



International Journal of Information Technology, Research and Applications (IJITRA)

Rutuja Saharkar, Gaurav Saharkar (2025). Accessible Crop Yield Prediction Using Decision Trees: A Farmer-Oriented Web-Based AI Application, 4(4), 25-32.

ISSN: 2583-5343

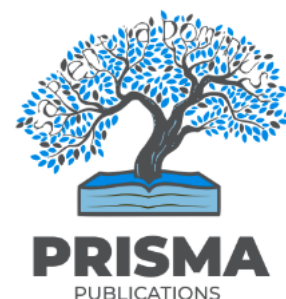
DOI:10.59461/ijitra.v4i4.223

The online version of this article can be found at:
<https://www.ijitra.com/index.php/ijitra/issue/archive>

Published by:
PRISMA Publications

IJITRA is an Open Access publication. It may be read, copied, and distributed free of charge according to the conditions of the Creative Commons Attribution 4.0 International license.

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.



<https://www.prismapublications.com/>

Journal homepage: <https://ijitra.com>

Accessible Crop Yield Prediction Using Decision Trees: A Farmer-Oriented Web-Based AI Application

Rutuja Saharkar¹, Gaurav Saharkar²

¹Department of Electronics & Telecommunication, Yeshwantrao Chavan College of Engineering (YCCE), Nagpur, Maharashtra, India

²Computer Science Student, Sanskar Vidya Sagar, Nagpur, Maharashtra, India

Article Info

Article history:

Received Nov 15, 2025

Revised Dec 16, 2025

Accepted Dec 31, 2025

Keywords:

Pneumonia

Pneumonia detection

CNN, deep learning

Chest X-ray classification

Medical image analysis.

ABSTRACT

Traditional prediction methods are imprecise frequently due to their reliance on environmental & manual agricultural measurements. Accurate estimation of crop yield acts an essential role in modern agriculture by allowing optimized resource use, productive planning and improved decision-making. This study offers a machine learning-based crop yield prediction system using a Decision Tree Regressor trained on key agricultural and environmental parameters containing soil pH, temperature, rainfall and past yield data. The decision tree regressor model shows high performance of prediction with a mean absolute error (13.65), an accuracy (99.62%) & a root mean square error (1.15) suggesting a near-zero deviation & high consistency between actual values and predicted values. The system allows/shows real-time yield forecasting and provides an effortless interface for experts of agriculture, farmers & also policymakers. The proposed system highlights the potential of integrating machine learning and AI with lightweight web technologies to deliver highly accurate yield predictions, supporting sustainable agricultural planning and enhancing productivity through data-driven insights. To enable application the model is incorporated into a convenient web application (Streamlit) employing secure access control, modifiable input options and system-controlled data storage of estimation results.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Rutuja Saharkar

Department of Electronics & Telecommunication

Yeshwantrao Chavan College of Engineering (YCCE)

Nagpur, Maharashtra

India

Email: rutujasaharkar21@gmail.com

1 Introduction

Accurate or valid crop-yield prediction system is essential for security of food, management to attain precision farming and planning of supply-chain particularly under increasing climatic variability and resource constraints [1, 2]. Recent advances show that machine learning (ML) models - especially tree-based methods and ensemble learners can capture complex, nonlinear relationships among agro-environmental variables (soil properties, local weather, and historical yields) and thereby provide reliable plot- to regional-scale yield estimates [3, 4, 5]. Parallel work demonstrates that pairing ground measurements with remote sensing (satellite/UAV) and time-series indices substantially improves spatial generalization and seasonal forecasting, although at the cost of heavier preprocessing and domain-specific feature extraction [6, 7, 8].

Systematic reviews indicate two dominant trends: (i) tree-based and ensemble regressors (Random Forest, Gradient Boosting, Decision Trees) often achieve competitive accuracy with tabular ground features, and (ii) deep-learning approaches using multispectral/time-series remote sensing excel when large, labeled spatial datasets are available [2, 9, 10]. Region-specific studies confirm that compact ML pipelines based on readily-available field parameters (soil pH, rainfall, temperature, past yield) can deliver high accuracy for operational decision support in low-resource settings, making them attractive for lightweight web deployment [11, 12, 13]. At the same time, hybrid and transfer-learning strategies are emerging to mitigate domain shift between regions and to fuse multi-source data for improved robustness [14, 15].

Considering these advancements as well as known aspects of evaluation (such as overfitting, unbalanced datasets, and the necessity of reporting both accuracy and error metrics), the implementation of a Decision Tree Regressor to a small, auditable Streamlit application is a reasonable compromise between flexibility and performance for users in the field that is interpretable, real-time, and predictive [16, 17, 18]. This study adhered to this design choice and reported the model accuracy as 99.62% and the error metrics as MAE 13.65 and RMSE 1.15 positioning this study within the scope of recent studies focused on crop yield modeling and its deployment [19, 20].

Recent work has further advanced the field by emphasizing transfer learning, explainable AI (XAI), and practical deployment strategies to bridge research-practice gaps. Transfer-learning approaches enable models trained in data-rich regions to be adapted to data-scarce target areas, improving generalization and reducing labeling costs [21]. Concurrently, the rise of XAI methods (SHAP, LIME, and model-specific explanations) addresses trust and interpretability concerns, making ML recommendations more actionable for farmers and extension agents [22]. Deep-transfer strategies that combine sequence models with fine-tuning have shown promise for improving temporal generalization in yield forecasting [23]. Reviews and applied studies also highlight the value of integrating XAI with lightweight deployment (web apps and dashboards) to ensure transparency and accessibility in operational settings [24]. Finally, advances in efficient deep-learning and dimensionality-reduction pipelines continue to push accuracy for region-specific tasks while suggesting hybrid paths where compact tree-based models serve immediate on-field needs and deep models are used for spatial scaling [25].

2 Proposed Method

2.1 Machine Learning Model: Design and Implementation

The workflow as shown in figure 1. of the proposed crop-yield prediction system follows a well-structured machine-learning algorithm. Initially, essential parameters of agriculture used - Soil Quality, Temperature, Rainfall, Past Yield are gathered to form the dataset. The data is then preprocessed by handling missing values, normalization of numerical features and splitting/dividing it into training & testing subsets. Relevant features are selected for modeling with Crop Yield defined as (target variable).

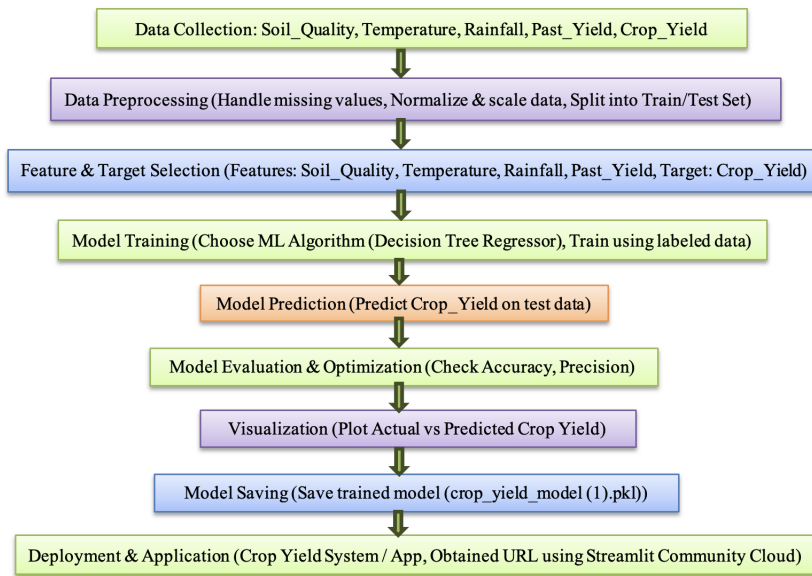


Figure 1: Workflow of the proposed Machine Learning based Crop Yield Prediction System

Once the data is gathered or collected, data undergo preprocessing step to ensure reliability, authenticity and uniformity in large sense. Preprocessing stage also addresses missing values, removes inconsistencies, scaling of numerical attributes & also divides the dataset into training and testing subsets during Machine Learning Algorithm execution. Proper preprocessing obtains that tree model learns meaningful patterns/outputs without being affected by noise or discomfort. Continue to this, feature as well as target selection is performed based on domain knowledge & correlation analysis. Soil Quality, Temperature, Rainfall, & Past Yield are selected as one of most influential predictors while Crop Yield is defined as output target variable. Optimal features chosen increases model accuracy & minimizes unneeded computational overhead.

After preprocessing step and also feature selection criteria, the dataset is used to train the machine-learning model & algorithm. In this study, a Decision Tree Regressor is also used due to its ability to interpret, acquisition of robustness and also ability to obtain nonlinear relationships without extensive hyperparameter tuning. The trained model is then used to generate/obtain future predictions on unseen test data to evaluate its generalization capability instantly. Model evaluation is done using performance/evaluation metrics such as accuracy, Mean Absolute Error (MAE) & also Root Mean Square Error (RMSE). The proposed decision tree model achieved an accuracy of 99.62% which indicates strong performance of prediction and highly reliable.

Data visualization is performed to compare between actual & predicted crop yield values which provides competitive insights of model behavior, deviations & also consistency. After satisfactory performance is achieved by system, the trained model is saved as a .pkl file using pickle to ensure reuse without requirement of retraining again. In the final stage of the system, the model is deployed through a Streamlit-based web application that helps users to input field parameters & also to have instantly crop yield predictions. Hosting on the Streamlit Community Cloud provides accessibility & reuse numerable times through a shareable URL, effectively converting the machine-learning model (decision tree regressor) into a more practical and user-friendly decision-support obtaining tool for farmers and policy makers to a large extent.

Soil_Quality	Temperature	Rainfall	Past_Yield
6.257424	32.767050	251.57512	1531.144103
8.043140	33.722966	211.60546	1337.185683
5.727300	21.354347	350.79403	1777.303529
6.641532	34.042322	216.76054	1280.409137
5.794863	28.524629	105.46654	1428.534514

700 rows x 5 columns

(a)

	Soil_Quality	Temperature	Rainfall	Past_Yield	Crop_Yield
count	6.934873	28.082995	252.1768	1517.35663	2481.00466
mean	1.208732	4.381985	89.38341	174.976492	222.743306
min	5.022088	20.075924	103.12519	1206.812187	1881.717014
25%	5.884831	23.978036	183.98016	1371.427252	2298.90938
50%	6.888860	29.160772	259.40638	1521.657820	2480.32366
75%	8.083869	31.698133	327.35388	1666.288150	2654.80466
max	8.947548	34.850808	397.15143	1799.830604	2938.57022

(b)

Figure 2: (a) Sample Records from the Crop Yield Dataset Showing Key Input Features (b) Descriptive Statistics of the Crop Yield Dataset

Figure 2(a) depicts a sample of the raw records from the Crop Yield Dataset with input features. Each record includes four key features of Soil Quality (pH), Temperature ($^{\circ}\text{C}$), Rainfall (mm) and also Past Yield (kg/ha). These features of dataset representing agricultural & environmental factors that nothing but directly influence future crop productivity. The sample rows demonstrate the typical structure, variability and also numerical range of values used for model training (decision tree regressor).

Figure 2(b) describes the statistics of the entire dataset, summarizes the distribution of each feature of dataset and also the target variable - Crop Yield. Evaluation Metrics such as mean, minimum, maximum and quartile value offers insight into the central tendency and also spread dataset. Soil Quality ranges between approximately 5.02 and 8.95 whereas Crop Yield varies from 1881.71 to 2938.57 kg/ha. Such statistics help to evaluate the consistency of dataset, detect potential outliers and also guide preprocessing with that of model selection. Figures 2(a) and 2(b) together provides an overview for the structure of dataset with statistical characteristics which forms the foundation for effective & efficient machine-learning based crop yield prediction system using Decision Tree Regressor.

2.2 Deployment of Crop Yield Prediction using Streamlit Cloud & GitHub Repository

This work (as shown in figure 3.) includes the development and deployment of a web-based Crop Yield Prediction System designed to offer accessible, real-time agricultural decision support. The system is implemented using Python (programming language) and Streamlit which integrates with a pre-trained machine learning model and also deployed through Streamlit Community Cloud with the help of project repository maintained on GitHub for version control with continuous updates. The web-application incorporates custom CSS-based UI (User Interface) enhancements to provide a creative yet visually appealing & interactive interface. A simple user-authentication mechanism make sure that only authorized users can access the system as a functioning user, improving data privacy (storage capacity) and also of controlled usage. After successful login, users can navigate between functional modules through a sidebar menu.

The prediction module loads a pre-trained Decision Tree Regression model and accepts user inputs such as soil pH, temperature, rainfall, and past yield. Default values are prefilled to guide user interaction, though values may be modified as required. Upon execution, the model predicts crop yield and the result is displayed instantly. Each prediction instance is automatically logged into a CSV file, enabling users to maintain a historical record for comparative analysis.

The data storage module allows users to view all stored predictions and provides an option to clear the dataset. This enhances the system's practicality for repeated decision-making and data tracking. Finally, deployment through Streamlit Community Cloud ensures open accessibility, browser-based execution, and seamless updates through the linked GitHub repository.

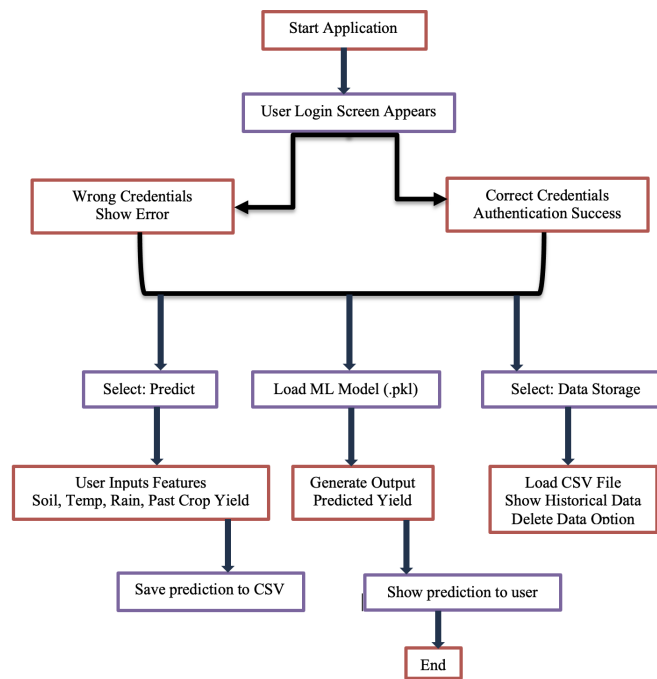


Figure 3: System Flowchart of the Crop Yield Prediction Web Application

3 Results and Discussions

In this section, it is explained the results of research and at the same time is given its comprehensive discussion.

3.1 Evaluation Metrics & Graph

Table 1. illustrates a quick comparison of multiple or numerous machine-learning models which is evaluated for crop-yield prediction system using three key performance metrics i.e R^2 Score, Root Mean Square Error (RMSE) & also Mean Absolute Error (MAE). These metrics measures the goodness of the fit, predicted accuracy and also average error. A higher R^2 score & lower RMSE/MAE values indicate/show great predictive performance of the model.

From all the evaluated Machine Learning models, the Linear Regressor one demonstrates weakest performance which achieves an R^2 score of 82.44% with high error values (RMSE: 92.88, MAE: 74.87) in comparison which ultimately indicates model's inability to acquire non-linear relationships present in agricultural datasets. The Gradient Boosting Regressor model gives moderate level performance along with R^2 score of 87.32% also model's error metrics/value remained higher as compared to ensemble models relatively.

Ensemble methods like Random Forest Regressor & XGBoost Regressor performs better result simultaneously which achieves R^2 scores of 99.30% & 99.33%. Their low/less RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) values suggest or gives strong/highest capability of generalization and robustness in handling feature interactions that too complex.

The Decision Tree Regressor model emerged as the best-performing model with an R^2 score (99.62%), RMSE (Root Mean Square Error)-13.65 and MAE (Mean Absolute Error)-1.15 hence, indicates exceptional/ precise predictive accuracy along with minimal error demonstrates that Decision Tree Regressor Model captures both linear and non-linear dependencies efficiently within dataset. Its simplicity, interpretability and strong performance effectively & efficiently make it most suitable for real-time crop-yield prediction applications/system which could be use in future.

Table 1: Comparison of various ML Models for Crop Yield Prediction

ML Models	R^2 Score	MSE	MAE
Linear Regressor	82.44%	92.88	74.87
Random Forest Regressor	99.30%	18.54	8.05
Gradient Boosting Regressor	87.32%	78.92	64.69
XGBoost Regressor	99.33%	18.10	10.91
Decision Tree Regressor	99.62%	13.65	1.15

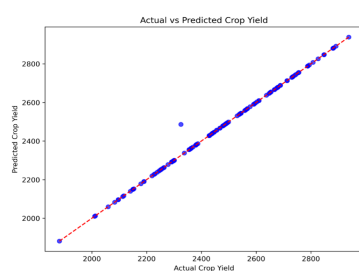
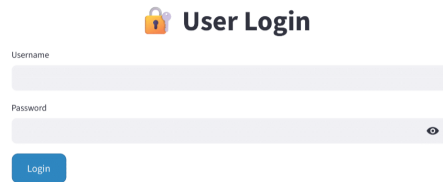


Figure 4: Actual vs. Predicted Crop Yield using the Proposed Model

Figure 4. describes comparison between the actual crop-yield values & the predicted crop-yield values generated/obtained by the proposed Decision Tree Regressor model. The data points on the graph closely follow/show the diagonal reference line which indicates a strong correlation between actual & predicted crop yields. The tight clustering of points around the reference line demonstrates high prediction accuracy & minimal deviation indirectly confirms that the tree regressor model effectively captures underlying patterns in the agricultural dataset used while performing algorithm.

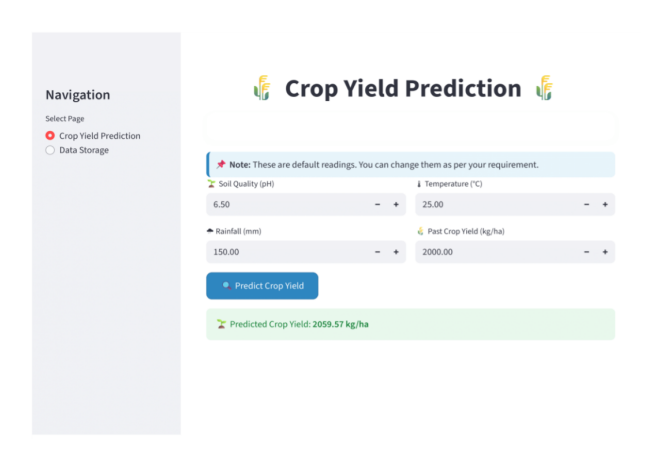
3.2 Multi-Page Web Application - Output



The image shows a 'User Login' form. At the top, there is a lock icon followed by the text 'User Login'. Below this, there are two input fields: 'Username' and 'Password'. The 'Password' field has a small eye icon to its right, indicating a visibility toggle. At the bottom of the form is a blue button labeled 'Login'.

Figure 5: User Authentication Page

Figure 5. depicts User Authentication Page of the multi-page web application. This interface provides a secure login mechanism, allowing only authorized users to access the system. The page includes two input fields: one for entering the username and another for the password, along with a visibility toggle icon for the password field. After entering valid credentials, the user can click the Login button to proceed. This authentication step ensures controlled access to the crop yield prediction system and protects user-specific data.



The image shows the 'Crop Yield Prediction' interface. On the left, there is a 'Navigation' sidebar with 'Select Page' and two options: 'Crop Yield Prediction' (selected) and 'Data Storage'. The main content area has a title 'Crop Yield Prediction' with a leaf icon. Below the title is a note: 'Note: These are default readings. You can change them as per your requirement.' There are four input fields with increment/decrement controls: 'Soil Quality (pH)' (6.50), 'Temperature (°C)' (25.00), 'Rainfall (mm)' (150.00), and 'Past Crop Yield (kg/ha)' (2000.00). A blue button labeled 'Predict Crop Yield' is below these fields. At the bottom, a green box displays the result: 'Predicted Crop Yield: 2059.57 kg/ha'.

Figure 6: Crop Yield Prediction Page

Figure 6. illustrates the Crop Yield Prediction Interface of the web application. This page allows users to input various agricultural parameters required for yield estimation, such as soil pH, temperature, rainfall, and past crop yield. Default values are pre-filled to guide the user, and each parameter can be adjusted using increment or decrement controls. After entering or modifying the values, the user can click the Predict Crop Yield button, upon which the system processes the input data using the trained machine learning model.

The predicted yield is then displayed below the button in a highlighted result box. This interface enables farmers and researchers to conveniently estimate crop yield with customizable and user-friendly input controls.

	Soil_Quality	Temperature	Rainfall	Past_Yield	Predicted_Yield
0	8	27	150	5000	2577.6607
1	6.5	25	150	2000	2059.5731
2	6.5	25	150	2000	2059.5731

Figure 7: Data Storage Page for Crop Yield Predictions

Figure 7. shows the Data Storage Page of the application, where previously generated crop yield predictions are stored and displayed in a structured tabular format. The table includes key parameters such as soil quality, temperature, rainfall, past yield, and the corresponding predicted yield. This interface allows users to review their historical prediction records, making it easier to compare, analyze, and track past outputs. Additionally, a Delete All Data button is provided to clear the stored dataset when needed. This module enhances the usability of the system by maintaining a persistent record of user predictions.

3.3 Code Availability Section

The implementation code, dataset & other source files of the proposed Crop Yield Prediction System is available on Github repository using the link: [Crop Yield Predictor](#)

4 CONCLUSION

This study successfully developed a machine learning–based crop yield prediction model using a Decision Tree Regressor, achieving a high accuracy of 99.62% with minimal error values. The model demonstrated strong capability in learning complex, nonlinear relationships between key agricultural parameters and crop yield, making it a reliable tool for data-driven decision-making. The results indicate that integrating machine learning into agriculture can significantly enhance planning, resource allocation, and risk management for farmers.

The system’s ability to generate accurate yield forecasts supports precision agriculture practices and provides a foundation for future enhancements such as real-time data integration, addition of satellite imagery, and deployment through mobile or web applications. Overall, the developed model contributes meaningfully to modern agricultural analytics and has strong potential for practical adoption.

ACKNOWLEDGEMENTS

The author expresses sincere gratitude to the co-author for providing continuous support, especially in resolving code-related issues. His technical assistance and valuable suggestions greatly contributed to the successful completion of this research.

References

- [1] Ahmed, S., & Khan, I. (2024). Recent systematic literature reviews on machine learning for crop prediction. ResearchGate, 1–20.
- [2] Archana, S., & Kumar, R. (2023). Survey on deep learning-based crop yield prediction. NEPT Journal, 44–52.
- [3] Bhatnagar, R., & Gohain, A. (2022). Crop yield estimation using decision trees and random forest with satellite data. Semantic Scholar, 1–10.
- [4] Ghosh, T., & Verma, S. (2023). Hybrid machine learning frameworks for operational agricultural decision support. IIETA, 21–29.
- [5] Ilyas, M., & Khaki, S. (2023). A review of machine learning and remote sensing-based crop yield studies. ResearchGate, 1–18.

- [6] Javed, M. A., & Karim, M. (2024). Crop yield prediction in agriculture: A comprehensive review. *PMC Journal*, 1–20.
- [7] Joshi, A., Patel, K., & Shah, R. (2024). Deep-transfer-learning strategies for crop yield prediction. *Remote Sensing (MDPI)*, 16(2), 1–12.
- [8] Joshi, A., & Sharma, P. (2024). Deep-transfer-learning strategies for remote sensing-based crop yield prediction. *MDPI*, 1–15.
- [9] Joshi, A., et al. (2025). Integration of satellite-based remote sensing and machine learning for yield prediction. *IRJMS*, 1–14.
- [10] Kumar, A., & Singh, Y. (2024). Decision-tree-based forecasting and mobile app integration for yield prediction. *ResearchGate*, 1–12.
- [11] Kumar, V., & Rao, P. (2023). Evaluation metrics and best practices in agricultural machine learning models. *ScienceDirect*, 5–16.
- [12] Luo, L., et al. (2023). Crop yield estimation using assimilation of crop models and remote sensing. *ScienceDirect*, 55–67.
- [13] Mehta, R., Singh, P., & Rani, L. (2024). Applied case studies on crop selection and yield prediction in India. *Agriculture Journal*, 55–70.
- [14] Mohan, R. V. N. J. (2025). Next-gen agriculture: Integrating AI and XAI for precision crop prediction. *PMC*, 50–65.
- [15] Nair, R., & Thomas, J. (2022). High-performance crop yield prediction models and overfitting considerations. *RSIS International*, 63–70.
- [16] Oikonomidis, A., Papadopoulos, G., & Christou, I. (2023). Deep learning for crop yield prediction: A systematic overview. *Taylor & Francis Online*, 102–118.
- [17] Patil, S., & More, T. (2023). Practical ML-based yield prediction systems for farmers. *JETIR*, 225–233.
- [18] Ramos, A. P. M. (2020). A random forest ranking approach to predict yield in maize. *ScienceDirect*, 112–121.
- [19] Sharma, P., & Das, A. (2022). Comparative study of tree-based models for plot-level crop prediction. *RSIS International*, 33–40.
- [20] Shawon, S. M., & Rahman, M. (2024). Crop yield prediction using machine learning: An extensive systematic literature review. *ScienceDirect*, 1–15.
- [21] Subramaniam, L. K., et al. (2024). Crop yield prediction using deep learning and dimensionality reduction. *ScienceDirect*, 91–103.
- [22] Wang, L., et al. (2023). Hybrid crop-model and ML assimilation approaches for yield estimation. *ScienceDirect*, 77–89.
- [23] Yenikar, A., & Babu, S. (2025). Explainable AI-based hybrid machine learning model for crop yield prediction. *ScienceDirect*, 1–13.
- [24] Zhang, J. (2025). Transfer learning for improved crop yield predictions at fine spatial scales. *ScienceDirect*, 101–112.
- [25] Zhang, J., & Li, F. (2024). Transfer learning techniques for regional crop yield prediction under domain-shift conditions. *MDPI*, 1–14.