IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 7, June 2025

Impact Factor: 7.67

XAI Enabled Forecasting of Crop Yields In India: Advancing Machine learning in Agriculture

Mrs. Ch. Srivatsa Alivelu Mangatayi[1], Singarapu Kalyani[2], Devavarapu Tanuj[3], G Bharath[4], Bandari Pavani[5]

Assistant Professor, Department Computer Science Engineering[1]

Students, Department Computer Science Engineering[2-5]

ACE Engineering College, Ghatkesar, India

Abstract: *Agricultural productivity plays a vital role in ensuring food security in India. However, traditional methods for crop selection and yield estimation often rely on guesswork and are susceptible to climatic uncertainties. This project leverages machine learning models to accurately predict crop yields and recommend suitable crops based on environmental and soil parameters. To improve transparency and trust in the system, Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME are integrated to interpret model predictions. By utilizing real-time data, the system enhances prediction accuracy and provides farmers and stakeholders with clear, data-driven insights for informed agricultural decision-making. The system is trained on diverse datasets including weather conditions, soil health metrics, and historical crop performance to ensure robustness across different regions. It supports adaptive learning to continuously update recommendations as new data becomes available. The integration of XAI not only boosts confidence in AI-driven outputs but also aids in identifying key factors affecting yield. This approach bridges the gap between advanced technology and grassroots-level farming, contributing to sustainable agriculture and better resource management[1]*

Keywords: Agricultural productivity, Crop selection, Yield estimation, Random forest, Environmental parameters, Soil parameters, SHAP (Shapley Additive exPlanations), LIME (Local Interpretable Modelagnostic Explanations)

I. INTRODUCTION

Agriculture is the backbone of India's economy, providing livelihoods to millions of farmers and playing a vital role in ensuring food security. Accurate crop yield prediction is essential for effective planning, resource allocation, and decision-making in the agricultural sector. Farmers, policymakers, and agricultural experts rely on these forecasts to make informed decisions about planting, harvesting, and resource management. Traditional methods of yield prediction often fall short in accuracy and adaptability, especially in the face of changing climatic conditions and diverse agricultural practices across India. With advancements in technology, Machine Learning (ML) has emerged as a powerful tool for forecasting crop yields by analyzing various factors such as weather conditions, soil quality, and rainfall patterns. While ML models provide improved accuracy over traditional methods, a significant challenge remains— the "black box" nature of many algorithms. This lack of transparency makes it difficult for end-users to understand how predictions are made, limiting trust and widespread adoption of these technologies in agriculture. To overcome this challenge, the proposed project introduces an Explainable Artificial Intelligence (XAI)- enabled system for forecasting crop yields in India. The primary objective of this project is to not only provide accurate yield predictions but also offer clear and understandable explanations of how these predictions are derived. By integrating XAI techniques, the system ensures that stakeholders, including farmers and agricultural policymakers, can comprehend and trust the model's outputs. The project focuses on developing a user-friendly interface where users can input relevant data to receive both predictions and their explanations. The XAI component helps identify and visualize the most influential factors affecting crop yields, enhancing transparency and enabling data-driven decisionmaking. By bridging the gap between complex machine learning models and practical agricultural applications, this project aims to improve agricultural productivity, promote sustainable farming practices, and build trust in AI-driven solutions. In conclusion, this project represents a significant step toward leveraging advanced technologies to address critical challenges in Indian agriculture. The integration of explainable AI into crop yield prediction not only enhances the model's reliability but also empowers users with actionable insights, contributing to a more sustainable and efficient agricultural ecosystem.

II. LITERATURE SURVEY

[1]Sharma, S., Rai, S., & Krishnan, N. C. (2020). Wheat Crop Yield Prediction Using Deep LSTM Model. *arXiv preprint arXiv:2011.01498*

This paper introduces a novel approach for early in-season wheat yield forecasting by employing a deep Long ShortTerm Memory (LSTM) model. Unlike traditional models that depend on handcrafted features, this study processes raw satellite imagery directly using LSTM, making the system more scalable and generalizable. The model was tested at the tehsil level across multiple states in India and achieved more than 50% improvement over existing baseline methods[1].

[2]Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2020). Crop Yield Prediction Integrating Genotype and Weather Variables Using Deep Learning. *arXiv preprint arXiv:2006.13847*

This study highlights the integration of genotype (genetic) data and weather variables using deep learning to predict crop yields more accurately. The model was trained on a comprehensive dataset encompassing various crop genotypes and corresponding climate conditions. By modeling the interaction between genotype and environmental factors, the research shows a notable improvement in yield prediction accuracy compared to conventional approaches[2].

[3]Fernandes, I., Vieira, C., Dias, K., & Fernandes, S. (2024).

Using Machine Learning to Combine Genetic and Environmental Data for Maize Grain Yield Predictions Across Multi-Environment Trials. Theoretical and Applied Genetics. This study explores the application of machine learning techniques to predict maize grain yields by combining genetic and environmental data. Conducted across multiple environmental trials, the research demonstrates that integrating these data sources can significantly improve the accuracy of yield predictions, offering valuable insights for breeding and agricultural practices[3]

[4]**Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z., & Tao, F. (2020). Wheat yield predictions at a county and field scale with deep learning, machine learning, and Google Earth Engine. *European Journal of Agronomy, 123*, 126204.**

This paper provides a comparative analysis of deep learning and traditional machine learning models for wheat yield forecasting using satellite imagery and weather data. The models were implemented via Google Earth Engine, allowing for scalable and cloud-based data processing. The deep learning approach, particularly using Convolutional Neural Networks (CNNs), showed superior performance at both county and field levels[4].

[5]Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronicsin Agriculture, 147*, 70–90.

This survey provides a comprehensive overview of deep learning applications across various agricultural domains, including crop yield prediction, plant disease detection, and livestock monitoring. The authors reviewed over 40 research papers, categorizing them based on application area and deep learning architecture. For yield prediction, the study notes that CNNs and RNNs (especially LSTMs) have been the most effective in analyzing image and time-series data, respectively[5].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

[6]Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science, 10*, 621.
This research investigates the use of deep neural networks (DNNs) to forecast corn yield by incorporating both historical weather data and satellite-derived vegetation indices. The DNN model achieved high predictive accuracy compared to regression and support vector machine models. The study utilized data spanning over 15 years and showed the DNN's strength in capturing non-linear relationships among variables[6].

III SYSTEM ARCHITECTURE

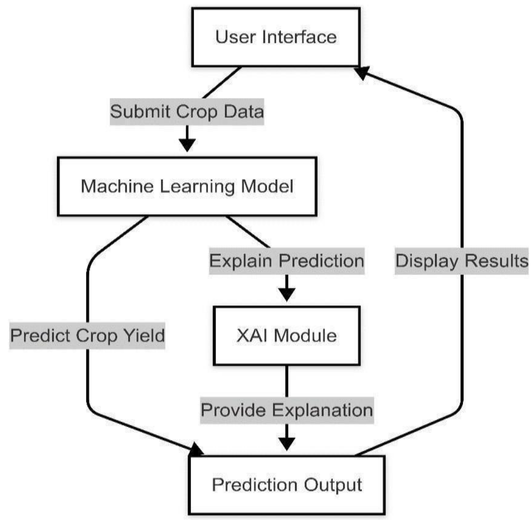

Fig 3.1:System Architecture

3.1.Random Forest Regressor

The Random Forest Regressor is widely used in crop yield forecasting to predict the amount of produce (such as quintals per hectare) based on multiple agricultural and environmental factors like soil characteristics, rainfall, temperature, and historical yields. It functions by constructing numerous decision trees and averaging their outputs to enhance prediction accuracy and reduce overfitting. In the context of Indian agriculture, where data can be noisy or incomplete, this model offers a highly robust solution for yield estimation. When integrated with explainability tools such as SHAP, farmers and agricultural planners can gain insights into how each input feature (e.g., rainfall or temperature) contributed to the yield prediction, thereby increasing transparency, building trust in the predictions, and enabling informed decisionmaking at both the micro and macro levels.

3.2 Random Forest Classifier

The Random Forest Classifier plays a crucial role in crop recommendation systems by identifying the most suitable crop to grow based on input features such as soil type, season, and regional climate conditions. It uses an ensemble of decision trees that vote on the best crop class (e.g., wheat, rice, maize), offering a more reliable and accurate recommendation system. In a country like India, where crop selection is often based on intuition or limited knowledge, this model provides datadriven support to maximize productivity and sustainability. When combined with XAI techniques, such as SHAP or LIME, the classifier's decisions become interpretable, helping farmers understand the rationale behind each recommendation—for instance, why rice is preferred in a high-rainfall region—thus encouraging the adoption of intelligent systems in agriculture**.**

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

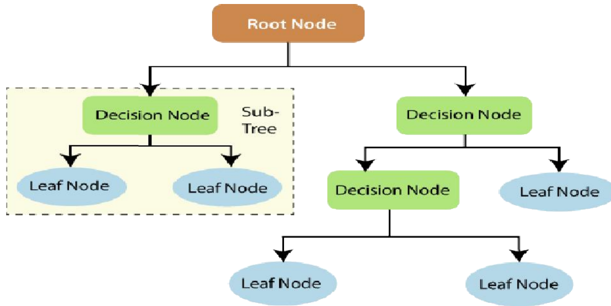

Fig 3.2 Random Forest

3.3 SHAP (Shapley Additive exPlanations)

SHAP is a powerful Explainable AI tool used to interpret the predictions made by complex machine learning models by quantifying the contribution of each input feature to a given prediction. In the domain of crop yield forecasting, SHAP helps both developers and end-users understand how factors such as temperature, soil moisture, and fertilizer use impact the predicted outcome. By offering both global explanations (overall feature importance across the dataset) and local explanations (specific to a single prediction), SHAP fosters transparency in model behavior. This is particularly valuable in Indian agriculture, where trust in AI systems is critical. With SHAP, farmers and stakeholders can visualize and understand why certain decisions were made, such as a lower predicted yield due to poor soil pH, thus enhancing confidence in the use of machine learning for critical agricultural decisions.

3.4 LIME

LIME is another Explainable AI method that focuses on generating simple, interpretable models around individual predictions to help users understand the output of more complex, black-box models. In agricultural applications like crop yield prediction or crop recommendation, LIME can take a specific input—such as a farm's environmental and soil data— and provide an easily understandable explanation for why the model produced a certain output. For example, it might show that low nitrogen levels and poor rainfall were key reasons for a low yield prediction for a particular field. This level of local interpretability is essential in the Indian context, where farmers need clear, actionable insights to make critical farming decisions. LIME thereby enhances the credibility and usability of machine learning models, fostering their integration into realworld agricultural practices

IV. METHODOLOGY

4.1 Input Data Collection : Two datasets are used:

Crop Recommendation Dataset (Crop_recommendation.csv) Contains 2,202 entries with soil and environmental parameters like Nitrogen, Phosphorus, Potassium, pH, Temperature, Humidity, Rainfall, and a target variable (Crop Label).

Crop Production Dataset (crop_production.csv) A larger dataset (15 MB) containing crop yield, area, season, temperature, humidity, rainfall, and total production volume, aggregated over different Indian states and years.

4.2.2 Data Preprocessing

Null values are identified and either imputed or dropped Crop names (categorical targets) are encoded for model compatibility.Duplicates are removed, and categorical fields like 'Season' are standardized.

4.3.3 Data Splitting

The pre processed data is split into training and testing sets using an 80:20 ratio. For the Crop Recommendation task, the input features are N, P, K, pH, temperature, humidity, and rainfall; the output is the crop label. For Yield Prediction, features include area, season, and environmental parameters; the target is yield (tonnes/hectare).

4.4.4 Model Selection and Training

Random Forest Classifier is trained on the crop recommendation dataset to identify the best crop for given soil and environmental parameters.Random Forest Regressor is trained on the crop production dataset to estimate expected yield based on environmental and geographical factors.Both models are evaluated using cross-validation techniques for robustness.

4.5.5 Model Evaluation

Classification metrics: Accuracy, Precision, Recall, F1-score.Regression metrics: R^2 Score, Mean Absolute Error (MAE), Mean Squared Error (MSE).Models are validated against the test dataset to ensure generalization capability.

Integration of Explainable AI(XAI) SHAP (Shapley Additive Explanations):

Offers global and local interpretability by explaining the contribution of each input feature to model outputs. Helps identify key drivers behind crop recommendations and yield predictions. LIME (Local Interpretable Model-Agnostic Explanations): Focuses on local fidelity, explaining individual predictions in a human-interpretable way, useful for farmers and stakeholders to understand "why" a specific crop or yield was recommended.

4.6 Deployment and User Interface

A simple frontend interface is built where users input: Soil data (N, P, K, pH) ,Environmental data (Temperature, Humidity, Rainfall) Backend processes this through trained models and returns recommended crop,Predicted Yield ,SHAP and LIME visualizations explaining model outputs.The system provides transparent, data-driven support to farmers for crop planning and yield forecasting.

V. RESULT

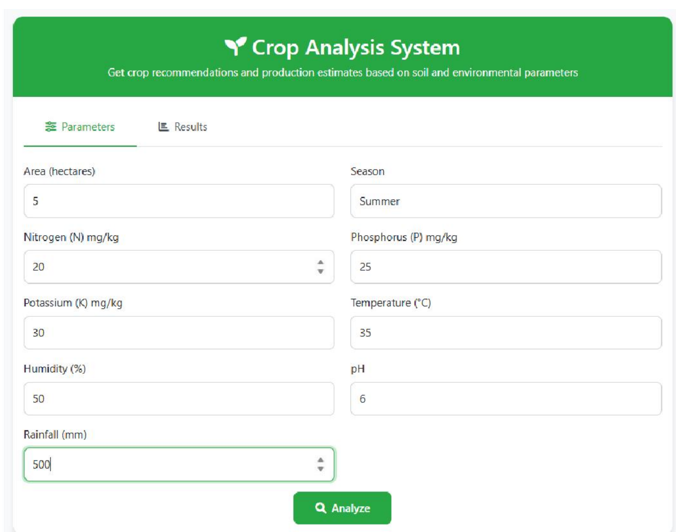

Fig 5.1 User providing input to the system

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 7, June 2025 Impact Factor: 7.67

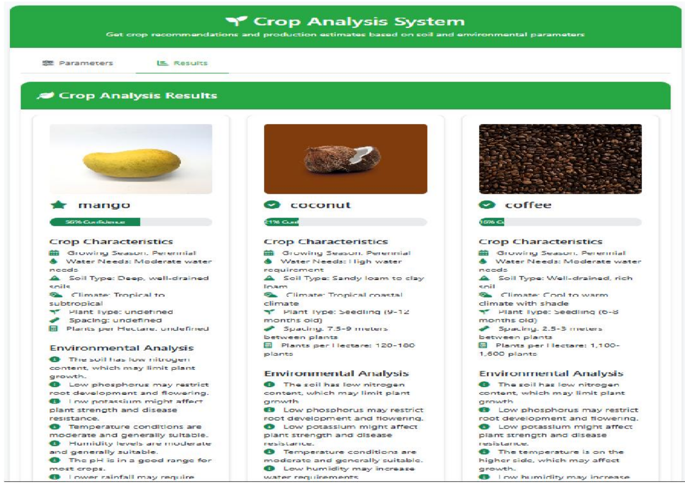

Fig 5.2 User Output

VI. CONCLUSION

Crop prediction using Machine Learning (ML) and Explainable AI (XAI) is transforming modern agriculture by enabling data-driven decision-making. Traditional farming methods rely heavily on experience and manual observations, which can be inaccurate and time-consuming. However, ML models can analyze vast amounts of data, including soil conditions, weather patterns, temperature, rainfall, and historical crop yields, to provide accurate predictions. These insights help farmers choose the right crops, optimize resource allocation, and improve overall agricultural productivity. Despite the effectiveness of ML models, their complexity often makes it difficult for farmers and stakeholders to trust their predictions. This is where Explainable AI (XAI) plays a crucial role. Techniques such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) provide transparency by explaining how different factors influence model decisions. By understanding why a model recommends a particular crop or yield estimate, farmers can make informed choices and take necessary actions to mitigate risks. Furthermore, integrating ML and XAI with real-time data from IoT sensors and satellite imagery enhances the accuracy and reliability of crop predictions. These technologies help monitor soil moisture, temperature, humidity, and other crucial parameters, allowing for dynamic adjustments based on changing environmental conditions. In conclusion, the combination of ML and XAI in crop prediction not only increases yield efficiency but also promotes sustainable farming practices. By making AI models interpretable and accessible, farmers and agricultural experts can leverage advanced technology to ensure food security and economic stability. With continuous advancements in AI and data analytics, future crop prediction systems will become even more precise, reliable, and widely adopted across the agricultural sector.

VII. ACKNOWLEDGMENT

We are also very thankful to MrsCh. Srivatsa Alivelu Mangatayi, Professor and, Department of Computer Science Engineering, ACE Engineering College, for her thoughtful guidance, advice, and valuable suggestions all through this project. We also appreciate our institution for the resources and support we received. Above all, we would like to extend our sincere appreciation to the editorial team of IJARSCT for allowing us to publish our work.

REFERENCES

[1] Sharma, S., Rai, S., & Krishnan, N. C. (2020). Wheat Crop Yield Prediction Using Deep LSTM Model. arXiv preprint arXiv:2011.01498.

[2] Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2020). Crop Yield Prediction Integrating Genotype and Weather Variables Using Deep Learning. bioRxiv.

[3] Fernandes, I., Vieira, C., Dias, K., & Fernandes, S. (2024). [Title of the Paper]. [Journal Name], [Volume(Issue)], [Page Numbers]. [DOI/Link if available].
[4] Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z., & Tao, F. (2020). Wheat Yield Predictions at a County and Field Scale with Deep Learning, Machine Learning, and Google Earth Engine. European Journal of Agronomy, 123, 126204.
[5] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis. Wiley.
[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[7] Molnar, C. (2020). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
[8] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
[9] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.
[10] Abdel-Ghaffar, E. A., & Daoudi, M. (2023). Personal authentication and cryptographic key generation based on electroencephalographic signals. Journal of King Saud University - Computer and Information Sciences, 35(5), Art. no. 101541.
[11].Ministry of Agriculture & Farmers Welfare, Government of India. (2024). Annual Crop Yield Reports. 12.Indian Meteorological Department. (2024). Weather and Crop Forecasting Reports.