Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 6 (2024) 799 - 805 e-ISSN: 2580-0760

Utilization of the Convolutional Neural Network Method for Detecting Banana Leaf Disease

Nita Helmawati[1*], Ema Utami[2]
[1,2]Magister of Informatics, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
[1]nita@students.amikom.ac.id, [2]ema.u@amikom.ac.id

Abstract

Banana leaf diseases such as Sigatoka, Cordana, and Pestalotiopsis pose a significant threat to banana productivity, with implications for food security and the global economy. Early detection of this disease is an important step to reduce its spread and maintain crop yield stability. This research utilizes the Convolutional Neural Network (CNN) method to detect banana leaf diseases based on image analysis of infected and healthy leaves. The dataset used includes 937 images consisting of four main categories, namely healthy leaves, Sigatoka, Cordana, and Pestalotiopsis. The dataset is processed through augmentation to increase data diversity and quality. The CNN model was applied for classification, with evaluation results reaching 92.85% accuracy, 95.73% recall, 93.52% precision, and 94.60% F1-score. This research contributes to the development of Artificial Intelligence-based technology for applications in the agricultural sector, especially in supporting farmers to detect banana leaf diseases quickly, accurately and efficiently. The research results also provide recommendations for exploring additional data augmentation and increasing dataset variety to improve model detection performance in the future. This shows CNN's potential to support sustainable agriculture in the modern era.

Keywords: *agricultural sector; artificial Intelligence; Banana leaf disease; CNN Model; technology development*

How to Cite: Nita Helmawati and Ema Utami, "Utilization of the Convolutional Neural Network Method for Detecting Banana Leaf Disease", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 8, no. 6, pp. 799 - 804, Dec. 2024.
DOI: https://doi.org/10.29207/resti.v8i6.6140

1. Introduction

Banana leaf disease is one of the most serious challenges faced by banana farmers in Indonesia. This disease not only has an impact on reducing the productivity of agricultural products but also has a significant economic impact on farmers' incomes, which ultimately affects global food security. With the high international market demand for bananas, the importance of early disease detection becomes increasingly crucial. Fast and accurate detection can help reduce the spread of infection, increase crop yields, and maintain production stability. Therefore, the development of real-time banana leaf disease detection tools is the main agenda in the modernization of the agricultural sector.

The traditional approach to detecting banana leaf diseases is usually done through visual observation by farmers who rely on personal experience [1]**.** However, this method tends to be less accurate and slow due to limited visual abilities and in-depth knowledge of the early signs of disease and this is where technology plays an important role [2]. With rapid progress in the field of Artificial Intelligence, opportunities to utilize machine learning methods such as CNN are increasingly wide open.

The CNN method was introduced by Yann LeCun in 1998 which offers a more efficient approach [3]. CNNs are able to extract features directly from raw images through convolution layers without requiring manual feature extraction [4]. This allows CNNs to recognize patterns in images automatically, making them very suitable for plant disease detection.

Previous research has tried to overcome this problem by using machine learning technology carried out by [5] uses a hybrid CNN-SVM model to detect disease Cordana leaf spot And Black Sigatoka with 90% accuracy, but in this study no search was carried out for recall, precision and F1-Score. Meanwhile, research conducted [6] achieved 84.73% accuracy, 84.80% precision, 84.73% recall, and 84.62% F1-Score. Although these results are quite good, there is still room for improvement, especially in increasing the level of

Received: 28-10-2024 | Accepted: 13-12-2024 | Published Online: 28-12-2024

accuracy and detection speed. By utilizing the CNN method with more advanced image processing techniques, it is hoped that more optimal results can be obtained, with higher accuracy and faster detection times. This research aims to overcome these shortcomings and contribute to increasing the efficiency of banana leaf disease detection so that it can help farmers identify diseases more precisely and quickly. As a result, agricultural productivity increases and economic stability and global food security can be well maintained.

2. Research Methods

Figure 1 is an overview of the research methods used to utilize the CNN method in detecting diseases on banana leaves.

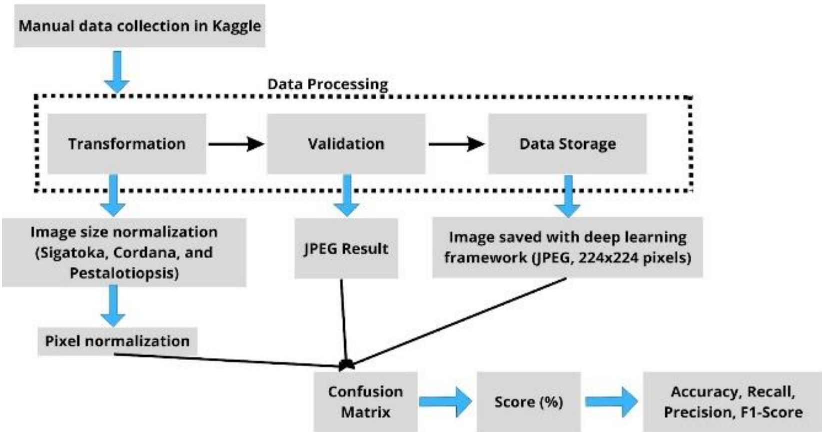


Figure 1. Research Methods

Data collection was carried out manually via the Kaggle platform, focusing on a dataset of images of banana leaves infected with various diseases. This dataset includes several disease categories, such as Sigatoka, Cordana, and Pestalotiopsis. Once downloaded, the data is saved in a suitable format for further analysis and model development.

The data processing stage involves several important steps. The first step is data transformation, which includes normalizing image sizes and pixel values to ensure uniformity. Next, validation is carried out to ensure the integrity and quality of the data before it is stored in a structured storage that is easy to access. The data is then saved in JPEG format for easy access and management. This structured storage allows data to be easily retrieved, processed, and reused at various stages of model development. With this approach, the processed data is ready to be used effectively to build more accurate and efficient models.

After the data has been processed, the CNN method is implemented simultaneously for banana leaf disease classification. In this process, method evaluation is carried out using several metrics, namely accuracy, recall, precision, and F1 Score, with the results calculated in the form of a score (%) [7]. This assessment aims to measure the model's performance in detecting and classifying banana leaf diseases accurately. Score calculations are carried out using Equation 1.

$$Accuracy = \frac{Total\ True\ Positives}{Total\ Dataset} \quad (1)$$

In Equation 1, accuracy is calculated by dividing the number of True Positives (TP) by the total dataset. TP is the number of data that is actually positive and successfully predicted positive by the model [8].

$$Recall = \frac{True\ Positives}{\text{True Positives + False Negatives}} \quad (2)$$

Recall is calculated in Equation 2 by dividing the number of TP by the total number of TP plus False Negatives (FN), where FN is the number of positive data that should have been predicted as positive but was predicted as negative [9].

$$Precision = \frac{True\ Positives}{\text{False Positives}} \quad (3)$$

In Equation 3, TP is the number of data that are TP and correctly predicted as positive by the model. Meanwhile, FP is the number of negative data that are incorrectly predicted as positive by the model [10].

$$F1 - Score = 2x\ \frac{\text{Precision x Recall}}{\text{Precision+Recall}} \quad (4)$$

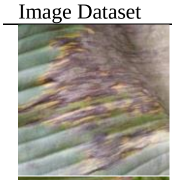
Precision in Equation 4 is calculated by dividing the number of TP by the total of TP plus False Positives (FP). FP is the number of negative data that is incorrectly predicted as positive by the model [11]. Precision measures the accuracy of the model's positive predictions, namely the proportion of correct positive predictions (TP) compared to all positive predictions (TP + FP). Meanwhile, Recall measures the model's ability to detect all existing positive examples, namely the proportion of TP compared to all positive examples that should exist (TP + FN) [12].

3. Results and Discussions

The result of this research is the development of a method for classifying banana leaf diseases using a dataset of images infected with Sigatoka, Cordana, and Pestalotiopsis diseases. This process includes data collection and implementation of the CNN method to increase detection accuracy.

This research uses a dataset consisting of images of banana leaves, specifically focusing on leaves infected with various diseases, namely Sigatoka, Cordana, and Pestalotiopsis, as well as healthy banana leaves. This dataset is publicly available and sourced from the Kaggle website. The images were taken in various real-world conditions using three different cell phone cameras [13]. This dataset provides a valuable resource for developing and testing machine learning models aimed at early detection and classification of banana leaf diseases. The composition of the data can be seen in Table 1.

Table 1. Dataset

Image Dataset	Types of Disease	Number of Datasets
	Sigatoka	473
	Cordana	162
	Pestalotiopsis	173
	Healthy	129

In this research, a verification process was carried out to ensure that all images used were not damaged or corrupted, and had been confirmed to have labels appropriate to the disease or health condition category. Thus, the processed dataset is completely the result of original data collection without modification, which directly reflects the image quality in field conditions[5]. Below is the image distribution in the banana leaf dataset which can be seen in Figure 2.

The following are the results of the data transformation process to ensure consistency and quality of the dataset:

Image size normalization: All images of banana leaves, both healthy and disease-infected (Sigatoka, Cordana, and Pestalotiopsis), were converted to uniform dimensions of 224x224 pixels. This step is carried out so that the image meets the input requirements of the deep learning model that will be used. Image size normalization also helps reduce dimensional variations that can affect model results.

Pixel value normalization: Each pixel in the image is normalized by dividing its maximum value by 255 so that the pixel value range is between 0 and 1. This aims to speed up the model training process because normalized values make the computing process more stable and efficient.

The process of normalizing a banana leaf image which is resized to 224x224 pixels can be seen in Figure 3

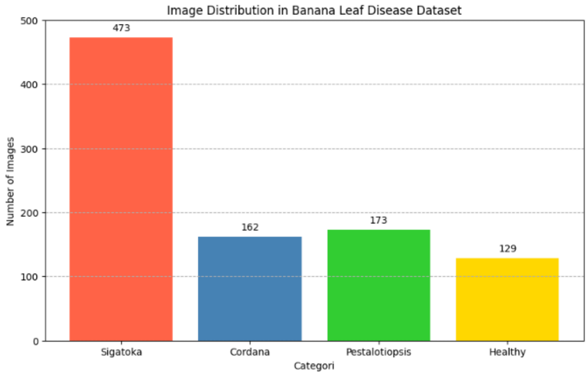


Figure 2. Data Processing

.

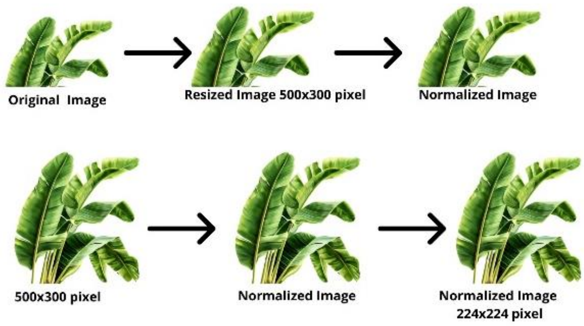


Figure 3. Image Normalization

Image format and accessibility checks have been checked to ensure that files open properly and are JPEG compliant. Next, validation is carried out which includes checking that each image has the correct label, namely the categories of healthy banana leaves and leaves infected with Sigatoka, Cordana, and Pestalotiopsis diseases [14]. Each image is checked to match the correct label so that no errors occur in the model training process.

The results of the data storage stage in processed datasets including images that have gone through cleaning, transformation and validation are stored in a structured manner and are easily accessible for model training purposes in the next stage. The images that have been selected and validated are stored in a directory organized by category, such as healthy, Sigatoka, Cordana and Pestalotiopsis which makes it easier to retrieve data when training the model [15]. All images are saved in a format compatible with the deep learning framework, namely JPEG format with normalized dimensions of 224x224 pixels. In addition, the dataset is stored in local storage with clear naming from sequence number 0 onwards to ensure that the dataset can be quickly accessed again and easily identified during the model development process or further analysis [16]. With this storage step, the dataset can be managed efficiently and ready to be used in banana leaf disease detection experiments.

After storing the data, this stage will carry out the implementation of the CNN method in classifying banana leaf diseases and there are evaluation matrices such as accuracy, precision, recall, and F1-score. This research uses a CNN method approach which provides advantages in increasing the accuracy of banana leaf disease detection [11], [17]. Table 2 describes the results of the classification of each category of disease on banana leaves.

Table 2. Confusion Matrix

Banana Leaf	TP	FP	FN	Datasets
Sigatoka	450	23	20	473
Cordana	140	15	7	162
Pestalotiopsis	160	13	3	173
Healthy	120	9	9	129
Total	870	60	39	937

The next step is to carry out the method classification process:

Total dataset $= 450 + 140 + 160 + 120 + 23 + 15 + 13 + 9 + 20 + 7 + 3 + 9 = 937$

Calculation of evaluation metrics such as accuracy, recall, precision, and F1-score is important in the classification process because each metric provides a different perspective on model performance.

Accuracy is the most commonly used metric in classification, as it shows how often a model makes correct predictions overall [18]. Accuracy is calculated as the ratio between the number of correct predictions to the total number of predictions made by the model as shown in Equation 1.

$$Accuracy = 450 + 140 + 160 + 120937 = 870937 = 0.9285$$

$$Accuracy = 92.85\% \quad (1)$$

Study [19] get 90% accuracy by using an augmented set with 399 images for each category of banana leaf disease. This accuracy result is 92.85% higher compared to this study. This indicates performance improvements caused by the use of augmentation techniques or differences in the classification methods used.

Recall as in Equation 2 is a metric for minimizing FN, namely cases where the model detects TP samples [19]. Recall shows how well the model is able to capture all positive cases in each category. Recall is relevant in scenarios such as banana leaf disease detection to ensure that the model does not miss existing diseases. If the model misses the true picture of Sigatoka disease (FN), it could lead to wrong actions in disease control. Therefore, recall provides important information about the model's ability to accurately detect each disease category.

$$Recall\ Sigatoka = \frac{450}{450+20} = 0.9573 \quad (2)$$

$$Recall\ Cordana = \frac{140}{140+7} = 0.9524$$

$$Recall\ Pestalotiopsis = \frac{160}{160+3} = 0.9815$$

$$Recall\ Healthy = \frac{120}{120+9} = 0.9302$$

$Calculating\ average\ recall$:

$$= \frac{(0.9573\ x\ 473)+(0.9524\ x\ 162)+(0.9815\ x\ 173)+(0.9302\ x\ 129)}{937} = 0.9573$$

Research recall [20] higher than in this study. The Gabor Extraction and RCNN methods used show strong performance in detecting various diseases on banana leaves.

Precision measures the proportion of correct positive predictions, that is, how accurate the model is in classifying a sample as positive as shown in Equation 3. This is especially useful for minimizing False Positives (FP), where the model incorrectly classifies negative samples as positive. In the context of banana leaf disease detection, precision is important to ensure that when the model detects a disease, it is most likely actually a disease. If precision is low, the model can produce too many FP, which means the model detects disease on leaves that are actually healthy. Therefore, precision is very important in ensuring the quality of positive predictions.

$$\text{Precision Sigatoka} = \frac{450}{450+23} = 0.9519 \quad (3)$$

$$\text{Precision Cordana} = \frac{140}{140+15} = 0.9032$$

$$\text{Precision Pestalotiopsis} = \frac{160}{160+13} = 0.9252$$

$$\text{Precision Healthy} = \frac{120}{120+9} = 0.9302$$

$Calculating\ average\ precision$:

$$= \frac{(0.9519 \text{ x } 473)+(0.9032 \text{ x } 162)+(0.9252 \text{ x } 173)+(0.9302 \text{ x } 129)}{937}$$

$$= 0.9352$$

Research result [21] has lower accuracy, precision, recall and F1-score compared to the results of this study, which indicates that the model used is more efficient in detecting disease compared to the DenseNet and Inception methods applied in this study.

F1-Score is the harmonic average of precision and recall. This metric provides a balance between the two, especially in situations where there is an imbalance between precision and recall [22]. F1-score is important because models have high precision but low recall (or vice versa). F1-score is important for the case of banana leaf disease classification to ensure that the model is not only good at detecting the disease (recall) but also precise in its precision [23]. By using the F1-score, you can see how well the balance between precision and recall is, and this metric provides a more holistic picture of model performance as shown in Equation 4.

$$F1 - Score\ Sigatoka = 2 \times \frac{0.9348 \text{ x } 0.9573}{0.9348 + 0.9573} = 0.9547 \quad (4)$$

$$F1 - Score\ Cordana = 2x \frac{0.9260 \text{ x } 0.9275}{0.9260 + 0.9275} = 0.9275$$

$$F1 - Score\ Pestalotiopsis = 2x \frac{0.9418 \text{ x } 0.9526}{0.9418 + 0.9526}$$

$$= 0.9526$$

$$F1 - Score\ Healthy = 2x \frac{0.9250 \text{ x } 0.9350}{0.9250 + 0.9350} = 0.9302$$

$$Calculate\ the\ average\ F1\ Score:$$

$$F1S = \frac{(0.9547x473)+(0.9275x162)+(0.9526x173)+(0.9302x129)}{937} = 0.9460$$

Accuracy and F1-score of research [20] higher than the results of this study which shows that the Gabor Extraction and RCNN-based approach succeeded in significantly improving disease detection performance.

Figure 4 presents the performance of the method in classification based on four evaluation matrices. Each bar on the graph represents the value of each matrix, with the y-axis showing the score or value of that metric in percentage form [24]. The model has an accuracy value of 92.85%, recall 95.73%, precision 93.52% and F1-score 94.60% which makes this model effective in detecting a high level of accuracy.

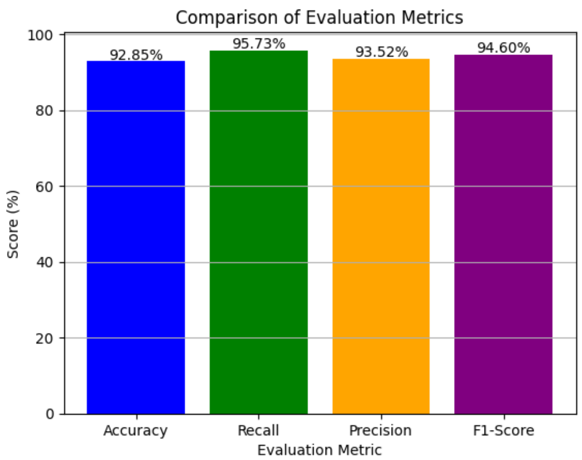


Figure 4. Comparison of Evaluation Metrics

This research succeeded in showing excellent results in classification using deep learning methods. Compared with previous research, as in [6] that uses model architecture ResNet50, ResNet101, And DenseNet201 with a superior precision value of 93%, finally this research shows advantages in the aspects of accuracy and F1-score. The CNN model developed in this research achieved an accuracy of 92.85%, higher than previous research with an accuracy of 90%. Apart from that, this model also produces an F1-score of 94.60%, slightly superior to previous research with an F1-score of 94%. Although there are small differences in precision and F1-score values, the results of this study prove that the CNN method is very effective for classification, with significant advantages in accuracy and F1-score. These results confirm the potential of CNNs in improving the performance of classification models and making a positive contribution to research in the field of deep learning.

4. Conclusions

The conclusion of this research shows that the CNN method produces a banana leaf disease classification model with an accuracy of 92.85%, recall of 95.73%, precision of 93.52%, and F1-score of 94.60%. This model proved effective in detecting and classifying banana leaf diseases, such as Sigatoka, Cordana, and Pestalotiopsis, as well as healthy leaves. The results of this research have great potential in agricultural applications to help farmers detect diseases quickly and accurately, which can increase agricultural yields and reduce losses.

References

[1] M. George, K. Anita Cherian, and D. Mathew, "Symptomatology of Sigatoka leaf spot disease in banana landraces and identification of its pathogen as Mycosphaerella eumusae," *Journal of the Saudi Society of Agricultural Sciences*, vol. 21, no. 4, pp. 278–287, May 2022, doi: 10.1016/j.jssas.2021.09.004.

[2] D. J. Maulana, S. Saadah, and P. E. Yunanto, "54-61 Data in Classifying Financial Distress Companies using SVM and Naïve Bayes," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 1, pp. 54–61, 2024, doi: 10.29207/resti.v8i1.5150.

[3] Y. Fonseca, C. Bautista, C. Pardo-Beainy, and C. Parra, "A plum selection system that uses a multi-class Convolutional Neural Network (CNN)," *J Agric Food Res*, vol. 14, Dec. 2023, doi: 10.1016/j.jafr.2023.100793.

[4] E. Correa, M. Garcia, G. Grosso, J. Huamantoma, and W. Ipanaque, "Design and implementation of a CNN architecture to classify images of banana leaves with diseases," in *2021 IEEE International Conference on Automation/24th Congress of the Chilean Association of Automatic Control, ICA-ACCA 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/ICAACCA51523.2021.9465178.

[5] V. Tanwar, B. Sharma, and V. Aanand, "Implementing a Hybrid CNN-SVM Model for Banana Leaf Disease Classification," in *2023 IEEE International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering, RMKMATE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/RMKMATE59243.2023.10369215.

[6] D. Rustandi, Sony Hartono Wijaya, Mushthofa, and Ratih Damayanti, "Anatomy Identification of Bamboo Stems with The Convolutional Neural Networks (CNN) Method," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 62–71, Feb. 2024, doi: 10.29207/resti.v8i1.5370.

[7] T. Thorat, B. K. Patle, and S. K. Kashyap, "Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming," *Smart Agricultural Technology*, vol. 3, Feb. 2023, doi: 10.1016/j.atech.2022.100114.

[8] T. Phattaraworamet, S. Sangsuriyun, P. Kutchomsri, and S. Chokphoemphun, "Image classification of lotus in Nong Han Chaloem Phrakiat Lotus Park using convolutional neural networks," *Artificial Intelligence in Agriculture*, vol. 11, pp. 23–33, Mar. 2024, doi: 10.1016/j.aiia.2023.12.003.

[9] M Mesran, Sitti Rachmawati Yahya, Fifto Nugroho, and Agus Perdana Windarto, "Investigating the Impact of ReLU

and Sigmoid Activation Functions on Animal Classification Using CNN Models," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 111–118, Feb. 2024, doi: 10.29207/resti.v8i1.5367.

[10] M. A. B. Bhuiyan, H. M. Abdullah, S. E. Arman, S. Saminur Rahman, and K. Al Mahmud, "BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases," *Smart Agricultural Technology*, vol. 4, Aug. 2023, doi: 10.1016/j.atech.2023.100214.

[11] Tejaswini, P. Rastogi, S. Dua, Manikanta, and V. Dagar, "Early Disease Detection in Plants using CNN," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 3468–3478. doi: 10.1016/j.procs.2024.04.327.

[12] M. Gomez Selvaraj *et al.*, "Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 110–124, Nov. 2020, doi: 10.1016/j.isprsjprs.2020.08.025.

[13] Md. A. B. B. H. M. A. S. I. T. T. C. Md. A. H. Shifat E. Arman, "BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning," 2024, doi: 10.1016.

[14] J. Deng, W. Huang, G. Zhou, Y. Hu, L. Li, and Y. Wang, "Identification of banana leaf disease based on KVA and GR-ARNet," *J Integr Agric*, vol. 23, no. 10, pp. 3554–3575, Oct. 2024, doi: 10.1016/j.jia.2023.11.037.

[15] Md. A. H. M. H. Md. M. I. G. M. S. H. Md Ripon Sheikh, "BananaSet: A dataset of banana varieties in Bangladesh," *www.elsevier.com/locate/dib*, 2024, doi: 10.17632/35gb4v72dr.4.

[16] S. Shetty and T. R. Mahesh, "SKGDC: Effective Segmentation Based Deep Learning Methodology for Banana Leaf, Fruit, and Stem Disease Prediction," *SN Comput Sci*, vol. 5, no. 6, Aug. 2024, doi: 10.1007/s42979-024-03031-9.

[17] Z. Zheng *et al.*, "An efficient and lightweight banana detection and localization system based on deep CNNs for agricultural robots," *Smart Agricultural Technology*, vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100550.

[18] D. Tribuana, Hazriani, and A. L. Arda, "Image Preprocessing Approaches Toward Better Learning Performance with CNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 1–9, Jan. 2024, doi: 10.29207/resti.v8i1.5417.

[19] A. Prasetyo and E. Utami, "Detection and Classification of Banana Leaf Diseases: Systematic Literature Review," *Telematika*, vol. 17, no. 2, pp. 128–141, Aug. 2024, doi: 10.35671/telematika.v17i2.2809.

[20] K. Seetharaman and T. Mahendran, "Leaf Disease Detection in Banana Plant using Gabor Extraction and Region-Based Convolution Neural Network (RCNN)," *Journal of The Institution of Engineers (India): Series A*, vol. 103, no. 2, pp. 501–507, Jun. 2022, doi: 10.1007/s40030-022-00628-2.

[21] Andreanov Ridhovan, Aries Suharso, and Chaerur Rozikin, "Disease Detection in Banana Leaf Plants using DenseNet and Inception Method," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 710–718, Oct. 2022, doi: 10.29207/resti.v6i5.4202.

[22] K. Seetharaman and T. Mahendran, "Leaf Disease Detection in Banana Plant using Gabor Extraction and Region-Based Convolution Neural Network (RCNN)," *Journal of The Institution of Engineers (India): Series A*, vol. 103, no. 2, pp. 501–507, Jun. 2022, doi: 10.1007/s40030-022-00628-2.

[23] H. A. Nugroho, S. Hasanah, and M. Yusuf, "Seismic Data Quality Analysis Based on Image Recognition Using Convolutional Neural Network," 2022.

[24] A. Kumar, N. Gaur, and A. Nanthaamornphong, "Machine learning RNNs, SVM and NN Algorithm for Massive-MIMO-OTFS 6G Waveform with Rician and Rayleigh channel," *Egyptian Informatics Journal*, vol. 27, Sep. 2024, doi: 10.1016/j.eij.2024.100531.