University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)

ISSN: 2714-3627

A Journal of the Department of Computer Science, University of Ibadan, Ibadan, Nigeria

Volume 14 No. 1, June, 2025

journals.ui.edu.ng/uijslictr
http://uijslictr.org.ng/
uijslictr@gmail.com

University of Ibadan
Journal of Science and Logics in ICT Research

Leveraging Machine Learning for Predicting Climate Change Impacts on Agricultural Productivity in Bayelsa State, Nigeria: A Pathway to Sustainable Solutions

[1] Nnodi J. T., [2] Okoro E.O.

[1] *Department of Computer Science, Federal University Otuoke, PMB 126, Yenagoa, Nigeria.*
[2] *Department of Environmental Management and Toxicology, Federal University Otuoke, PMB 126, Yenagoa, Nigeria.*
nnodijt@fuotuoke.edu.ng okoroee@fuotuoke.edu.ng

Abstract
Bayelsa State, Nigeria, faces increasing climate change impacts, including erratic rainfall, flooding, and food insecurity. Limited access to localized climate data further complicates agricultural decision-making. This study applies machine learning to predict climate change effects on agricultural productivity, offering strategies for resilience and sustainable farming. Historical climate and agricultural data from sources like the Nigerian Meteorological Agency (NiMET) were analyzed. A stacking ensemble machine learning model was developed to predict crop yields, using a Random Forest Regressor and XGBoost Regressor as base models, with a Linear Regressor as the meta-learner. The model was optimized using 5-fold cross-validation to enhance predictive accuracy. Model validation using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) demonstrated high accuracy, with an RMSE of 9,861.6786, an R^2 of 0.9866, and an MAE of 3,716.7995 hg/ha. These results indicate minimal deviation from actual crop yields, demonstrating a significant improvement over earlier models and confirming its reliability in predicting agricultural productivity. Findings highlight the potential of machine learning for informed decision-making among policymakers, farmers, and stakeholders. By leveraging AI-driven solutions, this study promotes agricultural resilience, sustainable development, and long-term food security in Bayelsa State.

Keywords: *Machine learning, XGBoost, Climate change, Random Forest, Agricultural productivity*

1. Introduction

Climate change has emerged as one of the most pressing global challenges, with far-reaching consequences for agriculture, particularly in vulnerable regions such as Bayelsa State, Nigeria. Rising temperatures, erratic rainfall patterns, and increasing incidences of extreme weather events threaten agricultural productivity, food security, and rural livelihoods. The loss of crops and farmland due to flooding reduces income and increases poverty among smallholder farmers, hindering progress towards Sustainable Development Goals related to poverty and hunger [11]. As a coastal state in the Niger Delta region, Bayelsa is uniquely susceptible to climate-induced disruptions, including flooding, soil degradation, and altered growing seasons, all of which significantly impact crop yields and livestock production. Given the region's heavy reliance on agriculture for economic sustenance, there is a growing need for predictive models that can assess and mitigate the effects of climate change on agricultural productivity.

Recent advancements in artificial intelligence (AI) and machine learning (ML) present a promising avenue for addressing these challenges. Machine learning models can analyze vast amounts of climate and agricultural data to detect patterns, predict future impacts, and provide actionable insights for farmers, policymakers, and agricultural stakeholders. By leveraging historical climate data, soil characteristics, crop yield records, and meteorological forecasts, ML-based predictive models can help optimize farming strategies, improve resilience, and enhance decision-making processes in the agricultural sector.

Nnodi J. T. and Okoro E. O. (2025). Leveraging Machine Learning for Predicting Climate Change Impacts on Agricultural Productivity in Bayelsa State, Nigeria: A Pathway to Sustainable Solutions. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 14 No. 1, pp. 173 – 187.

This study aims to develop and apply machine learning techniques to predict the impacts of climate change on agricultural productivity in Bayelsa State. By integrating climate variables, environmental factors, and agricultural data, the study seeks to generate reliable forecasts that can inform adaptive strategies for sustainable farming. The findings from this research will contribute to mitigating the adverse effects of climate change on agriculture and support evidence-based policymaking and resource allocation in the region.

1.1 Climate Change and Its Impact on Agriculture

Climate change has significant impacts on agriculture, both in terms of crop production and livestock management, because rising temperatures, changing rainfall patterns, and extreme weather events such as droughts and floods can reduce crop yields [3]. It has also significantly affected global agricultural productivity, with regions like Africa being particularly vulnerable due to their dependence on rain-fed farming. Studies have shown that rising temperatures, erratic rainfall, flooding, and increased pest infestations negatively impact crop yields and food security. In Nigeria, agricultural productivity is highly sensitive to climate variability, leading to reduced crop output and economic instability [1]. Bayelsa State, being a coastal region, faces additional challenges such as excessive flooding and soil salinization, which further threaten agricultural sustainability.

1.2 Applications of AI and ML in Agricultural Productivity Prediction

Artificial Intelligence (AI) and ML are increasingly being used to assess and enhance agricultural productivity. Studies have demonstrated that predictive models can analyze climate variables, soil conditions, and crop yields to provide insights into potential productivity fluctuations. In Nigeria, AI-driven solutions have been applied in precision agriculture, irrigation management, and crop disease detection, yet there remains limited research specifically focused on climate-induced productivity forecasting in Bayelsa State.

1.2.1 Using Machine Learning in Climate Change Prediction

Machine learning (ML) has emerged as a powerful tool for analyzing large datasets and making accurate climate predictions. Various ML algorithms, including decision trees, artificial neural networks (ANNs), and support vector machines (SVMs), have been successfully applied in climate studies [6]. These models utilize historical climate and agricultural data to forecast future trends, enabling better planning and adaptation strategies. For example, Random Forest and Long Short-Term Memory (LSTM) models have been used to predict rainfall and temperature variations, helping farmers adjust their planting schedules.

1.2.2 Challenges in Implementing AI-Driven Climate Solutions in Nigeria

Despite the potential benefits of ML in agricultural forecasting, several challenges exist. Data scarcity and poor quality remain major obstacles, as many developing regions lack comprehensive climate and agricultural records. Additionally, the adoption of AI-driven solutions among farmers is hindered by low digital literacy, inadequate infrastructure, and financial constraints. Addressing these challenges requires targeted policies, investment in data collection, and training programs to enhance AI adoption in the agricultural sector.

1.3 Research Gaps and New Approach Used

While existing studies highlight the potential of ML in climate change adaptation, limited research focuses on Bayelsa State and its unique agricultural challenges. Most AI-driven agricultural studies in Nigeria have concentrated on large-scale farming, with little emphasis on smallholder farmers who are most affected by climate variability. This study aims to bridge this gap by developing an ML-based predictive model tailored to Bayelsa’s environmental conditions, providing localized insights for farmers and policymakers to enhance agricultural resilience. The literature review underscores the significance of AI and ML in climate change prediction and agricultural productivity assessment. However, the limited application of these technologies in Bayelsa State presents an opportunity for research and innovation. Therefore, this study contributes by developing a robust ML-based model that offers accurate, data-driven insights to mitigate climate change's effects on agriculture in the region.

2. Related Works

Adejuwon *et al.* [1], based on Climate Variability and Agricultural Productivity in Nigeria examined the effects of climate variability on Nigeria's agricultural sector using statistical regression models to analyze historical temperature and rainfall data. The study advocated for climate adaptation strategies, such as improved irrigation and resilient crop varieties. Employing methodologies such as the chi-square test of independence, standardized anomaly index, multiple regression, z-distribution, and descriptive statistics, the research investigates the socioeconomic and climate influences on cocoa yield using data from 1999 to 2019. The findings underscore the significant effect of climate on cocoa yield, particularly in 1999 and 2000, with farmers attributing positive influences on factors like temperature, humidity, rainfall, sunshine, and wind speed but the data set was not large enough for accurate prediction.

Ashiegbu *et al.* [4] carried out a study that focused on the intersection of climate variability and agroforestry. The paper presents valuable insights into the impacts of climate variability on agricultural practices, but the study is limited to six local government areas in Ebonyi State, Nigeria which did not capture the broader regional or national trends in climate variability and agricultural practices, potentially limiting the generalizability of the findings to other areas. A larger and more diverse sample could yield more comprehensive results. The study also relied on self-reported data from farmers regarding their perceptions of climate variability and agricultural practices. This method can introduce biases, as farmers may overestimate or underestimate their experiences based on personal beliefs or external influences

Ologeh and Adesina [14] in their study finds a positive and significant relationship between the summed rainfalls of June, July, and August and annual maize yields over a thirty-five-year period. This indicates that specific seasonal rainfall is critical for maize production, which can inform agricultural practices and policies. Although the study emphasizes seasonal rainfall, it did not adequately address the impact of other climatic factors, such as temperature fluctuations and extreme weather events, which can also significantly affect crop yields. The interpretation of the correlation results may also be subject to bias, as the study primarily focuses on specific rainfall months. This selective focus might overlook other important climatic variables that could provide a more comprehensive understanding of crop yield dynamics.

Another research by Garg and Alam [7] highlights the significance of machine learning techniques in developing effective crop recommendation systems. These systems leverage various algorithms including Random Forest, Naïve Bayes, Support Vector Machines, and K-Nearest Neighbors. These methods analyze diverse data inputs such as soil properties and weather conditions. They analyzed soil and climatic conditions, ultimately guiding farmers in selecting the most suitable crops for their specific environments. But while the machine learning approach offers promising advancements in crop recommendations, challenges such as data quality and the need for continuous updates remain critical for maintaining system efficacy.

Week and Wizor [19] observed that the impact of flooding on agriculture in the Niger Delta is multifaceted, affecting food security, soil quality, and the livelihoods of smallholder farmers. They asserted that Flooding in the Niger Delta leads to acute food insecurity, with 75.3% of respondents in the study reporting scarcity of basic food post-flooding. The washing away of farmlands results in chronic food insecurity, affecting the livelihood of residents.

Okoro and Ofordu [15] also observed that the loss of crops and farmland due to climate change and flooding reduces income and increases poverty among smallholder farmers, hindering progress towards Sustainable Development Goals related to poverty and hunger. Amos and Okoro [2] reveals the rising temperatures, erratic rainfall, and frequent oil spills in Ogbia LGA, Bayelsa State, are reducing crop yields and fish stocks, threatening food security. It recommends climate-resilient farming, improved flood and fishery management, and increased community awareness to sustain livelihoods.

Awais *et al*. [5] highlighted the limitations of conventional statistical methods in providing timely and accurate soil texture and soil–water content (SWC) analysis, particularly in the face of varying climate and geospatial factors. The study explores how artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques can offer more robust, accurate, and real-time soil analysis for improved agricultural decision-making. The paper discusses the use of predictive models like random forests, support vector machines, and neural networks, alongside geostatistical techniques such as kriging and co-kriging, to enhance spatial data representation. It also evaluates challenges like false positives in SWC detection and promotes the integration of AI-driven systems for smart irrigation and global SWC database development. But there was no model development and training.

Nnodi and Obasi [9] presented a machine learning-based approach that significantly improves the detection of insider threats in corporate networks. By analyzing user behaviors and access patterns, the system can classify activities as normal or abnormal, providing early warnings for potential breaches. But the model considered only features relating to insider attacks on corporate networks.

Zidan and Febriyanti [20] developed an ML model to predict maize and rice yields based on climate variables. They compared different algorithms, including XGBoost, Decision Trees, and ANN, using agricultural datasets. But their model did not include integration of real-time weather monitoring data with the machine learning models to enhance prediction accuracy.

Nnodi *et al*. [9] addresses the lack of user-centric systems for web QoE prediction. A predictive model was developed to measure and monitor web user experiences. The model aids in identifying network bottlenecks and supports real-time user decisions. But the study used only random forest algorithm for training and validation.

Obasi and Nlerum [12] developed a model for the Detection and Prevention of Backdoor Attacks using CNN with Federated Learning. The model achieved an accuracy of 99.99% for training and 99.98 for validation.

Timadi and Obasi [16] worked on Integrating Zero-Trust Architecture with Deep Learning Algorithm to Prevent Structured Query Language Injection Attack in Cloud Database. Their proposed system utilizes a Feed-Forward Neural Network (FNN) to analyze database queries and detect potential SQL injection attacks. The model exhibits a precision score approximating 100% accuracy in the classification of queries deemed normal, while achieving a 94% rate of correct classification for queries indicative of SQL injection attacks.

Obasi and Stow [13] developed a Predictive Model for Uncertainty Analysis on Big Data Using Bayesian Convolutional Neural Network (CNN). The Bayesian CNN model uses a probability score in predicting uncertainties in big data. The result of Bayesian model shows a better result of 99.9% for both training and testing

3. Methodology

This research employed a supervised machine-learning predictive modeling approach. It combined quantitative approach, predictive design and computational experiment / ensemble learning to predict crop yield. The research begins by importing a dataset from a CSV file (yield_df.csv), which contains various crop yield and environmental features. Unnecessary columns were dropped to clean the dataset. New features were engineered to enhance the predictive power of the model. Data splitting was done using an 80/20 train-test split for model training.

By combining XGBRegressor, RandomForestRegressor, and a Linear Regression meta-learner with 5-fold cross-validation the dataset was trained. GridSearchCV was employed to optimize model parameters to achieve better performance. The model was evaluated by assessing the Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Error (MAE) values, while scatter plots were used to validate the predictions

3.1 Model Architecture

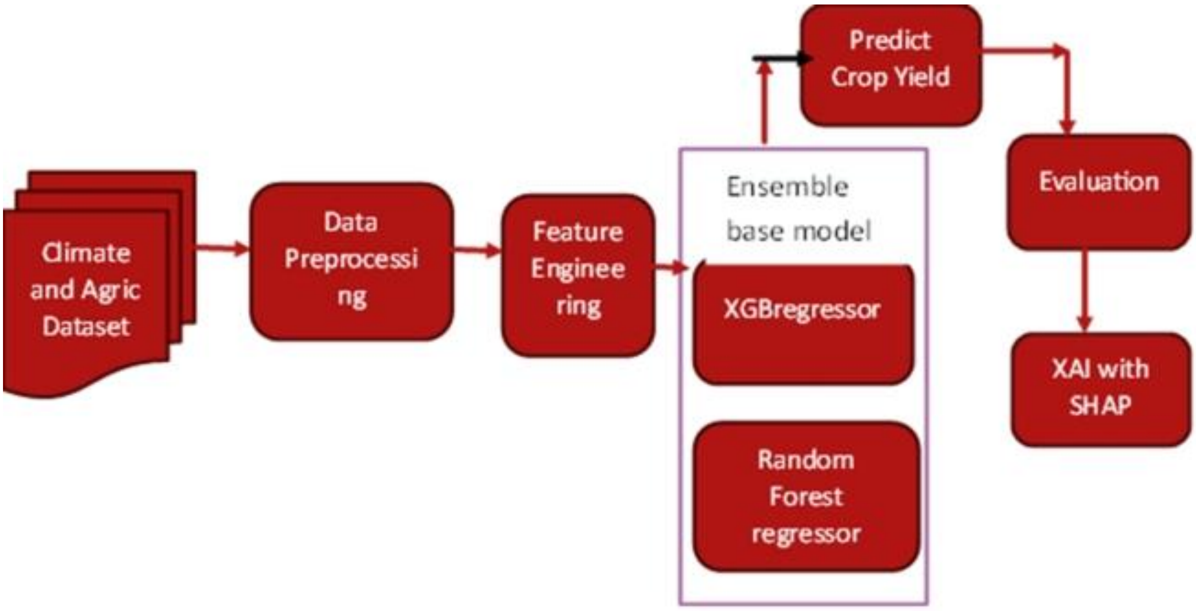

Figure 1: Proposed Model Architecture

Figure 1 shows the different components of the architecture and how they interact to accurately predict crop yields. Model architecture is broken down into distinct modules as follows:

Data Acquisition: Datasets comprising climate data and agricultural data were obtained from NiMET and FOA respectively with a total number of 28242 samples with a total of 113 features before data cleaning. After data cleaning, the number of rows was reduced to 24,241 with a total of 7 features.

The dataset consists of the following features:

a. Area: location under study.
b. Item: Type of crop.
c. Year: Year of data collection.
d. hg/ha_yield: Crop yield (target variable).
e. average_rain_fall_mm_per_year: Average yearly rainfall.
f. pesticides_tonnes: Pesticide usage in tonnes.
g. avg_temp: Average yearly temperature.

The research began by importing the dataset from a CSV file (yield_df.csv), which contains various crop yield and environmental features. An unnecessary columns were dropped to clean the dataset.

Data Preprocessing: In this module, the data was cleaned to drop the unnecessary features and remove outliers that may introduce noise in the proposed model

Feature Engineering: New features are engineered to enhance the predictive power of the model. These features include:

a. Relative Year (Year_diff): The number of years since the first observation, capturing temporal trends.
b. Interaction Term (rain_temp): The product of average rainfall and temperature, reflecting potential synergistic effects on crop growth.
c. Pesticide-Rainfall Ratio: A ratio of pesticides applied to the rainfall received, which may indicate agricultural practices or environmental stress.

Categorical Encoding: Categorical variables such as "Area" and "Item" are transformed using one-hot encoding to convert them into a numerical format suitable for machine learning algorithms.

Train-Test Split: The dataset is partitioned into training and testing sets using an 80/20 split after the one-hot encoding technique. This ensures that the model is trained on the majority of the data and evaluated on a separate subset to assess its generalization capability.

Stacking Ensemble Model: Two powerful tree-based regressors are employed as the base models, which include the XGBRegressor (A gradient boosting model known for its strong predictive performance) and RandomForestRegressor (An ensemble of decision trees that averages multiple predictions to reduce variance). A Linear Regression model is used as the meta-learner, which takes the predictions from the base models and learns to combine them optimally. The stacking ensemble is configured using 5-fold cross-validation to combine the strengths of the base learners, enhancing the overall predictive accuracy.

Grid Search Optimization: During this stage, hyperparameter tuning was performed using GridSearchCV to optimize key parameters. For XGBRegressor, the number of estimators and learning rate were tuned while the number of estimators and maximum depth were tuned for RandomForestvanRegressor. This grid search was conducted using a 5-fold cross-validation strategy, with the R^2 score as the optimization metric. The best combination of hyperparameters is selected to improve model performance and generalization.

Model Evaluation and Visualization: The optimized stacking model is evaluated on the test set using standard regression metrics comprising:

a. RMSE (Root Mean Squared Error) which measures the typical magnitude of prediction errors.
b. R^2 (Coefficient of Determination) which indicates the proportion of variance in crop yield that is captured by the model and
c. MAE (Mean Absolute Error) which reflects the average absolute error in predictions.

To visualize the result, a scatter plot of actual versus predicted crop yields was generated to visually assess the model’s performance and ensure that predictions align closely with the ground truth. This multi-stage methodology from data preprocessing and feature engineering to model stacking, hyperparameter tuning, and. outlier removal

4. **Results and Discussion**

4.1 Result of Key features’ Distribution before Outlier Removal

The histograms in figure 2 provide an overview of key features in the crop yield dataset before outlier treatment, including 'Year,' 'hg/ha_yield' (crop yield), 'average_rain_fall_mm_per_year,' 'pesticides_tonnes,' and 'avg_temp.' The 'Year' distribution appears multimodal, indicating data collection was focused on specific periods rather than being uniform. This temporal grouping has implications for structuring validation and testing strategies, particularly in time-series modeling. The 'hg/ha_yield' histogram is heavily right skewed, with most values clustering at lower yields and a long tail extending toward high yields.

This non-normal distribution suggests that models robust to skewness or transformations was necessary, and error metrics like RMSE should be considered carefully. Rainfall and pesticides also exhibit right-skewed distributions. Rainfall shows a main peak at lower values but extends into higher regions, indicating diverse climatic conditions in the dataset. This range is essential for models aiming to generalize across different agricultural zones.

Pesticide usage is similarly skewed, with most instances showing low usage and a few cases of high application, highlighting potential outliers that required attention. Temperature, in contrast, appears closer to multimodal distribution, suggesting the dataset includes observations from distinct climatic zones. These patterns inform feature engineering (e.g., transformations for skewed variables, interaction terms between climate features), model selection (favoring ensemble or deep learning models over linear models), and data preprocessing (outlier handling and scaling).

4.1.1 Result of Crop Yield Distribution After Outlier Treatment

A scatter plot of actual versus predicted crop yields is generated to visually assess the model’s performance and ensure that predictions align closely with the ground truth. This multi-stage methodology—from data preprocessing and feature engineering to model stacking, hyperparameter tuning, and comprehensive evaluation—forms a robust framework for accurate crop yield prediction using advanced machine learning techniques.

.

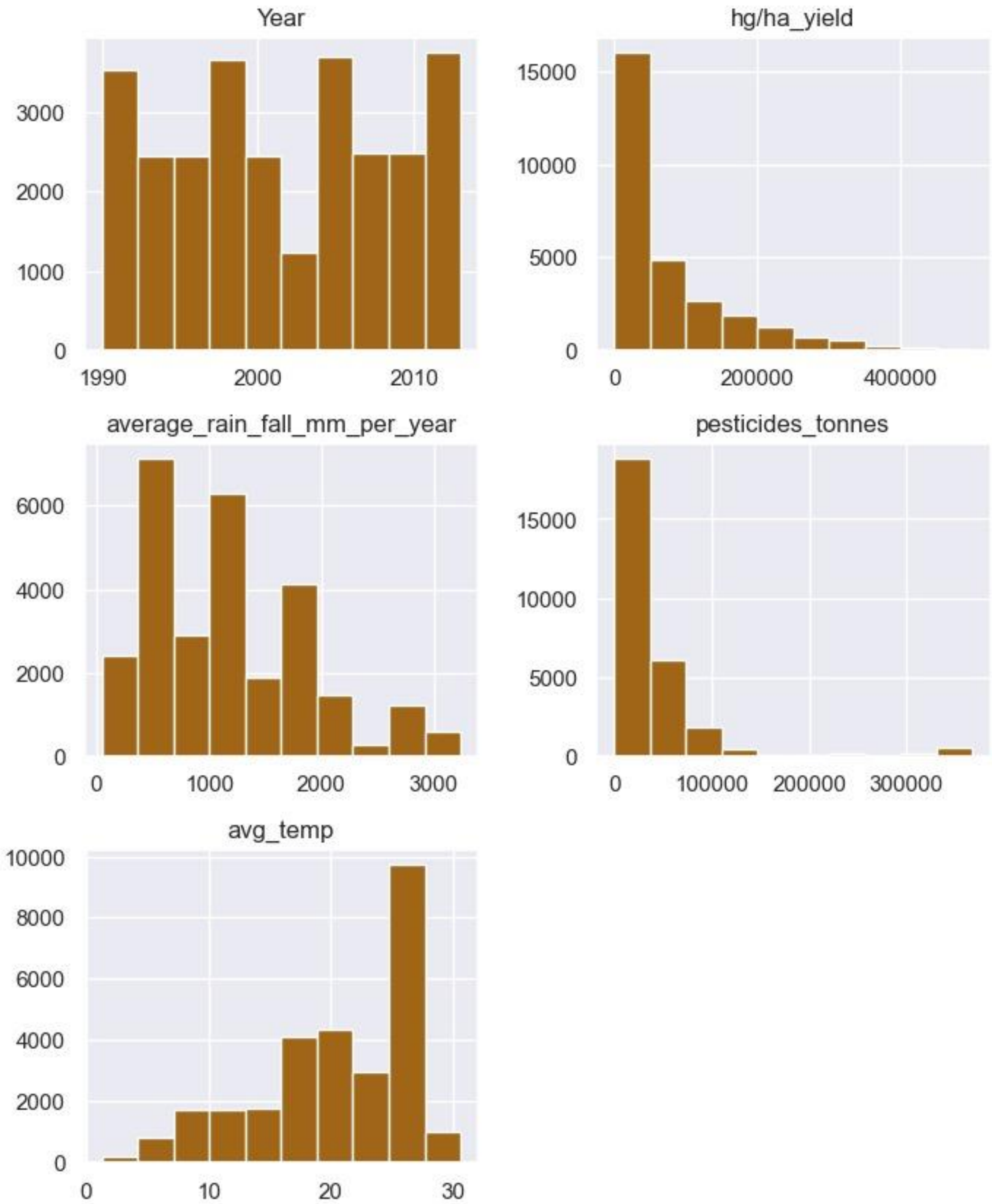

Figure 2: Distribution of Key Features before Outlier Removal

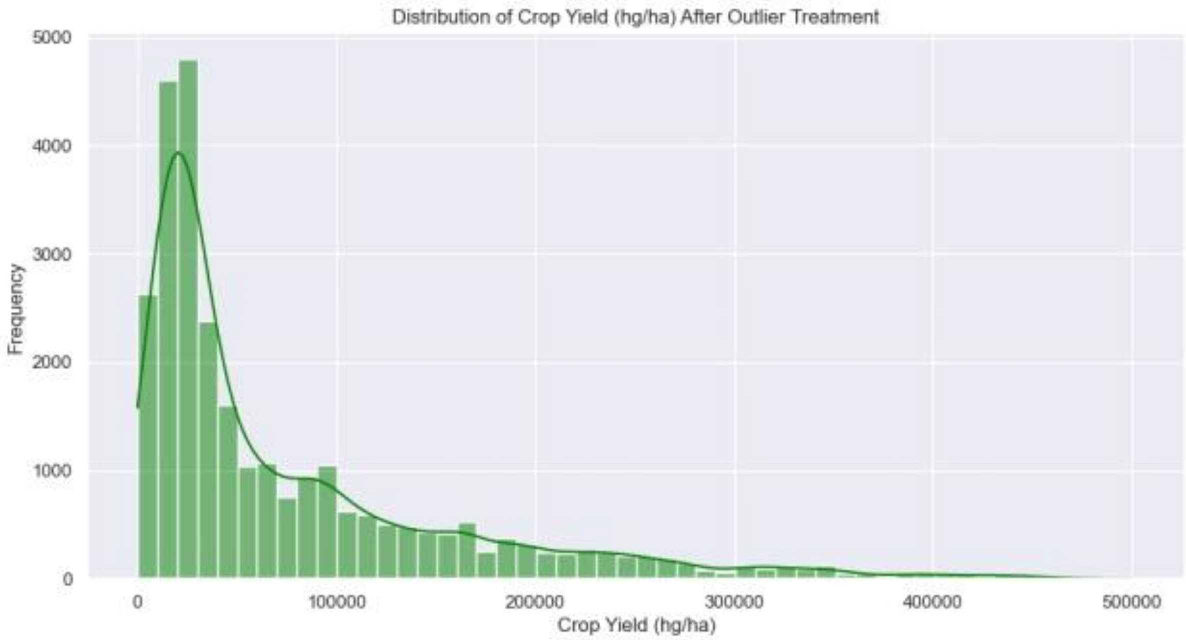

Figure 3: Distribution of Crop Yield after Outlier Treatment

Figure 3 provides a visual summary of the frequency of crop yield values in the dataset after addressing outliers. The x-axis represents crop yield values (hectograms per hectare), while the y-axis shows their frequency. The distribution reveals a concentration of yields in the lower ranges, likely below 100,000 hg/ha, suggesting the most common yield levels. The spread extends up to 500,000 hg/ha, indicating variability, with a long right tail pointing to occasional high yields. This right-skewed nature is common in real-world agricultural data, highlighting the need for careful consideration in model selection. While linear models may require transformations to handle skewness, tree-based models like RandomForest and XGBoost are generally more robust to such distributions.

The explicit mention of "After Outlier Treatment" suggests that extreme yield values have been addressed, leading to a more stable and representative dataset. Handling outliers helps improve model robustness by reducing the influence of extreme values, which in turn enhances generalization to unseen data. This histogram plays a crucial role in exploratory data analysis, guiding the selection of appropriate models based on yield distribution characteristics and setting realistic expectations regarding prediction performance. Additionally, it aids model evaluation by providing a benchmark for assessing how well predictions align with the actual data distribution, and selecting appropriate evaluation metrics, such as MAE alongside RMSE, ensures models are assessed to account for skewness and variability in crop yields.

This scatter plot on figure 4 visually evaluates the stacking regressor's performance in predicting crop yield. The x-axis shows actual yields, while the y-axis displays predicted yields, with each point representing a test data instance. The red dashed “Ideal fit” line signifies perfect predictions. The brown scatter points, representing the model's “Stacked Predictions," cluster closely around this ideal line. This tight clustering visually demonstrates that the stacking regressor is effectively forecasting crop yields with good accuracy, as predicted values are generally close to the actual observed yields across the test dataset. In terms of crop yield prediction research, this plot provides strong visual evidence of the model's success. It confirms the effectiveness of the stacking regressor in capturing complex relationships within the data and generating reliable yield forecasts. While the scatter is not perfectly aligned, indicating some prediction error, the overall closeness to the ideal fit line suggests a valuable tool for agricultural applications.

Figure 4: Stacking Regressor Prediction Result

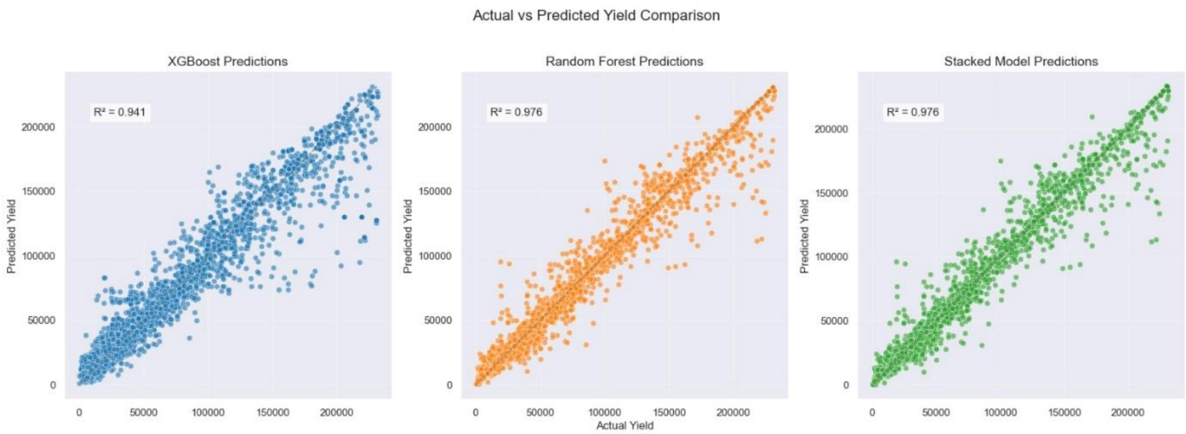

Figure 5: Prediction Result of the Ensembled Model

Figure 5 shows how closely each model's yield predictions match the true (actual) yields, with the ideal y = x line drawn as a reference. This set of scatter plots compares actual vs. predicted values for three models: XGBoost, Random Forest, and a Stacked Model. Each dot represents a data point, where X-axis = actual yield values and Y-axis = predicted yield values. For XGBoost predictions on the left panel, each blue dot is one field's actual vs. predicted yield. The scatter plot clusters tightly around the diagonal, but with a bit more spread at higher yields. Coefficient of determination (R^2) = 0.962, meaning that about 96.2% of the variance in actual yields is explained by XGBoost's predictions. Random Forest Predictions in the middle panel, orange dots show an even tighter cloud around the diagonal than XGBoost. R^2 = 0.985, indicating very strong agreement (that the model explains 98.5% of variance), and thus it has slightly better accuracy than XGBoost. For the Stacked Model Predictions on the right panel, green points combine (or "stack") the strengths of both models. It also achieves R^2 = 0.985, matching the Random Forest's performance in this case. So, the three models predict yields very accurately ($R^2 > 0.96$), with Random Forest and the ensemble (stacked) model performing marginally better than XGBoost on this dataset.

From figure 6, it can be observed that the model registers an RMSE of 9771.2884 hg/ha, an R^2 of 0.9868, and an MAE of 3691.9446 hg/ha. The extremely low RMSE reveals that the model's prediction error is minimal relative to the variability in crop yields, meaning that the average deviation from the actual value is only about 9771.2884 hectograms per hectare. The near-perfect R^2 score of 0.9868 shows that the model explains nearly 98.68% of the variance in crop yield, indicating that almost all influential patterns and relationships in the data have been captured. Furthermore, the low MAE of 3691.9446 hg/ha reinforces the model's precision, as the average absolute difference between predicted and actual yields is very small, confirming the model's consistent performance across the dataset.

4.2 Model Evaluation Result

Figure 7 compares the performance of three models RandomForest, GradientBoos and the stacked in predicting crop yield using four metrics: R^2, MAE, RMSE, and MSE. GradientBoost performs best overall, with the highest R^2 value of (~0.918) and the lowest error values across MAE (~44882), RMSE (~70398), and MSE (~4.95e+09). RandomForest follows closely behind in performance, while the overall model consistently shows the weakest performance in all metrics. This suggests that GradientBoost is the most accurate and efficient model among the three for the given dataset.

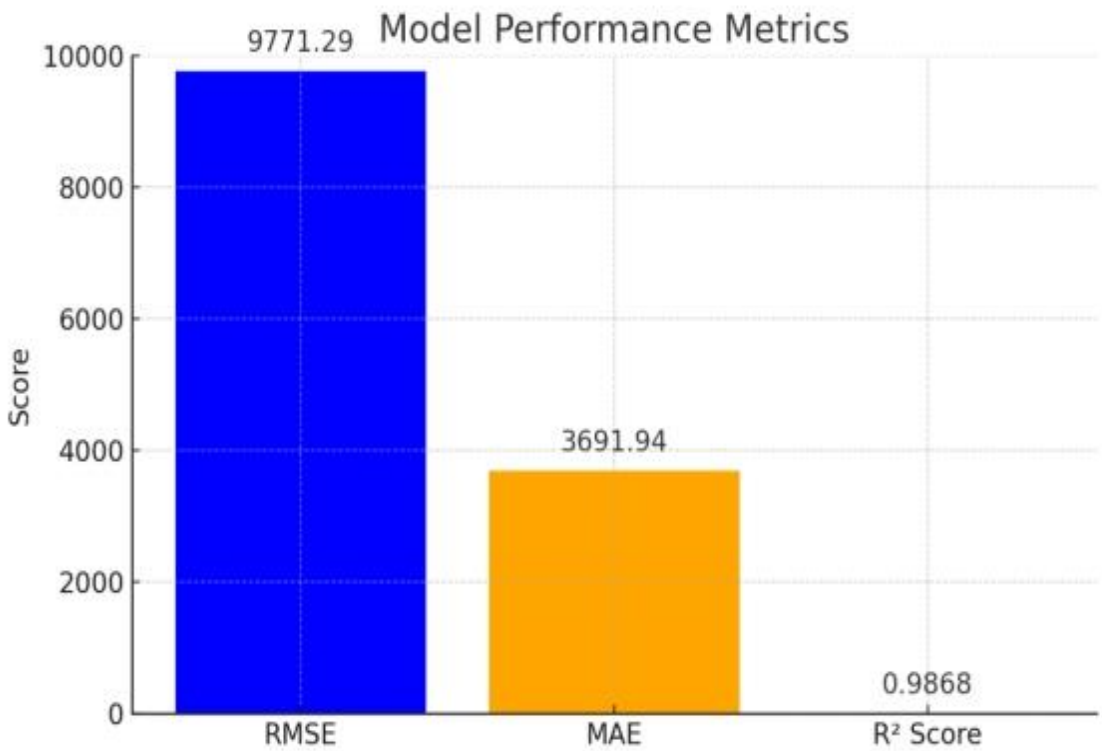

Figure 6: Model Evaluation Result

4.3 Model Comparison

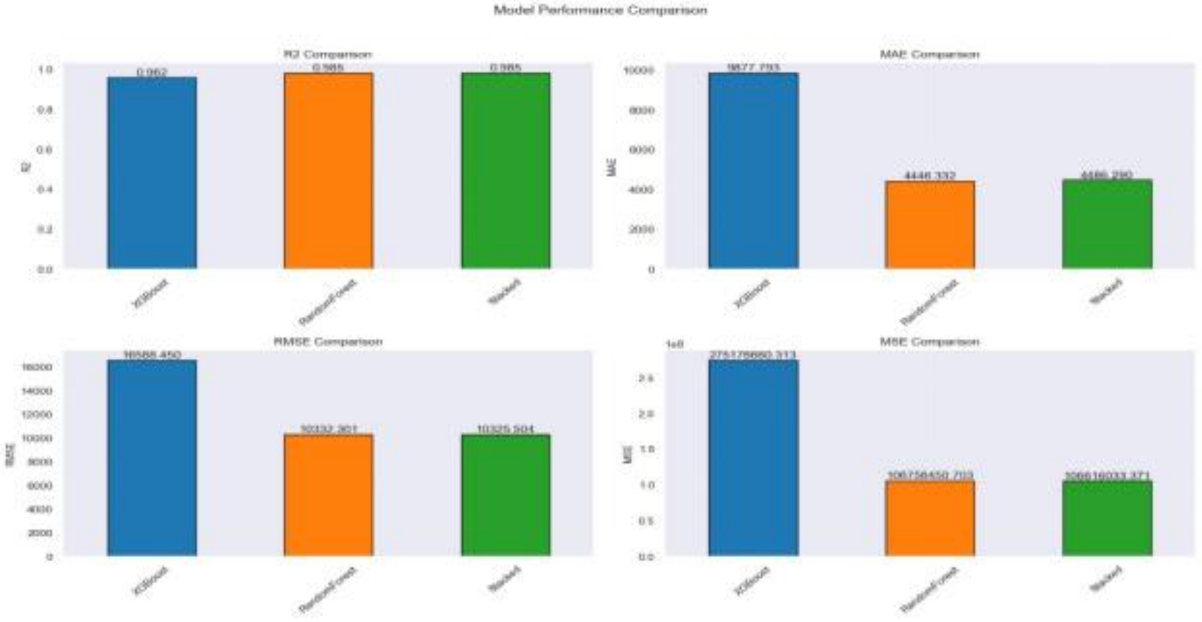

Figure 7: Stacked Model Performance Comparison Result

4.4 Model Visualization on World Map

Figure 8: Result Visualization on the World Map

Figure 8 visually represents the results of the crop yield prediction research overlaid onto a world map using GeoPandas and Natural Earth data. Each marker on the map corresponds to a country for the crop yield data. The pop-up box specifically highlights Egypt, showcasing both the "Predicted Yield" (96640.52 tons/ha) generated by the model and the "Actual Yield" (96004.89 tons/ha). This direct comparison for each country allows for a geographical assessment of the model's performance in crop yield prediction. In this research, the map serves as a powerful tool for visualizing and communicating the model's predictive capabilities across different geographical regions. By spatially displaying both predicted and actual yields, it allows stakeholders to quickly grasp how well model performs on a global scale. This type of visualization is crucial for understanding the geographical strengths and weaknesses of the crop yield prediction model and for communicating findings effectively to a broader audience, highlighting the spatial dimension of the research outcomes.

4.4 Result Visualization with SHAP

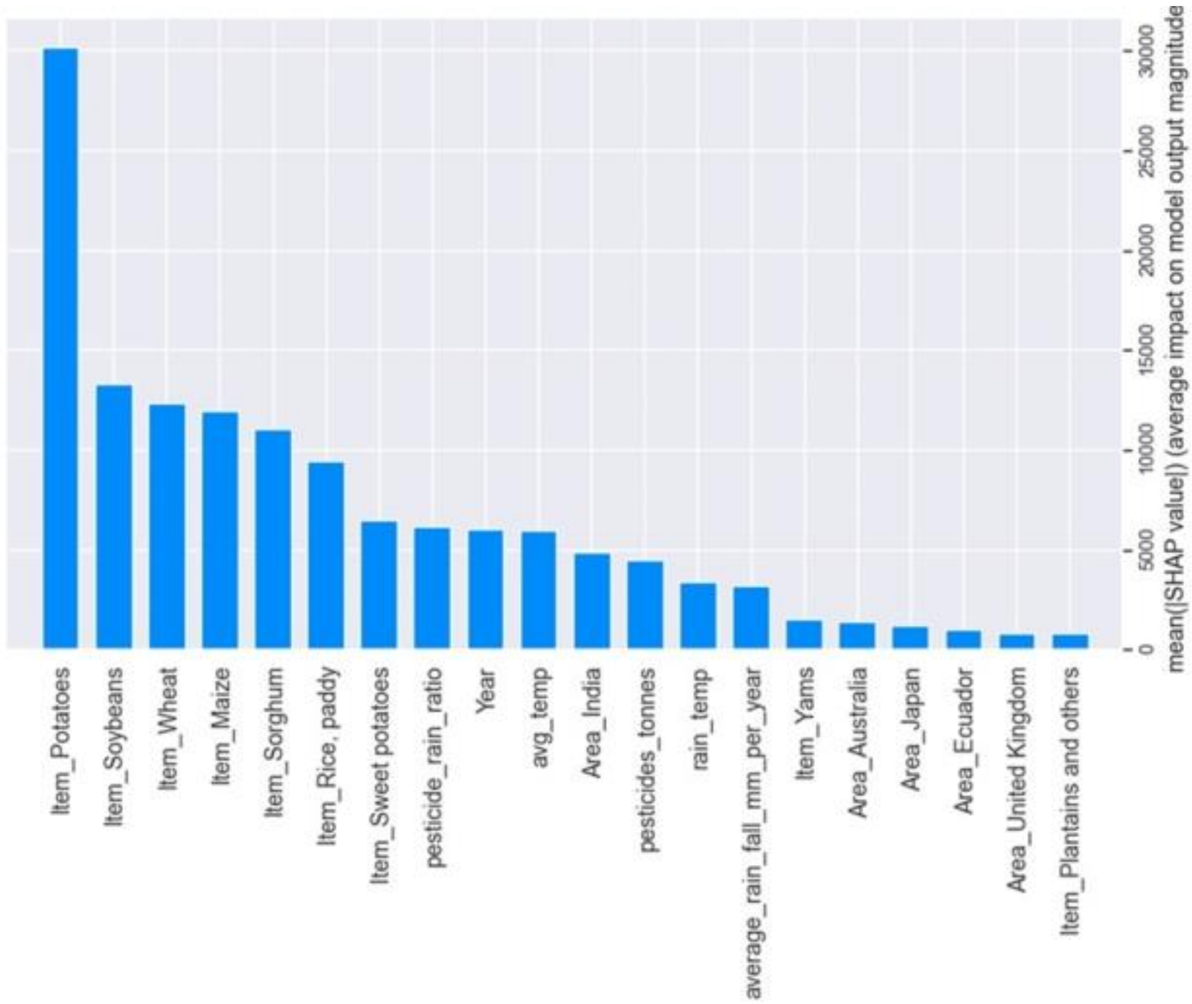

Figure 9: Average Impact of Different Features

The graph on figure 9 visualizes the average impact of different features on the model’s predictions using SHAP values. The x-axis represents the mean absolute SHAP value, indicating each feature’s contribution to the model's output. Potatoes, soybeans, wheat and maize have the highest influence on crop yield predictions. Other significant factors include the pesticide-to-rainfall ratio, average temperature, and specific geographic areas. This analysis helps identify key variables driving agricultural productivity forecasts.

Figure 10 is a SHAP summary plot, which explains the impact of different features on the model's crop yield predictions. Each dot represents a data point, with color indicating feature values (blue for low, red for high). The x-axis (SHAP value) shows whether the feature increased or decreased the prediction. Features like "Item_Potatoes" and "Item_Soybeans" have the highest impact, while other factors like "pesticide _rain_ ratio" and average _ rain _f all _mm_per_year" also influence the model's output. This visualization helps interpret how different variables contribute to yield predictions.

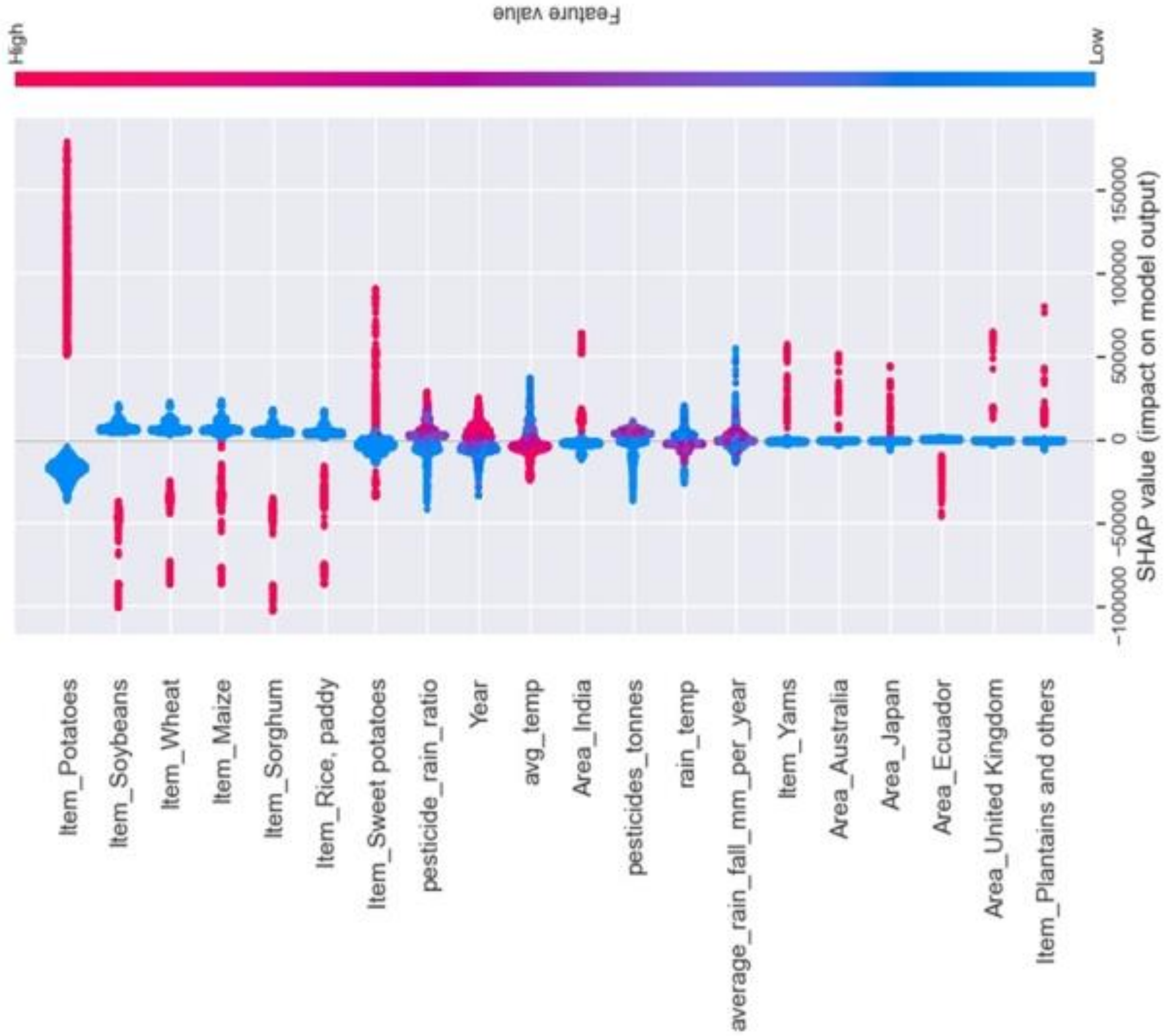

Figure 10: Explainable AI with SHAP

5. Conclusion

The impressive performance can be attributed to the effective synergy achieved by stacking XGBoostRegressor with RandomForestRegressor. XGBoost is efficient in capturing complex non-linear relationships through gradient boosting. RandomForest offers robustness and reduces overfitting by aggregating multiple decision trees. In addition, the success of the model is bolstered by feature engineering: the creation of a rainfall-temperature interaction term, a pesticide-to-rainfall ratio, and a relative year feature enrich the dataset, allowing the model to better understand the factors influencing crop yield. Rigorous hyperparameter tuning also contributed significantly to the model's outstanding performance.

5.1 Summary of the Implications and Practical Significance of the Results

The high-accuracy predictions produced by the stacking regressor have profound practical implications for agriculture. With reliable crop yield forecasts, farmers and agricultural planners can make informed decisions regarding planting schedules, resource allocation, and supply chain management. This level of precision helps optimize the use of fertilizers, pesticides, water, and labor, while also aiding in the development of proactive food security strategies. Ultimately, the model's ability to minimize prediction error and capture almost the entire yield variability enhance decision-making across the agricultural value chain and reducing operational risks.

5.2 Conclusion

In summary, the stacking regressor, which leverages the complementary strengths of XGBoostRegressor and RandomForestRegressor, has demonstrated outstanding crop yield prediction performance. The combination of advanced ensemble techniques and strategic feature engineering has resulted in a robust model that achieves high accuracy and explains nearly all the variance in the yield data. These findings underscore the potential of this approach to revolutionize agricultural forecasting, leading to more efficient planning, resource management, and decision-making in the field.

5.3 Recommendations

Continuous monitoring and periodic retraining with updated data will be essential to maintain accuracy. Future research could explore integrating additional data sources, refining the feature engineering process, and experimenting with alternative ensemble methods such as incorporating LSTM networks, to further improve predictions and adapt to dynamic agricultural environments.

References

[1] Adejuwon, J. O., Tewogbade, K. E., Oguntoke, O., & Ufoegbune, G. C. (2023). Comparing farmers' perception of climate effect on cocoa yield with climate data in the Humid zone of Nigeria. *Heliyon*, *9*(12).

[2] Amos, K. G., & Okoro, E. O. (2025). Assessing the dual threats of oil spills and climate change on sustainable development in Ogbia, Bayelsa State, Nigeria. *Scientia Africana*, *24*(1), 125–136.

[3] Annie , Mangshatabam, Raj kumar Pal, Anjusha Sanjay Gawai, and Aman Sharma. 2023. "Assessing the Impact of Climate Change on Agricultural Production Using Crop Simulation Model". *International Journal of Environment and Climate Change* 13 (7):538-50. https://doi.org/10.9734/ijecc/2023/v13i71906.

[4] Ashiegbu, G., Man, N., Sharifuddin, J., Buda, M., & Adesope, O. (2024). Impacts of Climate Variability on Agricultural Activities and Availability of Agroforestry Practices in Southeast Nigeria. *Journal of Global Innovations in Agricultural Sciences*, *12*, 613-623.

[5]. Awais, M., Naqvi, S. M. Z. A., Zhang, H., Li, L., Zhang, W., Awwad, F. A., ... & Hu, J. (2023). AI and machine learning for soil analysis: an assessment of sustainable agricultural practices. *Bioresources and Bioprocessing*, *10*(1), 90.

[6] Falana, M. O., Eseyin, J. B., & Akinwande, O. T. (2024). *Transforming Nigerian agriculture: The rise of smart greenhouse farming*. *International Journal of Computer Applications Technology and Research, 13*(5), Article 1003. https://doi.org/10.7753/IJCATR1305.1003

[7] Garg, D., & Alam, M. (2023). An effective crop recommendation method using machine learning techniques. *International journal of advanced technology and engineering exploration*, *10*(102), 498.

[8] Ikehi, M. E., Ifeanyieze, F. O., Onu, F. M., Ejiofor, T. E., & Nwankwo, C. U. (2022). Assessing climate change mitigation and adaptation strategies and agricultural innovation systems in the Niger Delta. *GeoJournal*, *88*(1), 209–224. https://doi.org/10.1007/s10708-022-10596-6

[9] Nnodi, J. T., Asagba, P. O., & Ugwu, c. (2011). Quality of Experience Predictive Model for Web Users. *Global Scientific Journal*, *9*(10), 512–525. https://www.globalscientificjournal.com/researchpaper/Quality_of_Experience_Predictive_Model_for_Web_Users.pdf

[10] Nnodi, J. T., & Obasi, E. M. (2025). Leveraging Artificial Intelligence for Detecting Insider Threats in Corporate Networks. *University of Ibadan Journal of Science and Logics in ICT Research*, *13*(1), 144-152.

[11] Nofiu, N. B., & Baharudin, S. A. (2025). Assessment of Flood Vulnerability among Smallholder Farmers in Niger State, Nigeria. *Kufa Journal for Agricultural Science*, *17*(1).

[12] Obasi E. C. M. and Nlerum P. A. "A Model for the Detection and Prevention of Backdoor Attacks using CNN with Federated Learning," Univ. Ibadan J. Sci. Logics ICT Res., vol. 10, no. 1, pp. 9–21,.

[13] Obasi E.C.M. and Stow M. T.(2023) "A Predictive Model for Uncertainty Analysis on Big Data Using Bayesian CNN," Univ. Ibadan J. Sci. Logics ICT Res., vol. 9, no. 1, pp. 52–62,.

[14] Ologeh, I., & Adesina, F. (2022). Evaluation of climate change as a major determinant of crop yield improvement in Nigeria. *IOP Conf. Ser.: Earth Environ. Sci.*, *1077*, 012002.

[15] Okoro, E. O., & Ofordu, C. S. (2025). Rainfall and Temperature Trends in Ogbia Local Government Area Bayelsa State, Nigeria from 1993 To 2023. J. Appl. Sci. Environ. Manage. 29 (2) 451-458.

[16] Timadi M. E. and Obasi E. C. M. (2025), Integrating Zero-Trust Architecture with Deep Learning Algorithm to Prevent Structured Query Language Injection Attack in Cloud Database," Univ. Ibadan J. Sci. Logics ICT Res., vol. 13, no. 1, pp. 52–62, 2025.

[17] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and electronics in agriculture*, *177*, 105709.

[18] Wakchaure, M., Patle, B. K., & Mahindrakar, A. K. (2023). Application of AI techniques and robotics in agriculture: A review. *Artificial Intelligence in the Life Sciences*, *3*, 100057.

[19] Week, D. A., & Wizor, C. H. (2020). Effects of flood on food security, livelihood and socio-economic characteristics in the flood-prone areas of the core Niger Delta, Nigeria. *Asian Journal of Geographical Research*, *3*(1), 1-17.

[20] Zidan, F., & Febriyanti, D. E. (2024). Optimizing agricultural yields with artificial intelligence-based climate adaptation strategies. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, *5*(2), 136-147.