KHATULISTIWA SMART : JOURNAL OF ARTIFICIAL INTELLIGENCE
e-ISSN : XXXX-XXXX
p-ISSN : XXXX-XXXX

Khatulistiwa Smart : Journal of Artificial Intelligence
Vol. 1 No. 1 (2024) : 26-34
Available online at : https://journal.literasikhatulistiwa.org/index.php/smart

Plant Disease Identification Using Deep Learning: A Systematic Literature Review

Mochammad Faid [1]
[1] Universitas Nurul Jadid, Probolinggo
mfaid@unuja.ac.id[1]

Doi
Received: 25 November 2024 *Accepted: 18 Desember 2024* *Published: 31 Desember 2024*

Abstract This research aims to analyze and summarize recent approaches in plant disease identification and classification using deep learning techniques. Through a systematic literature review, we evaluate the various methodologies, neural network architectures, and datasets used in recent studies in this field. Our findings show that the use of deep learning, especially by utilizing complex neural network architectures, has led to significant improvements in plant disease identification accuracy. One of the key findings is the highest accuracy achieved by the Inception Net CNN architecture-based Deep Learning method in detecting diseases in tomato plants, reaching 99.89%. These results confirm that deep learning approaches have great potential to optimize plant disease management and improve agricultural productivity globally.

Keywords *Plant diseases; Deep learning; Classification; Identification; Systematic literature review*

1. INTRODUCTION

The agricultural sector is the mainstay of people's livelihoods and serves as the primary source of income for the majority of Indonesia's population. Since April 1, 1969, through the implementation of the Five-Year Development Plan (Repelita), development focus has been directed toward this sector. Agriculture has been given top priority due to its dominance in the national economy. This is evident from its contribution to national income, provision of employment, enhancement of foreign exchange, and its role in maintaining food security. Additionally, the agricultural sector supports rural development and plays a role in environmental conservation. Given these positive aspects, the agricultural sector is continuously developed through various government programs aimed at increasing productivity and farmers' welfare. The support of modern technology and market access also plays a crucial role in the development of this sector (Rochaeni, 2023).

However, Indonesian farmers are often confronted with plant diseases that damage crops. Various plant diseases can emerge at different stages of plant development, disrupting growth and negatively affecting overall production. Plant diseases can be caused by various conditions during the plant development phases. Generally, the causes of plant diseases are categorized into two types: biotic and abiotic factors. Biotic factors include viruses, fungi, bacteria, mites, and snails that cause microbial infections in plants. Meanwhile, abiotic factors encompass water, temperature, light, and nutrient deficiencies that can hinder plant growth (Demilie, 2024).

Symptoms of various plant diseases can include spots, dead or dying tissues, blurry spores, lumps or swellings, and irregular fruit colors. These symptoms can lead to a decrease in both the quality and quantity of the harvest. The disease triangle involves three main components: a susceptible plant, an attacking pathogen, and environmental conditions that support the infection's development. If one of these components is absent or unsuitable, the infection may not occur or may not develop significantly. Understanding the interaction between these three components is crucial for effective plant disease control (Nizamani et al., 2023).

The detection and classification of plant diseases are essential applications of Deep Learning (DL), Machine Learning (ML), and computer vision in the agricultural industry. The primary goal is to develop algorithms and techniques that can automatically detect and classify plant diseases based on images of leaves or other plant parts. This will assist farmers in identifying and addressing diseases more effectively. After conducting in-depth studies and critical analyses of various recently developed ML and DL-based approaches for plant disease detection and classification, the authors have summarized several key challenges in this field. This summary provides opportunities for the research community to further investigate the causes of plant diseases. These approaches can significantly impact real-time systems for plant disease identification and diagnosis. Several factors and issues affecting the identification and classification of diseases have been summarized in this study (Demilie, 2024).

With technological advancements, particularly in artificial intelligence, the process of classifying plant diseases can be automated with high accuracy. This not only speeds up disease identification but also helps farmers take timely preventive or corrective actions. Moreover, the use of automatic classification methods can help minimize human errors and enhance overall agricultural productivity. Therefore, this research is highly relevant in supporting the development of sustainable and efficient agriculture (Demilie, 2024).

2. RESEARCH METHODS

The purpose of this study is to identify the trends and highest accuracy rates in the use of algorithms in image processing-based plant disease classification systems in the agricultural industry. This system is designed to detect, categorize, model, and overcome limitations associated with plant diseases, so as to improve crop management effectiveness and agricultural productivity (Mendalam, 2024). We conducted a systematic literature review (SLR) of various scientific sources to determine the best algorithms that can be used in plant disease classification. In this process, we evaluated the advantages and disadvantages of each algorithm, as well as examined their practical applications in the context of modern agriculture. This research is expected to provide valuable insights for the development of more accurate and efficient plant disease classification systems in the future.

2.1. SLR Technique

Figure 1 shows the PRISMA flow in this study. Based on Figure 1, a total of 90 articles were initially collected, but 40 articles made it through the identification stage, while 50 articles were excluded. At the identification stage, the 40 articles were then screened based on the inclusion and exclusion criteria, resulting in 14 articles advancing to the eligibility stage, while the remaining 26 articles were excluded. At the eligibility stage, 8 articles were finally included in the systematic review, while 6 articles did not make it to the final stage. This section presents information on the planning and selection criteria in selecting relevant papers published in the last 5 years, i.e. between 2019 and 2024, with the keywords plant diseases, plant foliar diseases, and algorithms.

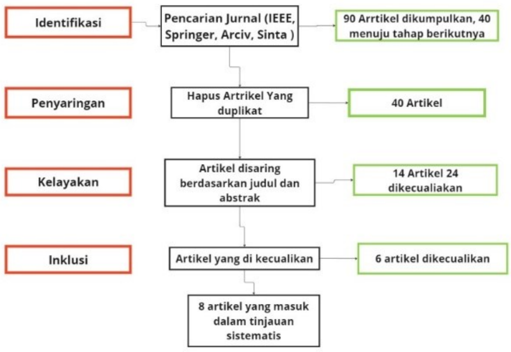

Figure 1. SLR Technique

2.2. Dataset

A major step in plant disease identification and categorization is to require a proper and sufficient dataset to identify each stage of an object (Ramanjot et al., 2023). Images can be taken manually using a camera device or obtained from open source repositories (Nagaraju & Chawla, 2020). Table 1 presents information about the datasets from each of the 15 selected scientific papers.

Table 1. Data Acquisition

Title	Plant types and classes	Data source	Total Image
(Shafik et al., 2023)	Pepper (Healthy and Unhealthy)	https://plantvillage	1855
(Shoaib et al., 2022)	Tomatoes (bacterial spot, blight, leaf fungus, leaf rot, partial spot, spider mite, two spot, spot mosaic virus, yellow leaf curl virus, Healthy)	https://plantvillage	18157
(Mendalam, 2024)	Tomatoes (bacterial leaf spot, septoria leaf spot, mosaic virus, yellow leaf curl virus, fungus, early year rot, late year rot, healthy leaves)	Smartphone, kamera	10337

(Ramanjot et al., 2023)	Medicinal Plants (Binahong, Betel Leaves, Bay Leaves, Life Connecting Leaves)	In-person and internet collection	600
(Nagaraju & Chawla, 2020)	Potatoes (Early rot, Late rot, Healthy)	https://plantvillage	2152
(Arshad et al., 2023)	Potatoes (Healthy, Late Blight, Early Blight)	https://plantvillage	5160
(Nishad et al., 2022)	Tomato (late blight, two-spotted spider mite, tomato yellow leaf curl virus, and tomato seha)	Kaggle	6800

2.3. Pre-processing

Image pre-processing is a data manipulation step used to clean image data before it is ready for use in further analysis. This technique is important to improve image quality and ensure that the processed data is more accurate (Zhao et al., 2022). Table 2 describes various pre-processing techniques applied by various researchers, covering methods such as normalization, noise removal, contrast enhancement, and segmentation. These techniques help prepare the images for the next stage of analysis and classification, ensuring more accurate and consistent results.

2.4. Augmentation

Data augmentation is a strategy used to increase the number of images in a data set with the aim of improving accuracy. In this section, we will explore various techniques used by various researchers to expand the size of the data set, thereby enriching the variety and representation in the processed data. Presented in table 2 are the augmentation processes of previous researchers.

Table 2. Augmentation

ref	Rotasi	Scaling	Pergeseran	crop	Flipping	Penambahan Noise	Perubahan Warna	Pra-pemrosesan
(Shafik et al., 2023)	√	√	-	√	-	√	√	Segmentasi KGDC
(Shoaib et al., 2022)	-	-	-	-	-	-	-	WGAN dan SRGAN
(Mendalam, 2024)	√	√	√	√	√	-	-	Menghilangkan bagroud foto
(Ramanjot et al., 2023)	√	√	-		√	-	-	-
(Nagaraju & Chawla, 2020)	√	√	-	-	-	-	√	Resezing 256×256
(Arshad et al., 2023)	√	√	√	-	-	-	-	-
(Nishad et al., 2022)	√	-	√	-	√	-	-	1. /255 RGB

2.5. Methods

To identify and categorize objects, computer vision-based systems follow a series of predefined steps, starting from image acquisition and followed by a number of image processing tasks such as scaling, filtering, segmentation, selection, and feature extraction. In this context, we conducted a review of a number of scientific works that explore plant disease identification and analysis by applying machine learning methods or deep learning-based techniques. Our review not only considered the technical aspects, but also highlighted various research parameters, including the size of the data set, the type of ML/DL algorithms used, the implementation of classifiers, the evaluation of performance metrics, and other relevant factors. In our analysis, we observed that a number of researchers managed to achieve the highest level of accuracy in their efforts. Presented in Table 3 is a comparison of previous research methods.

Table 3. Method comparison

Judul	Metode	Akurasi	Keterangan
(Shafik et al., 2023)	DSAtt-CMNetV3	96,8%	Berdasarkan beberapa eksperimen, DSAtt-CMNetV3 terbukti sangat efektif untuk klasifikasi penyakit daun lada dan proses segmentasinya ditingkatkan oleh KGDC. Metode Os-OA diperkenalkan untuk mengoptimalkan hyperparameter, yang terbukti sangat efektif dalam meningkatkan performa model ini.
(Shoaib et al., 2022)	CNN VGG16, CNN ResNet50, CNN DenseNet121	97,83 % 97,83 % 98,98 %	Dari beberapa experimen peneliti memakai 3 metode untuk menguji hasil terbaik dari deteksi penyakit daun tomat
(Mendalam, 2024)	CNN + Yolo7	98,8%	Meningkatkan deteknis penyakit pada daun tomat dengan menggunakan metode CNN yang ditingkatkan ke Yolo7
(Ramanjot et al., 2023)	CNN MobileNetV2	94,16%	Penelitian ini menggunakan teknologi deep learning dengan metode transfer learning menggunakan model MobileNetV2 untuk identifikasi daun tanaman obat. Hasilnya memberikan rekomendasi penting bagi peneliti baru yang ingin menerapkan teknologi deep learning dalam bidang botani, khususnya identifikasi tanaman obat. MobileNetV2 terbukti efisien dan akurat, menjadikannya pilihan yang direkomendasikan untuk penelitian lanjutan di bidang ini.
(Nagaraju & Chawla, 2020)	CNN (PLDPNet)	98.66%	Penelitian ini merekomendasikan menggunakan metode CNN (PLDPNet) dengan aumentasi tambahan Mosaik dan MixUp, segmentasi otomatis menggunakan U-Net

(Arshad et al., 2023)	VGG19 ResNet50	95.01% 67.01%	Penelitian ini merekomendasikan metode dengan 2 arsitektur VGG19 dan ResNet50 yang dimana hasil akurasi VGG19 lebih baik dari pada ResNet50.
(Nishad et al., 2022)	CNN (ResNet-101)	99%	Penelitian ini merekomendasikan untuk metode CNN (ResNet-101) karna tingkat akurasinya sangat tinggi yaitu Model yang dilatih dengan konfigurasi ini mencapai tingkat akurasi 99%

In this systematic literature review, a total of 8 selected articles were found focusing on plant leaf disease classification using various machine learning and deep learning. The study spans from 2021 to 2024 and covers various methodologies in detecting plant diseases. From the results of the literature read, the highest accuracy is with the title Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease using the Inception Net architecture CNN method with an accuracy value of 99.89%.

3. RESULTS AND DISCUSSION

In this section, we present the results obtained from our systematic literature review and discuss key findings regarding plant disease identification using deep learning techniques. The analysis is based on the selected studies and highlights the effectiveness of different methodologies, datasets, and neural network architectures.

3.1. Accuracy Analysis of Deep Learning Models

The findings from the reviewed literature indicate that deep learning approaches have significantly improved plant disease identification accuracy. The highest accuracy was achieved by the Inception Net-based CNN method, which reached 99.89% in detecting diseases in tomato plants. This demonstrates the potential of deep learning to enhance plant disease management and agricultural productivity.

Table 3 provides a comparative analysis of different deep learning models used for plant disease classification. Based on this comparison, the following insights were obtained:

a. The CNN (ResNet-101) model achieved 99% accuracy, making it one of the most effective architectures for plant disease classification.
b. CNN (PLDPNet) demonstrated high accuracy (98.66%) when combined with augmentation techniques like Mosaic and MixUp, as well as automatic segmentation using U-Net.
c. CNN + Yolo7 reached an accuracy of 98.8%, highlighting its effectiveness in real-time disease detection applications.
d. DSAtt-CMNetV3 yielded 96.8% accuracy, incorporating optimization techniques like Os-OA for improved hyperparameter tuning.
e. The MobileNetV2 model, known for its efficiency, achieved 94.16%, making it a suitable choice for lightweight applications in plant disease classification.

3.2. Dataset Considerations

The dataset used plays a crucial role in model performance. Our study found that:

a. The PlantVillage dataset was the most commonly used source for plant disease images, contributing significantly to training high-accuracy models.
b. Some researchers employed Kaggle datasets, smartphone-acquired images, or datasets collected from internet sources, providing a variety of real-world conditions for model training.
c. The total number of images in datasets varied widely, from 600 images (medicinal plants) to 18,157 images (tomato leaf diseases), impacting model robustness and generalizability.

3.3. Pre-processing and Data Augmentation Techniques

To enhance model performance, various pre-processing and augmentation techniques were applied:

a. Pre-processing techniques such as segmentation, resizing, and normalization were commonly employed to improve image quality
b. Data augmentation methods like rotation, scaling, cropping, flipping, noise addition, and color transformations were applied to increase dataset variability and enhance model generalization.
c. Advanced augmentation methods like WGAN and SRGAN were used to synthesize new images, further improving classification accuracy.

3.4. Challenges in Deep Learning-Based Plant Disease Identification

Despite achieving high accuracy, several challenges remain in the field of deep learning-based plant disease identification:

a. Data Imbalance: Many datasets have an unequal distribution of healthy and diseased samples, which may bias the model's learning process.
b. Real-World Variability: Factors like lighting conditions, image resolution, and different plant species make it difficult to generalize models across different datasets.
c. Computational Requirements: Deep learning models, particularly CNNs, require high computational resources, making them challenging to deploy on edge devices or low-power systems.
d. Lack of Standardized Benchmarks: There is a need for more standardized datasets and evaluation metrics to compare models fairly.

3.5. Implications for Agricultural Practices

The implementation of deep learning in plant disease detection has significant implications for agriculture:

a. Early Detection and Prevention: High-accuracy models enable real-time disease monitoring, helping farmers take preventive actions before major crop losses occur.
b. Automation and Scalability: Automated plant disease identification can reduce reliance on human expertise and improve scalability in large-scale farming.
c. Integration with IoT and Mobile Applications: The integration of deep learning with IoT-based sensors and mobile applications can enhance accessibility for farmers in remote areas.

3.6. Future Research Directions

To further enhance the accuracy and applicability of deep learning models for plant disease identification, future research should focus on:

a. Developing hybrid models that combine multiple architectures (e.g., CNN with transformers) for improved feature extraction.
b. Improving dataset diversity by collecting images from multiple environmental conditions and plant growth stages.
c. Deploying models on edge devices to enable real-time disease monitoring in resource-limited settings.
d. Exploring unsupervised learning techniques to handle cases where labeled data is scarce.

4. CONCLUSIONS

From our literature review, it is apparent that deep learning has become one of the dominant approaches in plant disease identification and classification. The use of complex neural network architectures, such as Inception Net CNN, has proven its superiority in disease detection accuracy. These findings provide important insights for the development of more sophisticated and efficient plant disease detection systems in the future. By continuing to explore the potential of this technology and increasing the availability of quality datasets, we can strengthen efforts to ensure agricultural prosperity and food security around the world.

5. REFERENCES

Arshad, F., Mateen, M., Hayat, S., Wardah, M., Al-Huda, Z., Gu, Y. H., & Al-antari, M. A. (2023). PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction. *Alexandria Engineering Journal*, *78*(June), 406–418. https://doi.org/10.1016/j.aej.2023.07.076

Demilie, W. B. (2024). Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-023-00863-9

Mendalam, P. (2024). *Machine Translated by Google Deteksi Efektif Penyakit Lengkeng Menggunakan*.SSS

Nagaraju, M., & Chawla, P. (2020). Systematic review of deep learning techniques in plant disease detection. *International Journal of System Assurance Engineering and Management*, *11*(3), 547–560. https://doi.org/10.1007/s13198-020-00972-1

Nishad, M. A. R., Mitu, M. A., & Jahan, N. (2022). Predicting and Classifying Potato Leaf Disease using K-means Segmentation Techniques and Deep Learning Networks. *Procedia Computer Science*, *212*(C), 220–229. https://doi.org/10.1016/j.procs.2022.11.006

Nizamani, M. M., Zhang, Q., Muhae-Ud-Din, G., & Wang, Y. (2023). High-throughput sequencing in plant disease management: a comprehensive review of benefits, challenges, and future perspectives. *Phytopathology Research*, *5*(1), 1–17.

Ramanjot, Mittal, U., Wadhawan, A., Singla, J., Jhanjhi, N. Z., Ghoniem, R. M., Ray, S. K., & Abdelmaboud, A. (2023). Plant Disease Detection and Classification: A Systematic Literature Review. *Sensors*, *23*(10). https://doi.org/10.3390/s23104769

Rochaeni, S. (2023). Pembangunan pertanian Indonesia. In *Graha Ilmu* (Nomor 2).

Shafik, W., Tufail, A., Namoun, A., De Silva, L. C., & Apong, R. A. A. H. M. (2023). A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends. *IEEE Access*, *11*(June), 59174–59203. https://doi.org/10.1109/ACCESS.2023.3284760

Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., & Park, S. H. (2022). Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. *Frontiers in Plant Science*, *13*(October), 1–18. https://doi.org/10.3389/fpls.2022.1031748

Zhao, Y., Chen, Z., Gao, X., Song, W., Xiong, Q., Hu, J., & Zhang, Z. (2022). Plant Disease Detection Using Generated Leaves Based on DoubleGAN. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *19*(3), 1817–1826. https://doi.org/10.1109/TCBB.2021.3056683