use a more technologically and data-driven approach to grow our crops; it has become essential in an agricultural world that is more dynamic and continuing to grow in complexity. The conventional farming practices are foundational yet, in most contexts, fail to provide the precision and adaptability that is expected from them by the emerging agricultural demands (Marapelli et al, 2023). The establishment of robust crop-management systems has quickly become one of the primary strategies for improving agriculture and making it more efficient and sustainable. This is a research paper that discusses a new agricultural system that was created to integrate various forms of data that, when combined, offer practical applications to farmers (Swarnkar et al, 2023). Using weather data, fertilizer guidelines, market prices, crop information, machinery tools, and soil types, this system will try to facilitate better decisions regarding crop management (Suguna et al, 2024). The system discussed herein acts as a centralized point from which farmers can harness a wide variety of information to enhance their agricultural outputs (Pukrongta et al, 2024). Due to its primary purpose, the system continues to provide viable crops based on the most favourable weather conditions for growing these crops. The system examines the weather trends and forecasts to provide appropriate crop recommendations based on what crops can be successfully grown at any time based on the prevailing climatic conditions. This approach improves crop yields while reducing cases of crop failures attributed to unfavorable weather conditions (Alazzai et al, 2024). Although the crop recommendation based on climatic conditions is an important feature of this system, it is also able to offer complete grower assistance with respect to fertilizer recommendations for all crops grown in the region. Fertilizers represent one of the most important factors contributing to the overall production capability of

a cropping system, but their effect will be entirely based on accurate and optimal use. The system generates crop-specific fertilizer recommendations that include the type and amount of fertilizer to use considering both the specific requirements of each crop and the climatic conditions that will exist at the time of planting. This, in return, will help optimize fertilizers, reduce their waste, and improve the health and yield of crops as a whole (Mana et al, 2024).

In addition, one of the other features of the system displays what the current are live prices of multiple seed varieties, based on what is currently available in the marketplace. Therefore, farmers can use this information to make good choices when purchasing seeds.

The foundation of the world's food security is the agriculture sector. More than 70% of rural households are reliant on agriculture. With over 60% of the workforce employed and contributing over 20% of the country's GDP, it is a significant sector of the Indian economy (Ma et al., 2023).

Even so, farmers will continually attempt to balance their profit and yield. In order to meet increasing demand for agriculture products, greater amounts of food will need to be produced or an increase in the yield from current crops will need to occur; however, it can be extremely challenging to optimize your profit from an uncertain market or fluctuating price for a commodity you grow. Many times farmers will not take advantage of opportunities for greater profit because they depend on their own experience and instinct to determine the proper time to harvest a crop; therefore, we propose to develop a machine learning (ML) based solution that will solve the aforementioned issues associated with determining when to harvest crops and reduce the difficulties experienced by farmers, resulting in increased profits for farmers. Although farmers utilize technological advancements in their farming practices, they still find it difficult to consistently produce a profit. Predictions were produced using supervised learning techniques on data on onion prices due to the study's conclusions that price fluctuations can cause financial troubles and that onion prices are highly unpredictable (Elavarasan & Vincent, 2020). Weather, fluctuating prices, and lack of current market dynamics can both greatly affect the earnings of farmers. For example, traditional harvest methods may use past averages or instinct for harvesting certain crops but do not account for the impact of weather and prices that are affected by so many different factors. Typically, prices for any agriculture product can be quite volatile, depending on many different reasons, such as weather, seasonality, world markets, and even unexpected events. If a farmer harvested their crop before they hit their optimal price, or harvested after and the crop spoiled or was of lower quality to achieve its optimal price, they will often see lower yields than they anticipated and a lower market value. Seasonal impacts that can significantly affect the quality and production of crops are temperature, sunlight, and moisture. To properly determine the correct harvest window, both professionals and farmers will need to understand these different variances. Eventually, if a farmer harvested their fruit or vegetable prior to it being optimally mature, lower yields and decreased market value will likely occur. A farmer that operates on a small scale often cannot access real-time local and regional market data. As such, it may be difficult to predict pricing of their crop and select the appropriate harvesting strategy to achieve their optimum price based on the present market conditions.

REVIEW OF LITERATURE

Patil et al. (2026) developed a Composite Sustainable Agricultural Index (CSAI) for Solapur district (2010–2023), combining EEI, SEI, and ESI components. The district showed moderate sustainability (CSAI 0.464–0.466), and ML models predicting the 22 underlying indicators achieved high accuracy (RMSE 0.0038, MAE 0.0028), supporting sustainable practices such as organic farming and agroforestry.

Krishna et al. (2026) built a machine learning framework for agricultural price forecasting using environmental, economic, and logistical factors. Comparing five algorithms (Linear Regression, SVM, AdaBoost, Random Forest, XGBoost) on R^2 , RMSE, and MAE, XGBoost performed best ($R^2 = 0.94$), confirming ML's value for reliable, sustainable price forecasting.

Sharma et al. (2025) used historical agricultural, district-level, and climate data from Maharashtra to predict maize yield with a Random Forest Regressor, applying one-hot encoding and feature scaling. Cross-validation confirmed strong predictive accuracy, showing that regional weather and practices significantly affect yield variability and that ML can improve resource allocation and reduce risk.

Wanjale et al. (2025) developed a web-based decision support system for millet selection in Maharashtra, using seasonal rainfall, soil characteristics, temperature, and past yield data. A voting classifier identified the best millet variety for a given farm, while Random Forest regression predicted expected yield, helping farmers use resources more efficiently under a variable climate.

Phatangare et al. (2024) developed an integrated ML decision-support system combining environmental variables (location, soil, weather) with economic variables (market prices) to improve farmer profitability. A Flask-based interactive web application delivered personalized crop and price recommendations, showing improved prediction accuracy and farmer decision-making.

Chakre et al. (2024) forecast daily crop prices across Indian cities using government data on seasonal trends and price variability. Comparing Random Forest, SVR, XGBoost, LightGBM, Linear Regression, and Neural

Networks, the study found ensemble and stacking models more accurate than any single model, improving agricultural market forecasting.

Jayaraman et al. (2023) examined how social capital affects performance in two Farmer Producer Companies (organic and conventional), using Self-Organizing Maps on membership survey data. Social ties, leadership, group participation, and geographic proximity emerged as significant contributors to group performance and supply-chain effectiveness.

Patil et al. (2023) applied machine learning classification and regression to predict crop yield and select crops, incorporating environmental factors such as humidity and soil quality often overlooked by traditional methods. Random Forest regression achieved strong yield prediction ($R^2 = 0.96$, MAE = 0.64), and Naïve Bayes achieved 99.39% accuracy for crop classification, demonstrating ML's value for decision-making and resource use.

Gupta et al. (2022) combined Relief-based feature selection with linear discriminant analysis and PSO-SVM, KNN, and RF models to improve crop yield prediction using weather, soil, water, and market variables. The framework supported more informed, risk-minimizing crop-planning decisions.

Bali and Singla (2022) reviewed ML and DL methods for crop yield prediction, emphasizing the value of accurate early forecasts. Hybrid models showed the greatest accuracy gains, with temperature and precipitation identified as key climatic drivers of yield alongside agronomic practices.

Research Gap

Although previous studies have demonstrated the effectiveness of machine learning techniques for crop yield prediction, agricultural sustainability assessment, and market price forecasting, most investigations have focused on individual objectives such as yield estimation, crop recommendation, or price prediction. Limited attention has been given to developing an integrated framework that simultaneously evaluates agricultural yield, crop quality, and profitability using climatic and soil variables. Furthermore, the majority of existing studies emphasize predictive performance without incorporating statistical validation to verify the significance of influencing factors or Explainable Artificial Intelligence (XAI) techniques to improve model transparency and interpretability. In addition, comprehensive analyses specifically addressing major crops across Maharashtra remain scarce. Therefore, this study addresses these research gaps by proposing a unified data-driven framework that combines statistical regression, multiple machine learning models, and SHAP-based explainable AI to identify the key determinants of agricultural productivity while providing transparent and reliable decision support for sustainable agriculture.

Agricultural yields can be impacted by many different rapidly changing factors including weather changes, soil characteristics, input use and cost, and the market. Farmers have access to traditional agricultural knowledge, but they frequently are unable to find timely integrated information they need for sound decision-making. Most existing data & information systems provide only a portion of what a farmer would need to know to make informed decisions; for example; most systems only provide either crop yield prediction, or an estimate of what price they could sell their product for based solely on use of one method to predict price, but do not integrate such predictions with weather conditions, fertilizer recommendations and/or crop selection and/or market trends. The result is that many farmers use poor crop estimates, inefficient resource use and lower profitability. Thus, there is a need to develop an integrated framework that is data-driven and looks at multiple agricultural parameters simultaneously to aid farmers in making good decisions, thereby increasing crop production and quality, and maximizing profits from crops grown.

RESEARCH METHODOLOGY

In this research, we will analyze data from more than one year about agricultural productions in Maharashtra. We have included monthly rainfall from the months of June to October and added up to the whole amount of rainfall throughout the whole year. Agricultural productivity for various commodities (bajra, jowar, rice, wheat, ragi, urad, tur, and gram) will also be included in our study. Average temperature and soil indicator (amount of nitrogen, phosphorous, and potassium) will be used to measure the success of producing each of the goods. By having all of the above-listed variables available for analysis, we can do a multifaceted analysis on the production, quality, and profit of agriculture in Maharashtra.

To increase the consistency and the dependability of the dataset, data preprocessing was carried out by removing irrelevant attributes and standardizing column names. Forward filling methods were chosen to address missing values. Extreme outliers were also handled through the removal of extreme outliers in crop yield (e.g., rice) using percentiles as filtering techniques. The removal of these outliers prevents distortion in the data for analysis and modeling purposes, which contributes to the overall accuracy and robustness of the results obtained through the remainder of the study.

By combining all major crop yields into a new composite indicator called Total Yield Index an integrated analysis across different crops can be accomplished more efficiently than individual crop assessments provide. Normalization was applied during visualization to overcome any potential issues caused by different scales between various types of crops before presenting results.

To examine the distributions and relationships of the variables, an exploratory analysis was performed using statistical summaries, correlation matrices, and graphical representations of the data using normalized bar charts, trend plots, and regression plots. The correlation analysis was used to identify significant relationships and assess the presence of multicollinearity between the variables, with emphasis on the components of the fertilizer.

Correlational analyses showed that Nitrogen, Phosphate, and Potash were all statistically correlated, prompting investigation into possible redundancy. Therefore, in order to avoid multi-collinearity effects and facilitate interpretation of our models, we retained only the representative variable, which was Nitrogen, as well as two key climatological variables: Rainfall and Temperature.

The hybrid modeling framework combined both statistical and machine-learning methods:

A linear regression model served as a baseline for understanding the linear relationships between dependent and independent variables.

The random forest regression model was used to identify the non-linear interaction of independent variables.

The XGBoost regression model was implemented to produce a boost in prediction accuracy.

The data was split into 80% for training and 20% for testing with features being standardised using feature scaling.

The use of the statsmodels library allowed for ordinary least squares (OLS) regression to be conducted alongside machine learning models as a means of assessing the statistical significance of each variable, clearly presenting the effects of each independent variable on each dependent variable through coefficients, p-values, and confidence intervals. This allows the findings to fit into the framework of traditional statistical research methodologies.

The performance of the model was evaluated using the following metrics:

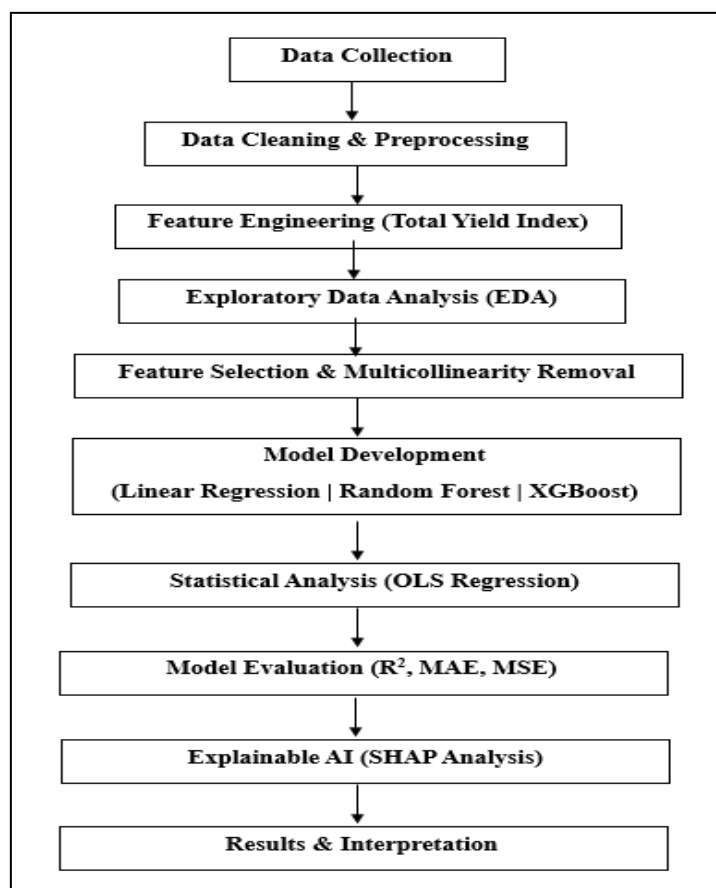
R² ("Coefficient of Determination")

MAE ("Mean Absolute Error")

MSE ("Mean Squared Error")

Various models were compared to determine which one would yield superior results in predicting yields.

SHAP, which stands for SHapley Additive Explanations, was used to give additional interpretability and allow for actionable insights to be made from the data. It does so by explaining how much each feature contributed to make a prediction for an entire dataset, enabling researchers to identify which variables were the key contributors to agricultural yield. By combining Explainable AI with Statistical Methods and Traditional Machine Learning, this provides the research community with a new way to look at the data they collect.



RESULTS & DISCUSSION

Descriptive Analysis of Crop Yield

The first study intended to understand how crop yields are distributed among the various major crops produced in Maharashtra. To help eliminate some of the scale differences between crops, normalization methods were used.

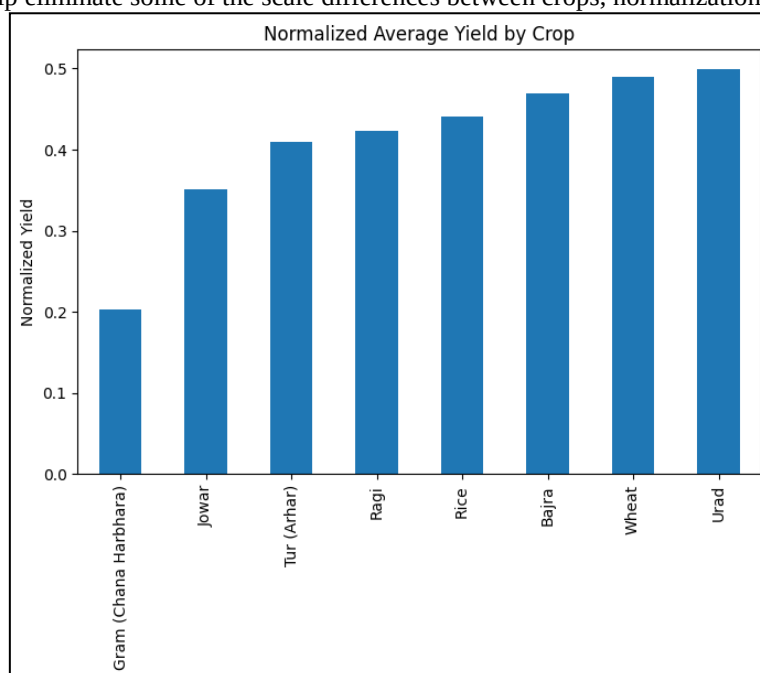


Figure 1 Normalized Average Yield of Major Crops in Maharashtra

The results show that, with respect to normalized yields, rice has the highest normalized yield of any crop, followed by wheat. Comparatively lower productivity is exhibited by crops such as millets and pulses including bajra, jowar, urad and tur. This strongly indicates that rice is the predominant crop being grown in this area but also raises questions regarding potential disparities between crop yields.

Table 1 Normalized Average Yield by Crop (approx., read from chart)

Crop	Normalized Yield
Gram (Chana Harbhara)	0.20
Jowar	0.35
Tur (Arhar)	0.41
Ragi	0.42
Rice	0.44
Bajra	0.47
Wheat	0.49
Urad	0.50

Temporal Trends in Crop Yield

An analytical approach was taken to ascertain the patterns of yield fluctuations of major crops across a span of several years.

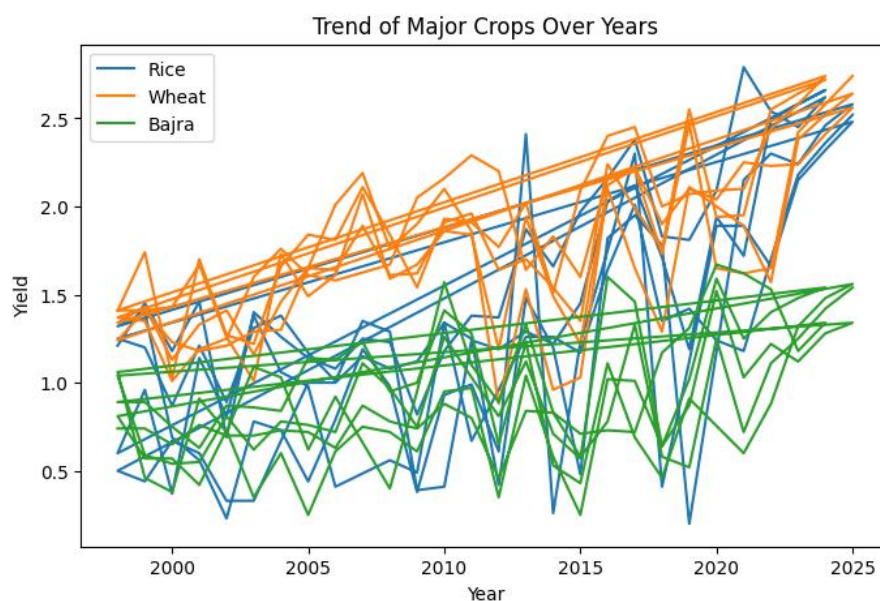


Figure 2 Temporal Trends of Major Crops Over Years

The analysis has shown that there is a moderate fluctuation in crop yield from year to year as a result of trends in the yield of the crop. For example, rice has shown to have increased variability from year to year compared to other crops such as wheat and bajra, which displays an increased susceptibility to environmental and climatic factors. In addition, it was found that there is no consistent annual upward trend in the yield of all crops, thus, indicating that yield improvements were not equivalent each year.

Table 2 Approximate Crop Yield Range Over Years (1998-2025, read from chart)

Crop	Approx. Yield Range	Overall Trend
Rice	~0.2 to 2.8	Rising, with high year-to-year fluctuation
Wheat	~1.0 to 2.75	Rising, generally at the upper range
Bajra	~0.3 to 1.6	Rising gradually, generally the lowest of the three

Correlation Analysis of Agricultural Factors

An analysis of the correlation matrix provides insights into the relationship between climatic and soil variables.

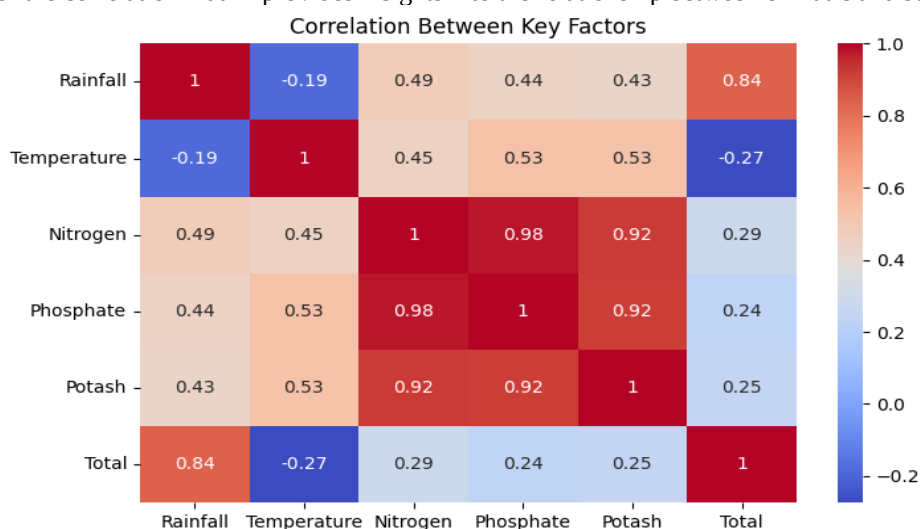


Figure 3 Correlation Matrix of Climatic and Soil Variables

The correlation analysis indicates that there are strong positive correlations among the major fertilizer elements (i.e., nitrogen, phosphorus, potassium), which suggests that there is an association due to multicollinearity. On a modest basis, total rainfall and the associations of the other yield-related variables with total rainfall—as well as the nitrogen and the association of the temperature with total yield—indicates that both temperature and nitrogen play an important role in agricultural production as well as yield.

Table 3 Correlation Coefficients Between Key Variables

Variable Pair	Correlation (r)
Rainfall - Temperature	-0.19
Rainfall - Nitrogen	0.49
Rainfall - Phosphate	0.44
Rainfall - Potash	0.43
Rainfall - Total Rainfall	0.84
Temperature - Nitrogen	0.45
Temperature - Phosphate	0.53
Temperature - Potash	0.53
Temperature - Total Rainfall	-0.27
Nitrogen - Phosphate	0.98
Nitrogen - Potash	0.92
Nitrogen - Total Rainfall	0.29
Phosphate - Potash	0.92
Phosphate - Total Rainfall	0.24
Potash - Total Rainfall	0.25

Impact of Rainfall on Crop Yield

According to results from the regression analysis, rainfall has a positive (albeit weak) correlation to total yield, thus indicating that even though rainfall is an important factor, it does not alone influence crop yields. Rather, other important factors that influence crop yield, including soil nutrients and temperature, will also significantly affect total yield.

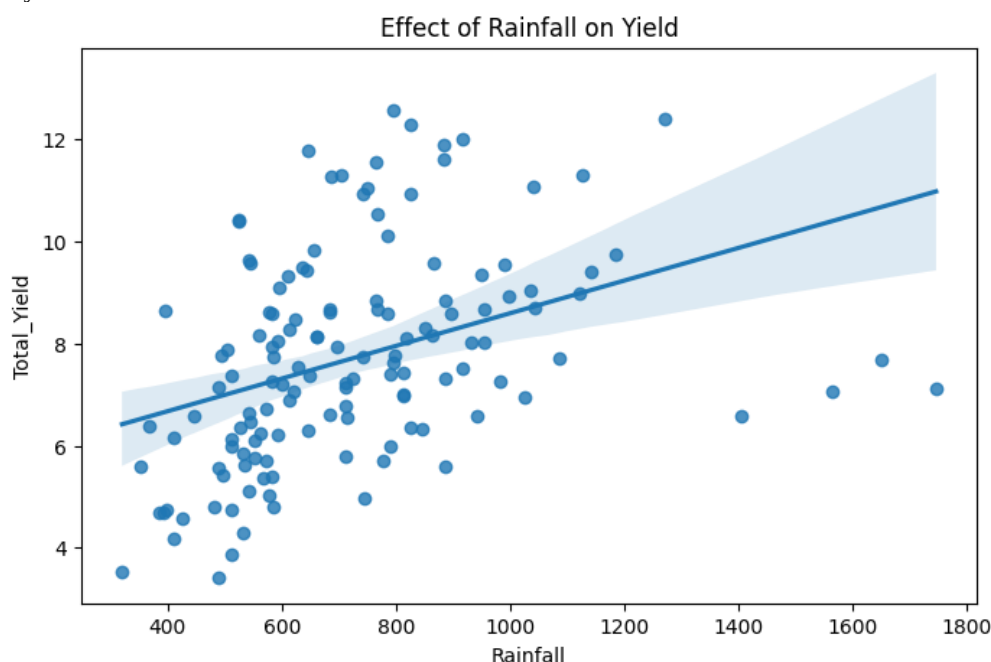


Figure 4 Effect of Rainfall on Total Crop Yield

The scatter plot shows a positive but weak linear relationship between total rainfall and total crop yield, with data points widely dispersed around the regression line. This confirms that rainfall alone explains only a limited share of yield variation, reinforcing the need to consider nitrogen, temperature, and other soil-climate variables jointly when modeling agricultural productivity.

Table 4 Rainfall - Yield Scatter Plot Data Summary

Metric	Value
Rainfall range plotted	approx. 300 to 1,800 mm
Total Yield range plotted	approx. 3.5 to 13
Regression coefficient (Rainfall, OLS)	0.0014
P-value (OLS)	0.232 (not significant)

MODEL PERFORMANCE COMPARISON

To evaluate predictive capability, multiple models were applied.

Table 5 Model Performance Comparison

Model	R ² Score
Linear Regression	0.249
Random Forest	0.327
XGBoost	0.038

With the highest R², the RF model excelled above all other models (0.327), demonstrating an ability to account for nonlinear associations within a data set. The linear regression model performed moderately compared with the XGBoost model which performed poorly; however, this was likely due, at least in part, to the small number of data points and the lack of complexity related to any given data feature.

Statistical Regression Analysis

The OLS regression model was used to assess statistical significance of variables.

Table 6 OLS Regression Results

Variable	Coefficient	P-value	Significance
Temperature	1.7369	0.000	significant
Nitrogen	0.0159	0.004	significant
Rainfall	0.0014	0.232	Not significant
Total Rainfall	0.0014	0.182	Not significant

The regression results indicate that temperature and nitrogen are statistically significant factors affecting crop yield ($p < 0.05$), whereas rainfall does not show significant influence. This highlights the importance of soil nutrients and climatic conditions over mere rainfall quantity.

Feature Importance Analysis

The analysis of the RF model feature importance was aimed at determining the main determinants of yield.

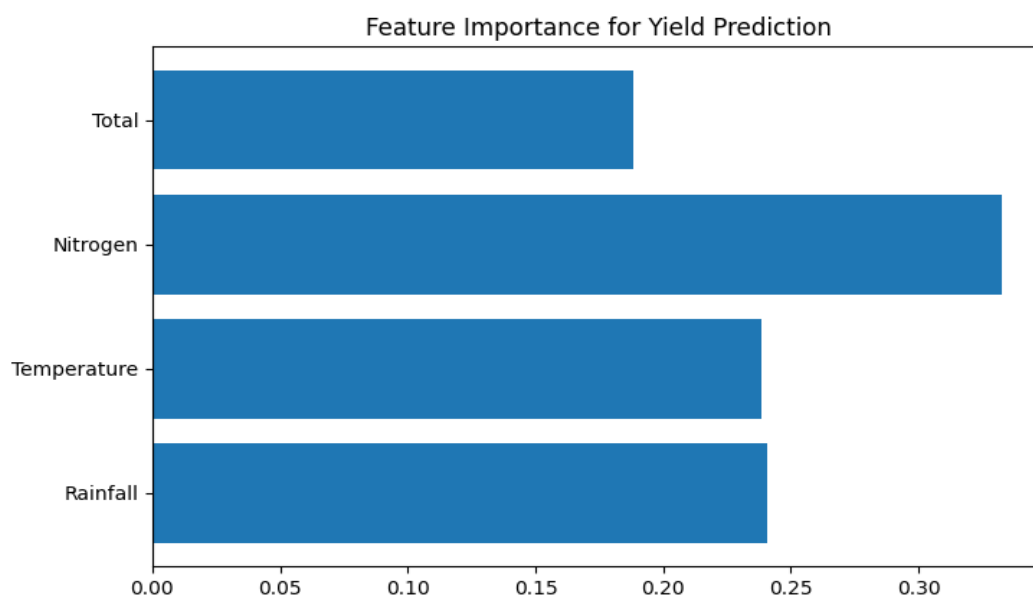


Figure 5 Feature Importance for Crop Yield Prediction

The overall finding of this study is that nitrogen and temperature had the greatest impact as compared to other variables in determining yield and that this result was confirmed through statistical analysis. The data also confirmed the importance of both soil fertility and climate to agricultural productivity.

Table 7 Feature Importance Values (approx., read from chart)

Feature	Importance (RF model)
Nitrogen	~0.33
Rainfall	~0.24
Temperature	~0.24
Total (Seasonal Rainfall Total)	~0.19

Explainable AI Insights (SHAP Analysis)

The analysis utilized SHAP for more interpretability of features within the model to find out how influential each feature is to model predictions, such as nitrogen being positively influential when having higher levels and with optimum temperature and less variability from rainfall than any other feature. This method of explainable AI also provides more trustworthiness to the model.

Table 8 SHAP Value Spread by Feature (approx., read from chart)

Feature	Approx. SHAP Range	Relative Impact
Temperature	-1.0 to +2.0	Highest
Nitrogen	-0.5 to +1.0	High
Rainfall	-0.5 to +0.5	Moderate
Total	-0.3 to +0.3	Lowest

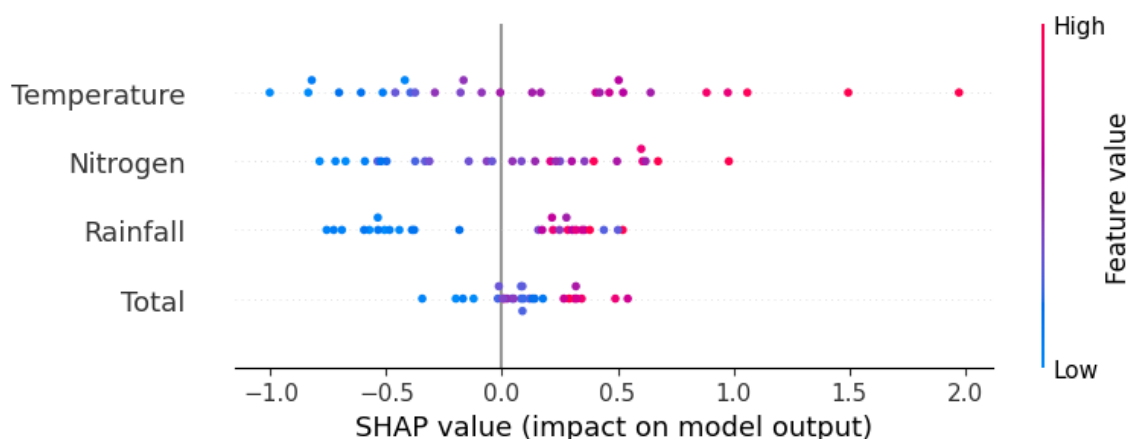


Figure 6 SHAP Summary Plot for Feature Contribution

DISCUSSION

This research shows that farmers in Maharashtra achieve higher yields due to various climatological and soil characteristics. Historically, rainfall was used as the prime factor affecting crop productivity, but the findings suggest temperature and nitrogen supply are of greater importance. The R^2 values across all study sites were low indicating that agro-ecosystems are very complex with additional unobserved factors having an effect on productivity, such as farm management, market forces and irrigation systems.

Providing a statistical and machine learning techniques approach combined with explainable artificial intelligence provides a means to effectively assess data for use within the agricultural sector. In conclusion, this research has advanced the ability of researchers to use multiple analytical techniques in conjunction with one another in order to achieve results that are more reliable, informative and interpretable compared to the traditional single analytical technique.

CONCLUSION & FUTURE SCOPE

A framework was developed to analyze yield and profitability in Maharashtra's agriculture sector utilizing data from multiple sources by using climate, soil, and crop data together. This analysis found that nitrogen and temperature were the primary environmental elements affecting crop yield, while rainfall alone had little direct influence. The model comparison showed that Random Forest models outperformed both boosting and linear models and demonstrated the necessity of capturing complex relationships within agricultural systems. The combination of using statistics, machine learning, and explainable AI (SHAP) provided both predictive power and transparency, thereby improving the validity of the results in practice. However, the relatively low predictive accuracy indicates that other contributing factors such as irrigation, market dynamics, and farm management techniques not contained in the dataset are affecting agricultural productivity. Future research can continue developing instruments that leverage real-time data, increase dataset size and representation across regions, and employ better predictive models through advanced deep learning or hybrid optimization in precision agriculture.

REFERENCES:

1. Marapelli, B., Anamalamudi, L., Potluri, C. S., Carie, A., & Anamalamudi, S. (2023, June). Enhancing Agricultural Decision-Making Through Machine Learning-Based Crop Yield Predictions. In *International Conference on Data Science and Network Engineering* (pp. 209-224). Singapore: Springer Nature Singapore.
2. Swarnkar, S. K., Dewangan, L., Dewangan, O., Prajapati, T. M., & Rabbi, F. (2023, September). AI-enabled crop health monitoring and nutrient management in smart agriculture. In *2023 6th international conference on contemporary computing and informatics (IC3I)* (Vol. 6, pp. 2679-2683). IEEE.
3. Suguna, K., Murali, P., Ayyasamy, P., & Obuli, O. (2024). Crop recommendation system using machine learning algorithm. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2, 1237-1242.
4. Pukrongta, N., Taparugssanagorn, A., & Sangpradit, K. (2024). Enhancing crop yield predictions with PEnsemble 4: IoT and ML-driven for precision agriculture. *Applied Sciences*, 14(8), 3313.
5. Alazzai, W. K., Obaid, M. K., Abood, B. S. Z., & Jasim, L. (2024). Smart agriculture solutions: Harnessing ai and iot for crop management. In *E3S Web of Conferences* (Vol. 477, p. 00057). EDP Sciences.
6. Mana, A. A., Allouhi, A., Hamrani, A., Rehman, S., El Jamaoui, I., & Jayachandran, K. (2024). Sustainable AI-based production agriculture: Exploring AI applications and implications in agricultural practices. *Smart Agricultural Technology*, 7, 100416.

7. Ma, X., Tong, J., Huang, W., & Lin, H. (2023). Characteristic mango price forecasting using combined deep-learning optimization model. *PloS one*, 18(4), e0283584.
8. Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE access*, 8, 86886-86901.
9. Patil, D. R., Deshmukh, M. S., Gebre, A. A., Thorat, P. S., & Sutar, S. S. (2026). Assessment of agricultural sustainability in Solapur District: A machine learning model for index prediction and indicators assessment. *Environment, Development and Sustainability*, 1-27.
10. Krishna, P. A., Narayana, G. V., Kotha, S. K., & Pattnayak, D. (2026, January). Machine Learning Based Agricultural Price Forecasting for Major Food Crops in India Using Environmental and Economic Factors. In *Biology and Life Sciences Forum* (Vol. 54, No. 1, p. 7). MDPI.
11. Sharma, N., Yadav, G., Ghule, V. R., Yache, Y. K. A., Solunke, P., Tati, R., & Khade, O. (2025, October). Predictive modeling of crop production in Maharashtra, India: A machine learning approach. In *AIP Conference Proceedings* (Vol. 3325, No. 1, p. 050025). AIP Publishing LLC.
12. Wanjale, K., Khatavkar, S., Khatal, A., Kharade, N., Kharate, O., Khedkar, S., & Khedekar, S. (2025, December). Data-Driven Millet Farming: A Machine Learning Approach to Crop Selection and Yield Prediction in Maharashtra. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-7). IEEE.
13. Phatangare, S., Laddha, A., Bambal, S., Borhade, B., & Atram, P. (2024, October). A Data-Driven Approach to Crop Yield and Market Price Prediction. In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 805-810). IEEE.
14. Chakre, S., Chaudhari, S., Chavhan, N., Gondhalekar, A., Sukte, C., & Magar, H. (2024, November). Data-Driven Crop Price Prediction: Machine Learning Approach for Agricultural Market Insights. In *International Conference on Data Science, Computation and Security* (pp. 271-279). Singapore: Springer Nature Singapore.
15. Jayaraman, A., Ramu, P., Rajan, S. C., & Thole, S. P. (2023). Data driven analysis of social capital in Farmer Producer Companies. *Heliyon*, 9(7).
16. Patil, P., Athavale, P., Bothara, M., Tambolkar, S., & More, A. (2023). Crop selection and yield prediction using machine learning approach. *Current Agriculture Research Journal*, 11(3).
17. Gupta, S., Geetha, A., Sankaran, K. S., Zamani, A. S., Ritonga, M., Raj, R., ... & Mohammed, H. S. (2022). Machine learning-and feature selection-enabled framework for accurate crop yield prediction. *Journal of Food Quality*, 2022(1), 6293985.
18. Bali, N., & Singla, A. (2022). Emerging trends in machine learning to predict crop yield and study its influential factors: A survey. *Archives of computational methods in engineering*, 29(1), 95-112.