International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 3, Issue 12, 2024

Development and Validation of an Ensemble Machine Learning Model for Enhanced Crop Yield Prediction

DOI: https://doi.org/10.5281/zenodo.14557299

Mustapha Ismail[1], Fauziyya Said Muhammad[2], Mohammed Mansur Ibrahim[3]

[1,2] Department of Computer Science, Gombe State University, Gombe, Nigeria
[3] Directorate of Information and Communication Technology, Federal University of Kashere, Gombe, Nigeria

Abstract

Accurate crop yield prediction is essential towards effective agricultural planning and food security for the growing population. This study aimed to develop and evaluate an ensemble machine learning model for crop yield prediction focusing on improving predictive accuracy and providing actionable insights for agricultural decision-making. The study utilized three machine learning algorithms – Decision Tree, Random Forest, and XGBoost. An ensemble approach using XGBoost was employed to combine the predictions of these algorithms, resulting in an R-squared (R^2) value of 0.99, MAE of 608.06 and MSE of 692453.82 showcasing the superior performance of the ensemble model compared to individual algorithms. The ensemble model's high accuracy demonstrates its potential for improving crop yield predictions. The model was further integrated into a user-friendly android application to assist farmers and agricultural stakeholders in making informed decisions.

Keywords: *Crop Yield, Random Forest, Decision Tree, XGBoost, Android Application.*

I. INTRODUCTION

Crop yield prediction plays a critical impact on the agricultural decision-making to enhance resource consumption and food supply chain. According to Hatfield et al., (2019), such technology equips farmers with accurate predictions that enable them to make informed planting decisions, optimize input management strategies, and reduce risks associated with the total demand from crop yields. By benefiting the individual farmer, crop yield prediction also benefits the broader food security at regional and global level, guiding policy interventions that can curb food shortages and associated malnutrition (Morton et al., 2017).

Historically, crop yield prediction has been approached with traditional statistics. Traditional statistical models for predicting crop yields often utilize regression analysis to correlate crop yields with different environmental factors, such as climatic variables, soil characteristics or agricultural practices. (Lobell et al., 2010; Tao et al., 2008). However, simple statistical models often fail to capture the complex non-linear relationships between crop yield and the factors that influence it (Khaki & Wang, 2019). Crop yield prediction plays a crucial role in ensuring food security, resource optimization, and risk management in agriculture. Accurate forecasts enable farmers, policymakers, and other stakeholders to make informed decisions regarding crop management practices, resource allocation, and market strategies. By leveraging machine learning techniques for yield prediction, the study provides a valuable tool for enhancing agricultural productivity and sustainability. (Satpathi et al., 2023).

Machine learning models, such as random forests, support vector machines, and artificial neural networks, have been widely applied to crop yield prediction tasks, often outperforming traditional statistical models (Everingham et al., 2016; Kamir et al., 2019). Individual Machine learning models may still suffer from limitations, such as overfitting, bias, and variance issues, which can lead to suboptimal performance (Zhou, 2012). To address these limitations and further improve predictive performance, ensemble learning and hybrid machine learning approaches have gained increasing attention in the field of crop yield prediction (Faith et al., 2021; Moradi et al., 2019).

Ensemble learning techniques have gained popularity in crop yield prediction due to their ability to improve prediction accuracy by combining multiple models (Polikar, 2012). Ensemble learning involves combining multiple individual machine learning models, each trained on the same data but with different initializations or configurations, to create a more robust and accurate predictor. Bagging, boosting, and stacking are common ensemble methods that leverage the diversity of base models to reduce prediction errors and enhance robustness (Dietterich, 2000). Ensemble models have been shown to outperform individual algorithms in various domains, including agriculture, by capturing complementary patterns and reducing overfitting (Sugiura et al., 2019).

Several studies have explored the application of machine learning techniques for crop yield prediction, however, there are limitations in terms of the methodologies and datasets used. Many studies have focused on individual machine learning algorithms, such as random forest, support vector machines, or neural networks. The performance of these individual models can be further improved by leveraging ensemble and hybrid approaches, which combine the strengths of multiple models and capture complex relationships in the data. Most existing research lacks the interpretability of the machine learning models used for crop yield prediction.

In this study, we leveraged the ensemble approach using three machine learning algorithms: Decision Tree, Random Forest, and XGBoost. By combining the predictions of these diverse models, we want to improve prediction accuracy and robustness thereby enhancing the reliability of crop yield forecasts. The best model was integrated into an Android application for easy accessibility.

II. LITERATURE REVIEW

Gulati and Jha (2020) explores using machine learning techniques like regression and ensemble methods to predict crop yields in India. The authors gathered agricultural data on crops, rainfall, temperature, and soil conditions from different Indian states. They tested algorithms including Linear Regression, Random Forest, Gradient Boosting and others on the dataset. The gradient boosting regressor achieved the highest accuracy with cross-validation score of 87.9% for predicting yield. Random Forest was most accurate (98.9%) for predicting crop production. However, only regression and ensemble methods were evaluated. Deep learning approaches could also be effective. Pavithra et al. (2022) proposed an ensemble algorithm approach for crop yield prediction, comparing multiple machine learning techniques including Random Forest, AdaBoost Classifier, Gradient Boosting Classifier, and K-Nearest Neighbors. Their system aims to help farmers select optimal crops based on soil and weather parameters like temperature, rainfall, nitrogen levels, etc. The authors used feature scaling techniques for preprocessing and achieved 99.4% accuracy with their ensemble model. They suggest their approach could help increase crop yields and farmer incomes through more precise agriculture practices.

The study by Madhusudhan (2022) proposes a crop yield prediction system using machine learning algorithms like random forest, decision trees, K-means clustering, Bayesian networks, support vector machines, linear regression, and artificial neural networks. Meteorological data like temperature, rainfall, and soil conditions are fed into models like random forest to predict optimal crops for given conditions. Of the algorithms tested, the random forest algorithm achieved the highest accuracy of over 90% in experiments using an Indian crop dataset. Limitations include reliance on historical data which may not fully capture future weather variability. Khan et al (2022) explore using machine learning techniques like random forests to predict crop yields based on historical data such as weather, soil conditions, and past crop yields. Experiments were conducted using crop data from Tamil Nadu, India. The random forest algorithm provided the highest accuracy (92.81%) for yield prediction compared to other methods like logistic regression and Naive Bayes.

Kalpana et al. (2023) applied Random Forest algorithm to predict crop yield by considering the factors that affect the climate change such as weather, humidity, temperature, moisture and historical yield. Data were collected from different sources and then dataset was created from it. The dataset was cleaned and analyzed; the yield of the crop was calculated based on temperature, rainfall and acre. According to the authors, a user-friendly website was designed for crop prediction where a user only needs to input the climate data for their preferred crop. Similarly, Factors like weather, soil, policies etc. influence Indian crop yields, requiring agricultural improvements for economic growth and food security (Jhajharia et al., 2023). They applied Random Forest, SVM, Gradient Descent, LSTM, and Lasso Regression to predict yields of 5 key crops in 33 Rajasthan districts. Random Forest performed best with 96.3% accuracy. However, remote sensing data could be added to potentially improve model performance further. Another study proposes using machine learning techniques to predict crop yields (Kamalesh & Ragaventhiran, 2023). The motivation is to help farmers make better decisions about which crops to plant to maximize income. The paper provides a literature review showing that many models have been proposed but have limitations like lower accuracy. The limitations of neural networks include higher relative error while supervised learning algorithms have trouble with non-linear input-output relationships. The authors recommend developing more robust models that can accurately predict yields based on weather, crop disease, growth stage, etc. They test some regression algorithms to compare accuracy in predicting yields given weather and soil conditions.

Ahmed et al. (2023) uses the machine learning approach for predicting crop yields in Nigeria using historical data. The authors developed models using three techniques – Support Vector Machine, Random Forest and Decision Tree Classifier. The models were trained on dataset of crop yields from different agricultural regions in Nigeria. Decision Tree Classifier performed best with the highest accuracy for predicting crop yields in the rainy season for South East region. Comparative evaluation found the Decision Tree Classifier had the lowest error rate. However, the dataset size and variety of crops analyzed is unclear. Manjunath and Palayyan (2023) proposes a hybrid machine learning (ML) approach for crop yield prediction by combining decision tree (DT), XGBoost, and random forest (RF) models. Linear regression, DT, RF, support vector machine, and XGBoost are evaluated on a crop dataset from Kaggle. DT, RF, and XGBoost have the highest R^2 scores indicating strong correlation with crop yield. The hybrid model outperforms individual models with the highest R^2 score of 0.9847. It also has superior accuracy of 98.6% compared to existing methods. An easy-to-use "Crop Yield Predictor" interface is developed to allow practical utilization of the prediction model by farmers and policymakers. Ed-Daoudi et al. (2023) explores how Machine Learning (ML) can enhance crop yield forecasting in Morocco. They analyze various environmental factors' impact on crop yields and incorporates them into ML models. The study compares the effectiveness of ML algorithms like Decision Trees, Random Forests, and Neural Networks against traditional statistical methods for crop prediction. Results showed that ML algorithms surpassed statistical models in yield prediction accuracy. ML approaches achieved mean squared error (MSE) values of 0.10-0.23 and coefficient of determination (R^2) values of 0.78-0.90, while statistical models had MSE values of 0.16-0.24 and R^2 values of 0.76-0.84. The Feed Forward Artificial Neural Network performed best, with the lowest MSE (0.10) and highest R^2 (0.90). Therefore, this study would utilize three machine learning algorithms – Decision Tree, Random Forest, and XGBoost together with an ensemble approach using XGBoost to combine the predictions of these algorithms.

III. METHODS AND MATERIAL

The dataset used is a secondary data obtained from two open-source repository; Food and Agriculture Organization and Climate Knowledge Portal of World Bank. The data is split into training and testing part; 80% training and 20% testing. Hyperparameter turning was done using k-fold cross-validation to improve the performance and accuracy of the model and appropriate evaluation techniques were used. Crop yield predictor was developed as an android application integrating the machine learning model for easy accessibility by farmers, stakeholders, and policy makers for decision making purposes. Figure 1 shows the architecture of the proposed model.

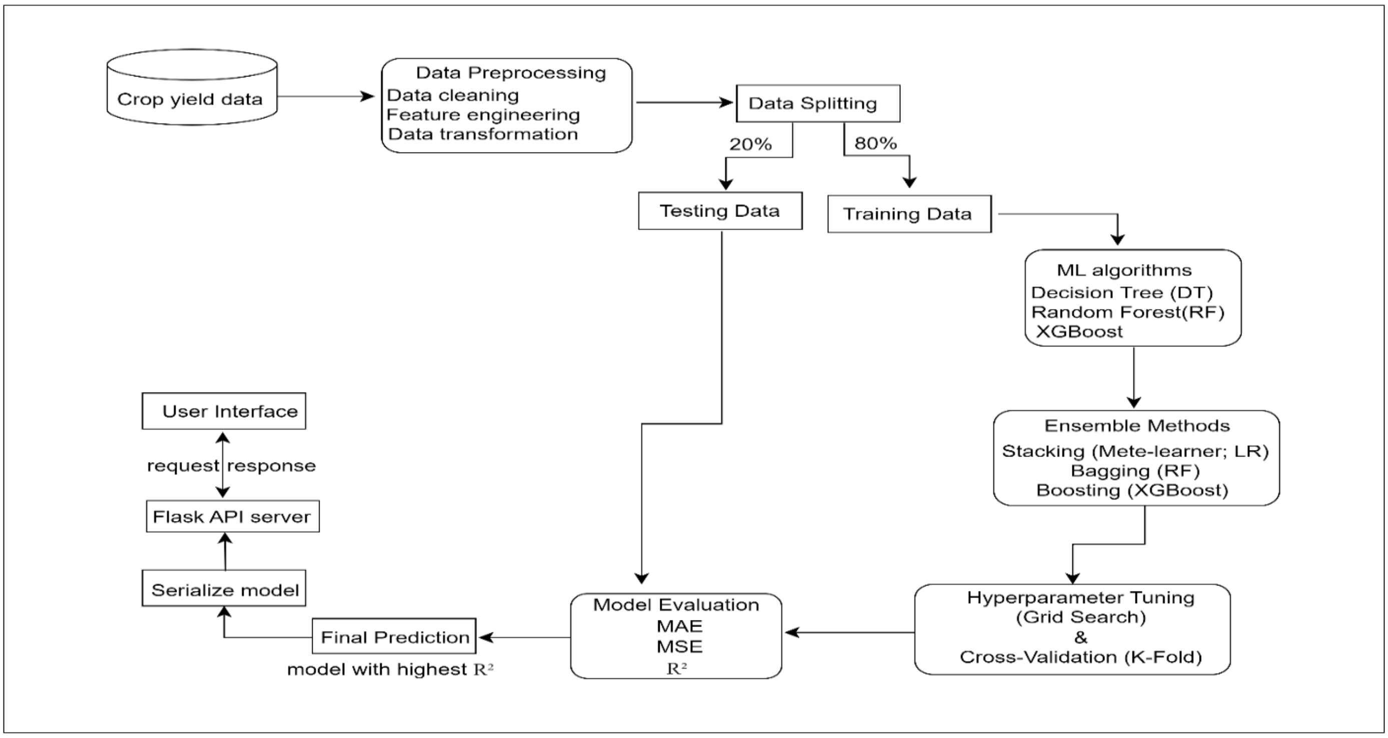

Fig 1 Architecture of Proposed Model

Table 1 Dataset Description

Features	Description	Dependent/Independent
Year	From 1990 – 2022	
Minimum Temperature	Average annual minimum surface temperature (World Bank)	Independent
Maximum Temperature	Average annual maximum surface temperature (World Bank)	
Item	37 unique crops cultivate in Nigeria (FAO)	
Precipitation	Annual precipitation/rainfall calculated in millimeters per year (FAO).	
Pesticides	Total annual pesticides used in Tones (FAO)	
Average Temperature	Calculated from Minimum and Maximum Temperature	
Yield	Total yield per year (hg/ha) (FAO).	Dependent

➢ *Dataset Description*

Table 1 shows the description of the dataset used. The data for this study were collected from two different open-source databases; Food and Agriculture Organization (FAO) and Climate Change Knowledge Portal by World Bank Group. The crop yield, Items and pesticides data are taken from FAO and the temperature and precipitation data are taken from the Climate Change Knowledge Portal. The data used is for the period of 32 years, from 1990 to 2022, and the pesticide data ranges from 1990 to 2021. The dataset contains 1200 samples and 8 columns.

➢ *Machine Learning Models*

Three machine learning algorithms were chosen; Decision Tree, Random Forest, and XGBoost based on their performance in the literature reviewed (Manjunath and Palayyan, 2023) and their ability to handle complex relationships. For the ensemble model, Stacking, bagging, and boosting were tested and the best was chosen among them.

- Decision trees are a form of supervised learning algorithm widely employed in machine learning for outcome prediction and modeling based on input data. This hierarchical structure consists of nodes and branches, where internal nodes evaluate attributes, branches represent attribute values, and leaf nodes provide final decisions or predictions. Decision trees are versatile, addressing both regression and classification problems, their ability to handle diverse data types and capture non-linear relationships aligns well with the multifaceted nature of agricultural data. However, careful tuning and pruning are necessary to ensure the model remains both accurate and interpretable. For parameter tuning, maximum depth of the tree (max_depth), the minimum samples required to split a node (min_samples_spit) and the minimum samples leaf (min_samples_leaf) were optimized using grid search with cross-validation. The final model used a max_depth of 10.
- Random Forest is an ensemble learning method applicable to both classification and regression tasks. Its core principle involves creating multiple decision trees and aggregating their predictions to achieve a more accurate and robust result. In the context of crop yield prediction, RF's ability to handle complex, non-linear relationships and provide feature importance makes it particularly suitable for agricultural applications, where numerous factors interact to influence yield outcomes. Key parameters such as the number of trees (n_estimators), maximum depth of the trees (max_depth), and minimum samples required to split a node (min_samples_split) were optimized using grid search with cross-validation. The mathematical representation of the RF model is:

$$RF(x) = (1/N) * \sum(DecisionTree\_i(x)) \quad (1)$$

Where:

RF(x) is the predicted outcome for input features x

N is the number of trees in the ensemble

DecisionTree_i(x) represents the i-th tree's prediction

- XGBoost (eXtreme Gradient Boosting) is a powerful and widely adopted machine learning algorithm known for its effectiveness in handling large datasets and achieving state-of-the-art performance in various tasks, including classification and regression. In agricultural applications, XGBoost's ability to handle complex data relationships, provide feature importance, and achieve high predictive accuracy makes it a valuable tool for crop yield prediction and other agricultural modeling tasks. However, users should be aware of the trade-offs between its power and complexity, and consider the interpretability needs of their specific application. Parameters such as learning rate (eta), maximum depth (max_depth), the number of boosting rounds (n_estimators), the fraction of samples used to train individual tree (subsample) and the fraction of features to be randomly sampled for each tree (colsample_bytree) were tuned using grid search with cross-validation.

➢ *Evaluation Metrics*

- Mean Squared Error (MSE) is calculated as the average of the squared differences between predicted and actual values. For a set of n predictions and corresponding observed values, MSE is computed as

$$MSE = (1/n) * \Sigma(Yi - \hat{Y}i)^2 \quad (2)$$

Where Yi represents observed values and $\hat{Y}i$ denotes predicted values. A lower MSE indicates better model performance, with zero representing a perfect model.

- Mean Absolute Error (MAE) measures the average magnitude of prediction errors without considering their direction. It is calculated as the sum of absolute differences between predicted and actual values, divided by the sample size:

$$MAE = (1/n) * \Sigma|Yi - \hat{Y}i|. \quad (3)$$

MAE provides a straightforward interpretation of the average prediction error in the original units of the target variable.

- The Coefficient of Determination (R^2) assesses how well variations in one variable can be explained by differences in another. It measures the strength of the linear relationship between variables, with values ranging from 0.0 to 1.0. An R^2 of 1.0 indicates perfect correlation, making it a reliable metric for evaluating a model's predictive power. R^2 helps quantify the proportion of variance in the dependent.
- Variable that is predictable from the independent variable(s).

$$R^2 = 1 - (SSres / SStot) \quad (4)$$

Where:

$SS_{res} = \Sigma(Y_i - \hat{Y}_i)^2$ (Sum of squares of residuals)

$SS_{tot} = \Sigma(Y_i - \bar{Y})^2$ (Total sum of squares)

Y_i = actual value

$\hat{Y}_i$ = predicted value

$\bar{Y}$ = mean of actual values

IV. RESULT AND DISCUSSION

Table 2 presents a comprehensive summary of performance metrics for all implemented models. The proposed ensemble boosting model demonstrates superior performance, exhibiting the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE), coupled with the highest coefficient of determination (R^2). This exceptional performance can be attributed to the model's ability to synergistically leverage the strengths of diverse individual algorithms, effectively mitigating bias and variance while circumventing overfitting issues. Furthermore, the ensemble approach excels in capturing intricate, non-linear relationships among the predictor variables.

The notably low MAE value achieved by the ensemble model is particularly significant, as it indicates a high degree of accuracy in yield prediction. MAE, being less sensitive to outliers compared to MSE, provides a robust measure of average model error in the original units of the target variable. This enhanced predictive capability has substantial implications for agricultural planning and resource allocation.

Table 2 Summary of Performance Metrics

Model	**MAE**	**MSE**	**R^2**
Decision Tree	8376.08	141072679.62	0.8835
Random Forest	4298.65	39420223.82	0.9674
XGBoost	2113.05	18351980.55	0.9848
Stacking	2382.56	19378093.72	0.9840
Bagging	2177.05	17946266.35	0.9852
Boosting	608.06	692453.82	0.9994

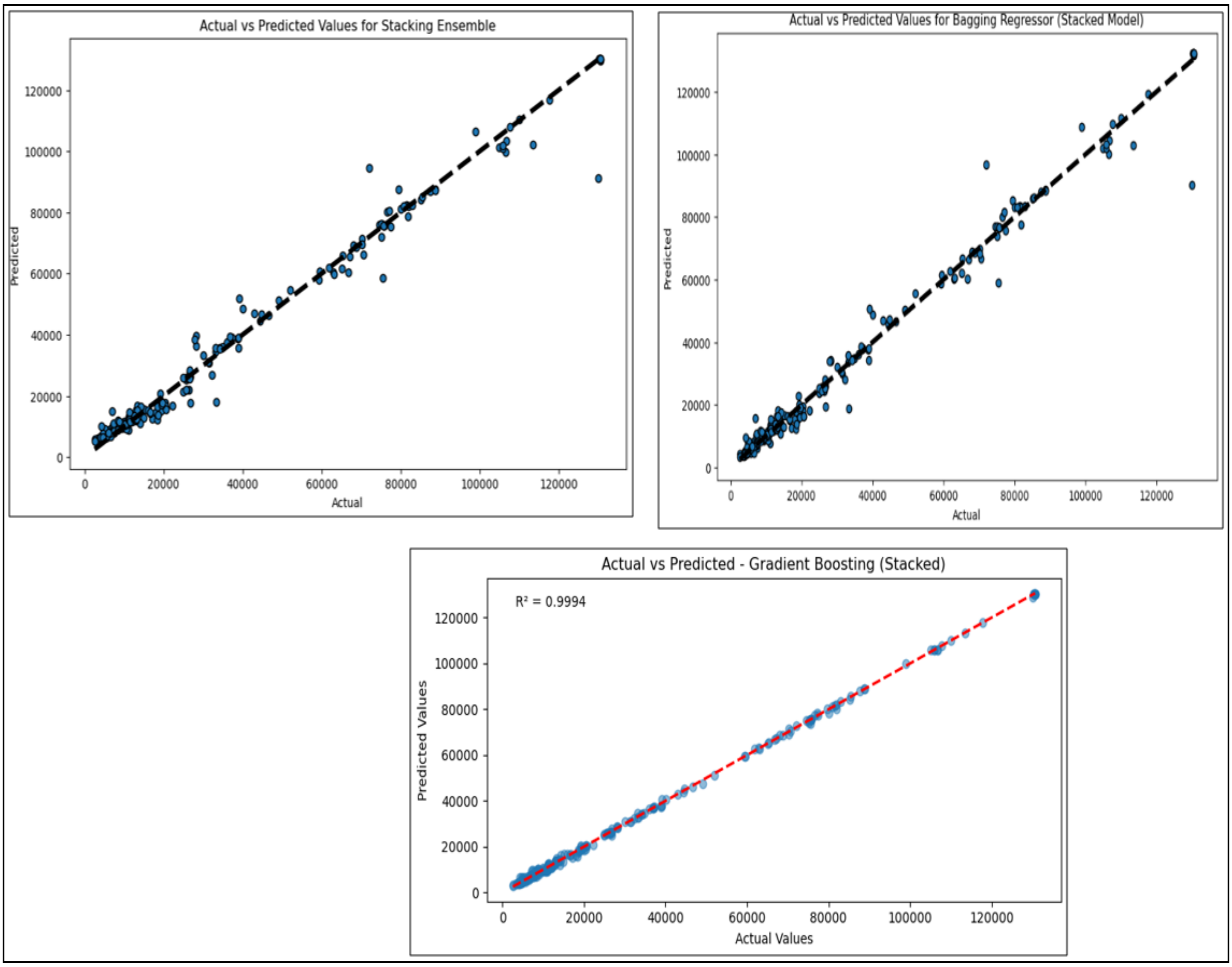

Fig 2 Predicted and Actual Yield Comparison of the Ensemble Models

➢ *Comparison of Cross-Validation Score with Initial Performance Metrics*

The comparison of the initial R^2 scores from train-test split with the mean scores from 5-fold cross validation is presented in the table below. All models show some level of overfitting as indicated by higher initial R^2 scores compared to the mean CV R^2 scores. DT shows the largest relative drop in performance with a 3.35% decrease in R^2 score, this indicates that DT is most prone to overfitting among the three algorithms. RT also shows a decrease of 2.74% in R^2 score, while this model also overfits, the ensemble nature of random forest helps in reducing the overfitting compared to DT. XGBoost shows the smallest relative decrease of 2.48%, indicating the best generalization among the three models. XGBoost's robust learning process, which includes regularization parameters (like colsample_bytree and subsample), helps mitigate overfitting.

Table 3 Comparison of CV R2 with Initial Performance Metrics

Model	**Initial R^2 Score**	**Mean CV R^2 Score**
Decision Tree	0.8835	0.85
Random Forest	0.9674	0.94
XGBoost	0.9848	0.96

➢ *Feature Importance Analysis*

Figure 3 visualizes the relative importance of each feature in predicting crop yield using random forest model. Precipitation is the most important feature in predicting crop yield according to the feature importance score (0.24). This indicates that the variations in precipitation levels have the greatest impact on crop yield predictions. Adequate water supply through precipitation is crucial for growth, affecting soil moisture, nutrient availability, and overall plant health. The second most importance feature is the year (0.18). this suggests that temporal factors, such as advancements in agricultural techniques, changes in crop varieties, and yearly climatic variations, significantly influence crop yield. This feature captures the trend and changes over time. Minimum temperature is the third most important feature (0.16). Temperature extremes, particularly low temperatures, can affect germination, growth rates, and susceptibility to diseases. This indicates that colder weather conditions might have a notable impact on crop yields. The use of pesticides is another significant feature (0.14). Effective pest control can enhance crop yield by reducing damage caused by pests and diseases. In addition, the important score suggests that while pesticides play a role, other factors like precipitation and minimum temperature have a more substantial impact. Average temperature is also an important factor (0.13). This feature likely captures the general climate conditions throughout the growing season, affecting the overall growth and health of crops. Maximum temperature (0.12), although the least important among the listed features, still plays a significant role. High temperatures can lead to heat stress, affecting photosynthesis and crop growth. However, its lower importance score relative to other features suggests that extreme high temperatures may not be as critical as precipitation or minimum temperature.

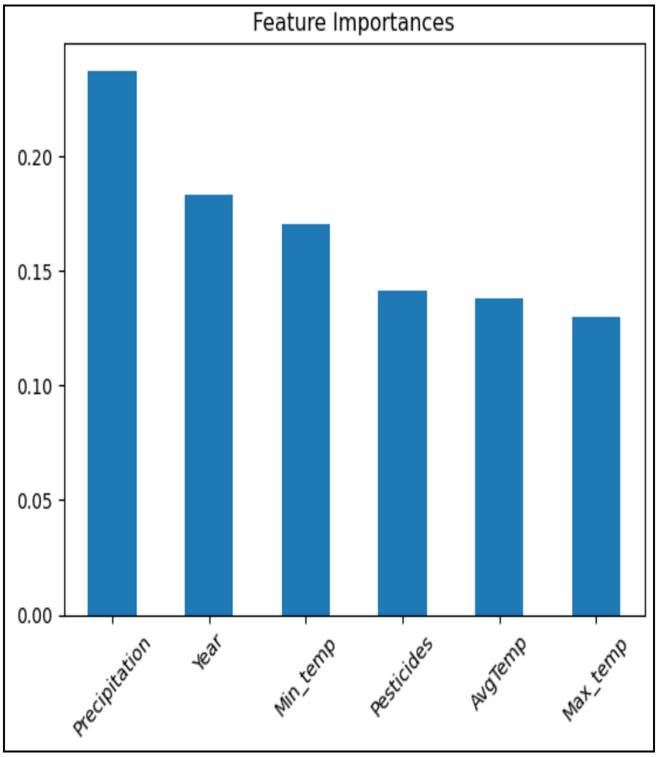

Fig 3 Feature Importance Plot

➢ *Model Selection Justification*

The boosting ensemble shows the best performance with an R^2 score of 99.9% coupled lower error metrics, therefore, the boosting ensemble is selected for further analysis and deployment in the android application.

The primary purpose of the Android application is to provide farmers and agricultural planners with an easy-to-use tool for predicting crop yields based on various environmental and management factors.

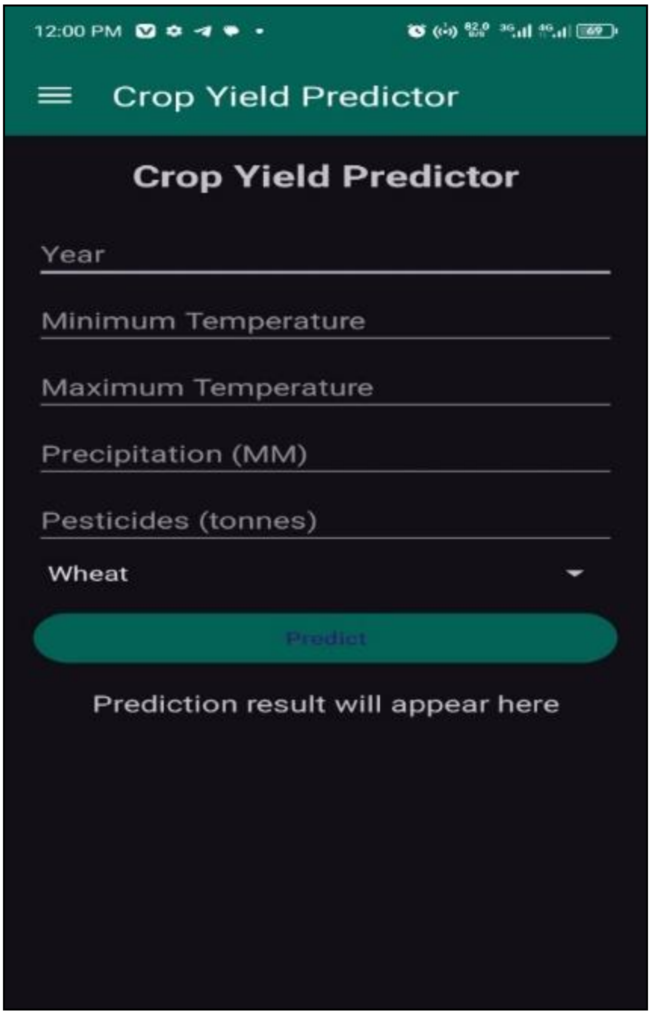

Fig 4 Crop Yield Predictor User Interface

V. CONCLUSION

The study successfully developed a crop yield prediction model using ensemble learning techniques, particularly boosting. The model's outstanding performance, characterized by a high R-squared value (0.99) and low error metrics (MAE of 608.06 and MSE of 692453.82), highlights the effectiveness of ensemble methods in capturing complex relationships within the dataset. The XGBoost algorithm is particularly noted for its robustness in crop yield prediction, effectively capturing underlying patterns without overfitting. Additionally, precipitation, year, and minimum temperature were identified as the most influential factors affecting crop yield, underscoring the vital role of climatic conditions in agricultural productivity. The developed Android application will serve as a practical tool for farmers and agricultural planners, facilitating real-time crop yield predictions and informed decision-making for optimizing resource allocation and management strategies.

Future studies should explore additional features such as soil quality, pest prevalence, and economic factors to further improve the accuracy and robustness of crop yield prediction models. While the current model performed exceptionally well on the provided dataset, future research should validate its generalization capabilities across different geographical regions and crop types. The findings should be communicated to policymakers to support the development of data-driven agricultural policies that promote sustainable farming practices and optimize crop yields.

REFERENCES

[1]. Ahmed, A., Adewumi, S. E., Yemi-Peters, V., & Peters, V. Y. (2023). Seasonal crop yield prediction in Nigeria using machine learning technique. *Journal of Applied Artificial Intelligence*, 4(1), 9-20. https://doi.org/10.48185/jaai.v4i1.728

[2]. Chandol, M. K., Mohanraj, E., Umaselvi, M., & Sankar, K. (2021). Enhancement of agriculture-based crop yield prediction using R tool and machine learning. *Turkish Online Journal of Qualitative Inquiry*, 12(7), 5155-5165.

[3]. Dietterich, T. G. (2000). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg.

[4]. Ed-daoudi, R., Alaoui, A., Ettaki, B., & Zerouaoui, J. (2023). Improving crop yieldpredictions in Morocco using machine learning algorithms. *Journal of Ecological Engineering*, 24(6), 392-400. https://doi.org/10.12911/22998993/162769

[5]. Everingham, Y., Smyth, C., & Inman-Bamber, N. (2016). Ensemble data mining approaches to forecast regional sugarcane crop production. Field Crops Research, 200, 25-32.https://doi.org/10.1016/j.fcr.2016.10.005

[6]. Faith, N. O., Uribio, P. O., & Nyongesa, H. O. (2021). Ensemble machine learning approach for crop yield prediction. *Journal of Agricultural Informatics*, 12(1), 29-39. https://doi.org/10.17700/jai.2021.12.1.642

[7]. Gulati, P., & Jha, S. K. (2020). Efficient crop yield prediction in India using machine learning techniques. *International Journal of Engineering Research and Technology*, 8(10), 24-26

[8]. Hatfield, J. L., Boote, K. J., Kimball, B. A., Ziska, L. H., Izaurralde, R. C., Ort, D., ... & Nakicenovic, N. (2019). Climate impacts on agriculture: Implications for crop production. *Agronomy*, 9(6), 1-37.

[9]. Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. Procedia Computer Science, 218, 406-417. https://doi.org/10.1016/j.procs.2023.01.023

[10]. Kamalesh, M., & Ragaventhiran, J. (2023). Crop yield prediction using machine learning. *International Research Journal of Innovations in Engineering and Technology, 7*(5), 297–299. https://doi.org/10.47001/IRJIET/2023.705041

[11]. Kamir, E., Waldner, F., & Hochman, Z. (2019). Estimating wheat yields in Australia using climate records, satellite image time-series and machine learning methods. ISPRS Journal of Photogrammetry and Remote Sensing, 160, 124-135. https://doi.org/10.1016/j.isprsjprs.2019.12.012

[12]. Kalpana, P., Prem, I. A., Mary, S. J. R., & Rani, R. S. A. V. (2023). Crop yield prediction using machine learning. REST Journal on Data Analytics and Artificial Intelligence, 2(1), 16-20. https://doi.org/10.46632/jdaai/2/1/3

[13]. Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621. https://doi.org/10.3389/fpls.2019.00621

[14]. Khan, P. A., Hussain, M. S., Ali, M. M., & Khan, M. Z. A. (2022). Crop yield prediction using machine learning algorithms. *Journal of Emerging Technologies and Innovative Research*, 13(5), 125-133.

[15]. Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150(11), 1443-1452.

[16]. Manjunath, C. M., & Palayyan, B. P. (2023). An efficient crop yield prediction framework using hybrid machine learning model. *Revue d'Intelligence Artificielle*, *37*(4), 1057–1067. https://doi.org/10.18280/ria.370428

[17]. Manivasagam, M. A., Sumalatha, P., Likitha, A., Pravallika, V., Satish, K. V., & Sreeram, S. (2022). An efficient crop yield prediction using machine learning. *International Journal of Research in Engineering, Science and Management*, 5(3), 106-111.

[18]. Morton, J. F., Schroeder, R. A., & Cook, B. I. (2017). Climate trends and global crop production since 1980. *Science Advances*, 3(5), e1601202.

[19]. Moradi, R., Arabi, B., & Taki, O. (2019). Crop yield prediction using hybrid approach of staged-clustering and kernel simple regression. Computers and Electronics in Agriculture, 167, 105037. https://doi.org/10.1016/j.compag.2019.105037

[20]. Pavithra, S. R., Priya Darshini, G., Maheswari, M., & Malathi, A. (2022). An ensemble algorithm for crop yield prediction. International Journal of Advance Research and Innovative Ideas in Education, 8(3), 4647-4656.

[21]. Polikar, R. (2012). Ensemble learning. In Ensemble machine learning (pp. 1-34). Springer, Boston, MA.

[22]. Tao, F., Yokozawa, M., Liu, J., & Zhang, Z. (2008). Climate–crop yield relationships at provincial scales in China and the impacts of recent climate trends. *Climate Research*, 38(1), 83-94. https://doi.org/10.3354/cr00771

[23]. Satpathi, A., Setiya, P., Das, B., Nain, A. S., Jha, P. K., Singh, S., & Singh, S. (2023). Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India. *Sustainability*, *15*(3), 2786. https://doi.org/10.3390/su15032786

[24]. Sugiura, R., Hirafuji, M., & Nakamura, R. (2019). Comparative analysis of machine learning techniques for prediction of rice yield. Computers and Electronics in Agriculture, 158, 417-426.

[25]. Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. Chapman and Hall/CRC. https://doi.org/10.1201/b12207