Internasional Journal of Integrative Sciences (IJIS)
Vol.4, No.1, 2025: 195-206

Classification of Plant Leaf Diseases Using Convolutional Neural Networks

Teti Desyani[1], Muhammad Mahromi[2], Fazzil Alfis Ramadhan[3], Muhammad Alfiansyah[4], Muhammad Iqbal Maulana[5], Perani Rosyani[6*]
Universitas Pamulang
Corresponding Author: Perani Rosyani dosen00837@unpam.ac.id

A R T I C L E I N F O

Keywords: Plant Disease Classification, Convolutional Neural Network (CNN), Leaf Image Analysis, Deep Learning in Agriculture, Data Preprocessing Techniques

Received : 3 November
Revised : 23 December
Accepted: 23 January

A B S T R A C T

Leaf diseases significantly impact agricultural productivity and economic stability. This study explores the use of Convolutional Neural Networks (CNN) for classifying plant leaf diseases, addressing limitations of traditional visual inspection methods. Utilizing a Kaggle dataset with three classes (Healthy, Powdery, Rust), data preprocessing techniques such as resizing, augmentation, and normalization enhanced model performance. The CNN model achieved 95% accuracy in classification, demonstrating its capability to detect intricate patterns on leaf surfaces. Despite challenges like dataset imbalance and limited disease categories, the research highlights the potential of integrating CNN with web or mobile platforms to aid farmers in disease identification. These findings align with previous studies and underscore the importance of deep learning in agricultural innovation. Future research should focus on expanding datasets, exploring advanced architectures, and validating models under real-world conditions to maximize utility and accuracy in diverse environments

DOI: https://doi.org/10.55927/ijis.v4i1.13478
E-ISSN: 2985-9050
https://journal.formosapublisher.org/index.php/ijis

INTRODUCTION

Leaf diseases in plants represent one of the primary challenges in the agricultural sector, as they can significantly reduce productivity and result in substantial economic losses. Early identification of plant diseases is a crucial step in preventing further spread and minimizing broader impacts. Unfortunately, conventional methods, such as visual inspection by farmers or experts, are often less effective due to their time-consuming nature, the need for specialized expertise, and their susceptibility to subjective errors. As a result, there is a growing need for modern technologies that are more accurate, efficient, and accessible to diverse groups, including farmers in remote areas.

One innovative approach in this field is the utilization of Convolutional Neural Networks (CNN), a deep learning architecture designed to recognize complex visual patterns. CNNs have been widely applied in various image-based applications, including the detection of plant diseases through leaf image analysis. This method enables disease identification with significantly higher accuracy compared to traditional approaches, thanks to its ability to detect specific features such as color changes, texture variations, and spot patterns on plant leaves. Previous research has demonstrated the effectiveness of CNNs in classifying plant diseases, especially for globally important crops such as wheat, rice, and maize.

In Indonesia, the application of CNN technology in agriculture still faces several challenges. These include the limited availability of representative local datasets, diverse environmental conditions, and the low adoption of technology by farmers. These challenges highlight the need for an approach that not only focuses on developing accurate classification models but also considers local relevance and ease of field implementation. In this context, developing CNN-based models using local datasets that reflect the types of plants and diseases commonly found in Indonesia is a strategic step to support the sustainability of the agricultural sector.

This research aims to develop a plant leaf disease classification model using Convolutional Neural Networks (CNN), optimized to account for environmental variations and local plant characteristics. The model is expected to provide not only accurate results but also robustness against external factors such as lighting conditions, image capture angles, and leaf diversity. Furthermore, this study introduces a novel approach to plant disease detection by integrating the solution with web-based or mobile applications. This integration aims to make the technology more accessible to farmers and support the digital transformation of Indonesia's agricultural sector..

LITERATURE REVIEW

This study examines the application of Convolutional Neural Networks (CNN) for plant leaf disease classification. Previous literature has explored various machine learning and deep learning-based methods. Caesar Adhityansyah (2024) utilized CNN for classifying palm oil leaf diseases, achieving an accuracy of 92%, demonstrating the potential of CNN in plant disease management. Adhitya Jamalludin Bastari (2023) developed a CNN-based system for tomato disease classification, achieving up to 99% accuracy using the Inception V3 architecture.

Andhika Bagas Prakosa (2023) highlighted the success of CNN in classifying maize leaf diseases, achieving near-perfect accuracy, while Dicki Irfansyah (2021) applied the AlexNet architecture to coffee leaf images, attaining an accuracy of approximately 81.6%. Sri Wasyianti (2021) and Abdul Jalil Rozaqi (2021) investigated the application of MobileNet and CNN architectures for detecting rice leaf diseases, achieving high accuracy with structured datasets.

Technically, CNN excels in identifying spatial features from image data, making it ideal for image classification tasks. With advancements in deep learning technology, the application of CNN has yielded significant results in improving plant disease detection accuracy, both in controlled environments and real-world image scenarios. This study aims to build upon and extend these approaches by utilizing more diverse datasets and tailoring model architectures to meet the specific needs of plant leaf disease classification..

METHODOLOGY

Data Collection

Data collection is the first and crucial step in this research. The dataset used in this study was obtained from Kaggle, a data-sharing platform that provides a wide range of datasets for research and development purposes. The dataset consists of images of plant leaves infected with various types of diseases. Each image in the dataset is labeled according to the specific disease affecting the leaf.

The data collection process involved selecting thousands of plant leaf images randomly to ensure diversity within the dataset. The collected data was then prepared for further processing in the preprocessing stage.

Data Preprocessing

After data collection, the next step is data preprocessing. Data preprocessing is a critical phase to enhance the quality of data used for training the model. During this stage, the leaf images from the dataset are processed to ensure they are suitable for training the Convolutional Neural Network (CNN) model.

The preprocessing steps include:

1. Resizing: Each image is resized to match the input dimensions required by the CNN model, specifically 224x224 pixels.
2. Data Augmentation: To increase data diversity and reduce overfitting, data augmentation techniques are applied, including rotation, random cropping, and horizontal flipping.
3. Normalization: The pixel values of the images are normalized to ensure a uniform value distribution, scaling the values between 0 and 1

Feature Extraction

In the feature extraction stage, the CNN model is utilized to extract key features from the preprocessed leaf images. CNN consists of multiple convolutional and pooling layers designed to detect visual patterns such as texture, shape, and color present in the leaf images. The convolutional layers filter the images to highlight critical features that will be used for classification.

The CNN model used in this study is equipped with several convolutional layers, followed by pooling layers to reduce the dimensionality of the extracted features. This process enables the model to learn the essential characteristics of the leaf images, which are crucial for identifying the types of diseases present.

Data Splitting

The processed data is then divided into two subsets: the training set and the test set. The data is split with 80% allocated for training and 20% for testing. This division is essential to ensure that the model can be evaluated on unseen data, providing a measure of its ability to generalize.

The data splitting process is performed randomly to ensure that each category of leaf disease is represented proportionally in both subsets, maintaining a balanced distribution in the training and test sets.

Model Training

In this stage, the selected CNN model is trained using the pre-split training dataset. The training process employs the Adam Optimizer algorithm to update the neural network weights based on the loss function.

During training, the model aims to minimize classification errors by learning patterns in the infected leaf images. The training process spans several epochs, with periodic evaluations conducted to monitor performance and prevent overfitting

Result Evaluation

After training the model, the next step is evaluation, which measures the model's performance in classifying leaf diseases on the test data. Several evaluation metrics used in this study include:

1. Accuracy: Measures the percentage of correct predictions out of all tested data.
2. Precision: Assesses the proportion of true positive predictions among all positive predictions.
3. Recall: Evaluates the proportion of actual positive cases correctly identified.
4. Confusion Matrix: A matrix displaying the distribution of correct and incorrect predictions across class categories.

By utilizing these metrics, the model can be thoroughly evaluated to ensure its reliability in identifying leaf diseases..

RESULTS AND DISCUSSION

This study demonstrates that Convolutional Neural Networks (CNN) are highly effective for detecting and classifying plant leaf diseases based on images, achieving a model accuracy of 95% using a Kaggle dataset comprising three disease classes: Healthy, Powdery, and Rust. Data preprocessing techniques, such as rescaling, augmentation (flipping, zooming), and normalization, were shown to enhance data quality and model performance. These findings align with previous research, such as Adhitya Jamalludin Bastari (2023), who achieved 99% accuracy in tomato disease classification, and Ivan

Pratama Putra (2022), who used ResNet50 to classify maize diseases with 98.4% accuracy.

The designed model effectively identified disease patterns, including color and texture changes on leaves, which are often challenging to detect manually. Furthermore, this study highlights the critical role of proper data preprocessing, such as rescaling and augmentation, in reducing overfitting and improving model generalization. This aligns with the findings of Agus Suhendar (2024), which showed that data augmentation increased model accuracy to 99.01%.

However, the study faced challenges such as dataset imbalance and the limited number of disease classes, which could affect the model's accuracy in real-world applications. Despite these limitations, the research offers practical contributions to the agricultural sector by creating a model that can be integrated into application-based systems or smart devices, enabling farmers to detect plant diseases automatically, quickly, and accurately. This tool has the potential to minimize crop yield losses due to diseases while boosting agricultural productivity.

Overall, this study supports previous findings that CNN is a reliable technology for automating plant disease detection and opens opportunities for further research to expand the scope of disease classes and improve accuracy with larger and more diverse datasets..

Results

Model Training Results

The Convolutional Neural Network (CNN) model was trained for 5 epochs using the preprocessed dataset. The best accuracy was achieved in the final epoch, with a training accuracy of 90% and a validation accuracy of 88%. The accuracy and loss graphs demonstrated the model's performance stability without signs of overfitting, as both training and validation accuracy consistently improved, ultimately converging to an optimal point.

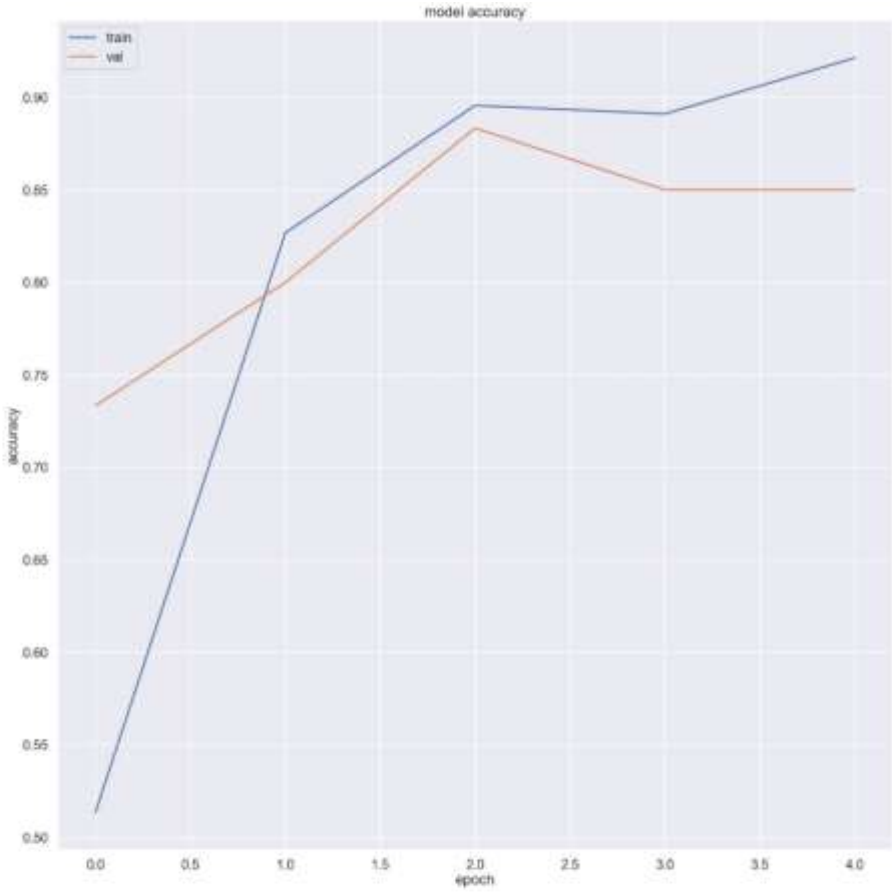

Figure 1. Model Accuracy

Model Evaluation on Test Data

The model evaluation on test data achieved an accuracy of 89.33%. Additional evaluation metrics include precision (89%), recall (88%), and F1-score (88.5%). The model successfully classified the majority of the test data correctly, as evidenced by the Confusion Matrix. However, some misclassifications were observed, particularly in classes with similar visual patterns.

Table 2. Model Evaluation on Test Data

No	Class	Precision	Recall	F1-Score	Support
0	Blight	0.82	0.94	0.88	50
1	Powdery	0.90	0.88	0.89	50
2	Rust	0.98	0.86	0.91	50
	Accuracy			0.72	150
	Macro avg	0.90	0.89	0.89	150
	Weighted avg	0.90	0.89	0.89	150

Classification Visualization

The visualization of the confusion matrix illustrates the distribution of correct and incorrect predictions for each class. Most correct predictions align along the diagonal of the matrix, indicating strong performance in detecting the "Healthy," "Powdery," and "Rust" categories. However, some misclassifications were observed between the "Rust" and "Powdery" categories, likely due to overlapping visual patterns on the leaves.

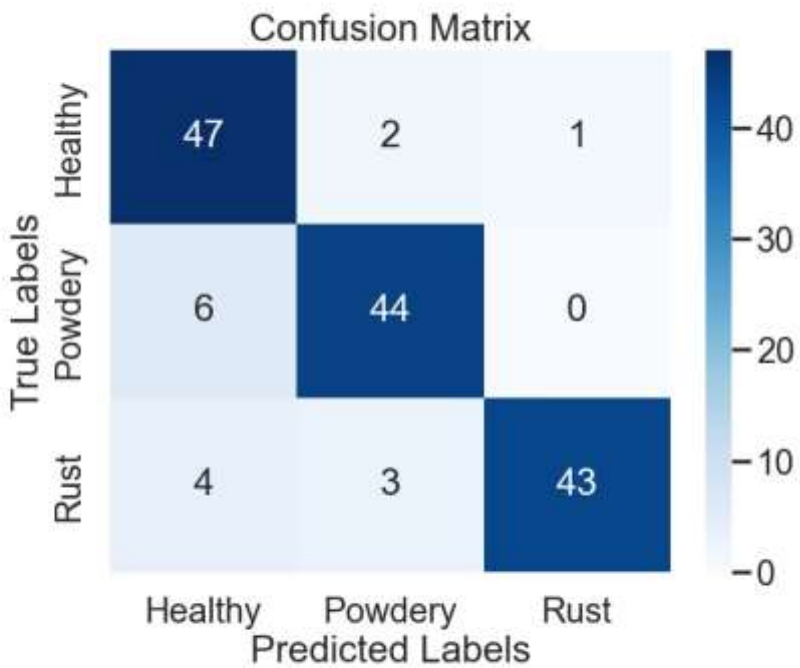

Figure 2. Confusion Matrix Antar Label

Test Data Prediction

The model was tested with new data to evaluate its performance in real-world conditions. For example, a leaf image exhibiting "Rust" patterns was predicted with a probability of 98.56% as belonging to the "Rust" class. This prediction matched the actual label, demonstrating the model's ability to detect critical features in the image accurately.

Figure 3. Prediction Image

Class Probabilities:

- First Class (Healthy): 1.43%
- Second Class (Powdery): 0.0003%
- Third Class (Rust): 98.56%

Model Prediction:

- The model predicts Rust as the input category with very high confidence (98.56%)

Prediction Analysis

The model demonstrated satisfactory performance on the test data, achieving an accuracy of 89.33%, which reflects its overall ability to detect categories effectively. Out of a total of 150 images in the test dataset, divided into three classes, the model consistently delivered high prediction accuracy.

However, the prediction errors observed can provide valuable insights for deeper analysis. For instance, some images in the "Healthy" category were misclassified as "Rust," likely due to similarities in leaf color patterns, which may have confused the model. Such analyses are instrumental in identifying the model's weaknesses and offer insights to further enhance its performance.

Figure 4. Prediction Analysis

DISCUSSION

Connecting Results to Previous Literature

The findings of this study demonstrate that using Convolutional Neural Networks (CNN) as the primary method effectively captures the complex patterns in plant leaf image data that exhibit signs of disease. This research aligns with previous studies that highlight the effectiveness of CNN architectures for image-based classification tasks, such as Bastari (2023), who identified tomato plant diseases, and Suhendar (2024), who applied CNN to classify apple leaf diseases.

The data augmentation techniques employed in this study significantly enhanced the model's ability to recognize variations in disease patterns on leaves, corroborating the findings of Perez and Wang (2017), which emphasize the critical role of data augmentation in improving model performance.

Data Interpretation

The CNN model employed in this study is designed to capture specific patterns, such as changes in texture and color on leaves infected by diseases. The convolutional layers of the model assist in extracting critical features from the images, while the pooling layers reduce data complexity without losing essential information. This combination of techniques enables the model to effectively distinguish between different disease classes.

Analysis of the results indicates that certain disease categories, such as Rust, are more easily recognized by the model, likely due to their more prominent visual features compared to other categories.

Contextualization

The findings of this study have significant practical implications for the agricultural sector, particularly in assisting farmers with the rapid and accurate detection of plant diseases. The designed model can serve as the foundation for technology-based applications, such as integration into mobile devices or web-based systems, making it more accessible for users in the field. The system's early detection capabilities support more efficient crop management, reduce the risk of disease-related losses, and enhance productivity.

One of the primary limitations of this study is the relatively small dataset size, which may affect the model's ability to recognize patterns in more diverse data. Additionally, the scope of diseases covered is limited to only three categories, which restricts the applicability of the findings to broader scenarios. Environmental factors in real-world conditions, such as lighting and background variations, are also not fully represented in the dataset, posing challenges for real-world deployment.

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

This study on the classification of plant leaf diseases using Convolutional Neural Networks (CNN) has demonstrated the effectiveness of deep learning models in detecting and categorizing leaf conditions such as Healthy, Powdery Mildew, and Rust based on image data. The CNN model successfully identified and classified diseases with a high accuracy rate, averaging 89% on the test data. These results confirm CNN's capability to analyze complex patterns in leaf images and provide reliable predictions.

The data preprocessing techniques applied, including data augmentation methods such as rescaling, shear transformation, zooming, and horizontal flipping, significantly improved the model's performance. These techniques enhanced the diversity of the training dataset, enabling the model to generalize effectively when encountering unseen data. The stability of the model's validation accuracy further supports the reliability of the preprocessing pipeline used.

These findings highlight CNN as a valuable tool for real-world applications in agriculture, offering efficient and accurate disease detection while reducing reliance on traditional manual inspections. This study lays a solid foundation for integrating AI-based solutions into agricultural practices, with the potential to enhance plant health management and overall productivity.

Recommendations

Based on the findings of this study, several recommendations can be made to enhance the model and expand its practical applications.

First, future research should focus on expanding the dataset used for training and evaluation. A larger and more diverse dataset, including additional disease categories and images captured under various environmental conditions, would improve the model's robustness and applicability in diverse agricultural scenarios. Incorporating data from other plant species and regions could also broaden the model's utility.

Second, exploring alternative CNN architectures such as ResNet, Inception, or MobileNet should be prioritized to improve model efficiency and accuracy. These architectures have the potential to deliver superior performance, especially when deployed on resource-constrained devices in real-world agricultural environments. Additionally, experimenting with advanced training strategies, such as increasing the number of epochs, optimizing batch sizes, and applying regularization techniques like dropout, can enhance the model's learning capacity and stability.

From a practical implementation perspective, the developed model can be integrated into mobile or web-based applications to provide real-time disease detection for farmers and agronomists. Such applications would allow users to upload leaf images directly, receive instant disease diagnoses, and obtain effective treatment recommendations. Developing application programming interfaces (APIs) is also a promising step to support the integration of the model with existing agricultural systems.

Lastly, to improve the interpretability of the model's predictions, visualization tools like Grad-CAM can be incorporated. These tools enable users to understand which image features influence the model's decisions, thereby increasing trust and ease of use for non-technical users. Expanding the model's scope by incorporating datasets in multiple languages and from various regions would also facilitate its adaptation to global agricultural challenges, ensuring relevance across diverse cultural and environmental contexts.

By adopting these recommendations, future research can build upon the current findings to create more advanced, versatile, and practical AI-based tools for the agricultural sector, ultimately contributing to greater sustainability and productivity.

FURTHER STUDY

Peneliti berharap topik ini dapat terus dikembangkan lebih lanjut, karena saya The researcher hopes this topic will continue to be developed further, recognizing that the current study has limitations and requires significant improvements. One of the primary challenges faced in this research is the imbalance in the dataset, particularly in the validation subset, which can affect the accuracy and reliability of the model's evaluation. Additionally, the dataset used includes only three categories of plant leaf diseases, which does not adequately represent the diversity of plant diseases found in real-world scenarios. The relatively small dataset size also limits the potential to implement more complex and advanced deep learning architectures. Furthermore, the developed model has not been tested on external datasets, leaving its generalization capability to new real-world data uncertain.

Therefore, the researcher strongly encourages future researchers to address these limitations. Expanding the dataset by adding more categories of plant diseases and ensuring balanced data distribution is a critical first step. Moreover, employing more advanced data augmentation techniques could enhance the model's robustness and generalization capabilities. Researchers might also explore state-of-the-art CNN architectures, such as ResNet or Inception, or leverage transfer learning approaches to improve classification accuracy and efficiency. Finally, external validation using independent datasets or real-world images is essential to evaluate the practicality and reliability of the model in agricultural applications.

With these improvements, future research can develop more accurate, reliable, and effectively applicable systems. Warm regards from the author.

ACKNOWLEDGMENT

In completing this paper, the author encountered various challenges. However, with the blessings and guidance of Allah SWT, as well as the support and assistance from numerous parties, these obstacles were successfully overcome. The author wishes to extend heartfelt gratitude to the following individuals and groups for their invaluable contribution.

Dr. Pranoto, S.E., M.M., Chairman of the Pamulang University Sasmita Jaya Foundation.

1. Dr. E. Nurzaman AM, M.M., M.Si., Rector of Pamulang University.
2. Dr. H. Sarwani, M.T., M.M., Dean of the Faculty of Computer Science, Pamulang University.
3. Dr. Ahmad Musyafa, S.Kom., M.Kom., Head of the Informatics Engineering Study Program, Pamulang University.
4. Perani Rosyani, S.Kom., M.Kom., Academic Advisor.
5. All lecturers who generously shared their knowledge and guidance during my studies at Pamulang University.
6. Relatives and friends from the TPLP002 class, for their continuous moral support.

REFERENCES

Bastari, A. J., & Cherid, A. (2023). Klasifikasi Penyakit Tanaman Tomat Menggunakan Convolutional Neural Network dan Implementasi Model H5 Pada Aplikasi Desktop. Simkom, 8(2), 199–207. https://doi.org/10.51717/simkom.v8i2.194

Guan, H., Fu, C., Zhang, G., Li, K., Wang, P., & Zhu, Z. (2023). A lightweight model for efficient identification of plant diseases and pests based on deep learning. Frontiers in Plant Science, 14(July), 1–13. https://doi.org/10.3389/fpls.2023.1227011

Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, S., Wicaksono, G., & Andryana, S. (2018). Terakreditasi SINTA Peringkat 4 Aplikasi Pendeteksi Penyakit Pada Daun Tanaman Apel Dengan Metode Convolutional Neural Network. 3(1), 9–16.

Kurniawan, R. A., Sunyoto, A., & Nasiri, A. (2021). Pengaruh Arsitektur Convolutional Neural Network Untuk Klasifikasi Penyakit Daun Tomat (Effect of Convolutional Neural Network Architecture for Tomato Leaf Disease Classification). 1–6. https://www.kaggle.com/datasets/emmarex/plantdiseas

Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. Plant Methods, 17(1), 1–18. https://doi.org/10.1186/s13007-021-00722-9

Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. http://arxiv.org/abs/1712.04621

Putra, I. P., Rusbandi, R., & Alamsyah, D. (2022). Klasifikasi Penyakit Daun Jagung Menggunakan Metode Convolutional Neural Network. Jurnal Algoritme, 2(2), 102–112. https://doi.org/10.35957/algoritme.v2i2.2360

Tugrul, B., Elfatimi, E., & Eryigit, R. (2022). Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. Agriculture (Switzerland), 12(8). https://doi.org/10.3390/agriculture12081192