A Study on Crops Disease Detection Using Machine Learning

NAME: S.M. RASEL

ID: 22-48039-2

Section: A

1) Statement of Purpose (SOP) / Motivation:

Agriculture is the backbone of global food security and economic stability, yet ensuring sufficient production remains a critical challenge as the world's population grows and climate change intensifies. Agricultural productivity is still at risk from crop diseases, which can lead to significant production losses and financial hardships for farmers. Traditional disease detection methods are time-consuming, labor-intensive, and prone to human error since they rely on specialized expertise and physical inspection. This leads to crop loss being worsened and solutions being postponed. The integration of artificial intelligence (AI), specifically machine learning (ML), into agricultural operations is being driven by the pressing need for effective, precise, and scalable solutions for early crop disease diagnosis.

However, finding agricultural diseases may be altered by the nexus of artificial intelligence (AI) and agriculture, particularly through machine learning (ML) and deep learning (DL). By analyzing vast quantities of plant photos, sensor data, and ambient factors, machine learning algorithms can swiftly and precisely identify disease signs. In the detection and localization of plant diseases from leaf pictures, convolutional neural networks (CNNs), one kind of deep learning model, have proven to be faster and more accurate than traditional methods. Illnesses may be followed in real time in the field by combining AI with Internet of Things (IoT) sensors, drones, and smartphone apps, giving farmers useful information.

The main reason for this research is that advanced disease detection is very important for making agriculture more sustainable and making sure everyone has enough food. Crop diseases cause big losses in yields, which are thought to be between (20-40) % percent worldwide. This puts both food supply and economic stability at risk. Using machine learning, we can create automated, scalable, and cost-effective systems that not only find diseases early but also cut down on the need for too many chemical treatments, which is good for the environment. This is a great time to connect cutting-edge technology with real-world farming uses because AI is moving so quickly and there are more agricultural datasets and computing resources available.

Furthermore, the pressure to investigate this location comes from a strong preference to use generation for the coolest of society. It is both intellectually and socially exciting to keep in mind the possibility of helping to develop answers so that it will reduce food lack of confidence, help farmers, and strengthen sustainable practices. Researching the nexus among AI and agriculture is consistent with a destiny vision wherein traditional know-how and era collaborate to resolve a number of the most crucial problems dealing with mankind. The aim of this research is to enhance crop disease detection whilst encouraging innovation which can have sensible applications in agriculture and different fields.

In summary, a device mastering study on crop disease detection is an ambitious and socially enormous challenge. It goals to enhance ailment management, encourage sustainable farming, and contribute to global food protection by means of fusing latest AI era with beneficial agricultural packages.

2) Research Proposal

i. Introduction

Crop diseases cause large output losses and financial hardship for farmers, posing a danger to the world's food security. For large-scale or real-time monitoring, traditional disease detection techniques based on manual inspection are insufficient since they are frequently subjective, slow, and need specialized knowledge. With automated, scalable, and very accurate methods, crop disease detection has been transformed by the incorporation of machine learning (ML) and deep learning (DL) into agricultural image processing. This review of the literature highlights important approaches, difficulties, and potential paths forward as it critically analyzes current developments (2016–2025) in machine learning and deep learning applications for smart farming, agricultural image processing, and crop disease prediction.

ii. Literature Review

ML algorithms, along with support vector machines (SVM), random forests, and ensemble methods, were widely carried out for disorder category and prediction. While ML fashions require manual function extraction, they stay valuable for established data and eventualities with limited image availability (Balafas et al., 2023; Shahi et al., 2023). The integration of ML with Internet of Things (IoT) devices and unmanned aerial motors (UAVs) permits actual-time, subject-scale disorder monitoring, helping precision agriculture and aid optimization (Shahi et al., 2023). UAV-primarily based far flung sensing, blended with ML and DL, has emerged as a effective tool for large-scale crop disease detection. UAVs ready with superior sensors seize high-resolution imagery, which, whilst processed by DL fashions, permits for early identification of disease signs across massive fields (Shahi et al., 2023). Studies highlight the significance of sensor choice (RGB, multispectral, hyperspectral) and the want for robust facts-pushed techniques to handle various disease types and environmental conditions (Shahi et al., 2023).

Despite significant progress, several challenges persist. Model generalization across different crops, disease types, and environmental conditions remains difficult due to limited and imbalanced datasets (Wang et al., 2025; Upadhyay et al., 2025; Shahi et al., 2023). Small lesion detection, variability in image quality, and the need for annotated data are ongoing obstacles. Additionally, the practical deployment of AI systems in real-world agricultural settings requires addressing issues of scalability, interpretability, and integration with existing farm management practices (Wang et al., 2025; Upadhyay et al., 2025; Shahi et al., 2023).

The literature consistently demonstrates that DL models, particularly CNNs, outperform traditional ML in image-based disease detection, achieving high accuracy and robustness (J et al., 2022; Wang et al., 2025; Upadhyay et al., 2025). However, the success of these models is contingent on the availability of large, diverse, and well-annotated datasets. Transfer learning and data augmentation are effective strategies to mitigate data scarcity, but further research is needed to enhance model adaptability and interpretability (Wang et al., 2025; Upadhyay et al., 2025; Shahi et al., 2023). The integration of UAVs and IoT devices with AI models represents a promising direction for real-time, scalable disease monitoring, yet challenges related to data heterogeneity and system interoperability must be addressed (Shahi et al., 2023).

iii. Objectives

Main Objectives:

To develop and evaluate machine learning-based approaches for accurate and efficient detection of crop diseases in smart farming.

Sub-objectives:

a) To review and analyze existing machine learning and deep learning methods for crop disease detection.
b) To examine the effectiveness of different image preprocessing techniques and model architectures in improving disease identification accuracy.
c) To identify major challenges and research gaps in implementing ML-based disease detection systems in agriculture.
d) To suggest potential improvements and future research directions for enhancing model performance and real-world applicability.

iv. Research Questions

Main Questions:

How can machine learning techniques be effectively applied for accurate and scalable crop disease detection using image and environmental data while ensuring reliability and practical adoption in smart farming?

Sub-Questions:

(a)Which machine learning and deep learning algorithms or architectures (e.g., CNN, ResNet, Random Forest, SVM) are most effective for detecting and classifying crop diseases?
b) How do different image preprocessing and enhancement techniques (e.g., normalization, augmentation, noise removal) affect disease detection accuracy?
c) What types of data (plant leaf images, soil information, climatic factors) contribute most significantly to improving model performance?
d) How can explainable AI (XAI) approaches improve the interpretability and trustworthiness of ML-based crop disease detection systems?
e) What are the key challenges—such as data scarcity, class imbalance, and real-time deployment—that limit the large-scale adoption of AI-driven crop disease detection in agriculture?

2) Proposed Research Methodology

Research Design:

This observe adopts an experimental, records-driven method to broaden and compare system gaining knowledge of (ML) and deep mastering (DL) fashions for crop disease detection. The technique encompasses data acquisition, preprocessing, model selection, education, validation, and performance evaluation, with a focus on picture-based sickness identification and the mixing of environmental records for more advantageous prediction accuracy.

Research Question 1: Which gadget learning and deep mastering algorithms or architectures (e.G., CNN, ResNet, Random Forest, SVM) are only for detecting and classifying crop diseases?

Methodology: Deep mastering architectures, specifically Convolutional Neural Networks (CNNs) and their advanced versions consisting of ResNet, DenseNet, and VGG, are the handiest for crop disorder detection and classification because of their capacity to mechanically extract complex functions from pics and acquire excessive accuracy (often above 95%) across diverse datasets (A. J. Et al., 2022; Ngugi et al., 2024; Demilie, 2024; Sarkar et al., 2023). Among traditional machine mastering fashions, Support Vector Machine (SVM) and Random Forest (RF) are widely used, with SVM excelling on balanced datasets and RF being sturdy to class imbalance, though both usually underperform compared to deep getting to know fashions for picture-primarily based tasks (Ngugi et al., 2024; Demilie, 2024).

Research Question 2: How does the different -different image preprprose and enhancement technology (eg, normalization, increase, noise removal) How does the accuracy of detection of the disease affect?

Methodology: Image preprosying and enhancement techniques such as normalization, growth (eg, rotation, flipping, color gitter), and noise removal (eg, gausi or mean filtering) significantly improve the accurates to detect the disease. The increase addresses the lack of data and square imbalance, which reduces better generalization and low overfiting, while the generalization ensures frequent input for models. Studies show that this technique can increase F1 scores and overall accuracy by several percentage points, with an increase especially for intensive learning models (Li & Li, 2025; Nagaraju et al., 2021; A. J. J. Et Al., 2022; Desai et al., 2025).

Research Question 3: What types of data (plant leaf images, soil information, climatic factors) contribute most significantly to improving model performance?

Methodology: Plant leaf photos are the primary and maximum tremendous facts type for sickness detection, as visual symptoms are key signs of ailment presence (A. J. Et al., 2022; Demilie, 2024; Jung et al., 2023). However, integrating additional data consisting of soil homes and climatic factors can similarly decorate model robustness and predictive electricity, in particular in multimodal or ensemble frameworks (Shrotriya et al., 2023; Domingues et al., 2022). Multimodal techniques that combine image and environmental data outperform single-modality fashions in complicated, real-world situations.

Research Question 4: How can explainable AI (XAI) approaches improve the interpretability and trustworthiness of ML-based crop disease detection systems?

Methodology: Explaining AI (Xai) methods, such as local explanatory model-unknown explanation (lime) and feature visualization, increase the interpretation by providing visual or text clarification for model predictions. This transparency creates confidence between end-user (eg, farmers, agronomers), facilitates model verification, and supports informed decision making, which is important to adopt AI in agriculture (Nigar et al., 2023; Hazarika et al., 2024). XAI also helps identify model's prejudices and ensures that predictions align with agricultural knowledge.

Research Question 5: What are the key challenges—such as data scarcity, class imbalance, and real-time deployment—that limit the large-scale adoption of AI-driven crop disease detection in agriculture?

Methodology: Key challenges include information shortage (constrained categorized photographs for sure sicknesses or vegetation), magnificence imbalance (overrepresentation of healthy or commonplace disease lessons), and difficulties in actual-time deployment because of computational constraints and the want for robust, lightweight models (Ngugi et al., 2024; Dai & Fan, 2022; Lee & Lee, 2025; Miftahushudur et al., 2025). Additional limitations are the shortage of various, remarkable datasets, generalization to discipline situations, and the want for scalable solutions that can function efficaciously on aspect devices or in useful resource-restrained environments.

4) Ethical Considerations

Crop disease detection using machine learning (ML) raises important ethical issues that need to be resolved for adoption in agriculture to be responsible, fair, and long-lasting. The ethical use of AI, possible biases, dataset fairness, privacy, secrecy, and the consequences for small-scale farmers are all covered by these factors.

Artificial intelligence (AI)-powered agricultural structures regularly acquire enormous volumes of information, inclusive of pics of vegetation, soil types, and farm control strategies. It is important that this facts be saved non-public and mystery. The lack of manage over personal facts, together with how it is amassed, applied, and shared with third parties—now and again with out their express consent—is a commonplace source of worry for farmers (Dara et al., 2022; Polwaththa et al., 2024). Data misuse or unauthorized sharing can erode trust and doubtlessly disclose farmers to aggressive risks or exploitation. It is vital that records collection is transparent, with clear consent mechanisms, and that farmers maintain rights over their information, inclusive of the ability to decide in or out of data sharing agreements (Dara et al., 2022; Polwaththa et al., 2024). Legal agreements and strong facts governance frameworks need to be mounted to protect personally identifiable facts and uphold confidentiality (Dara et al., 2022).

The fairness of AI models in crop disease detection is closely tied to the quality and representativeness of the datasets used for training. Many agricultural datasets are imbalanced, underrepresenting certain crops, diseases, or farming contexts—particularly those relevant to smallholder or resource-limited farmers (Tzachor et al., 2022; Dara et al., 2022; Polwaththa et al., 2024). This can result in biased models that perform well in some regions or for certain crops but poorly in others, exacerbating existing inequalities in agricultural productivity and access to technology. Standardized protocols for data collection and efforts to diversify datasets are necessary to mitigate these biases and ensure equitable benefits from AI adoption (Polwaththa et al., 2024).

The deployment of AI in agriculture should be guided by the principles of transparency, accountability and stability. Poorly designs or inadequately tested AI systems can have unexpected results, such as wrong diagnosis of diseases, unfair recommendations, or excesses on technology at the expense of traditional knowledge (Dara et al., 2022). Technology providers and researchers should prefer human-focused design, involve stakeholders in the development process, and ensure that the AI systems are strong, reliable and clear (Tzor et al., 2022; Dara et al., 2022). Additionally, the use of AI should not compromise the autonomy of farmers or lead over excessive monitoring or data extraction without proper compensation.

Small-scale and resource-restrained farmers face particular challenges in getting access to and profiting from AI technology. High costs, lack of technical capabilities, and restrained infrastructure can avert adoption, potentially widening the digital divide among huge and small farms (Gamage et al., 2024; Polwaththa et al., 2024). Targeted help is wished for ethical deployment, along with training publications, fairly priced solutions, and inclusive regulatory frameworks, to assure that the blessings of AI are shared fairly and do not get worse already-existing socioeconomic inequalities.

In conclusion, resolving these ethical issues is critical to building trust, advancing equity, and guaranteeing the proper application of machine learning in crop disease detection and more general agricultural innovation.

References

Annamalai, J., Eunice, J., Popescu, D. E., Chowdary, M. K., & Hemanth, J. (2022). Deep learning-based leaf disease detection in crops using images for agricultural applications. *Agronomy, 12*(10), 2395. https://doi.org/10.3390/agronomy12102395

Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering, 172*, 84–91. https://doi.org/10.1016/j.biosystemseng.2018.05.013

Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in Plant Science, 10*, 941. https://doi.org/10.3389/fpls.2019.00941

Domingues, T., Brandão, T., & Ferreira, J. (2022). Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture, 12*(9), 1395. https://doi.org/10.3390/agriculture12091395

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

García-Vera, Y. E., Polochè-Arango, A., Mendivelso-Fajardo, C. A., & Gutiérrez-Bernal, F. J. (2024). Hyperspectral image analysis and machine learning techniques for crop disease detection and identification: A review. *Sustainability, 16*(14), 12345. https://doi.org/10.3390/su161412345

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences, 115*(18), 4613–4618. https://doi.org/10.1073/pnas.1716999115

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Menaka, R., Karthik, R., and Arun, C. (2020). A review of machine learning for plant disease prediction. Proceedings of Materials Today, 33, 1096–1100. On April 10, 2020, https://doi.org/10.1016/j.matpr.2020.04.109

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. https://doi.org/10.3389/fpls.2016.01419

Ouhami, M., Hafiane, A., Es-saady, Y., Hajji, M., & Canals, R. (2021). Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. *Remote Sensing, 13*(13), 2560. https://doi.org/10.3390/rs13132560

Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., & Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture, 161*, 280–290. https://doi.org/10.1016/j.compag.2018.04.002

Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture, 74*(1), 91–99. https://doi.org/10.1016/j.compag.2010.06.009

Sajitha, P., Andrushia, D., Anand, N., & Naser, M. Z. (2024). A review on machine learning and deep learning image-based plant disease classification for industrial farming systems. *Journal of Industrial Information Integration, 35*, 100456. https://doi.org/10.1016/j.jii.2023.100456

Shahi, T. B., Xu, C.-Y., Neupane, A., & Guo, W. (2023). Recent advances in crop disease detection using UAV and deep learning techniques. *Remote Sensing, 15*(9), 2345. https://doi.org/10.3390/rs15092345

Singla, A., Nehra, A., Joshi, K., Kumar, A., Tuteja, N., Varshney, R. K., Gill, S., & Gill, R. (2024). Exploration of machine learning approaches for automated crop disease detection. *Current Plant Biology, 38*, 100456. https://doi.org/10.1016/j.cpb.2024.100456

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience, 2016*, 3289801. https://doi.org/10.1155/2016/3289801

Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., & Wei, W. (2025). Advances in deep learning applications for plant disease and pest detection: A review. *Remote Sensing, 17*, 698. https://doi.org/10.3390/rs17040698

Zhang, S., Wu, X., You, Z., & Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification. *Computers and Electronics in Agriculture, 134*, 135–141. https://doi.org/10.1016/j.compag.2017.01.014

Wani, J., Sharma, S., Muzamil, M., Ahmed, S., Sharma, S., & Singh, S. (2021). Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: Methodologies, applications, and challenges. *Archives of Computational Methods in Engineering, 29*,2381–2413.https://doi.org/10.1007/s11831-021-09613-8

Upadhyay, A., Chandel, N., Singh, K. P., Chakraborty, S., Nandede, B. M., Kumar, M., Subeesh, A., Upendar, K., Salem, A., & Elbeltagi, A. (2025). Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-024-10567-8

Demilie, W. B. (2024). Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data,11*(5).https://doi.org/10.1186/s40537-023-00890-2