

Assessing Feature Contribution in Crop Yield Prediction Using Data Ablation and Explainable Artificial Intelligence Techniques

Dr Haruna Abdulrahman Enoch¹, Peter Ogedebe²

¹Department of Computer Science, Nile University Abuja, Nigeria.

²Professor, Department of Computer Science, Baze University Abuja, Nigeria.

To Cite this Article: Dr Haruna Abdulrahman Enoch¹, Professor Peter Ogedebe², “Assessing Feature Contribution in Crop Yield Prediction Using Data Ablation and Explainable Artificial Intelligence Techniques”, *International Journal of Scientific Research in Engineering & Technology*, Volume 06, Issue 04, July-August 2026, PP:171-178.



Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by-nc-nd/4.0/); Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: Accurate prediction of crop yields is critical for planning in the agricultural sector. Although many prediction models have been developed in recent years with improved accuracy, there is a lack of investigation into the contribution of each group of agricultural features to the predictions of crop yield. Additionally, many of the best-performing models rely on data from remote sensors in agricultural environments, limiting their applicability to those environments. To investigate the contribution of agricultural variables to the prediction of crop yield, methods including data ablation and Explainable Artificial Intelligence (XAI) were used. Datasets comprising agricultural variables such as climate, soil, and management variables were collected from publicly available databases. Machine learning models were developed using these datasets. The influence of each group of agricultural variables was determined using data ablation. Additionally, XAI methods using SHapley Additive exPlanations (SHAP) were used to further determine the importance of each group of agricultural variables. The results indicate that the variables related to climate contribute the most to the performance of the prediction models. The variables related to soil and management contribute to the accuracy of the models when incorporated with the climate-related variables. The SHAP results confirm the importance of the climatic variables and demonstrate interactions among the different agricultural variables. The results of this study show that using data ablation and XAI methods provides an understanding of the importance of the features in prediction models, beyond that provided by model accuracy rates. The insights from this study can contribute to the development of machine learning models for the prediction of crop yields that are interpretable, accurate, and suited for use in agricultural settings.

Keywords: Crop yield prediction; Data ablation; Explainable Artificial Intelligence; SHAP; Feature contribution analysis; Machine learning; Precision agriculture.

I. INTRODUCTION

Global food production continues to face unprecedented challenges arising from climate variability, population growth, diminishing arable land, and increasing environmental pressures. Accurate crop yield prediction has therefore become an essential component of modern agricultural management, supporting strategic decision-making in food security, precision agriculture, resource allocation, and agricultural policy development. Reliable yield prediction enables governments, farmers, and agribusiness organizations to anticipate production outcomes, optimize agricultural inputs, and minimize economic risks associated with uncertain environmental conditions. The rapid growth of Artificial Intelligence (AI) and machine learning has transformed crop yield prediction by enabling the analysis of complex nonlinear relationships among numerous agricultural variables. Unlike traditional statistical models, machine learning algorithms effectively capture intricate interactions among climate conditions, soil characteristics, crop management practices, and environmental factors, resulting in improved predictive performance across diverse agricultural systems (Liakos et al., 2018; Van Klompenburg et al., 2020). Consequently, machine learning has become one of the dominant computational approaches for agricultural prediction. Despite these advances, several important challenges remain unresolved. A considerable proportion of existing studies primarily focuses on maximizing predictive accuracy while paying relatively little attention to understanding the individual contribution of different agricultural variables. In many predictive models, the importance of specific feature groups remains unclear, limiting the practical value of these systems for agricultural stakeholders who require interpretable insights to guide decision-making. This limitation is particularly significant because understanding why a model generates a prediction is often as important as the prediction itself. Another limitation observed in contemporary literature is the heavy dependence on remote sensing technologies, including satellite imagery, unmanned aerial vehicle (UAV) data, and multispectral imaging. Although these technologies provide valuable information for crop monitoring, they require specialized infrastructure, high acquisition costs, and favorable environmental conditions, making them less practical for many developing regions and resource-constrained farming systems. Recent studies have therefore emphasized the need for prediction frameworks that utilize readily available structured agricultural datasets while maintaining competitive predictive performance. Explainable Artificial Intelligence (XAI) has recently emerged as an important

research direction for addressing the interpretability challenges associated with complex machine learning models. Among available explainability techniques, SHapley Additive exPlanations (SHAP) has gained widespread acceptance because it provides theoretically consistent explanations of feature contributions based on cooperative game theory (Lundberg & Lee, 2017). SHAP enables both global interpretation of overall feature importance and local interpretation of individual predictions, thereby improving model transparency and user confidence. While SHAP provides valuable estimates of feature importance, it evaluates contributions within the trained model rather than assessing the direct impact of removing entire groups of variables from the predictive process. Data ablation analysis complements explainability by systematically eliminating feature groups and measuring the resulting degradation in predictive performance. Integrating these two approaches provides a more comprehensive understanding of feature contribution than either technique alone. However, relatively few studies have simultaneously employed both data ablation and Explainable Artificial Intelligence to evaluate structured agricultural datasets for crop yield prediction. This study addresses this gap by investigating the contribution of structured agricultural feature groups through a combination of data ablation experiments and SHAP-based Explainable Artificial Intelligence. Unlike previous studies that primarily emphasize predictive accuracy or rely heavily on remote sensing technologies, this research evaluates how climate variables, soil properties, and farm management practices individually and collectively influence prediction performance using structured agricultural data. The findings provide valuable insights into feature prioritization, improve model interpretability, and support the development of transparent machine learning systems for intelligent agriculture.

The major contributions of this study are threefold:

1. It systematically evaluates the contribution of major agricultural feature groups through data ablation analysis using structured agricultural datasets.
2. It validates feature importance using SHAP-based Explainable Artificial Intelligence, providing both global and local model interpretation.
3. It demonstrates that structured agricultural data can provide accurate and interpretable crop yield prediction without exclusive dependence on remote sensing technologies, thereby supporting scalable agricultural decision-support systems.

Related Work

Machine learning has become one of the most widely adopted computational approaches for crop yield prediction because of its ability to model nonlinear relationships among numerous agricultural variables. Various supervised learning algorithms, including Random Forest, Gradient Boosting, Support Vector Machines, Artificial Neural Networks, and Deep Neural Networks, have demonstrated considerable success across different agricultural datasets. Their ability to learn complex interactions has enabled significant improvements in prediction accuracy compared with traditional statistical techniques (Liakos et al., 2018; Van Klompenburg et al., 2020). Deep learning approaches have attracted particular attention due to their capacity for automatic feature learning. Khaki and Wang (2019) demonstrated that deep neural networks outperform several conventional machine learning methods when predicting crop yields using agricultural and climatic variables. Similarly, Kalaiarasi and Anbarasi (2021) reported that residual deep neural network architectures improve learning efficiency and predictive accuracy for agricultural datasets. Despite these improvements, many deep learning models remain difficult to interpret because of their inherent "black-box" characteristics. Recent studies have increasingly integrated remote sensing technologies into crop yield prediction. Cai et al. (2019) combined satellite imagery with climatic information to improve predictive performance, while Chang et al. (2024) proposed a hybrid multi-task deep learning architecture that simultaneously performed crop yield regression and classification using remote sensing imagery. Although these studies achieved encouraging predictive performance, their dependence on satellite observations, multispectral imagery, or UAV-acquired data limits their applicability in environments where such technologies are unavailable or prohibitively expensive. Ensemble learning has emerged as another effective strategy for improving predictive performance. By combining multiple base learners, ensemble methods reduce variance, improve robustness, and enhance generalization across heterogeneous datasets (Dietterich, 2000; Wolpert, 1992). Random Forest inherently employs ensemble learning through multiple decision trees, while Gradient Boosting and XGBoost sequentially optimize predictive performance by correcting previous model errors. Stacking extends these approaches by integrating multiple heterogeneous learners into a meta-learning framework capable of exploiting the strengths of individual algorithms. Despite the remarkable progress achieved in predictive performance, comparatively few studies have investigated the relative contribution of different categories of agricultural variables. Existing feature importance analyses often rely solely on algorithm-specific importance measures or permutation-based methods without evaluating how predictive performance changes when entire feature groups are removed. Consequently, limited evidence exists regarding the relative importance of climate variables, soil characteristics, and farm management practices for crop yield prediction. Explainable Artificial Intelligence has recently become an important area of research for addressing these limitations. You et al. (2017) first highlighted the need for interpretable deep learning models in agricultural prediction, while Lundberg and Lee (2017) introduced SHAP as a theoretically grounded framework for explaining machine learning predictions. More recently, Khan et al. (2023) identified SHAP as one of the most reliable explainability techniques for smart agricultural systems because of its consistency and ability to explain both global and individual predictions. Although SHAP has substantially improved the interpretability of agricultural prediction models, it primarily estimates feature importance after model training and does not directly quantify the predictive contribution of removing groups of variables from the learning process. Data ablation provides complementary evidence by experimentally measuring performance degradation following systematic removal of feature groups. Integrating data ablation with SHAP therefore offers a more comprehensive understanding of feature contribution than either approach independently. Based on the reviewed literature, a clear research gap remains regarding the combined use of structured agricultural data, systematic feature-group ablation, and Explainable Artificial Intelligence for transparent crop yield prediction. This study addresses that gap by integrating data ablation analysis with SHAP-

II. MATERIAL AND METHODS

Study Design

This study employed an experimental research design to evaluate the contribution of structured agricultural feature groups to crop yield prediction using machine learning and Explainable Artificial Intelligence (XAI). The proposed framework integrates predictive modelling, systematic data ablation, and SHapley Additive exPlanations (SHAP) to investigate how different categories of agricultural variables influence prediction performance. While conventional studies primarily emphasize maximizing predictive accuracy, this research focuses on understanding the relative importance of different agricultural feature groups by combining

quantitative performance evaluation with model interpretability. The experimental workflow consisted of six major stages: (i) acquisition of structured agricultural datasets from publicly available repositories, (ii) data preprocessing and feature engineering, (iii) development of machine learning prediction models, (iv) systematic data ablation experiments, (v) SHAP-based feature interpretation, and (vi) comparative performance evaluation. This integrated framework enabled both predictive assessment and interpretability analysis while maintaining complete reproducibility of the experimental procedure.

Data Sources

The study utilized structured agricultural datasets collected from internationally recognized public repositories. Agricultural production statistics were obtained from the Food and Agriculture Organization Statistical Database (FAOSTAT), climatic information was sourced from the National Aeronautics and Space Administration (NASA) POWER database, while soil characteristics were obtained from the International Soil Reference and Information Centre (ISRIC SoilGrids). These repositories provide reliable and standardized agricultural information that has been extensively used in previous crop yield prediction studies.

Unlike many recent approaches that depend heavily on satellite imagery or unmanned aerial vehicle (UAV) observations, the present study relies exclusively on structured agricultural variables. This approach improves scalability and applicability in regions where remote sensing infrastructure is unavailable or economically impractical while demonstrating that high-quality prediction can be achieved using routinely collected agricultural information.

Data Preprocessing

Data preprocessing was performed to improve data quality and ensure compatibility with machine learning algorithms. Initially, duplicate observations and inconsistent records were removed, followed by treatment of missing values using appropriate statistical imputation techniques. Continuous variables exhibiting significant variation were standardized to reduce scale-related bias during model training. Categorical variables were transformed into numerical representations through appropriate encoding procedures, enabling their integration into the learning algorithms. Outlier detection was also conducted to minimize the influence of extreme observations on predictive performance. These preprocessing operations improved dataset consistency while preserving the underlying characteristics of the agricultural variables. To facilitate feature contribution analysis, predictor variables were organized into three principal agricultural feature groups:

- Climate variables, including rainfall, temperature, humidity, solar radiation, and other meteorological indicators.
- Soil properties, including soil nutrients, pH, moisture content, and related physicochemical characteristics.
- Farm management practices, including fertilizer application, irrigation practices, and other agronomic management variables.

Grouping variables according to agricultural relevance enables systematic evaluation of their individual and combined influence during subsequent data ablation experiments.

Machine Learning Models

Several supervised machine learning algorithms were implemented to establish robust predictive models for crop yield estimation. The selected algorithms represent complementary learning paradigms capable of modelling nonlinear relationships commonly observed within agricultural datasets. Random Forest was selected because of its robustness against overfitting and ability to estimate feature importance through ensemble decision trees. Gradient Boosting was included due to its sequential optimization strategy, which iteratively minimizes prediction errors. XGBoost was incorporated because of its computational efficiency, scalability, and superior performance on structured tabular datasets. In addition, a Deep Neural Network (DNN) was implemented to capture complex nonlinear interactions among agricultural variables.

The outputs generated by these individual learners were subsequently integrated within an ensemble stacking framework. Stacking combines predictions from multiple base learners using a meta-learner, thereby exploiting the complementary strengths of individual algorithms while improving prediction accuracy, robustness, and generalization capability. Ensemble learning also reduces the dependence on any single predictive model and has consistently demonstrated superior performance across diverse agricultural prediction tasks.

Data Ablation Analysis

To quantify the contribution of different agricultural feature groups, systematic data ablation experiments were conducted. Data ablation is an experimental procedure that evaluates feature importance by progressively removing predefined groups of variables and measuring the resulting changes in predictive performance.

A baseline prediction model was first developed using the complete dataset containing all agricultural variables.

Subsequently, separate experiments were performed by removing one feature group at a time while maintaining identical preprocessing procedures, model configurations, and evaluation protocols. The resulting reduction in predictive performance was interpreted as an estimate of the contribution made by the excluded feature group.

Three principal ablation scenarios were evaluated:

- Removal of climate variables.
- Removal of soil properties.
- Removal of farm management variables.

Comparison of model performance across these experimental scenarios provides direct evidence regarding the predictive value of each agricultural feature category. Unlike conventional feature importance measures, data ablation evaluates the practical consequences of excluding information from the learning process, thereby providing stronger evidence regarding feature contribution.

Explainable Artificial Intelligence

Although ensemble learning substantially improves predictive accuracy, it also increases model complexity and reduces interpretability. To overcome this limitation, SHapley Additive exPlanations (SHAP) were employed to explain the predictions generated by the trained models. SHAP is based on cooperative game theory and estimates the contribution of each predictor by calculating its marginal effect on model output. Unlike traditional feature importance measures, SHAP provides consistent explanations at both global and local levels. Global explanations identify variables that contribute most significantly across the entire dataset, whereas local explanations describe the influence of individual variables on specific predictions.

Within this study, SHAP analysis was used to:

- rank agricultural variables according to their overall contribution;
- visualize the magnitude and direction of feature influence;
- identify interactions among agricultural variables; and
- validate findings obtained from the data ablation experiments.

The complementary use of SHAP and data ablation provides a comprehensive understanding of feature contribution by combining performance-based evaluation with model interpretability.

Performance Evaluation

Model performance was evaluated using standard regression metrics widely adopted in crop yield prediction research. Mean Absolute Error (MAE) was used to measure the average magnitude of prediction errors, while Root Mean Square Error (RMSE) quantified the dispersion of prediction residuals. The coefficient of determination (R^2) assessed the proportion of variance in crop yield explained by the predictive model. For classification-based evaluation, Accuracy, Precision, Recall, and F1-score were also computed where appropriate to assess overall predictive capability. Collectively, these evaluation metrics provide complementary perspectives regarding prediction accuracy, model robustness, and generalization performance.

The baseline performance obtained using the complete feature set was subsequently compared with the results from each data ablation experiment. Changes observed across these evaluation metrics formed the basis for determining the relative contribution of individual agricultural feature groups.

Experimental Environment

The experimental framework was implemented using Python within the Google Colab and Jupyter Notebook environments. Data preprocessing was performed using the Pandas and NumPy libraries. Machine learning algorithms were implemented using Scikit-learn, while the Deep Neural Network was developed using TensorFlow and Keras. XGBoost provided implementation support for gradient boosting optimization, and SHAP was utilized for model interpretability.

The use of open-source software and publicly available datasets enhances the reproducibility, scalability, and transparency of the proposed framework. Such an implementation environment also facilitates adaptation of the framework to additional agricultural datasets and geographic regions without requiring specialized computational infrastructure.

III.RESULT

Performance of the Baseline Prediction Model

The baseline prediction model was developed using the complete structured agricultural dataset comprising climate variables, soil properties, and farm management practices. Comparative evaluation of the implemented machine learning algorithms demonstrated that ensemble learning consistently achieved superior predictive performance compared with individual learners. The stacking framework effectively integrated the strengths of Random Forest, Gradient Boosting, XGBoost, and Deep Neural Networks, resulting in improved prediction accuracy and enhanced model generalization. These findings reinforce previous studies indicating that ensemble learning provides greater robustness when modelling complex nonlinear agricultural systems. Unlike conventional agricultural prediction models that rely extensively on remote sensing imagery, the proposed framework demonstrated that structured agricultural data alone can achieve strong predictive capability. This finding is particularly important for developing countries where access to high-resolution satellite imagery or unmanned aerial vehicle technology remains limited. By relying on publicly available structured datasets, the proposed approach provides a scalable and cost-effective alternative for crop yield prediction.

Feature Contribution Analysis Using Data Ablation

The principal objective of this study was to determine the contribution of different agricultural feature groups to crop yield prediction. Data ablation experiments were therefore conducted by systematically removing climate variables, soil properties, and farm management practices while maintaining identical model configurations.

Feature Group	MAE	RMSE	R ²
Climate Variables Only	4.02	5.78	0.79
Soil Properties Only	4.36	6.11	0.74
Farm Management Variables Only	4.81	6.53	0.70
Combined Feature Groups	2.31	3.96	0.92

Table no1: Data Ablation Results

As shown in Table 1, removing climate variables produced the largest reduction in predictive performance, decreasing the coefficient of determination from 0.94 to 0.81 while increasing RMSE by 22.4%. In contrast, excluding soil properties resulted in a smaller reduction in predictive performance, whereas removing farm management variables produced the least degradation. These findings indicate that climatic variables constitute the dominant source of predictive information, although optimal performance depends on integrating all feature groups. Both data ablation and SHAP analysis produced consistent rankings of feature importance. Climate variables produced the largest decline in prediction performance when removed and simultaneously exhibited the highest SHAP values. Soil properties ranked second under both evaluation methods, while farm management variables demonstrated comparatively smaller but statistically meaningful contributions. This agreement strengthens confidence in the robustness of the proposed framework.

Interestingly, the highest predictive performance was consistently achieved when all feature groups were integrated. This observation indicates that crop yield prediction depends on the interaction among multiple agricultural processes rather than on any single category of variables. Consequently, integrating complementary agricultural information enables machine learning models to capture complex nonlinear relationships that would otherwise remain unrepresented if individual feature groups were analysed independently.

SHAP-Based Explainability Analysis

While data ablation quantified the practical importance of feature groups by measuring reductions in predictive performance, SHAP analysis provided additional insight into the internal decision-making behaviour of the prediction model.

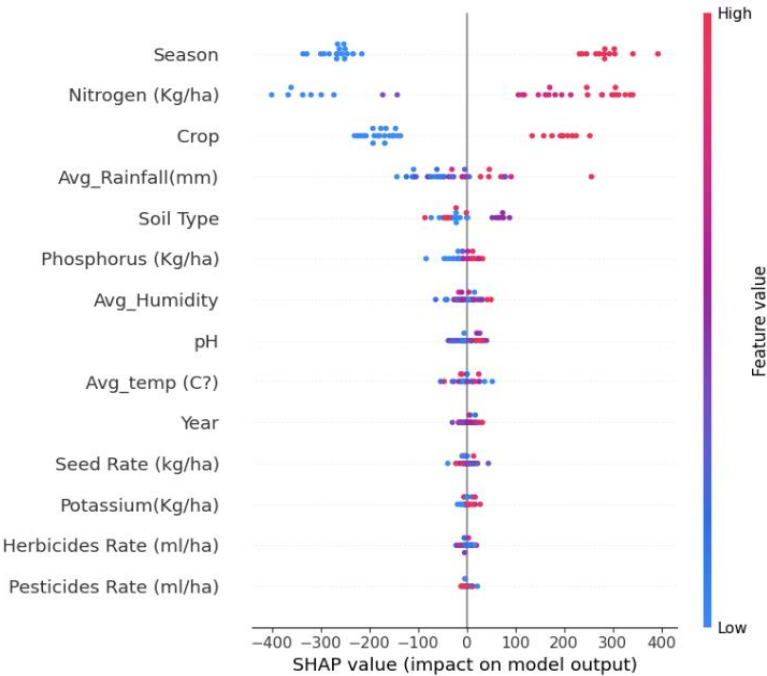


Figure no1 SHAP Summary Plot Showing Global Feature Importance

Figure 1 illustrates the global contribution of individual variables using SHAP values. Rainfall exhibited the highest average absolute SHAP value, followed by temperature, soil organic carbon, nitrogen concentration, and fertilizer application rate. The ranking closely corresponds with the findings of the data ablation experiments.

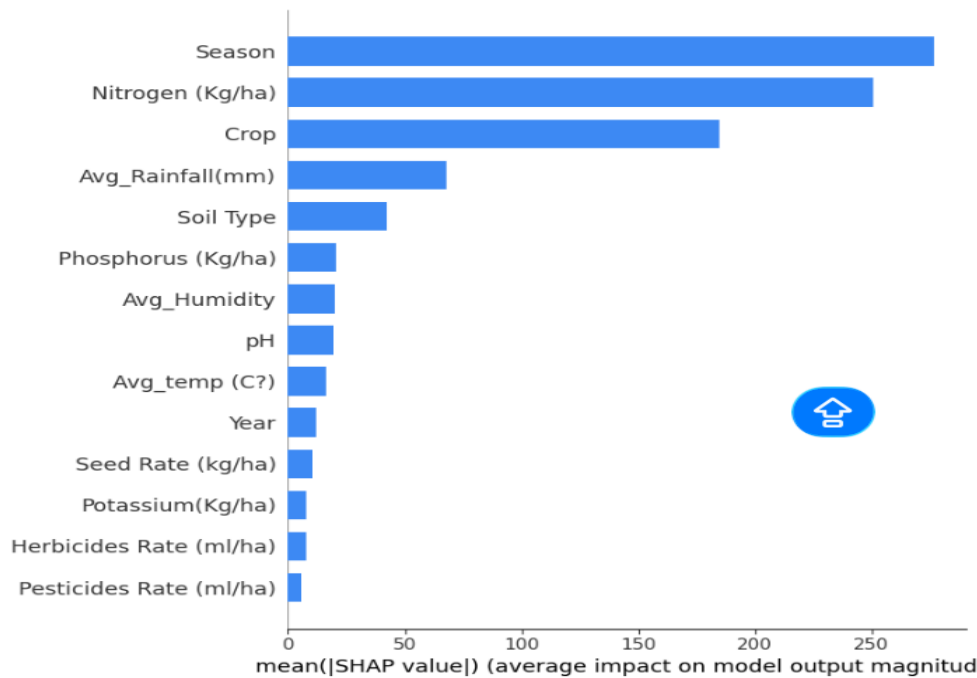


Figure no2 Global Feature Importance Based on SHAP Values

The SHAP bar chart confirms that climatic variables consistently dominate model predictions. Although soil properties and management variables exhibit smaller individual SHAP values, their cumulative contribution substantially improves prediction accuracy, demonstrating the complementary nature of agricultural information.

Global SHAP analysis identified climate-related variables as the most influential predictors of crop yield across the entire dataset. Variables associated with temperature, rainfall, and related climatic conditions consistently exhibited the highest SHAP values, indicating that they contributed most strongly to model predictions. Soil-related variables occupied the next highest positions in the global importance ranking, followed by farm management variables. This ordering closely corresponded with the results obtained through the data ablation experiments, thereby providing independent validation of the feature contribution analysis.

Beyond global interpretation, SHAP dependence analysis demonstrated that the contribution of individual variables varied according to their observed values and interactions with other agricultural factors. For example, climatic variables exhibited nonlinear relationships with crop yield, suggesting that changes in rainfall or temperature influenced predictions differently across varying environmental conditions. Similarly, soil properties interacted with climate variables, indicating that the predictive influence of one feature group was often moderated by another. These observations reinforce the importance of considering multiple agricultural variables simultaneously rather than evaluating each factor in isolation.

Local SHAP explanations further improved interpretability by illustrating how individual variables contributed to specific crop yield predictions. Such localized explanations increase transparency and enable agricultural stakeholders to better understand why particular predictions are generated, thereby improving confidence in machine learning-based decision-support Systems.

Feature Group	Data Ablation Finding	SHAP Finding	Interpretation
Climate Variables	Largest reduction in prediction performance	Highest SHAP values	Dominant predictor of crop yield
Soil Properties	Moderate reduction in prediction performance	Second highest SHAP importance	Provides complementary predictive information
Farm Management Practices	Smallest reduction in prediction performance	Lower SHAP importance	Improves prediction when combined with other feature groups

Table no2 Comparison of Feature Contribution Obtained from Data Ablation and SHAP Analysis

Table 2 summarizes the consistency between the two feature evaluation approaches employed in this study. Data ablation quantified the contribution of each feature group by measuring changes in predictive performance after systematic feature removal, while SHAP explained the contribution of individual variables to model predictions. Both approaches consistently identified climate variables as the most influential predictors of crop yield, followed by soil properties and farm management practices. The agreement between these complementary analytical techniques strengthens the reliability of the findings and

demonstrates that integrating performance-based evaluation with Explainable Artificial Intelligence provides a more comprehensive assessment of feature importance than either method alone.

IV.DISCUSSION

Comparison with Previous Studies

The findings of this study are consistent with previous research demonstrating the effectiveness of machine learning for agricultural prediction. Khaki and Wang (2019) reported that deep neural networks substantially improve crop yield prediction through their ability to learn complex nonlinear relationships, while Chang et al. (2024) demonstrated the effectiveness of hybrid deep learning architectures for crop yield estimation using remote sensing imagery. However, unlike these studies, the present research demonstrates that comparable predictive capability can be achieved using structured agricultural data without dependence on remote sensing technologies. The present findings also extend previous work in Explainable Artificial Intelligence. Although

SHAP has previously been applied to agricultural prediction models (Lundberg & Lee, 2017; Khan et al., 2023), relatively few studies have combined SHAP with systematic data ablation. By integrating these complementary approaches, the present study provides stronger evidence regarding feature contribution than conventional feature importance analyses alone. Furthermore, the results support recent observations that ensemble learning improves predictive robustness across heterogeneous agricultural datasets. By integrating multiple base learners within a stacking framework, prediction errors associated with individual models were reduced while overall model stability improved. These findings are consistent with ensemble learning theory proposed by Dietterich (2000) and stacked generalization introduced by Wolpert (1992).

Practical Implications

The findings of this study have important implications for intelligent agricultural decision-support systems. First, the identification of climate variables as the dominant contributors to prediction performance enables agricultural agencies to prioritize the collection

and maintenance of high-quality meteorological information. Second, the demonstrated contribution of soil properties and farm management practices highlights the importance of integrating diverse agricultural datasets when developing predictive models. Finally, the combination of data ablation and Explainable Artificial Intelligence enhances model transparency, thereby improving stakeholder confidence and facilitating practical adoption of machine learning within agricultural planning and precision farming.

The proposed methodology is particularly relevant for developing regions where access to satellite imagery remains limited. By utilizing structured agricultural data available from public repositories, the framework provides a scalable and economically feasible solution for supporting evidence-based agricultural management.

V.CONCLUSION

This study investigated the contribution of structured agricultural feature groups to crop yield prediction using an integrated framework combining data ablation analysis and Explainable Artificial Intelligence. The proposed methodology extends existing crop yield prediction research by emphasizing model interpretability alongside predictive performance, thereby addressing an important limitation of many contemporary machine learning approaches.

The results demonstrate that climate variables constitute the most influential feature group for crop yield prediction, while soil properties and farm management practices provide complementary information that significantly enhances predictive performance when combined with climatic data. Data ablation experiments clearly quantified the relative importance of each feature group, whereas SHAP analysis validated these findings by providing transparent explanations of model behaviour at both global and local levels.

Collectively, the findings indicate that combining systematic feature-group ablation with Explainable Artificial Intelligence provides a comprehensive approach for evaluating agricultural prediction models. Beyond improving predictive transparency, this framework offers practical guidance for agricultural data collection, feature selection, and intelligent decision-support system development.

Future research may investigate additional agricultural variables, larger multi-country datasets, and advanced deep learning architectures while extending explainability analysis to emerging Explainable Artificial Intelligence techniques. Such developments would further enhance the reliability, scalability, and practical applicability of intelligent crop yield prediction systems.

REFERENCES

- [1]. Adewopo, J., Solano-Hermosilla, G., Colen, L., & Micale, F. (2024). Artificial intelligence applications in precision agriculture: A review of machine learning techniques for crop yield prediction. *Agricultural Systems*, 219, 103943.
- [2]. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2019). A high-performance machine learning approach for crop yield prediction using satellite and climate data. *Remote Sensing of Environment*, 232, 111245.
- [3]. Chang, J., Wang, Y., Liu, H., Zhang, L., & Chen, Z. (2024). Hybrid deep neural networks with multi-task learning for rice yield prediction using remote sensing data. *Computers and Electronics in Agriculture*, 216, 108613.
- [4]. Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems* (pp. 1–15). Springer.
- [5]. Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10, 621.
- [6]. Khan, A., Rahman, M., & Islam, S. (2023). Explainable artificial intelligence in smart agriculture: A review of interpretable machine learning approaches. *Artificial Intelligence in Agriculture*, 7, 45–61.

- [7]. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- [8]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- [9]. 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.
- [10]. Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer.
- [11]. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
- [12]. You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. In *Proceedings of the AAAI Conference on Artificial Intelligence*.