Identifying and Categorizing Plant Diseases Through Deep Learning Techniques - A Survey

A. THAMARAISELVI[1], S. LESA2, C. POONGODI[3]
[1]Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu
[2]Student, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu
[3]Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tamil Nadu

Abstract— Identifying and classifying plant diseases is important to maintain healthy crops and increase agricultural productivity. Early detection and identification of plant diseases from leaf images using deep learning is an important and challenging aspect of agricultural research. There is a need for such research in India because agriculture is one of the major sources of income, accounting for 17% of the gross domestic product (GDP). A good harvest and increased yield can contribute to farmers and the national economy. Traditional diagnostic methods rely on manual inspection and laboratory tests, which are time-consuming and labor-intensive. With the rapid development of deep learning, especially convolutional neural networks (CNN), the ability to improve and refine diagnostic procedures has increased. This study investigates the use of deep learning technology in the identification and classification of leaf diseases. Deep learning models are used to analyze image data in large pages and identify some visual patterns to accurately classify various diseases. The paper also discusses the problems encountered when training deep learning models, such as inconsistent datasets and overloading, and proposes ways to solve these problems. This work demonstrates the potential of deep learning as a powerful tool for monitoring and controlling agricultural diseases and provides an excellent solution for disease detection and control.

Indexed Terms- Plant Disease, Deep Learning, CNN, Transfer Learning, Image Segmentation, Feature Extraction.

I. INTRODUCTION

The rapid development of machine learning (ML), especially deep learning (DL), has led to major disruptions in many areas, such as agriculture. While machine learning is a part of artificial intelligence (AI), which involves training algorithms to recognize patterns in data, deep learning is a more advanced branch of ML that uses multilayered neural networks. Another is to build complex models. In agriculture, deep learning has emerged as a powerful tool for traditionally manual and time-consuming tasks. Plant diseases that affect global crop yields and food security are often detected by laboratory technicians, which often requires expensive and time-consuming tests. While useful in some cases, these methods are not always applicable, especially in large-scale farming, where early detection is important to reduce crop yields. Timely and accurate identification of plant diseases is important for good management, but it is still a major challenge, especially in regions with little knowledge of agriculture. To address these issues, recent research has focused on developing automated systems for plant disease detection using deep learning. Deep learning, especially convolutional neural networks (CNNs), has been incredibly successful in image classification, including medical applications, facial recognition, and more recently agricultural applications. By training the model on a large database of plant images, the deep learning program can learn important features of various plant diseases, thereby identifying and describing the appearance pain of the disease. This research paper provides a comprehensive overview of the techniques and methods used for plant disease identification and classification through deep learning. It reviews various deep learning models used for this problem, from early experiments on simple neural networks to more advanced CNN-based architectures. This paper also explores the fundamental issues in developing this model, such as big data, high-quality data acquisition for heterogeneous classes in the training data, and improving the interpretation and generalization of predictive models. The survey also highlights the various plants and organisms focused on in current research, as well as the tools and platforms used to facilitate design and deployment. The main goal of this research is to fully understand the current state of plant disease diagnosis through in-depth study and to gain an understanding of how this technology can be improved and integrated into real-life agriculture

II. LITERATURE REVIEW

[1] Vasileios balafas, Scientific review of machine learning (ML) and deep learning (DL) for disease detection and classification in precision agriculture. It proposes a framework that divides research into two categories: distribution (identification of pathogens targeting healthy plants) and detection (pathology of diseased leaves). The paper also provides an overview of the materials used in these projects, including their styles and types. A computational study examined five object detection algorithms and eighteen classification algorithms on the PlantDoc dataset. The results show that YOLOv5 performs well in object detection, while ResNet50 and MobileNetv2 provide the best balance between accuracy and efficiency in classification. This article provides insights into plant disease diagnosis using machine learning and deep learning.

[2] Diana Susan Joseph, Focusing on creating datasets for the automatic identification of plant diseases in food grains like rice, wheat, and maize, using real-life images of diseased plants. These images, sourced from various datasets and online platforms, were pre-processed to remove irrelevant items. Initially, each disease class had only 100 images, which was insufficient for deep learning models, so data augmentation techniques like rotation, flipping, and zooming were applied to increase the dataset size. The dataset was then split into 80% for training, 10% for validation, and 10% for testing.Several fine-tuned convolutional neural network (CNN) models, including MRW-CNN, were trained and evaluated using performance metrics like accuracy, F1-score, precision, and recall. The results showed high accuracy, particularly when MRW-CNN was trained from scratch. The study emphasizes the value of using real-life images for practical applications, helping farmers detect diseases early and take timely action to minimize crop losses.

[3] Muhammad Bammad Saleem, Here emphasizes the importance of early detection of plant diseases and how Deep Learning (DL) has improved accuracy compared to traditional Machine Learning (ML). It reviews various DL models used for plant disease detection and the visualizations that help identify disease symptoms. The review also discusses performance metrics to evaluate these models. Despite advancements, the study identifies gaps in detecting diseases before symptoms appear. One major issue is the use of datasets like PlantVillage, which has plain backgrounds that don't reflect real-world conditions. More practical datasets, simulating complex environments with background noise and environmental factors, are needed to improve disease detection accuracy.

[4] Emmanuel Moupojou, The FAO estimates that food production must increase by 70% by 2050, while a third of food is lost due to plant diseases. To help address this, deep learning models have been developed to detect crop diseases early. These models are often trained on datasets like PlantVillage and PlantDoc. However, PlantVillage's laboratory images with simple backgrounds perform poorly on real-world field images. PlantDoc, with 2,569 field images, has some misclassifications due to lack of expert annotation. To overcome these limitations, the Field Plant dataset was created, with 5,170 field images and 8,629 annotated leaves across 27 disease classes. It outperforms PlantDoc in classification tasks, but challenges remain with real-world images. The study suggests using model ensembling and image segmentation for improved accuracy.

[5] K. P. Asha Rani, Plant disease detection using 38 transfer learning models, evaluating their performance on three datasets: Sunflower, Cauliflower, and Agri-ImageNet (a plant-related subset of ImageNet). The aim is to identify the best models for disease detection. Factors such as accuracy, dataset characteristics, hyperparameters, overfitting, and model complexity were considered. The models were trained with early stopping to prevent overfitting.InceptionResNetV2 was used as the benchmark, but EfficientNetV2B2 and EfficientNetV2B3 outperformed others across the

datasets. Some models, including VGG-16, VGG-19, Inception, NasNet, and ResNet, showed lower performance, possibly due to their complex architectures or suboptimal fine-tuning.The EfficientNetV2 models were preferred over EfficientNetB models for their balance of accuracy and lower computational demands. Similarly, ConvNeXT models, though similar in performance, were less efficient. The system also predicts the pathogen type responsible for the disease, providing audio outputs with actionable remedies for farmers.

[6] Mohammed Saeed Alzahrani, Focusing on the early detection of tomato leaf diseases, which pose a significant threat to crop yield and quality. Using computer vision and deep learning, the research aims to improve disease classification, with transfer learning enhancing the model's efficiency and cost-effectiveness. Early identification of tomato diseases is crucial for managing and preventing economic damage. The study compares three deep learning models—DenseNet169, ResNet50V2, and ViT (Vision Transformer)—for diagnosing tomato diseases. The dataset used for training and testing includes both healthy and diseased tomato images. DenseNet121 achieved the highest accuracy, with a training accuracy of 99.88% and testing accuracy of 99%. ResNet50V2 and ViT also showed good performance, with testing accuracies of 95.60% and 98%, respectively. These results highlight the potential of deep learning for accurate and efficient disease detection, which can help manage tomato diseases early, improving both yield and quality. The study also demonstrated that ensemble models are effective due to their quick training times and exceptional performance. While this research offers a simple, cost-effective approach for diagnosing tomato leaf diseases, it hasn't yet been integrated into a mobile application. Future improvements include incorporating advanced AI and IoT technology to enhance the model's capabilities.

[7] Mahrin Tasfe, Emphasizing the importance of automated early detection and classification of paddy diseases to improve crop management, reduce pesticide use, and prevent disease spread. It highlights the role of deep learning (DL) models in classifying paddy diseases, discussing their applications, strategies for performance improvement, and data augmentation techniques. The study also explores datasets used in this field and identifies gaps, challenges, and open issues in current research. It stresses the significance of computer vision for effective disease management in the context of population growth, climate change, and limitations of manual diagnosis. The paper provides an overview of common paddy diseases, their symptoms, and causes, while offering insights into open-access datasets. It aims to guide future research and address challenges in precision agriculture.

[8] Meenakshi Aggarwal, Exploring the use of federated learning (FL) for classifying rice leaf diseases, addressing data privacy concerns in traditional machine learning models. FL enables decentralized training on local devices, maintaining privacy. The paper introduces a federated transfer learning (F-TL) framework for four rice diseases: bacterial blight, brown spot, blast, and tungro. It evaluates deep learning models, including CNN and transfer learning models, with MobileNetV2 and EfficientNetB3 performing the best, achieving up to 99% accuracy. The federated approach outperforms traditional models in accuracy, loss, and resource efficiency. The study concludes that F-TL ensures data privacy while delivering strong performance, with MobileNetV2 chosen for its lightweight nature. Future improvements may include hyperparameter optimization, advanced averaging techniques, and encryption methods for better privacy and performance.

[9] Zhichao Chen, Addressing the challenge of detecting tomato leaf diseases, which impact crop yields and economic outcomes. Timely diagnosis is crucial, and deep learning has significantly improved disease detection. However, model effectiveness depends on high-quality training data. To solve the class imbalance issue, the paper proposes a cycle-consistent generative adversarial network (CyTrGAN) based on Transformer models for generating synthetic diseased tomato leaf images. The model combines Transformer architecture for global dependencies and a densely connected CNN for local features, enhancing disease classification. The proposed model achieved high accuracy: 99.45% on PlantVillage, 98.30% on AI Challenger, and 95.4% on a private dataset, outperforming previous models. CyTrGAN generates better synthetic images, improving model

robustness. The study also highlights dataset limitations, such as simplified environments and imbalanced categories, which can lead to overfitting. The authors suggest creating larger, more diverse datasets to improve the model's real-world applicability.

[10] Weihao su, A Vision Transformer (ViT) model for detecting and classifying surface defects on green plums, addressing the challenge of identifying multiple and minor defects that traditional methods miss. The model identifies defects like scars, flaws, rain spots, and rot, using a dataset of 2799 images categorized into 18 types of defects. The ViT model achieves an accuracy of 96.21%, outperforming other deep learning models like VGG16 and ResNet. While effective, the model faces challenges with training speed. This method, applicable to other fruits, enhances fruit classification and improves the value of minimally defective produce. However, improvements can be made by using rotating conveyors for real-time defect detection and by exploring hyperspectral imaging for internal defect identification to further enhance food safety and quality.

III. PERFORMANCE METRICS

Study	Model(s) Used	Dataset(s)	Accuracy (%)	Comments
[1]	YOLOv5, ResNet50, MobileNetv2	PlantDoc	98%	YOLOv5 excelled in object detection; ResNet50 and MobileNetv2 showed the best balance for classification.
[2]	MRW-CNN	Rice, wheat, maize (augmented real-life images)	High accuracy, particularly for MRW-CNN	Fine-tuned CNN models showed high performance; MRW-CNN trained from scratch was highly accurate.
[3]	Various Deep Learning models	PlantVillage, others	ACC 98%	Identified issues with dataset quality (e.g., plain backgrounds) affecting model performance.
[4]	Model ensembling, Image segmentation	PlantVillage, PlantDoc, Field Plant	Outperformed PlantDoc	Field Plant dataset (5,170 images) outperformed PlantDoc in classification tasks; challenges with real-world images.
[5]	EfficientNetV2B2, EfficientNetV2B3, ConvNeXT	Sunflower, Cauliflower, Agri-ImageNet	Best performance from EfficientNetV2B3	EfficientNetV2 models preferred for their balance of accuracy and computational efficiency.
[6]	DenseNet169, ResNet50V2, ViT (Vision Transformer)	Tomato leaf images	DenseNet121: 99.88% training, 99% testing	DenseNet121 achieved the highest accuracy for tomato leaf disease detection.
[7]	Various Deep Learning models	Paddy disease datasets	98.76%	Discussed the importance of DL for paddy disease detection; emphasized challenges like data augmentation.

[8]	MobileNetV2, EfficientNetB3 (F-TL)	Rice leaf diseases	Up to 99%	Federated Learning models performed well with MobileNetV2 being the most efficient.
[9]	CyTrGAN (Transformer + CNN)	PlantVillage, AI Challenger, Private dataset	99.45% (PlantVillage), 98.30% (AI Challenger), 95.4% (private)	CyTrGAN outperformed other models, generating synthetic images for better model robustness.
[10]	Vision Transformer (ViT)	Green plums (18 types of defects)	96.21%	ViT model outperformed other models like VGG16 and ResNet in defect detection.

IV. SCOPE FOR IMPROVEMENT

Transfer learning, vision transformer these algorithms can handle sequence data and it could help in time series data. VIT has revolutionized image classification by replacing traditional convolutions with self-attention mechanisms. It can handle long-range dependencies and spatial relationships, which makes them ideal for complex plant disease images. In CNN,Models like ResNet101, ResNet152, may improve performance. These algorithms and models can produce output with high accuracy.

CONCLUSION

Deep learning has revolutionized the field of plant disease identification and classification, offering significant improvements in accuracy, speed, and scalability compared to traditional methods. Convolutional Neural Networks (CNNs) have proven to be highly effective in extracting features from plant images, enabling the detection of even subtle signs of disease. Transfer learning has further enhanced these models by leveraging pre-trained networks, reducing the need for large labeled datasets and speeding up the deployment of disease detection systems. Additionally, the integration of advanced techniques like Generative Adversarial Networks (GANs) for data augmentation and hybrid models that combine CNNs with other machine learning algorithms has further boosted performance. With continuous advancements in deep learning and the availability of large, annotated plant disease datasets, these models are becoming more robust, allowing for real-time and on-field monitoring of plant health. This has the potential to greatly enhance crop yields management, reduce losses, and ensure food security. As the technology evolves, the combination of deep learning with other emerging technologies like Internet of Things sensors, drones, and satellite imaging will likely transform plant disease detection into a highly efficient, automated, and cost-effective process, making it a vital tool for modern agriculture.

REFERENCES

[1] Vasileios balafas, emmanouil karantoumanis, Malamatilouta and Nikolaos ploskas, Machine Learning and Deep Learning for Plant Disease Classification and Detection, IEEE Access (Volume: 11), 16 October 2023 , 10.1109/ACCESS.2023.3324722 , https://ieeexplore.ieee.org/document/10286031.

[2] Diana susan joseph , (student member, ieee), Pranav m. pawar ,and Kaustubh chakradeo, Real-Time Plant Disease Dataset Development and Detection of Plant Disease Using Deep Learning, IEEE Access (Volume: 12), 25 January 2024, 10.1109/ACCESS.2024.3358333, https://ieeexplore.ieee.org/document/10414062

[3] Muhammad Hammad Saleem, Johan Potgieter and Khalid Mahmood Arif , Plant Disease Detection and Classification by Deep Learning, MDPI *Plants* 2019, *8*(11), 468, https://doi.org/10.3390/plants8110468

[4] Emmanuel moupojou, Appolinaire tagne, Florent retraint ,Anicet tadonkemwa, Dongmo wilfried,

Hyppolite tapamo and Marcellin nkenlifack, FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning, IEEE Access (Volume: 11), 29 March 2023 10.1109/ACCESS.2023.3263042, https://ieeexplore.ieee.org/document/10086516

[5] k. p. Asha rani and. Gowrishankar, Pathogen-Based Classification of Plant Diseases: A Deep Transfer Learning Approach for Intelligent Support Systems, IEEE Access (Volume: 11), 09 June 2023, 10.1109/ACCESS.2023.3284680, https://ieeexplore.ieee.org/document/10147120

[6] Mohammed Saeed Alzahrani and Fawaz Waselallah Alsaade, Transform and Deep Learning Algorithms for the Early Detection and Recognition of Tomato Leaf Disease, MDPI *Agronomy* 2023, *13*(5), 1184, 22 April 2023, https://www.mdpi.com/2073-4395/13/5/1184

[7] Mahrin tasfe , Akm nivrito , Fadi al machot ,Mohib ullah , (member, ieee), and Habib ullah, Deep Learning Based Models for Paddy Disease Identification and Classification: A Systematic Survey, IEEE Access (Volume: 12), 27 June 2024, 10.1109/ACCESS.2024.3419708, https://ieeexplore.ieee.org/abstract/document/10574294

[8] Meenakshi Aggarwal , Vikas Khullar , Nitin Goyal , Rama Gautam , Fahad Alblehai , Magdy Elghatwary and Aman Singh, Federated Transfer Learning for Rice-Leaf Disease Classification across Multiclient Cross-Silo Datasets,MDPI *Agronomy* 2023, *13*(10), 2483, 27 September 2023, https://doi.org/10.3390/agronomy13102483

[9] Zhichao Chen, Guoqiang Wang , Tao Lv and Xu Zhang, Using a Hybrid Convolutional Neural Network with a Transformer Model for Tomato Leaf Disease Detection, MDPI *Agronomy* 2024, *14*(4), 673, 26 March 2024, https://doi.org/10.3390/agronomy14040673

[10] 10. Weihao Su, Yutu Yang , Chenxin Zhou, Zilong Zhuang and Ying Liu, Multiple Defect Classification Method for Green Plum Surfaces Based on Vision Transformer, MDPI *Forests* 2023, *14*(7), 1323, 28 June 2023, https://doi.org/10.3390/f14071323

[11] Jian P, Chen K, Cheng W (2021) GAN-Based One-Class Classification for Remote-Sensing Image Change Detection. IEEE Geosci Remote Sens Lett https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10560510

[12] Abdollahi A et al (2020) Building footprint extraction from high resolution aerial images using Generative Adversarial Network(GAN) architecture. IEEE Access 8:209517–209527 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9260150

[13] Le VNT et al (2020) A novel method for detecting morphologically similar crops and weeds based on the combination of contour masks and filtered Local Binary Pattern operators. GigaScience9(3): giaa017 https://academic.oup.com/gigascience/article/9/3/giaa017/5780256

[14] Albahli S et al (2021) An improved faster-RCNN model for handwritten character recognition. Arab J Sci Eng 1–15 https://www.researchgate.net/publication/350508860_An_improved_faster-RCNN_model_for_handwritten_character_recognition

[15] Albahli S et al (2021) Recognition and Detection of Diabetic Retinopathy Using Densenet-65 Based Faster-RCNN. ComputMater Contin 67:1333–1351 https://www.techscience.com/cmc/v67n2/41339

[16] Chen J, Yin H, Zhang D (2020) A self-adaptive classification method for plant disease detection using GMDH-Logistic model. Sustain Compute Inform Syst 28:100415 https://www.sciencedirect.com/science/article/abs/pii/S2210537920301426

[17] Atila Ü et al (2021) Plant leaf disease classification using efficient net deep learning model. Eco Inform 61:101182 https://www.sciencedirect.com/science/article/abs/pii/S1574954120301321

[18] Solano-Rojas B, Villalón-Fonseca R, Marín-Raventós G (2020) Alzheimer’s Disease Early Detection Using a Low Cost Three-dimensional Densenet-121 Architecture. In: International Conference on Smart Homes and Health

Telematics. Springer https://link.springer.com/chapter/10.1007/978-3-030-51517-1_1

[19] Agarwal M et al (2020) ToLeD: Tomato leaf disease detection using convolution neural network. Procedia Comput Sci 167:293–301 https://www.sciencedirect.com/science/article/pii/S1877050920306906

[20] Argüeso D et al (2020) Few-Shot Learning approach for plant disease classification using images taken in the field. Comput Electron Agric 175:105542 https://www.sciencedirect.com/science/article/abs/pii/S0168169920302544