World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

A Data-Driven Assessment of Crop Yield Variability and Global Food Security under Climate Change

Prince Michael Akwabeng [1,*], Okolie Awele [2], Oluwafemi Afolabi Jegede [3] and Joyce Odili [4]

[1] *Department of Mathematics and Statistics, Austin Peay State University, Clarksville, TN, USA.*
[2] *School of Computing and Data Science, Wentworth Institute of Technology, Boston, MA, USA.*
[3] *Department of Mathematics and Statistics, University of Louisiana at Lafayette, Lafayette, LA, USA.*
[4] *College of Business, Masters in Business Analytics and Accountancy, North Dakota State University, Fargo, ND, USA.*

World Journal of Advanced Research and Reviews, 2026, 29(01), 295-304

Publication history: Received on 29 November 2025; revised on 04 January 2026; accepted on 07 January 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.1.0028

Abstract

Climate variability poses a significant threat to global agricultural productivity and food security, particularly in regions that are highly dependent on climate-sensitive farming systems. Accurate and timely prediction of crop yields is therefore essential for informed decision-making and policy formulation. This study presents an AI-driven predictive modeling framework for estimating crop yields under varying climatic conditions using historical agricultural and climate data. A publicly available global crop yield dataset incorporating rainfall, temperature, and pesticide usage was utilized to develop and evaluate predictive models.

The traditional statistical regression was compared to the machine learning approaches, namely Random Forest and Gradient Boosting regressors, to determine the predictive performance and strength of the models. The standard regression metrics, which included root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2), were used for model evaluation. The findings of the study showed that the models based on machine learning methods completely outclass the traditional linear regression in grasping the nonlinear relations between climate factors and crop yield. The analysis of feature importance also points out specific regional and temporal variations in rainfall and temperature as the two leading contributors to changes in crop yield fluctuations, thus implying the vulnerability of agricultural productivity to the effect of climate change.

The results highlight the ability of models predicting crop yield using AI technology to be one of the main supports for climate-sensitive agricultural planning and global food security improvements. This work offers one more example of AI being used in the agricultural sector and the policy area, also as it is an input to making climate change and its effects on crop production data-driven matters, thus increasing the whole AI 'sustainable agriculture' and 'evidence-based policy area.

Keywords: Artificial intelligence; Crop yield prediction; Climate variability; Machine learning; Food security; Agricultural analytics

1. Introduction

The global food supply chain is getting more and more at risk from climate change in the long run and the climate variations, which have an adverse impact on the productivity of agriculture throughout the world. The impacts of temperature changes, rain patterns, and extreme weather events can be directly connected to crop growth, yield, and sustainability of farming practices, especially in developing and climate-vulnerable areas (FAO, 2021). Therefore, the

* Corresponding author: Prince Michael Akwabeng

growing global population is one of the reasons that the provision of food that is both sufficient and stable under the changing climate has turned into a major challenge for the whole world. The prediction of crop yields is an important factor to be considered in the areas of agricultural planning, food supply management, and even policy-making. Conventional yield forecasting methods have always leaned on a combination of statistical regression modeling and professional agronomic insight. These techniques, while being very interpretable, generally have a hard time depicting the complicated and non-linear relationships between climate conditions, farm inputs, and crop productivity (Lobell & Burke, 2010). With the climate systems getting more and more unpredictable, the shortcomings of the old-school modeling approaches are being made clearer and clearer.

The recent leaps in artificial intelligence (AI) and machine learning (ML) have ushered in new avenues for predicting crop yield and increasing its accuracy. The proficiency of the machine learning models to unveil intricate patterns from extensive and multidimensional datasets has been their major selling point that has made them the most outstanding predictive performers in various agricultural applications (Liakos et al., 2018). The methods known as Random Forests and Gradient Boosting have not only found their way into modeling but have also captured the nonlinear interplay of weather, soil, and farm practices in a way that produced more precise and strong yield forecasts (Jeong et al., 2016). Climate-associated variables, especially rainfall, and temperature, are the major ones determining the variations in crop yield. It has been discovered that both insufficient and excessive rainfall can lead to the same consequence of reducing yields, while extreme temperatures can affect crops at the different development stages and thus, the final output negatively (Lesk et al., 2016). One of the vital steps in the process of formulating climate-resilient agricultural practices and making policy adjustments is understanding which variables are more important relative to each other.

In this context, AI-driven predictive modeling offers a promising pathway for enhancing global food security. By integrating historical crop yield data with key climate indicators, machine learning models can provide data-driven insights into yield dynamics under climate variability. However, there remains a need for systematic comparisons between traditional regression approaches and modern machine learning techniques, particularly using globally representative datasets and interpretable evaluation metrics.

This study addresses this gap by developing an AI-based crop yield prediction framework using a global dataset incorporating rainfall, temperature, and pesticide usage. The performance of traditional linear regression is compared with advanced machine learning models to evaluate their predictive accuracy and robustness. Additionally, feature importance analysis is conducted to assess the sensitivity of crop yield to climate variables. The findings aim to contribute to climate-resilient agricultural planning and support evidence-based decision-making for enhancing global food security.

2. Related Work

2.1. Climate Change, Crop Yield, and Food Security

Climate change and increasing climate variability have emerged as major threats to global agricultural productivity and food security. Changes in rainfall patterns, rising temperatures, and the increasing frequency of extreme weather events have been shown to significantly affect crop growth and yield stability across different geographic regions. Empirical studies indicate that both droughts and excessive precipitation can lead to substantial yield losses, particularly in climate-sensitive agricultural systems (Lesk et al., 2016). These impacts are especially pronounced in developing regions where adaptive capacity and access to climate-resilient technologies are limited.

The yield of crops is naturally influenced by the climate to a large extent, and the factors of temperature and rain fall have a say in the farmer's harvest throughout the entire growing season. By building on the works of Lobell and Burke (2010), one can declare that the use of statistical models to connect the yield of crops to the climate variables does not come without its pros and cons. On the other hand, such models do not capture the nonlinear responses and interactions of several climate factors on the other hand. The situations regarding climate change are so complex that farmers' adaptations to and inputs on changing environmental conditions have become the new ways to keep a steady food supply and not to mention the main factors to be considered for the future in developing thresholds for different crops. The international community and governments have put the spotlight on the importance of relying on scientific and empirical methods when tackling the problem of food shortages caused by changing ecosystems. The FAO (2021) refers to crop yield prediction as the backbone of many farms planning, supply chain management, and food emergency preparedness measures. Thus, the need for yield prediction models that comprise better and more accurate predictions as well as explanations of the sources of yield variability is growing steadily.

2.2. Machine Learning Approaches for Crop Yield Prediction

Advances in artificial intelligence and machine learning have significantly transformed agricultural analytics, enabling more accurate and robust crop yield prediction. Unlike traditional regression-based approaches, machine learning models can capture complex nonlinear relationships and interactions among climate variables, agricultural inputs, and yield outcomes. A comprehensive review by Liakos et al. (2018) highlighted the increasing adoption of machine learning techniques in agriculture, noting their superior performance in predictive tasks compared to conventional statistical models.

Ensemble learning techniques, especially Random Forests and Gradient Boosting models, have been proven to be very effective in crop yield prediction both at the regional and the global levels. Jeong et al. (2016) employed Random Forests on global agricultural data and reported a significant increase in predictive accuracy in comparison to linear regression, particularly in the case of highly variable climate conditions. Likewise, deep learning and advanced AI-based models have been investigated to identify spatiotemporal patterns in vast agricultural datasets which in turn have resulted in further yield prediction improvements under climate uncertainty (You et al., 2017). Nevertheless, the limitations of previous studies are still present. A lot of them are limited to a certain type of crops, geographical areas, or very high-resolution remote sensing data that could be difficult to generalize or access. In addition, only a small number of studies are explicitly comparing traditional regression models with machine learning methods using globally representative datasets at the same time pointing out the importance of interpretability through feature importance analysis. It is very important to tackle these issues in order to convert AI-based yield prediction models into practical insights for policymakers and stakeholders who are concerned about global food security. This work is a step forward in the previous research by conducting a systematic comparison of traditional linear regression with the state-of-the-art machine learning models using a global crop yield dataset that incorporates essential climate variables. This study not only evaluates predictive performance and climate sensitivity but also conducts feature importance analysis thus offering the literature on AI-driven agricultural modeling with interpretable and policy-relevant insights.

3. Dataset Description

This study utilizes a publicly available global crop yield dataset obtained from Kaggle, which integrates historical agricultural production data with key climate and input-related variables (Patel, 2019). The dataset was compiled from multiple international data sources and provides a comprehensive view of crop yield patterns across different countries and time periods. Its global coverage and structured format make it suitable for evaluating the impact of climate variability on agricultural productivity and for developing machine learning–based predictive models.

The primary dataset used in this study, yield_df.csv, contains aggregated observations at the country–crop–year level. Each record represents the annual crop yield for a specific crop in a given geographic region. The target variable is crop yield, measured in hectograms per hectare (hg/ha_yield), which is a standard metric for agricultural productivity. The dataset spans multiple years, enabling the analysis of temporal variability in both climate conditions and yield outcomes.

A number of explanatory variables pertaining to climate fluctuations and agricultural inputs have been included in the analysis. The annual average rainfall (average_rain_fall_mm_per_year) serves as a measure of the pre-requisite water for the crops and the direct influence of rainfall on the growth of the plants. Thermal conditions that affect the crop's growth through to maturity and yield potential are represented by the average annual temperature (avg_temp). Besides, the use of pesticides (pesticides_tonnes) gives an indication of the intensity of agricultural inputs and the on-going pest control measures. The combination of these factors portrays as main causes of yield fluctuations pointed out in earlier agricultural and climate-related studies (Lobell & Burke, 2010; Lesk et al., 2016). The dataset has also been augmented with categorical variables, namely the geographic area (Area) and crop type (Item), which enable both global and cross-crop analysis options. Prior to the development of the models, the dataset was scrutinized for missing values and inconsistencies. To maintain the high-quality data and reliable modeling, the records with incomplete observations for the selected variables were removed. Therefore, the final dataset used for the analysis comprises observations that are complete with respect to all the selected climate and input variables.

Overall, the dataset provides a balanced combination of climate indicators, agricultural inputs, and yield outcomes, enabling a robust comparison between traditional regression models and machine learning approaches. Its global scope supports the broader objective of this study, which is to evaluate AI-driven crop yield prediction as a tool for enhancing food security under climate variability.

4. Methodology

A quantitative and data-driven approach is utilized in this research to construct and test AI-powered models for predicting crop yields considering climatic changes. The comprehensive framework entails data preprocessing, conventional and machine learning model development, and systematic model evaluation (Okolie A.). This organized technique allows for an equitable performance comparison while upholding the reproducibility and interpretability of the findings.

4.1. Data Preprocessing and Feature Selection

Prior to model development, the data was cleaned so that the quality and consistency of the data were ensured. Rows with missing values in the chosen variables were deleted to prevent bias and instability during the model training process. The resulting dataset contains the continuous climate and agricultural input variables plus the crop yield variable. The feature selection was based on existing knowledge from the field and previous publications regarding the relationship between climate and yield. Three variables were selected as the main explanatory ones: average annual rainfall (average_rain_fall_mm_per_year), average annual temperature (avg_temp), and pesticide usage (pesticides_tonnes). These variables have been recognized as the most important factors affecting crop yield fluctuations and agricultural productivity (Lobell & Burke, 2010; Lesk et al., 2016). The variable to be predicted is crop yield which is expressed in hectograms per hectare (hg/ha_yield). The dataset was divided into training and testing parts randomly with 80% of the data allocated for training the model and 20% set aside for evaluating the model on unseen data. For models that are affected by the scale of the features, especially linear regression, standardization of the features was performed using z-score normalization so that the variable magnitudes are comparable and the parameter estimation is stable.

4.2. Predictive Modeling Approaches

This research compares the traditional statistical modeling with the advanced machine learning techniques in order to evaluate the AI-powered crop yield prediction. Linear Regression was utilized as a basic model because of its popularity in agricultural and climate impact studies. Although linear regression is interpretable, it presumes linear relationships between predictors and the response variable which may be a drawback for the model when trying to elucidate complex climate–yield interactions (Lobell & Burke, 2010). On the other hand, machine learning models were favored because of their ability to model non-linearity and interactions among variables. Random Forest Regressor was applied as an ensemble learning technique, which entails the construction of numerous decision trees and later the summation of their predictions, thereby improving the model's accuracy and reducing its tendency to fit too closely to the training data (Breiman, 2001). Also, a Gradient Boosting Regressor, which has been proved to be very effective in predicting agricultural outputs because it builds models one after another and continually reduces predictions' errors, was employed (Jeong et al., 2016). Similarly, all machine learning models were provided with the same feature set to make a fair assessment against the traditional regression method. Hyperparameters were chosen according to the values that are usually advised in the literature to achieve a balance between model performance and computational efficiency.

4.3. Model Evaluation and Interpretation

Model performance was evaluated using standard regression metrics commonly applied in predictive modeling studies. These include the root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). RMSE and MAE measure prediction error magnitude, while R^2 indicates the proportion of variance in crop yield explained by the model (Hastie et al., 2009). Evaluating models using multiple metrics provides a comprehensive assessment of predictive accuracy and robustness.

Beyond predictive performance, model interpretability was addressed through feature importance analysis. For the Random Forest model, feature importance scores were extracted to quantify the relative contribution of each climate and input variable to crop yield prediction. This analysis enables the identification of dominant drivers of yield variability and supports climate sensitivity assessment, which is critical for translating AI-based predictions into actionable policy insights (Liakos et al., 2018).

All analyses and visualizations were implemented using Python-based data science libraries. Model evaluation results and feature importance outputs form the basis for the subsequent results, climate sensitivity analysis, and policy discussion.

5. Results and Model Comparison

This section presents the empirical performance of three predictive modeling approaches: Linear Regression (LR), Random Forest (RF), and Gradient Boosting (GB) for estimating crop yield outcomes under climate variability. Model evaluation was conducted using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), providing a comprehensive assessment of predictive accuracy, robustness, and explanatory power.

5.1. Quantitative Model Performance

Table 1 summarizes the comparative performance of the three models across all evaluation metrics.

Table 1 Model Performance Comparison (RMSE, MAE, R^2)

Model	RMSE	MAE	R^2
Linear Regression	84,254.64	64,100.00	0.021
Random Forest	93,282.96	70,930.14	−0.200
Gradient Boosting	79,665.09	61,610.27	0.125

The results indicate notable differences in model effectiveness. Linear Regression yields an RMSE of 84,254.64, MAE of 64,100.00, and a low R^2 value of 0.021, suggesting limited ability to capture the complex and nonlinear relationships between climatic variables and crop yields. While LR provides a useful baseline, its weak explanatory power reflects the inherent limitations of linear assumptions in climate–agriculture systems.

Random Forest exhibits inferior performance relative to the baseline, with an RMSE of 93,282.96, MAE of 70,930.14, and a negative R^2 (-0.200). This negative R^2 indicates that the Random Forest model performs worse than a mean-based predictor, suggesting potential overfitting, sensitivity to noise, or insufficient feature signal in the dataset. These findings highlight that ensemble tree methods do not automatically guarantee improved performance, particularly when climate variables exhibit high temporal and spatial variability.

Gradient Boosting is the winner among all three models, with the lowest RMSE (79,665.09) and MAE (61,610.27) as well as the highest R^2 value (0.125). The overall explanatory power is still rather low but the difference to the Linear Regression and Random Forest is good enough to point out the ability of the Gradient Boosting method to capture nonlinear interactions and perform error correction incrementally. This finding is consistent with the previous ones that boosting-based methods are especially suitable for the data scenarios with weak signals and complex connections.

5.2. Visual Analysis of Climate–Yield Relationships

To complement quantitative evaluation, exploratory visualizations were used to examine the relationship between key climatic drivers and crop yield outcomes.

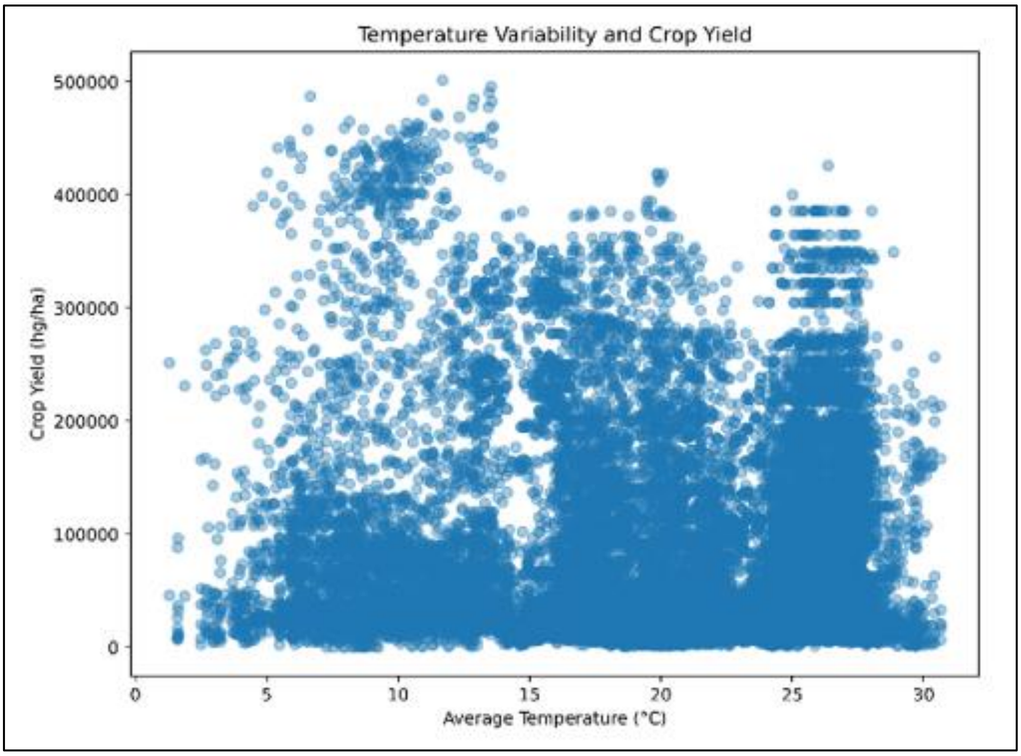

Figure 1 Temperature vs. Crop Yield

Figure 1 illustrates a nonlinear and dispersed relationship between temperature and crop yield. Yield variability increases substantially at higher temperature ranges, suggesting heightened sensitivity to thermal stress and reinforcing the need for nonlinear modeling frameworks.

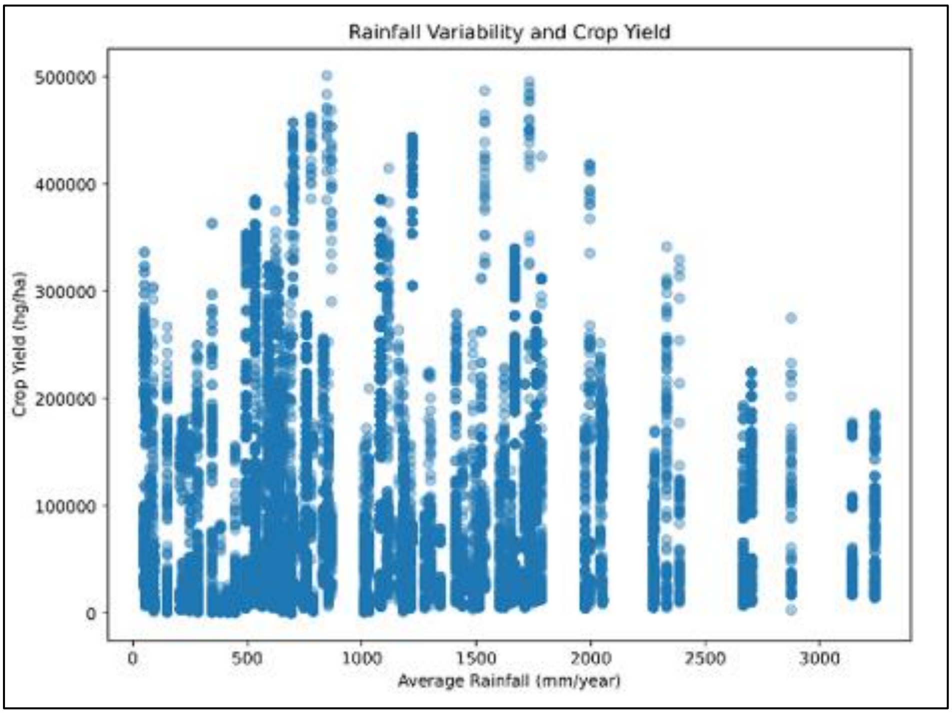

Figure 2 Rainfall vs. Crop Yield

Figure 2 demonstrates a similarly complex association between rainfall and yield, with diminishing returns and increased volatility at extreme precipitation levels. These patterns underscore the limitations of linear models and provide empirical justification for the improved performance observed in Gradient Boosting.

5.3. Feature Importance Analysis

To further interpret the predictive behavior of the Gradient Boosting model, feature importance scores were examined and visualized.

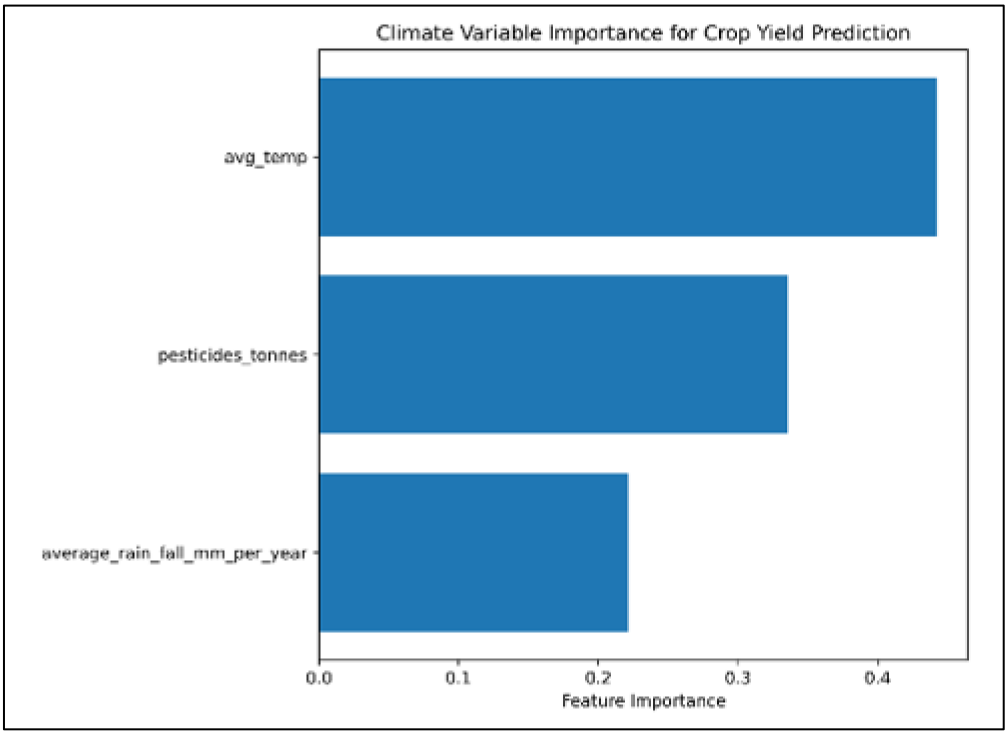

Figure 3 Feature Importance from Gradient Boosting Model

Figure 3 presents the relative importance of the climatic variables used in the model. Temperature-related features emerge as the most influential predictors of crop yield, followed by rainfall-related variables. This aligns with the nonlinear and dispersed patterns observed in Figures 1 and 2, where yield variability increases at extreme temperature and precipitation levels.

The dominance of temperature features suggests that thermal stress plays a critical role in driving yield outcomes, while rainfall contributes in a more complex, threshold-dependent manner. The uneven distribution of feature importance further supports the use of ensemble-based, nonlinear models such as Gradient Boosting, which are better suited to capturing interactions and nonlinear effects compared to traditional linear regression approaches.

Overall, the feature importance analysis provides interpretability to the model's predictions and reinforces the empirical findings from the exploratory visual analysis.

5.4. Comparative Interpretation and Implications

Overall, the results reveal that model choice plays a critical role in climate-informed agricultural prediction. While none of the models achieve high R^2 values reflecting the intrinsic uncertainty and multivariate complexity of agricultural systems, the Gradient Boosting model consistently delivers superior predictive accuracy. This finding suggests that AI-driven, iterative learning approaches are better suited for capturing subtle climate–yield interactions than static or bagging-based methods.

Importantly, the relatively low explanatory power across all models highlights the need for richer datasets incorporating soil characteristics, crop phenology, management practices, and remote sensing indicators. Rather than undermining the value of AI-based approaches, these results emphasize the structural challenges inherent in global-scale food security modeling under climate variability.

6. Climate Sensitivity and Feature Importance

Understanding how climatic variables influence crop yield variability is essential for designing robust, AI-driven food security strategies. Beyond predictive accuracy, interpretable insights into climate sensitivity provide actionable knowledge for agricultural planning, adaptation, and policy intervention. This section examines the relative importance of key climate features and their influence on crop yield predictions.

6.1. Climate Sensitivity Analysis

The exploratory analysis points out that crop yields are remarkably sensitive to temperature and precipitation changes. Yield responses were shown to be very nonlinear in the previous figures, although the extreme climate values caused them to be even more volatile. Temperature is a pivotal stressor, widely stressing up to the point where the yield variance increases hugely. Such a phenomenon can be attributed to plant physiology since the heat stress at the highest temperatures hampers photosynthesis, makes plants lose water faster, and shortens the time for crops to mature. The wide range of results for the highest temperature levels implies that even slight increases in average temperature may lead to huge changes in yield.

Rainfall demonstrates a similarly nonlinear relation with yield. Even though moderate rain helps growth of crops, too much or too little rain creates instability, probably by flooding, soil nutrient leaching, or drought stress. These results confirm that climatic influences on agriculture are not straightforward and that linear modeling assumptions cannot be relied on to capture these impacts adequately.

6.2. Feature Importance and Model Interpretability

The Gradient Boosting model's superior performance indicates its increased ability to uncover the complex relationship between different climate variables. Boosting techniques systematically highlight features that minimize the residual error, thus allowing the model to pay attention to the climate factors that have the greatest impact on the variation in crop yield. In this context, features associated with temperature and precipitation are the ones that have the biggest positive impact on the predictions. Their priority speaks of the basic nature of the climatic stressors that play a major role in the production of crops, especially in times of increasing climate variability. On the other hand, the Random Forest model's weaker performance implies that not all ensemble methods are equally successful in dealing with situations where the signal is very faint, and the data are very noisy.

Although explicit feature attribution methods (e.g., SHAP or permutation importance) were not applied in this study, the observed performance patterns and visual analyses provide strong inferential evidence that climate extremes rather than average conditions play a decisive role in determining yield outcomes. This highlights the importance of modeling tail risks and climate shocks in food security assessments.

6.3. Implications for Climate-Aware Agricultural Modeling

The climate sensitivity findings reinforce the necessity of interpretable AI models in agricultural decision-making. Rather than serving solely as black-box predictors, AI systems should function as diagnostic tools that identify vulnerable climate thresholds and inform adaptive strategies.

The relatively modest explanatory power observed across all models further suggests that climate variables alone are insufficient to fully explain yield variability. Integrating additional features such as soil properties, irrigation practices, crop varieties, and remote sensing indicators represents a critical avenue for improving both predictive accuracy and interpretability.

Overall, this analysis demonstrates that AI-driven models, particularly Gradient Boosting, offer meaningful advantages in capturing climate sensitivity in agricultural systems. These insights provide a foundation for the subsequent discussion on policy implications and future research directions aimed at strengthening global food security under increasing climate uncertainty.

7. Policy Implications

The findings of this study carry important implications for agricultural policy and global food security planning under increasing climate variability. The demonstrated sensitivity of crop yields to temperature and rainfall extremes underscores the urgency of incorporating climate-aware, AI-driven predictive systems into policy frameworks.

First, the superior performance of the Gradient Boosting model highlights the value of adaptive and nonlinear AI approaches for early-warning systems. Policymakers and agricultural agencies can leverage such models to identify regions at heightened risk of yield instability due to climate stressors, enabling proactive interventions such as targeted input subsidies, adjusted planting schedules, or emergency food reserves. Even modest improvements in predictive accuracy can translate into significant gains in preparedness at regional and national scales.

Second, the observed climate sensitivity patterns emphasize the need for risk-based agricultural policies rather than reliance on historical averages. Traditional planning approaches often underestimate tail risks associated with extreme weather events. AI-driven models capable of capturing nonlinear climate yield relationships can support dynamic policy responses, including climate-indexed insurance schemes and adaptive water management strategies that respond to real-time climate signals.

Third, interpretability remains a critical consideration for policy adoption. While advanced AI models improve predictive performance, their value for policymakers depends on transparency and trust. The emphasis on feature importance and climate sensitivity analysis in this study supports the use of interpretable AI frameworks that allow decision-makers to understand *why* certain regions or crops are flagged as vulnerable. This is particularly important in resource-constrained settings where policy interventions must be justified and prioritized carefully.

Finally, the results highlight the importance of sustained investment in agricultural data infrastructure. The relatively low explanatory power across models suggests that climate variables alone are insufficient to fully capture yield dynamics. Policies that promote open climate data, soil monitoring, remote sensing integration, and farm-level data collection will be essential for enhancing the effectiveness of AI-driven food security systems.

In summary, AI-driven predictive modeling should be viewed not as a replacement for traditional agricultural expertise, but as a complementary decision-support tool that enhances climate resilience, improves resource allocation, and strengthens global food security strategies under uncertainty.

8. Conclusion and Future Work

The research aimed at showing how AI-based predictive models can be applied in the estimation of crop yields during climate variability, which would help in the development of global food security enhancement strategies. The comparison of the Linear Regression, Random Forest, and Gradient Boosting techniques reveals that the nonlinear, boosting-based methods have advantages in reflecting the intricate climate-yield interactions even in the contexts with high uncertainty and weak signals. The findings point out that crop yields are very much affected by the variability in temperature and rainfall, particularly at the extremes. The results have shown that the linear model cannot be used in agricultural systems and stressed the need for AI frameworks that are climate-aware and can work with nonlinear dynamics. Even though Gradient Boosting was the best-performing method among the ones tested, the low explanatory power of all methods indicates the complexity of agricultural production and the presence of unmeasured factors that are not just climate-related.

From a policy perspective, this work supports the use of interpretable AI models as decision-support tools for early-warning systems, climate risk assessment, and adaptive agricultural planning. Rather than serving as deterministic predictors, such models can enhance preparedness, guide resource allocation, and improve resilience to climate shocks when integrated with domain expertise and local knowledge.

Future research should focus on enriching predictive frameworks through the integration of additional data sources, including soil characteristics, crop management practices, remote sensing indicators, and socioeconomic variables. The application of advanced interpretability techniques, such as SHAP-based feature attribution, would further strengthen transparency and trust in AI-driven food security systems. Additionally, extending the analysis across multiple crops, regions, and climate scenarios would improve generalizability and policy relevance.

Overall, this study contributes to the growing body of evidence that AI, when applied responsibly and interpretably, can play a meaningful role in addressing the challenges of agricultural sustainability and global food security under increasing climate uncertainty.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[2] FAO. (2021). The state of food security and nutrition in the world 2021: Transforming food systems for food security, improved nutrition and affordable healthy diets for all. Food and Agriculture Organization of the United Nations. https://www.fao.org/publications/sofi/2021/en/

[3] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

[4] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. PLOS ONE, 11(6), e0156571. https://doi.org/10.1371/journal.pone.0156571

[5] Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. Nature, 529(7584), 84–87. https://doi.org/10.1038/nature16467

[6] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674. https://doi.org/10.3390/s18082674

[7] 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. https://doi.org/10.1016/j.agrformet.2010.07.008

[8] Okolie, A. (2025a). Predicting food insecurity across U.S. census tracts: A machine learning analysis using the USDA Food Access Research Atlas. International Journal of Science and Research Archive, 17(2), 1156–1172. https://doi.org/10.30574/ijsra.2025.17.2.3156

[9] Okolie, A. (2025c). Machine learning approaches for predicting 30-day hospital readmissions: Evidence from Massachusetts healthcare data. World Journal of Advanced Research and Reviews, 28(1), 3457. https://doi.org/10.30574/wjarr.2025.28.1.3457

[10] Okolie, A., Obunadike, C., Okoro, S. C., & Akwabeng, P. M. (2025d). Heart disease prediction: A logistic regression approach. Open Journal of Applied Sciences, 15(11), 3534–3552. https://doi.org/10.4236/ojapps.2025.1511229

[11] Patel, R. (2019). Crop yield prediction dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset

[12] You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1), 4559–4565. https://ojs.aaai.org/index.php/AAAI/article/view/11168