Grenze International Journal of Engineering and Technology, January Issue

Plant Disease Prediction using Machine Learning: A Comprehensive Review

Harshal Shriram Patil[1] and Jitrndra Saxena[2]
[1]Research Scholar, SAGE University, Indore,India
Email: patil.harshal.s@gmail.com
[2]Professor, Institute of Advance Computing, SAGE University, Indore, India
Email: jsaxena531@gmail.com

***Abstract*— Plant diseases have severely affected agricultural yields, causing significant economic losses. To address this, advanced technologies like Machine Learning (ML), Deep Learning (DL), and the Internet of Things (IoT) are being leveraged for early disease detection. This study evaluates ML and DL techniques, including Convolutional Neural Networks (CNNs), support vector machine (SVM), Logistic regression (LR) and transfer learning techniques for plant disease prediction. While DL models excel in processing complex image data, they demand extensive labelled datasets and computational power. Traditional ML approaches, though less resource-heavy, often face limitations in feature extraction and accuracy. The findings highlight the trade-offs between model performance and resource requirements, offering insights for optimizing disease detection in agriculture.**

I. INTRODUCTION

Agriculture is essential for maintaining global food security, yet farmers worldwide encounter persistent challenges such as erratic weather, temperature shifts, soil degradation, and biotic threats like pests, fungi, bacteria, and viruses. These issues lead to substantial yield losses, higher production expenses, and, in many cases, financial hardship for farmers. Recent advancements in technology including Machine Learning (ML), Deep Learning (DL), Edge Computing, and the Internet of Everything (IoE) offer transformative solutions to modernize agricultural practices.

One promising application is the early identification and prediction of plant diseases using automated image analysis. Since leaves often show the first visible signs of infection, ML and DL models can process leaf images to detect abnormalities quickly. This enables farmers to take timely, targeted action, reducing crop damage while boosting productivity and sustainability.

This study presents a tech-driven framework for early plant disease detection using leaf image datasets. By implementing cutting-edge deep learning techniques and edge computing capabilities, the system provides real-time disease alerts and practical recommendations. Such an approach empowers farmers to make informed decisions, mitigates reliance on unpredictable environmental factors, and helps optimize yields while minimizing risks and costs.

Plant diseases are among the leading causes of diminished agricultural output and substantial economic losses globally. These diseases can be classified into bacterial, viral, and fungal infections, each presenting distinct challenges to crop viability.

Grenze ID: 01.GIJET.12.1.612

Viral infections predominantly harm fruit and vegetable crops, frequently causing stunted development, malformed produce, and severe reductions in yield. Fungal pathogens, in contrast, target multiple plant structures—such as leaves, stems, roots, and fruits—leading to blight, mildew, and rot, which impair photosynthesis and nutrient absorption. Bacterial infections, including soft rot, vascular wilt, necrosis, and tumors, damage plant tissues and disrupt water transport systems, further weakening crops.

Recent advancements in machine learning (ML) and deep learning (DL) have demonstrated significant potential in automating disease identification and classification through image-based analysis. Nevertheless, several obstacles limit the accuracy and adaptability of these models. A major challenge lies in acquiring high-quality, diverse datasets. Precise and uniform data collection is critical, particularly for images of leaves, stems, and fruits, which exhibit the earliest disease symptoms. High-resolution imaging is essential to capture fine details, while pre-processing techniques—such as noise reduction, contrast enhancement, and background removal—are necessary to improve feature extraction.

Another key challenge is image segmentation and classification, especially when symptoms overlap or when different diseases exhibit similar visual traits. Variability in disease manifestations across plant species, environmental conditions, and geographic regions further complicates model training and reduces generalizability. To enhance the robustness of disease detection systems, it is imperative to incorporate comprehensive datasets that account for diverse climates, soil conditions, and crop varieties.

This section provides an introduction to plant disease prediction and examines the key challenges associated with it. The subsequent section delves into the deep learning techniques employed for plant disease detection and classification.

II. Deep Learning for Plat Disease Prediction

Traditional machine learning (ML) techniques for plant disease detection depend on manual feature extraction, requiring experts to identify key visual traits like color, texture, and lesion patterns. These features train classifiers such as SVM, k-NN, or Decision Trees. While useful in controlled settings, such methods face limitations in scalability, generalizability, and adaptability to diverse crops, environments, and disease strains. Manual feature selection also introduces bias, reducing accuracy and reproducibility in real-world agricultural applications.

Deep learning (DL), particularly Convolutional Neural Networks (CNNs), addresses these challenges by autonomously learning discriminative features from raw image data, bypassing the need for manual engineering. DL models excel at detecting intricate disease patterns across varying lighting conditions, backgrounds, and plant species. Their superior accuracy, robustness, and adaptability make them ideal for large-scale, real-time disease detection. As a result, DL has revolutionized precision agriculture, enabling advanced disease identification, severity assessment, and data-driven farming decisions.

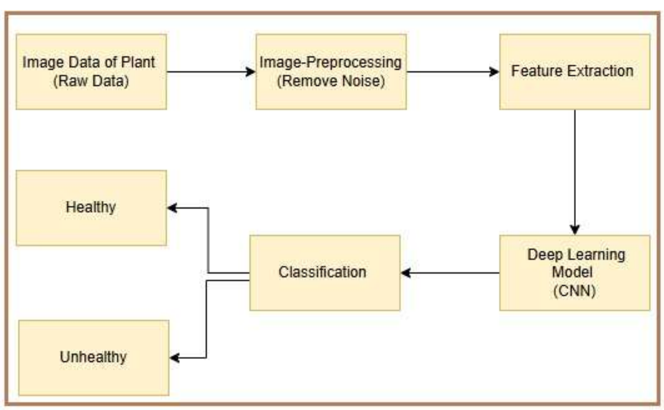

Figure 1: Deep learning Model flow for plant disease prediction

The above figure 01 demonstrate that a deep learning framework designed for predicting plant diseases using image-based data. Initially, raw images of plants including leaves and stems are gathered as input. These images undergo pre-processing to eliminate noise and improve quality through methods such as normalization, resizing, and filtering. Subsequently, discriminative features like texture, color, and shape are extracted to minimize data redundancy and emphasize key disease-related patterns. A Convolutional Neural Network (CNN) is then employed for automated feature learning and classification, leveraging its ability to detect hierarchical spatial features in visual data. The processed features are fed into a classification layer, which predicts whether the plant is healthy or diseased. This end-to-end automated system enhances the precision and speed of plant disease diagnosis, supporting smart agricultural practices by enabling early detection, minimizing reliance on manual inspection, and optimizing crop yield through timely interventions.

III. LITERATURE SURVEY

This section presents a critical review of existing literature on the effectiveness of deep learning (DL) and machine learning (ML) techniques in the early detection of plant diseases. A comprehensive analysis of peer-reviewed studies was conducted to assess the impact of ML-based approaches on agricultural productivity. Relevant research articles were sourced from reputable databases, including Scopus-indexed journals, to ensure the inclusion of high-quality scholarly work.

Midhunraj P. and others presents a comprehensive review of machine learning (ML) and deep learning (DL) techniques for detecting and classifying plant diseases, which significantly impact agricultural productivity. It analyzes various research methodologies and evaluates numerous plant disease datasets, such as Plant Village. The methodology involves comparing traditional ML algorithms with advanced DL models like CNNs. The study highlights advancements, challenges, and future directions in plant disease recognition systems [1]. The author addresses the limitations of traditional and deep learning methods in large-scale plant disease identification. It introduces a few-shot learning approach using a pre-trained ImageNet model fine-tuned with PlantCLEF2022 for feature extraction. The CNN classifier is replaced with an SVM for improved performance. Results show high accuracy, especially on Plant Village and PDD271 datasets [2]. Chouhan S and others study focuses on classifying six medicinal plant species using real-time images and deep learning via IoT. Due to similar leaf traits, traditional identification is challenging. A CNN-based model was developed, achieving 99% accuracy, outperforming benchmark deep learning and machine learning models. The research highlights CNN's effectiveness and suggests future work on automated plant growth estimation [3]. Hassan S and others presents a novel approach for plant disease detection by combining deep CNN-based feature extraction with multiple machine learning classifiers. It highlights that SVM and Logistic Regression outperform several pre-trained deep learning models in accuracy, precision, and recall. This lightweight yet efficient model offers an effective solution for plant disease identification using fewer parameters [4]. Naralasetti V. and others introduces a plant disease prediction framework using transfer learning with VGG16 for feature extraction and a deep neural network for classification. The model shows remarkable accuracy (96.56%) on the PlantVillage dataset. The approach effectively captures intricate disease features and outperforms traditional methods in accuracy, F1-score, and Kappa score, highlighting its strength in early disease detection [5]. Kini A. and others presents a transfer learning-based approach for early-stage black pepper leaf disease detection using ConvNets. By fine-tuning pre-trained models like ResNet18 on a curated leaf dataset, the model achieves 99.67% accuracy. The system shows promise for agricultural applications by enabling timely and accurate disease detection, even at early infection stages, with minimal validation loss [6]. Thiagarajan J and others emphasizes early detection of banana plant diseases using machine learning. It critiques CNN limitations in scale and rotation invariance and proposes alternative hybrid models: ANN with SIFT and ANN with HOG-LBP. These models enhance disease detection by improving feature extraction and recognition capabilities, offering promising accuracy and encouraging future exploration using video-based datasets [7]. Ramamoorthy R. and others presents an enhanced deep learning-based approach using convolutional neural networks (CNNs) and TensorFlow for detecting plant leaf diseases and suggesting treatments. By automating disease recognition and linking symptoms to specific causes, the model improves disease management. With 95% accuracy, it outperforms conventional models, offering a practical solution for sustainable agriculture [8].

Islam M. and others proposed a deep learning-based cotton disease prediction model using fine-tuned transfer learning algorithms (VGG-16, VGG-19, Inception-V3, and Xception). They used a publicly available cotton leaf dataset to evaluate model performance, with the Xception model achieving the highest accuracy of 98.70%. A smart web application was developed for real-time disease prediction to help farmers in agriculture. Their approach aims to automate cotton leaf disease diagnosis and improve agricultural productivity in Bangladesh [9].

Sajitha P and others present a comprehensive review of machine learning and deep learning techniques for image-based plant disease classification. The study highlights the use of various algorithms such as CNNs, SVM, and decision trees across diverse publicly available plant disease datasets. It analyzes key developments and challenges in implementing automated detection systems for industrial farming. The review emphasizes the need for high-quality datasets, robust models, and scalable systems for future advancements in smart agriculture

[10]. Ahmed and Yadav (2023) systematically analyzed machine learning and deep learning techniques for early plant disease diagnosis. They used the Plant Village dataset consisting of 17 diseases across 12 crop species. The study applied SVM, GLCM, CNN, and ANN models to detect and classify plant diseases efficiently. K-means clustering was also employed for real-time image-based detection. Results highlighted CNN and ANN's high accuracy, showcasing the potential for AI-powered precision agriculture in large-scale farming [11]. The authors developed and compared machine learning models for predicting Cotton Leaf Curl Disease (CLCuD) using 9 years of ecological and disease data from four major cotton-growing locations in India. They tested MLR, Bootstrap Forest (BSF), Boosted Tree (BST), and ANN on ~8000 plants per season. BSF achieved the highest R^2 (training: 0.81, validation: 0.80), outperforming ANN (R^2: 0.79) and BST (R^2: 0.59), supporting early disease prediction and intervention planning [12]. Omaye J and others conducted a systematic cross-comparative review of 113 studies (2013–2023) on machine learning techniques for detecting diseases in apple, cassava, cotton, and potato plants. The study highlights the success of CNNs and deep learning in early disease identification using image datasets. While ML shows great potential in prevention, diagnosis, and monitoring, the review points out a significant gap in research related to plant recovery prediction and broader disease management integration [13]. Isinkaye and others conducted a comprehensive review of deep learning and content-based filtering methods for enhancing plant disease detection and treatment recommendations. They emphasized the importance of integrating these technologies to ensure accurate disease identification and improved agricultural outcomes. The study analysed existing techniques, identified research gaps, and proposed future directions to bridge limitations. It advocates a fusion of CNNs and recommender systems as a promising approach toward sustainable agriculture and global food security [14]. Khan and others explored deep transfer learning for fine-grained classification of maize leaf diseases. They applied four advanced models—VGGNet, Inception V3, ResNet50, and InceptionResNetV2—to tackle the challenges posed by subtle disease features. ResNet50 outperformed others with 87.51% validation accuracy and 99.80% recall. The dataset consisted of annotated images of maize leaves showing nuanced disease patterns. The study highlights transfer learning's robustness and generalization ability, proposing it as a reliable tool for agricultural disease diagnostics [15].

IV. COMPARATIVE ANALYSIS

TABLE I. COMPARATIVE ANALYSIS

Authors & their work	Methodology used	Analysis	Limitations
P. K. Midhunraj, K. S. Thivya, and M. Anand conducted a comprehensive review on plant disease detection and classification using Machine Learning (ML) and Deep Learning (DL) techniques [1].	The study reviews and compares various ML and DL techniques such as Support Vector Machines (SVM), Decision Trees (DT), and Convolutional Neural Networks (CNNs). It also evaluates multiple plant disease datasets.	The paper analyzes the effectiveness of ML and DL models in detecting and classifying plant diseases. It highlights trends, performance metrics, and the shift from traditional ML to more accurate DL methods.	The review notes challenges in current systems such as limited dataset availability, variability in image quality, and the complexity of real-time implementation in diverse agricultural environments.
Uskaner Hepsağ proposed a few-shot learning strategy for efficient plant disease identification [2].	Utilized a pre-trained CNN model on ImageNet, fine-tuned with PlantCLEF2022; feature embeddings extracted; classification head replaced with SVM; tested on PlantVillage (38 classes) and PDD271 (271 classes) datasets.	Achieved 88.4% accuracy at 10-shots and 75.5% at 5-shots on PlantVillage; 74.2% at 5-shots, 67.5% at 3-shots, and 56.3% at single-shot on PDD271, outperforming previous few-shot approaches.	Performance declines with fewer shots; PDD271 is complex with more classes, showing lower accuracy; reliance on similarity of fine-tuning dataset (PlantCLEF2022) to target data.
S. Chouhan and others proposed a deep learning approach to classify different medicinal plant species [3].	A Convolutional Neural Network (CNN) model was developed to classify six plant species using real-time images via Internet of Things (IoT).	The CNN model achieved 99% accuracy, outperforming all compared machine learning models	study limited to only six plant species; does not yet address plant growth estimation, which is identified as future work.
Hasan S. and others proposed a deep CNN-based model for plant disease detection and tested it with various ML classifiers [4].	Proposed a deep CNN-based model for plant disease detection and tested it with various ML classifiers.	SVM and Logistic Regression achieved better performance in accuracy, precision, and recall compared to some deep learning models, using fewer parameters.	Limited dataset diversity; performance evaluation on more plant types and real-field conditions is not addressed.
Naralasetti V and others proposed a deep learning-based framework for plant disease prediction using transfer learning [5].	Deep feature extraction using VGG16 and Classification using custom deep neural network.	Achieved 96.56% accuracy; superior performance in F1-score and Kappa score. The model effectively captures fine-grained disease features.	Performance validated only on a benchmark dataset

Kini A. and others proposed a ConvNet based transfer learning model for early-stage black pepper leaf disease detection [6].	Used transfer learning with pre-trained CNNs (InceptionV3, GoogleNet, SqueezeNet, ResNet18) on a real-time annotated black pepper leaf image dataset	ResNet18 achieved the highest accuracy of 99.67%, outperforming other models. The method demonstrated strong performance in validation accuracy and low loss.	Results are dataset-specific. Model generalization to different environmental conditions and pepper varieties not tested.
Thiagarajan J and others analyzed banana plant disease detection using machine learning and proposed two novel hybrid models [7].	Evaluated CNN limitations and proposed two alternative approaches: (1) ANN with Scale-Invariant Feature Transform (SIFT), and (2) ANN with combined HOG (Histogram of Oriented Gradients) and LBP (Local Binary Patterns).	The hybrid ANN models show improved performance in feature extraction and disease classification compared to traditional CNNs. The models address pattern complexity and local features more effectively.	CNN lacks rotation and scale invariance. Integration with feature extraction methods is limited. Real-time implementation and testing on video data remain unexplored.
Ramamoorthy R proposed a deep learning model for plant disease detection and treatment suggestions [8].	Used TensorFlow's object detection API and CNN-based topologies for leaf disease identification and cause-symptom correlation.	The model achieved 95% accuracy, offering reliable disease detection and helpful treatment insights, improving practical utility for farmers.	Limited validation on diverse crop types.
Islam M and others proposed a DL model for cotton disease detection [9].	Fine-tuning transfer learning using VGG-16, VGG-19, Inception-V3, and Xception on a public cotton dataset	Xception model achieved highest accuracy (98.70%). Model performance evaluated and deployed in a web-based application for real-time usage.	Generalization to other crops or unseen conditions not validated; real-world environmental factors not addressed.
Sajitha P and others reviewed ML and DL models for plant disease classification [10]	Literature review of image-based plant disease detection using ML/DL algorithms	Discussed successes of existing systems, sources of plant datasets, comparative analysis of ML and DL approaches in industrial farming contexts.	Lacks experimental validation and dependent on dataset quality
Ahmed and analysed AI-based models for plant disease detection [11]	Used Plant Village dataset. Use ML models SVM, GLCM and DL models CNN, ANN; K-means clustering for real-time image analysis	CNN and ANN provided superior performance in disease classification across multiple crop species and disease types	Real-time deployment challenges. Limited generalization beyond Plant Village dataset performance under varying environmental conditions not tested.
Sain S. developed ML-based models for predicting CLCuD [12].	Used MLR, Bootstrap Forest (BSF), Boosted Tree (BST), and Artificial Neural Network (ANN) 9-year dataset with GDD and disease scores	BSF achieved the best prediction accuracy ($R^2 = 0.80$), followed by ANN ($R^2 = 0.79$) validated using 75:25 split incorporated GUI for field use	Limited to Indian agro-climatic zones. Models may need retraining for different cultivars environments.
Omaye J and others conducted a review of ML applications in disease detection for apple, cassava, cotton, and potato [13]	Systematic literature review (113 papers from 2013–2023) focused on ML techniques including CNNs, SVMs, and other deep learning models applied to plant image datasets	Deep learning models like CNNs excelled in identifying plant diseases across all four crop	Lack of research in recovery prediction and complete disease management
Isinkaye F and others reviewed DL & filtering for plant disease detection and treatment [14].	Systematic literature review on deep learning (e.g., CNNs) and content-based filtering approaches for plant health diagnosis and recommendation systems	Highlights the effectiveness of CNNs in disease detection and the potential of content-based filtering in personalizing treatment suggestions	Lack of integration between detection and recommendation systems
Khan I and others proposed a deep transfer learning approach for maize leaf disease classification [15].	Used four transfer learning models: VGGNet, Inception V3, ResNet50, InceptionResNetV2 on maize disease image data	ResNet50 achieved the best performance (accuracy 87.51%, precision 90.33%, recall 99.80%), proving its effectiveness in nuanced disease detection	Limited to maize. May require retraining for different crops.

The table 01 systematically compares existing plant disease prediction approaches, revealing key limitations. Current models face challenges with inadequate training data quality and diversity, and and limited generalizability, as most deep learning solutions target specific crops or diseases.

A critical research gap exists in handling real-world agricultural complexities existing systems detect single diseases, whereas field conditions often present co-infections and disease interactions that complicate diagnosis. Bridging this laboratory-to-field implementation gap remains crucial for developing robust, scalable solutions in agricultural AI.

V. CONCLUSION

The increasing prevalence of plant diseases poses significant risks to agricultural productivity and global food supply chains. This research conducted a comprehensive evaluation of machine learning such as SVM, LR and deep learning approaches CNN, ANN and transfer learning for automated disease identification. Our analysis revealed that while deep learning architectures deliver superior classification accuracy, they necessitate substantial computational resources and extensive training datasets. Conversely, conventional machine learning algorithms demonstrate greater computational efficiency but with reduced predictive performance. These findings emphasize the need for context-specific model selection based on available infrastructure and data resources

REFERENCES

[1] Midhunraj P. K., Thivya K. S., & Anand M. (2024) "An analysis of plant diseases on detection and classification: From machine learning to deep learning techniques" Multimedia Tools and Applications, 83, 48659–48682.

[2] Uskaner Hepsag P. (2024)" Efficient plant disease identification using few-shot learning: A transfer learning approach "Multimedia Tools and Applications" 83, 58293–58308.

[3] Chouhan S. S., Singh U. P., Sharma U., Jain S. (2024). Classification of different plant species using deep learning and machine learning algorithms "Wireless Personal Communications", 136, 2275–2298.

[4] Hassan S. M., Maji A. K. (2022). Deep feature-based plant disease identification using machine learning classifier. Innovations in Systems and Software Engineering, S.I.: Multifaceted Intelligent Computing Systems (MICS).

[5] Naralasetti V, Bodapati J. D. (2024)" Enhancing plant leaf disease prediction through advanced deep feature representations: A transfer learning approach" Journal of the Institution of Engineers (India): Series B, 105(3), 469–482.

[6] Kini A. S., Prema K. V.,Pai, S. N." Early stage black pepper leaf disease prediction based on transfer learning using ConvNets" Scientific Reports, (2024) 14:1404

[7] Thiagarajan J. D., Kulkarni S. V., Jadhav S. A., Waghe A. A., Raja S. P., Rajagopal S., Poddar H., & Subramaniam, S. (2024). Analysis of banana plant health using machine learning techniques." Scientific Reports", 14, 15041

[8] Ramamoorthy R., Kumar E. S., Naidu R. C. A.,Shruthi K. (2023) "Reliable and accurate plant leaf disease detection with treatment suggestions using enhanced deep learning techniques." SN Computer Science, 4(158).

[9] Islam M. M., Talukder M. A., Sarker M. R. A., Uddin M. A., Akhter A., Sharmin S., Mamun M. S. A., Debnath S. K. (2023) "A deep learning model for cotton disease prediction using fine tuning with smart web application in agriculture" Intelligent Systems with Applications, 20, 200278.

[10] Sajitha P., Andrushia A. 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, 38, 100572.

[11] Ahmed I.,Yadav P. K. (2023). A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases. Sustainable Operations and Computers, 4, 96–104.

[12] Sain S. K., Monga D., Paul D., Prakash A. H., Kumar P., Prasad Y. G., Kumar A., Mohan M. (2024) "Cotton leaf curl disease (CLCuD) prediction modeling in upland cotton under different ecological conditions using machine learning tools", Ecological Informatics, 81, 102648

[13] Omaye J. D., Ogbuju E., Ataguba G., Jaiyeoba O., Aneke J., Oladipo F. (2024) "Cross-comparative review of machine learning for plant disease detection: Apple, cassava, cotton and potato plants", Artificial Intelligence in Agriculture, 12, 127–151.

[14] Isinkaye F. O., Singh P. K., Olusanya M. O. (2024) "Deep learning and content-based filtering techniques for improving plant disease identification and treatment recommendations: A comprehensive review" Heliyon, 10, e29583.

[15] Khan I., Sohail S. S., Madsen D., Khare B. K. (2024) "Deep transfer learning for fine-grained maize leaf disease classification" Journal of Agriculture and Food Research, 16, 101148.