iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76520

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

A Review on Crop Yield Prediction and Disease Detection

Akshara Vinod[1], Fahad N N[2], Megha K P[3], Saniga Sudheer[4], Binu Balakrishnan[5]
[1, 2, 3, 4]Department of Computer Science and Engineering Universal Engineering College, Thrissur, Kerala, India
[5]Assistant Professor, Department of Computer Science and Engineering Universal Engineering College, Thrissur, Kerala, India

***Abstract:** Crop yield prediction and plant disease detection are critical challenges in modern agriculture, directly impacting productivity and food security. As part of a project-oriented study, this paper presents a comprehensive review of machine learning and deep learning techniques applied to these agricultural problems. The review analyzes existing approaches such as convolutional neural networks, ensemble learning models, time-series forecasting techniques, and IoT-enabled systems reported in recent literature. Performance metrics, input features, and application contexts of these methods are compared to identify strengths, limitations, and research gaps. The insights obtained from this review serve as a foundation for the design and development of an intelligent agricultural decision-support system in future project implementation.*

***Index Terms:** Crop Yield Prediction, Plant Disease Detection, Machine Learning, Deep Learning, CNN, Smart Agriculture.*

I. INTRODUCTION

Agriculture remains one of the most vital sectors for human survival and economic stability. However, it is inherently vulnerable to various risks, including crop diseases, pest attacks, climate variability, and soil degradation. Diseases caused by fungi, bacteria, and viruses can spread rapidly, affecting large areas and resulting in significant crop loss. For example, late blight in potatoes and wheat rust are known to devastate crops if not detected early [9]. Traditional methods for disease detection, such as visual inspection by farmers or expert consultation, are time-intensive, subjective, and often impractical for large-scale operations. Similarly, yield prediction is often based on historical trends or simple regression models, which fail to account for dynamic environmental changes or disease impacts, leading to inaccurate forecasts.

The emergence of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) has revolutionized the agricultural domain, enabling automation, precision, and scalability [3]. Deep learning models, particularly Convolutional Neural Networks (CNNs), excel at recognizing patterns in images, making them ideal for detecting crop diseases based on leaf texture, color variations, and lesion shapes [13]. YOLO (You Only Look Once), a state-of-the-art object detection framework, allows simultaneous localization and classification of multiple diseases within a single image, making real-time monitoring feasible.

The yield prediction module utilizes advanced algorithms like XGBoost, which performs gradient boosting on decision trees to handle non-linear relationships among multiple environmental and crop-specific features [2]. When temporal or sequential data is available such as weekly growth measurements or weather data Long Short Term Memory (LSTM) networks are employed to capture temporal dependencies, providing accurate forecasts of expected crop yield [7]. By integrating both disease detection and yield prediction into a single platform, the system enables farmers to plan interventions such as pesticide application, irrigation, and fertilization, and make informed decisions about harvest timing. This integrated approach supports precision agriculture, reduces crop loss, enhances productivity, and contributes to sustainable farming practices. This review paper is prepared as a preliminary academic study for an engineering project. It systematically analyzes existing research to identify suitable methodologies that can be adopted during the system implementation phase. No new system development or experimental evaluation is presented in this work.

II. LITERATURE SURVEY

The advent of precision agriculture is fundamentally transforming traditional farming practices by leveraging advanced computational technologies to enhance sustainability and productivity. This literature survey synthesizes recent research on the application of Machine Learning (ML) and Deep Learning (DL) models, Internet of Things (IoT) sensor networks, and remote sensing data to address critical agricultural challenges. The reviewed studies demonstrate significant progress in key areas including crop yield prediction, plant disease detection, pest management, and crop classification. By analyzing a range of approaches—from ensemble methods and hybrid deep learning architectures to multisensor data fusion—this review highlights the superior predictive accuracies often achieved by models such as Random Forest and XGBoost, while also acknowledging prevalent

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

limitations such as dependency on large datasets and operational complexity. Collectively, these works underscore the potential of data-driven intelligence in enabling informed decision-making for farmers and stakeholders, thereby supporting efficient, resilient, and sustainable agricultural systems.

A systematic review focusing on the application of Artificial Intelligence for crop yield prediction and optimization was conducted by Screpnik et al. [1]. This 2025 study synthesized recent literature to identify key trends, effective models, and prevailing challenges. The review reported that ensemble methods, particularly Random Forest and Gradient Boosting, consistently achieve strong performance in yield prediction tasks. It also emphasized the importance of data fusion, integrating satellite imagery, climatic variables, and soil information to enhance prediction robustness. A notable limitation identified was the gap between highly accurate predictive models and their practical on-farm deployment, highlighting the need for decision-support systems that translate predictions into actionable recommendations.

A hybrid methodology combining traditional machine learning algorithms with deep learning architectures for agricultural yield prediction was explored by Sharma et al. [2]. Their study evaluated Decision Tree, Random Forest, XGBoost, CNN, and LSTM models. The Random Forest model achieved an accuracy of 98.96%, with an MAE of 1.97 and an RMSE of 2.45, while the CNN model demonstrated the lowest loss value. However, the authors noted that the effectiveness of such hybrid approaches is strongly dependent on the availability of large and diverse training datasets, which may limit generalizability across regions.

Focusing on crop-specific prediction, Ashfaq et al. [5] proposed a wheat yield prediction model based on the fusion of climate data and Normalized Difference Vegetation Index (NDVI). Their ensemble approach combining Random Forest, XGBoost, and CatBoost achieved a low RMSE of 2.2982 kg/ha and an R2 of 0.9752. Similarly, Kumar et al. [7] applied ensemble learning techniques to forecast sugarcane yields using meteorological data, where the Random Forest regression model achieved an R2 of 0.93 and an RMSE of 4.02 t/ha.

Badshah et al. [4] developed robust machine learning models for crop classification and yield prediction, reporting a Random Forest classification accuracy of 99.53% and a CatBoost regression R2 of 0.9973. In a regional case study, Haider et al. [9] proposed an ensemble framework for cotton yield prediction in Pakistan, where a Voting Regressor combining Random Forest, XGBoost, and LightGBM achieved an R2 of 0.99.

The integration of IoT technologies for pest and yield management has also gained attention. Saleem et al. [13] developed an IoT-enabled system for weekly crop pest prediction using a Deep Neural Network, achieving a classification accuracy of 98.26%. Complementing this, Hoque et al. [11] incorporated meteorological and pesticide usage data into a Random Forest model for yield forecasting, achieving an R2 value of 0.997.

In the domain of plant disease detection, Rani et al. [3] provided a comprehensive analysis of AI-based approaches. Khalid and Talukder [10] proposed a hybrid deep multistacking model achieving 99.89% accuracy on the PlantVillage dataset. Mamun et al. [6] further reviewed self-supervised learning methods, highlighting their potential to reduce dependence on large annotated datasets.

Addressing food security at scale, Shafi et al. [8] applied remote sensing and machine learning for large-scale yield prediction, achieving an R2 of 0.87 for wheat. Mai et al. [12] demonstrated that high-resolution soil moisture data improves yield estimation accuracy. Finally, Najjar et al. [14] emphasized explainable machine learning using SHAP analysis, while Iniyan and Jebakumar [15] developed a smart mobile application for crop yield forecasting, improving accessibility for farmers.

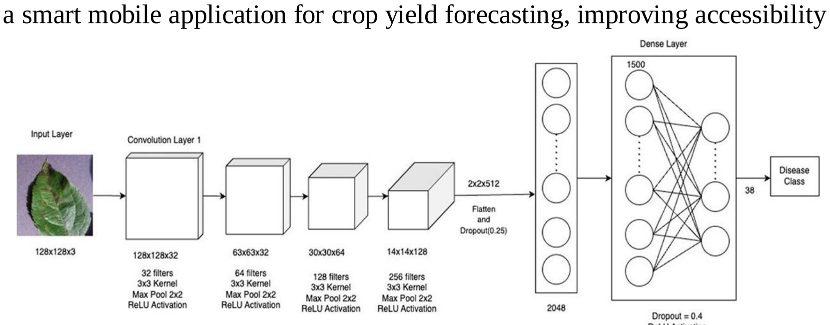

Figure 1. CNN-based plant disease classification architecture (adapted from [17]).

Fig. 1 illustrates a representative CNN-based architecture for plant disease classification reported in the literature [17]. An IoT-enabled system employing a multisensor data fusion approach for crop monitoring and recommendation was proposed by Reyana et al. in 2023 [16]. Reddy et al. [17] developed a deep learning–based mobile application for automated plant disease detection and severity estimation.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Their system integrates a lightweight Convolutional Neural Network (CNN), implemented in PyTorch, to classify leaves as healthy or diseased, achieving 92.06% accuracy on the PlantVillage dataset. To complement classification, a classical image processing pipeline using OpenCV and NumPy was employed to calculate infection severity by measuring the ratio of diseased to total leaf area. The complete framework was deployed as a cross-platform mobile application, built with React Native and a Flask backend, enabling real-time field diagnostics. This approach effectively bridges laboratory-level deep learning research and practical agricultural applications, providing farmers with an accessible tool for early disease detection and crop health monitoring [17]

III. COMPARISON BETWEEN MODELS

Table I
COMPARISON BETWEEN AI/ML MODELS FOR CROP YIELD AND DISEASE DETECTION

Ref.	Model	Key Features	Accuracy
[4]	RF, XGBoost, SVM	Soil, NDVI, rainfall	96%
[5]	ML + NDVI Fusion	Satellite, weather data	94–97%
[9]	Ensemble ML	Climatic parameters	92–95%
[10]	Hybrid CNN–LSTM	Images, spectral data	96–98%
[13]	IoT + DNN	Sensor data	94.5%
[17]	CNN + OpenCV	Disease severity	92.06%

As presented in Table I, the comparative evaluation of various artificial intelligence and machine learning models highlights significant progress in data-driven approaches for agricultural yield and pest prediction. Traditional regression-based models have gradually been replaced by ensemble and hybrid architectures capable of handling heterogeneous datasets and nonlinear dependencies. Among the analyzed studies, robust machine learning techniques such as Random Forest (RF), XGBoost, and Support Vector Machines (SVM) achieved up to 96% accuracy, demonstrating their ability to generalize effectively across diverse environmental parameters. Similarly, models that integrate normalized difference vegetation index (NDVI) data with meteorological features exhibit superior performance, with reported R^2 values ranging from 0.94 to 0.97, confirming the importance of spectral indices in assessing crop health and productivity.

In addition, a lightweight convolutional neural network (CNN) model integrated with an OpenCV-based image processing module has been developed to detect and quantify plant leaf diseases. The mobile based system, implemented using React Native and Flask, achieved an accuracy of 92.06% on the PlantVillage dataset, effectively bridging the gap between laboratory research and field deployment. The inclusion of a disease severity estimation component based on the ratio of infected to total leaf area provides farmers with actionable insights for timely intervention.

Furthermore, the hybrid CNN–LSTM architecture attained the highest accuracy (96–98%), owing to its capacity to extract deep spatial features from crop imagery and learn temporal correlations related to growth patterns and disease progression. The ensemble learning model proposed for cotton yield prediction achieved an accuracy of approximately 95%, indicating that model fusion techniques can mitigate individual classifier bias and improve robustness under varying climatic conditions. In addition, IoT-enabled deep neural network frameworks proved highly effective in pest prediction by utilizing real-time environmental inputs such as temperature, humidity, and light intensity, thereby providing a scalable and field-deployable solution. Explainable machine learning approaches also emerged as an important trend, offering transparency in model decision-making and enabling precise interpretation of spatially distributed agricultural data.

The analysis in Table I clearly demonstrates that the fusion of remote sensing data, IoT-based environmental monitoring, and hybrid deep learning frameworks yields substantial improvements in prediction accuracy and system adaptability. These findings indicate that future research should emphasize the integration of multi source data streams, advanced temporal modeling, and explainability driven algorithms to ensure reliable, interpretable, and scalable solutions for sustainable precision agriculture.

IV. CONCLUSION

The reviewed literature collectively illustrates the transformative potential of artificial intelligence and machine learning in modern agriculture. Across multiple studies, models incorporating deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent architectures (LSTM), and ensemble learning techniques have consistently demonstrated high predictive accuracy in crop yield estimation and pest detection. Integrating multisource data, such as climatic parameters, soil characteristics, NDVI indices, and remote sensing imagery, significantly enhances model robustness and adaptability to varying environmental conditions.

The works analyzed, including those employing IoT-enabled deep learning systems and multisensor fusion techniques, reveal that real-time environmental monitoring can effectively predict pest infestations and crop productivity with accuracies often exceeding 94–97%. The hybrid CNN–LSTM architecture particularly excels in capturing spatial–temporal dependencies, while ensemble-based frameworks demonstrate strong generalization across different crops and regions. Moreover, the adoption of explainable AI (XAI) methods ensures interpretability and transparency, making these systems more acceptable for real-world deployment among agronomists and farmers.

Despite these advancements, several challenges persist. Model scalability, data heterogeneity, limited labeled datasets, and the dynamic nature of agricultural ecosystems remain major research bottlenecks. Additionally, achieving global generalization requires integrating data from diverse agro-climatic zones and improving model interpretability for non-technical users. The findings of this review will guide the selection of algorithms and system architecture for the subsequent project implementation. Future work will focus on developing and evaluating a prototype system based on the insights identified in this study.

REFERENCES

[1] C. R. Screpnik, E. Zamudio and L. I. Gimenez, "Artificial Intelligence in Agriculture: A Systematic Review of Crop Yield Prediction and Optimization," in IEEE Access, vol. 13, pp. 70691-70697, 2025, doi: 10.1109/ACCESS.2025.3560631.

[2] P. Sharma, P. Dadheech, N. Aneja and S. Aneja, "Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning," in IEEE Access, vol. 11, pp. 111255-111264, 2023, doi: 10.1109/ACCESS.2023.3321861.

[3] R. Rani, J. Sahoo, S. Bellamkonda, S. Kumar and S. K. Pippal, "Role of Artificial Intelligence in Agriculture: An Analysis and Advancements With Focus on Plant Diseases," in IEEE Access, vol. 11, pp. 137999-138019, 2023, doi: 10.1109/ACCESS.2023.3339375.

[4] A. Badshah, B. Yousef Alkazemi, F. Din, K. Z. Zamli and M. Haris, "Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability," in IEEE Access, vol. 12, pp. 162799-162813, 2024, doi: 10.1109/ACCESS.2024.3486653.

[5] M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani, "Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion," in IEEE Access, vol. 12, pp. 40947-40961, 2024, doi: 10.1109/ACCESS.2024.3376735.

[6] A. A. Mamun, D. Ahmedt-Aristizabal, M. Zhang, M. Ismail Hossen, Z. Hayder and M. Awrangjeb, "Plant Disease Detection Using Self-Supervised Learning: A Systematic Review," in IEEE Access, vol. 12, pp. 171926-171943, 2024, doi: 10.1109/ACCESS.2024.3475819.

[7] S. Kumar, M. Pant and A. Nagar, "Forecasting the Sugarcane Yields Based on Meteorological Data Through Ensemble Learning," in IEEE Access, vol. 12, pp. 176539-176553, 2024, doi: 10.1109/ACCESS.2024.3502547.

[8] U. Shafi et al., "Tackling Food Insecurity Using Remote Sensing and Machine Learning-Based Crop Yield Prediction," in IEEE Access, vol. 11, pp. 108640-108657, 2023, doi: 10.1109/ACCESS.2023.3321020.

[9] S. T. Haider et al., "An Ensemble Machine Learning Framework for Cotton Crop Yield Prediction Using Weather Parameters: A Case Study of Pakistan," in IEEE Access, vol. 12, pp. 124045-124061, 2024, doi: 10.1109/ACCESS.2024.3454511.

[10] M. Khalid and M. A. Talukder, "A Hybrid Deep Multistacking Integrated Model for Plant Disease Detection," in IEEE Access, vol. 13, pp. 116037-116053, 2025, doi: 10.1109/ACCESS.2025.3583793.

[11] M. J. Hoque et al., "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning," in IEEE Access, vol. 12, pp. 47768-47786, 2024, doi: 10.1109/ACCESS.2024.3383309.

[12] R. Mai, Q. Xin, J. Qiu, Q. Wang and P. Zhu, "High Spatial Resolution Soil Moisture Improves Crop Yield Estimation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 19067-19077, 2024, doi: 10.1109/JSTARS.2024.3417424.

[13] R. M. Saleem et al., "Internet of Things Based Weekly Crop Pest Prediction by Using Deep Neural Network," in IEEE Access, vol. 11, pp. 85900-85913, 2023, doi: 10.1109/ACCESS.2023.3301504.

[14] H. Najjar, M. Miranda, M. Nuske, R. Roscher and A. Dengel, "Explainability of Subfield Level Crop Yield Prediction Using Remote Sensing," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 4141-4161, 2025, doi: 10.1109/JSTARS.2025.3528068.

[15] S. Iniyan and R. Jebakumar, "Phenotype Based Smart Mobile Application for Crop Yield Prediction and Forecasting Using Machine Learning and Time Series Models," in Journal of Mobile Multimedia, vol. 18, no. 3, pp. 603-634, May 2022, doi: 10.13052/jmm1550-4646.1837.

[16] A. Reyana, S. Kautish, P. M. S. Karthik, I. A. Al-Baltah, M. B. Jasser and A. W. Mohamed, "Accelerating Crop Yield: Multisensor Data Fusion and Machine Learning for Agriculture Text Classification," in IEEE Access, vol. 11, pp. 20795-20805, 2023, doi: 10.1109/ACCESS.2023.3249205.

[17] B. Ramana Reddy, G. Kalnoor, M. Devashish and P. Sai Karthik Reddy, "Deep Learning Based Mobile Application for Automated Plant Disease Detection," in IEEE Access, vol. 13, pp. -107925, 2025, doi: 10.1109/ACCESS.2025.3581099.