IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.20, No.1, January 2026, pp. 37~46
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.110529

Leaf Disease Detection Model in Gayo Coffee Plantations Using Deep Learning

Rahmad Hidayat [1*], Rizki Ananda[1], Muhammad Reza Zulman[1]
Hari Toha Hidayat[2], Ilham Safar[2], Riwanul Nasron[2]

[1]Program Studi Teknik Informatika, Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA

[2]Program Studi Teknologi Rekayasa Multimedia, Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA

Corresponding Author: rahmad_hidayat@pnl.ac.id

Abstrak

Kopi merupakan salah satu komoditas perkebunan tropis terpenting yang sangat mendukung dan membantu perekonomian Dataran Tinggi Gayo. Serangan berbagai penyakit menyebabkan penurunan produktifitas maupun kualitas kopi Gayo secara signifikan. Penelitian ini mengembangkan model deteksi penyakit daun pada tanaman kopi memanfaatkan metode Convolutional Neural Network (CNN). Model yang dikembangkan pada penelitian ini menggunakan dua dataset. Dataset pertama yaitu Penyakit Daun Kopi Gayo (PDKG) merupakan data citra daun yang sehat dan terserang penyakit diakuisisi pada perkebunan kopi Gayo. Citra hasil akuisisi dalam dataset PDKG selanjutnya dipreposesing untuk meningkatkan kualitas citranya. Hasil pelatihan dan pengujian model pada dataset PDKG menunjukan akurasi sebesar 0.91. Pada dataset public Coffee Leaf Diseases (CLD) model memperoleh akurasi sebesar 0.95 meningkat sebesar 7,1% dibandingkan dengan studi sebelumnya. Model hasil penelitian ini dapat membantu petani kopi lokal di dataran tinggi Gayo dalam mendeteksi penyakit daun sejak dini serta mengelola kesehatan tanaman secara lebih efisien dan akurat.

Kata kunci*—Deteksi Penyakit Daun, Tanaman Kopi Gayo, CNN*

Abstract

Coffee is one of the most important tropical plantation commodities, significantly supporting the economy of the Gayo Highlands. Attacks of various diseases can significantly reduce the productivity and quality of Gayo coffee. This study developed a leaf-disease detection model for coffee plants using a Convolutional Neural Network (CNN). The model developed in this study used two datasets. The first dataset, the Gayo Coffee Leaf Disease (PDKG), comprises 900 images of healthy and diseased leaves collected from Gayo coffee plantations. The acquired images in the PDKG dataset were then preprocessed to improve their image quality. The results of model training and testing on the PDKG dataset showed an accuracy of 0.91. On the public Coffee Leaf Diseases (CLD) dataset, the model achieved an accuracy of 0.95, a 7.1% increase over previous studies. The resulting model can help local coffee farmers in the Gayo Highlands detect leaf diseases early and manage plant health more efficiently and accurately.

Keywords—*Leaf Disease Detection, Gayo Coffee Plants, CNN*

Received August 20th,2025; Revised January 27th, 2026; Accepted January 30th, 2026

1. INTRODUCTION

Coffee stands among the world's most vital tropical plantation crops and plays a key role in Indonesia's economy. As the fourth-largest coffee producer globally, after Brazil, Vietnam, and Colombia, Indonesia produced around 774.6 thousand tons in 2021. The production comprised mainly Robusta (81.2%) and Arabica (18.8%) varieties. This output marked a 1.62% increase over the previous year's 762.20 thousand tons and is expected to keep rising. Beyond providing livelihoods for millions of Indonesian farmers, coffee is deeply woven into the cultural and ecological fabric of major producing regions such as Sumatra, Java, and Sulawesi [1-4]. Arabica coffee is prized over Robusta for its complex flavor, distinctive acidity, and unique aroma, which contribute to its higher price in international markets. Within Indonesia, Aceh province stands out as the leading producer of Arabica coffee, responsible for nearly 80% of the nation's Arabica output—about 125,000 tons, or roughly 18.8% of total national coffee production. Arabica varieties flourish best at elevations between 1,000 and 2,000 meters above sea level in cool climates like those of the Gayo Highlands, Aceh [5]. Renowned for its herbal notes, low caffeine content, and lingering aftertaste, Gayo Arabica coffee is especially favored in European and American markets.

Nearly all of Indonesia's Arabica coffee production comes from the Gayo Highlands, spanning three regencies in Aceh Province: Central Aceh, Bener Meriah, and Gayo Lues. This region's coffee has received international recognition through Geographical Indication (GI) certification since 2010. Data from the Central Statistics Agency (BPS) show that, in 2011, the total cultivated area reached about 95,000 hectares. Aceh Tengah Regency accounted for 48,500 hectares, yielding an average of 0.68 tons per hectare annually. Bener Meriah Regency contributed 39,000 hectares with an average yield of 0.78 tons per hectare, while Gayo Lues Regency had 7,000 hectares producing 0.5 tons per hectare [6]. In Central Aceh Regency alone, there are 34,576 coffee farmers, each managing an average of 1.4 hectares [7]. Gayo coffee refers to a coffee variety whose beans are ground to produce a widely enjoyed beverage. Rich in nutrients, Gayo coffee contains thousands of naturally occurring compounds, including carbohydrates, nitrogenous substances, lipids, minerals, vitamins, phenolics, and alkaloids. However, coffee plants cultivated in the Gayo Highlands face persistent challenges from various diseases that threaten both productivity and quality. Disease outbreaks can diminish coffee quality, leading to lower market prices. Furthermore, despite ongoing expansion of cultivation areas, annual productivity has not shown significant growth [8-9].

In 2018, the Baitul Qiradh Baburrayan Coffee Cooperative saw a drop in coffee exports, recording only 800 tons compared to 1,065 tons in 2017 [9]. According to the Aceh Central Statistics Agency, overall coffee production in Bener Meriah Regency has increased. However, this rise does not align with the significant expansion of cultivated land—from 87,533 hectares in 2010 to 97,796 hectares in 2014. During this period, Gayo coffee production grew only modestly, from 41,027 tons in 2010 to 42,079 tons in 2014. These trends suggest that declining productivity, likely caused by diseases and pests affecting Gayo coffee plants, is a major factor [10-13]. Diseases can affect various parts of the coffee plant, including the leaves, stems, and roots. Leaf health is commonly assessed by examining color, as it closely reflects chlorophyll content—an essential pigment for plant processes [14]. For Gayo coffee farmers, it is crucial to identify specific criteria for assessing plant health and recognizing signs of disease. Accurate identification of healthy and diseased plants requires clear criteria, particularly through the examination of coffee leaves for symptoms such as nematode infestation, leaf rust, or leaf spots. This enables farmers to anticipate and manage disease outbreaks more effectively. While information on coffee leaf diseases is widely available, the main challenge is to develop tools that are rapid, accurate, and user-friendly for disease detection.

This study introduces an automated approach for detecting coffee leaf diseases using Convolutional Neural Networks (CNNs) and transfer learning to provide fast and accurate identification for Gayo coffee farmers. The process starts with the collection of images showing both healthy and diseased coffee leaves—including leaf rust, leaf spot, and nematode symptoms—using digital cameras or smartphones in field environments. This strategy ensures that the dataset reflects actual variations in lighting, background, and leaf orientation. Images are then uniformly resized and normalized to optimize the training process. To further increase the robustness of the dataset and reduce overfitting, data augmentation techniques such as rotation, flipping, zooming, and brightness adjustment are applied [15]. The core of the detection system is the NASNetMobile architecture, a lightweight CNN pre-trained on the ImageNet dataset, selected for its balance of classification accuracy and computational efficiency. This makes it especially well-suited for implementation on smartphones and other devices with limited resources. Model performance is evaluated using metrics like accuracy, precision, recall, and F1-score to ensure reliable differentiation between healthy and diseased leaves. The trained model is integrated into a user-friendly, lightweight application, empowering farmers to capture coffee leaf images and receive instant disease diagnoses. Ultimately, this solution delivers a practical, efficient, and precise tool for early detection, enabling proactive disease management and supporting higher coffee productivity.

2. METHODS

This research presents a disease detection model for coffee leaves, structured around two primary stages: training and testing. The training stage starts with preprocessing to enhance image quality. To increase the diversity of the dataset, image augmentation is applied, followed by feature extraction and training of a deep learning convolutional neural network (CNN). During testing, images are preprocessed and their features extracted using the same CNN architecture. These extracted features are then classified by the model developed during training.

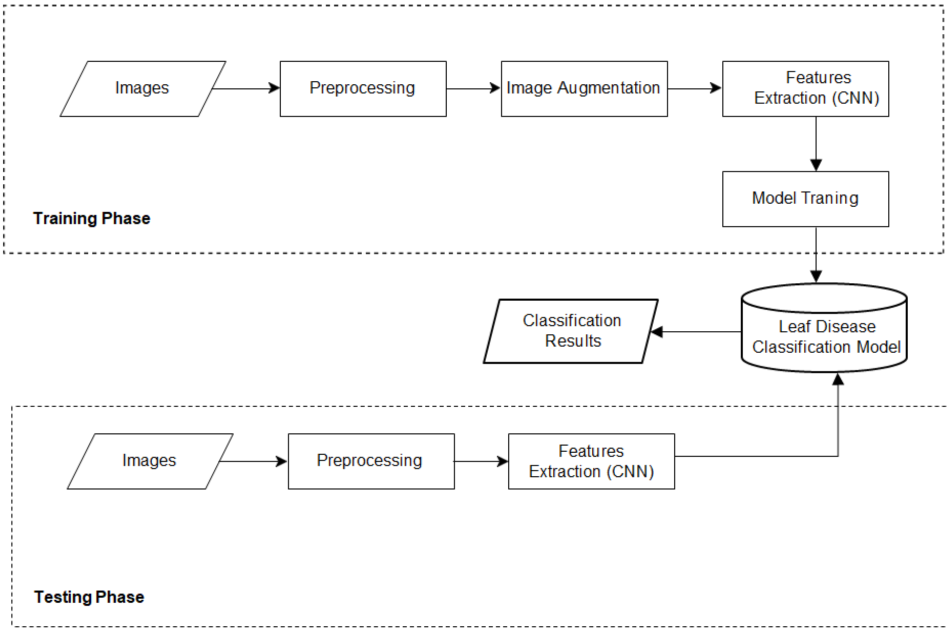


Figure 1. Proposed Method

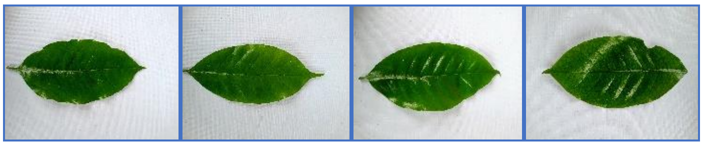
2.1 Datasets

In this study, we used two datasets, namely : Gayo Coffee Leaves Diseases (GCLD) dan dataset Coffee Leaf Diseases (CLD).

- Gayo Coffee Leaves Diseases (GCLD) Dataset

  The GCL dataset consists of coffee leaf images collected directly from a plantation in Paya Tungel Village, Jagong Jeget District, Central Aceh Regency. It contains 900 images, each sized at 256 × 256 pixels, categorized into three groups: leaf rust (LR), leaf spot (LS), and healthy leaves (HL). To ensure diversity and representativeness, the collection process considered varying visual conditions of the leaves. An expert in coffee plant diseases from the Central Aceh Regency Plantation Office verified the class labels for all images. The dataset was split into 80% for training (720 images) and 20% for testing (180 images), with an even distribution among the classes: 300 healthy leaves, 300 with leaf rust, and 300 with leaf spot. Table 1 presents sample images from each class in the GCLD dataset.

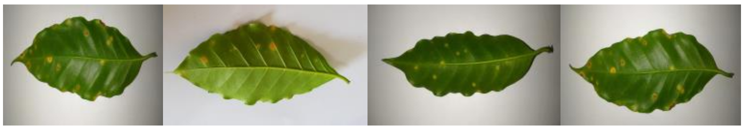
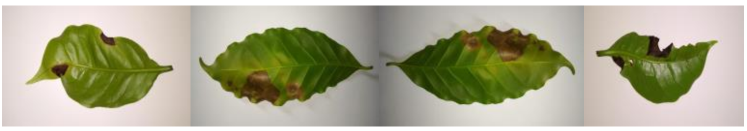
Table 1. Example of Images in GCLD Dataset

No.	**Images**	**Class**
1		LR
2		LS
3		HL

- Coffee Leaf Diseases (CLD) Dataset

  The Coffee Leaf Diseases (CLD) dataset comprises coffee leaf images sourced from publicly available online repositories [16].

Table 2. Example of Images in CLD Dataset

No	**Images**	**Class**
1		LR
2		LS
3		HL

It contains 1,562 images of varying resolutions, classified into three categories: healthy leaves (HL), leaf rust (LR), and leaf spot (LS). The dataset includes images captured under diverse lighting, angles, and backgrounds, ensuring substantial visual variety. Prior to use, all images were preprocessed—resized to 256 × 256 pixels and relabeled to confirm accurate class assignment. Label verification was based on visual references of disease symptoms described in agricultural literature. The CLD dataset was split into 80% for training (1,249 images) and 20% for testing (313 images), with the following class distribution: 466 healthy leaves, 548 leaf rust, and 548 leaf spot images. Table 2 provides sample images from each class in the CLD dataset.

2.2 Prepocessing

The preprocessing stage is designed to enhance image quality and prepare data for model training [17]. By refining images, this step helps the model more effectively identify important visual patterns—especially in coffee leaf images collected under diverse field lighting conditions—thereby optimizing overall recognition performance.

(a) Coffee Leaf Image Before Resizing

(b) Coffee Leaf Image After Resizing

Figure 2. Coffee Leaf Image Beforeand After Resizing

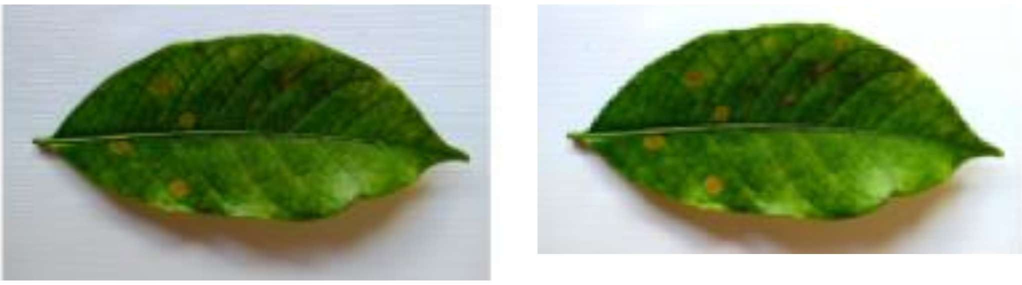
- Resize
  All images in the dataset were resized to 128 × 128 pixels to ensure consistency and match the model's input requirements. Standardizing image dimensions allows the model to process data more efficiently and maintain uniformity across the dataset. Figure 2.a illustrates a Gayo coffee leaf prior to resizing, while Figure 2.b displays the image after this adjustment.

(a) Coffee Leaf Image Before Rescaling

(b) Coffee Leaf Image After Rescaling

Figure 3. Coffee Leaf Image Before and After Rescaling

- Rescale (Pixel Values Normalization)
  Image pixel values, originally spanning 0-255, were normalized to 0–1. This step helps speed up training, improves model stability, and reduces the impact of large pixel values on weight updates. Figure 3 presents Gayo coffee leaves before and after normalization.
- Contrast Adjustment
  Contrast adjustment is performed by changing the image brightness within a range of 0.8 to 1.2. This approach helps counteract uneven lighting that often occurs during plantation image acquisition, enabling the model to more effectively identify important features of the leaf. Figure 4 illustrates the leaf images before and after contrast adjustment.



(a) Coffee Leaf Image Before Contrast Adjustment   (b) Coffee Leaf Image After Contrast Adjustment

Figure 4. Coffee Leaf Image Before and After Contrast Adjustment

- Image Augmentation
  Augmentation is used to expand and diversify the training dataset. This process includes rotation, translation (horizontal and vertical shifts), zoom, and horizontal flipping. By applying these techniques, the model learns to recognize objects under different conditions and perspectives, improving its generalization to new data. Figure 5 displays examples of the augmented images.

2.3 Convolutional Neural Network (CNN)

Following data preprocessing, a Convolutional Neural Network (CNN) is implemented for detecting leaf diseases in Gayo coffee plants. This method leverages transfer learning by employing the NASNetMobile architecture, which has been pre-trained on the ImageNet dataset.

- Transfer Learning
  The implemented CNN model is based on the NASNetMobile architecture, which is pre-trained on the ImageNet dataset [18]. This lightweight model is well-suited for image classification tasks on devices with limited resources. In this setup, the NASNetMobile layers are frozen (non-trainable) and serve exclusively as feature extractors, without further retraining.
- CNN Model Architecture
  The CNN architecture used in this study is composed of several key layers, with NASNetMobile serving as the base model (excluding its output layer by setting include_top=False) [19]. The architecture features a Conv2D layer with 32 filters, a 3×3 kernel size, ReLU activation, and 'same' padding. To prevent overfitting, a MaxPooling2D layer with a 2×2 pool size and a Dropout layer with a rate of 0.5 are included. The Flatten layer transforms the convolutional outputs into a one-dimensional vector. Finally, a Dense output layer with 3 neurons and a Softmax activation function is used, matching the number of leaf disease classes.

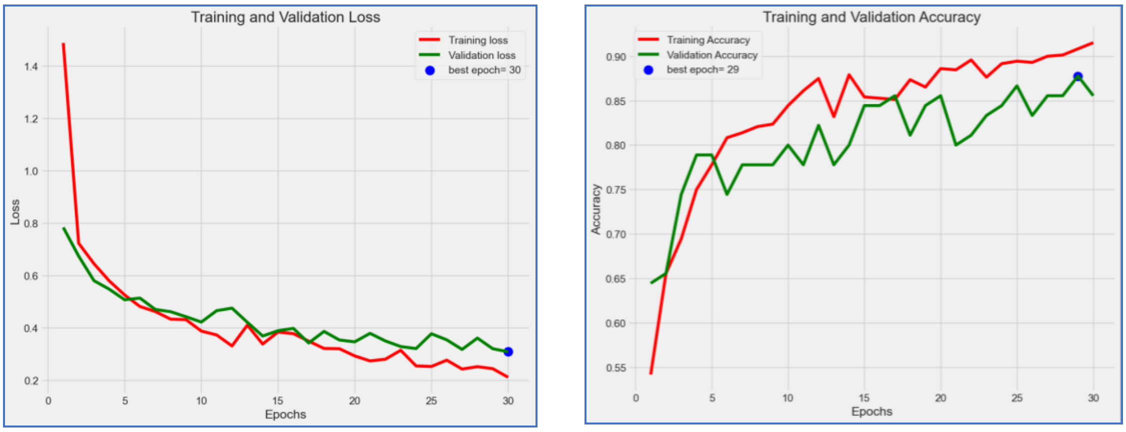
- Model Compilation
  The model was compiled with the categorical cross-entropy loss function, suitable for multi-class classification tasks with categorical outputs. Adam optimizer was selected, set to a learning rate of 0.001. Model performance was evaluated using accuracy, reflecting how well the model correctly classifies images.
- Training process
  Training was conducted over 30 epochs with a batch size of 64. Data augmentation was applied exclusively to the training set to enhance generalization, while validation and test sets underwent only rescaling. Both training and validation utilized the flow_from_dataframe method.

Figure 5. Coffee Leaf Image After Augmentation

3. RESULTS AND DISCUSSION

3.1 The Experiments on the GCLD dataset

Throughout training, both accuracy and loss metrics for the training and validation sets were monitored to track model performance.



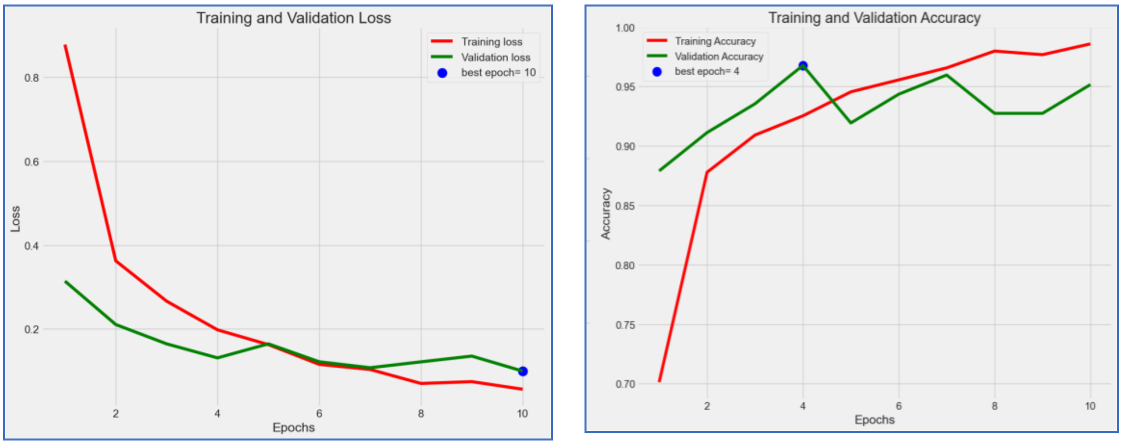
a. Loss b. Accuracy

Figure 5. The loss and accuracy of the model on the GCLD Dataset

The data revealed a consistent improvement in accuracy and a reduction in loss with each epoch, with no clear signs of overfitting. Final evaluation on the test set yielded a loss of 0.22 and an accuracy of 0.91. The low loss value reflects the model's effectiveness in minimizing prediction errors, while the high accuracy underscores its reliability in classifying Gayo coffee leaf images. Given the dataset's visual complexity and the presence of three classes (leaf rust, leaf spot, and healthy leaves), these results highlight the strong and dependable performance of the transfer learning-based CNN using the NASNetMobile architecture. Figure 5 presents the corresponding loss and accuracy on GLCD Dataset.

3.2 Benchmarking using CLD Dataset

This study benchmarks the proposed model against the one developed by Shetty et al. (2024) [20], which utilized the VGG16 architecture for coffee leaf disease classification on the CLD dataset and reported a test accuracy of 0.88. In contrast, the model in this work employs NASNetMobile with additional layers (Conv2D, MaxPooling2D, Dropout, and Dense) and achieves a test accuracy of 0.95 on the CLD dataset. This 7.06% improvement over Shetty et al.'s results demonstrates the benefits of using NASNetMobile and an optimized training strategy. These findings suggest that the improved model architecture and training techniques adopted in this study lead to superior performance. The loss and accuracy curves for the CLD dataset are shown in Figure 6.

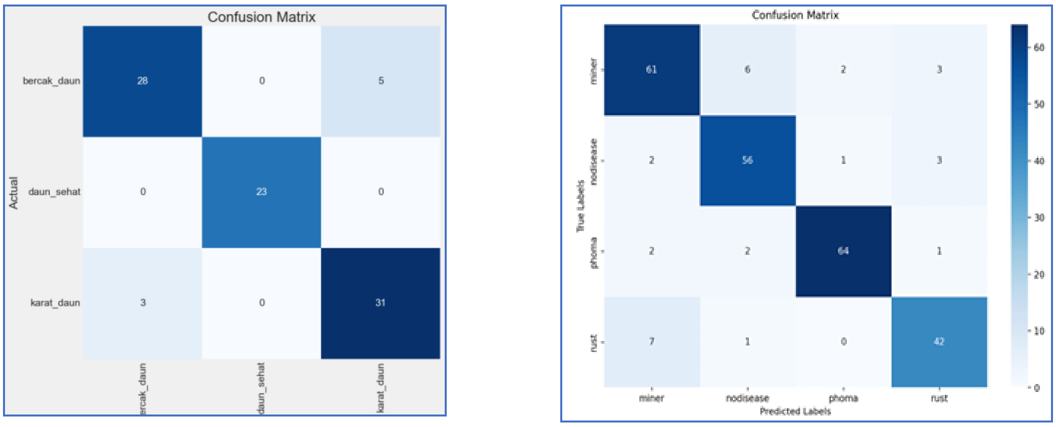


a. Loss b. Accuracy

Figure 6. The accuracy and loss of the model on the CLD Dataset

The confusion matrix analysis reveals that the proposed model achieved higher classification accuracy across all three target classes compared to previous studies. For the leaf spot class, the model accurately identified most samples and had fewer misclassifications than the Shetty et al. (2024) model, which frequently confused leaf spot with leaf rust. In the healthy leaf class, the proposed model delivered more consistent predictions and produced fewer false positives, demonstrating an enhanced ability to distinguish healthy leaves from diseased ones. Regarding the leaf rust class, both models correctly classified most samples, but the proposed model had a lower misclassification rate. Overall, the confusion matrix indicates that the proposed model not only achieved a greater number of correct predictions in each class but also maintained a more balanced distribution of errors. Figure 7 shows the confusion matrix heatmap on the CLD dataset for the model by Shetty et al. (2024).

The classification report shows that the proposed model performs consistently across all classes. In the LS class, the precision value of 0.90 and recall of 0.85 are slightly higher than those of the Shetty et al. (2004) model, which has a precision of 0.88 and a recall of 0.83. In the HL class, the proposed model achieved a score of 1 in precision, recall, and F1-score.



a. Confusion Matrix of the Proposed Model

b. Confusion Matrix of Shetty et al. (2024)

Figure 7. Confusion Matrix heatmap on the CLD dataset obtained

Meanwhile, in the LR class, the proposed model achieved a precision of 0.86 and a recall of 0.91, slightly outperforming the Shetty model, which had a precision of 0.84 and a recall of approximately 0.89. Although this difference is relatively small, it still suggests that the model can reduce cross-prediction errors between leaf rust and leaf spot.

Table 3. The Proposed Model Classification Report on the CLD dataset

Class	Precision	Recall	F1- Score	Support
LS	0.90	0.85	0.88	33
HL	1.00	1.00	1.00	23
LR	0.86	0.91	0.89	34

Overall, the proposed model shows improvements in all metrics compared to the previous research model, both in the ability to recognize each class and in maintaining a balance between precision and recall. This has a direct impact on improving overall accuracy and classification accuracy on the test data. The proposed model classification report in detail on the CLD dataset is shown in Table 3.

4. CONCLUSIONS

The proposed Gayo coffee leaf disease detection model exhibited robust classification capabilities. On the GCLD dataset, it achieved an accuracy of 0.91, while testing on the CLD dataset reached 0.95—representing a 7.1% improvement over previous studies. The results also indicated that factors such as lighting and shooting distance have a notable impact on detection accuracy, as demonstrated by the more consistent image quality and performance on the CLD dataset. To further enhance model performance and generalizability, expanding the size and diversity of training datasets is recommended. Additionally, optimizing hyperparameters like learning rate, batch size, and number of epochs can help achieve optimal results. Comparing alternative CNN architectures, including MobileNetV2, EfficientNet, or ResNet, may also identify the most effective approach.

REFERENCES

[1] Putri, A. Y. P., & Sodik, A. (2019). Identifikasi penyakit tanaman kopi arabika dengan metode K-Nearest Neighbor (K-NN). Seminar Nasional Sains dan Teknologi Terapan, VII, 759–764.

[2] Badan Pusat Statistik Nasional. (2022). Jumlah produksi kopi di Indonesia (2017–2021). Jakarta: Badan Pusat Statistik.

[3] Auliansyah, G., Fachuruddin, F., & Yunus, Y. (2019). Evaluasi kesesuaian lahan pada

tanaman kopi arabika (Coffea arabica L.) organik menggunakan Sistem Informasi Geografis (SIG) di Kecamatan Pegasing Kabupaten Aceh Tengah. Jurnal Ilmiah Mahasiswa Pertanian, 4(2), 329–338.

[4] Badan Pusat Statistik. (2015). Data perkembangan luas areal dan produksi tanaman tahunan dan semusim komoditi perkebunan 7 tahun terakhir (2009–2015) Kabupaten Aceh Tengah. Dinas Perkebunan dan Kehutanan Kabupaten Aceh Tengah.

[5] Dishutbun. (2012). Statistik Perkebunan dan Kehutanan Tahun 2011. Dinas Perkebunan dan Kehutanan Kabupaten Aceh Tengah, Takengon, Indonesia.

[6] Sugiarti, L. (2017, Juni). Identifikasi hama dan penyakit pada tanaman kopi di kebun percobaan Fakultas Pertanian Universitas Winaya Mukti. Jurnal Agro Wiralodra, 1, 16–22.

[7] Sufi, S. (2019). Implementasi program pemberdayaan petani kopi: Pada Koperasi Baitul Qiradh Baburrayyan Kota Takengon. Lhokseumawe: Unimal Press.

[8] Badan Pusat Statistik Aceh. (2014). Jumlah produksi kopi di Kabupaten Bener Meriah (2010–2014). Bener Meriah.

[9] Dharmadewi, A. (2020, September). Analisis kandungan klorofil pada beberapa jenis sayuran hijau sebagai alternatif bahan dasar food supplement. Jurnal Emasains: Jurnal Edukasi Matematika dan Sains, 9, 171–176.

[10] Mahfud, M. C. (2012). Teknologi dan strategi pengendalian penyakit karat daun untuk meningkatkan produksi kopi nasional. Malang: Balai Pengkajian Teknologi Pertanian Jawa Timur.

[11] Sutoyo, T., dkk. (2009). Teori pengolahan citra digital. Yogyakarta: Andi.

[12] Esgario, J. G. M., Krohling, R. A., & Ventura, J. A. (2019). Deep learning for classification and severity estimation of coffee leaf biotic stress. arXiv. https://arxiv.org/abs/1907.11561

[13] Montalbo, F. J. P., & Hernandez, A. A. (2020). Classifying barako coffee leaf diseases using deep convolutional models. International Journal of Advances in Intelligent Informatics, 6(2), 197–209. https://doi.org/10.26555/ijain.v6i2.495

[14] Sutoyo, T., Mulyanto, E., Suhartono, V., & Nurhayati, D. O. W. (2009). Teori pengolahan citra digital. Yogyakarta: Andi; Semarang: UDINUS.

[15] Cmcbinus. (2023). Pengolahan citra digital: Konsep dan teknik. https://binus.ac.id/malang/2023/07/pengolahan-citra-digital-konsep-dan-teknik/

[16] Rahmad Hidayat, Agus Harjoko, and Aina Musdholifah, "A Robust Image Retrieval Method Using Multi-Hierarchical Agglomerative Clustering and Davis-Bouldin Index," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 441–453, Apr. 2022, doi: 10.22266/ijies2022.0430.40.

[17] M. H. Ali, T. Mahmud, M. T. Aziz, Md. F. B. A. Aziz, M. S. Hossain, and K. Andersson, "Leveraging Transfer Learning for Efficient Classification of Coffee Leaf Diseases," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India: IEEE, Jun. 2024, pp. 1–6. doi: 10.1109/ICCCNT61001.2024.10725973.

[18] R. Hidayat, A. Harjoko, and A. K. Sari, "Content Based Image Retrieval Berdasarkan Fitur Low Level: Literature Review," *J. Buana Inform.*, vol. 8, no. 2, Apr. 2017, doi: 10.24002/jbi.v8i2.1077.

[19] R. Karthik, R. Aswin, K. S. Geetha, and K. Suganthi, "An Explainable Deep Learning Network With Transformer and Custom CNN for Bean Leaf Disease Classification," *IEEE Access*, vol. 13, pp. 38562–38573, 2025, doi: 10.1109/ACCESS.2025.3546017.

[20] A. Shetty, "Coffee Leaf Disease Detection Using CNN," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 8, pp. 439–444, Aug. 2024, doi: 10.22214/ijraset.2024.63940.