Original Research Article

Advancing Coffee Leaf Rust Disease Management: A Deep Learning Approach for Accurate Detection and Classification using Convolutional Neural Networks

ABSTRACT

Coffee Leaf Rust (CLR), caused by the fungus *Hemileia vastatrix*, poses a severe threat to global coffee production. Timely detection is critical for effective control measures. This study employs Convolutional Neural Networks (CNNs) to enhance CLR detection accuracy. Traditionally, this task relies on expert assessment. DL emerges as a promising approach, capable of autonomously extracting salient features. Our model, trained on a diverse dataset, accurately identifies CLR. Using 1365 meticulously curated images, the model undergoes rigorous preprocessing and augmentation. The DL-based approach achieves remarkable accuracy (98.89%), precision (99.00%), recall (98.07%), and an F1 score of (98.55%). These outcomes establish the CNN model as a proficient system for precise, real-time CLR diagnosis. This study contributes to the creation of an efficient system, safeguarding coffee orchard vitality and productivity.

Keywords: Coffee Leaf Rust, Early Detection, Convolutional Neural Networks, Accuracy, Management.

1. INTRODUCTION

Coffee, one of the most popular and widely consumed beverages globally, is an essential agricultural commodity that supports the livelihoods of millions of farmers (Voora *et al.,* 2019) [18]. The diseases affecting coffee plants are a critical factor severely limiting coffee's productivity. Biotic stresses, such as leaf miner, rust, phoma, and Cercospora leaf spot, damage coffee plants and cause defoliation and a reduction in photosynthesis, thus reducing the production and quality of the product (Esgario *et al.,* 2022) [2]. Thus, identifying and measuring plant diseases is highly important in phytopathology. It is essential to understand both causal agents and the severity of the symptoms for effective pest and disease management (Kranz *et al.,* 1988) [5]. If not treated appropriately, these diseases can cause significant leaf damage and crop fatality (Sabrina and Maki, 2022) [15]. However, the prevalence of Coffee Leaf Rust (CLR) disease has emerged as a significant challenge, jeopardizing coffee production and sustainability. Coffee Leaf Rust (CLR) is caused by the fungus *Hemileia vastatrix,* which affects the leaves of coffee plants. This leads to defoliation, reduced photosynthesis, and lower yields. The disease is characterized by shapeless lesions that result in early leaf loss, significantly impacting productivity. When conditions favor the disease, about 35% of a coffee plantation is typically affected, and during prolonged dry seasons with a high occurrence of the fungus, damage can exceed 50% (Kushalappa *et al.,* 1984) [6]. Early recognition of CLR is crucial to prevent losses and enhance the productivity and quality of coffee crops (Suhartono *et al.,* 2013) [17]. Traditionally, identifying CLR has relied on manual inspection and visual assessment by trained experts, which is time-consuming, subjective, and prone to human error.

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have provided new opportunities for automated disease detection in various crops, offering precise and efficient solutions. Artificial intelligence, particularly in the form of deep learning (DL), is causing a transformative shift in agriculture. By mimicking human cognitive processes, it facilitates more efficient farming practices with notably enhanced outcomes (Kouadio *et al.,* 2018) [4]. DL, a potent branch of machine learning, has demonstrated superior performance in various domains compared to conventional methods (Shah *et al.,* 2022; Mendieta *et al.,* 2019) [16,9]. Within agriculture, DL has proven exceptionally adept at identifying plant diseases through visual assessment (Montalbo and Hernandez, 2020) [10]. Researchers have been utilizing convolutional neural networks (CNNs), a specific type of DL algorithm, for tasks like classifying plant species and detecting diseases (Dutta and Rana, 2021; Montalbo *et al.,* 2020) [1,11]. These CNNs excel at recognizing distinctive features and acquiring robust learning from data, rendering them a highly promising tool for applications in agriculture (Lee *et al.,* 2020) [7]. In this manuscript, we aim to investigate the efficacy of employing deep learning technology, specifically a CNN-based approach, for the detection of CLR disease in coffee leaf rust detection. Our study builds upon the vast dataset of annotated coffee leaf images, encompassing both healthy leaves and leaves affected by CLR disease. By leveraging the power of CNNs, we aim to develop a robust and reliable model capable of accurately distinguishing between healthy and infected coffee leaves. The proposed research will follow a systematic methodology, comprising data preprocessing, model training, and evaluation stages. Subsequently, a CNN architecture will be designed, optimized, and trained on the dataset to extract meaningful features and learn discriminative patterns associated with CLR disease. The outcomes of this study hold great promise for the coffee industry, offering an automated and reliable tool for early detection of CLR disease. By providing timely information to farmers, they can implement targeted interventions, such as fungicide applications or removal of infected plants, thus minimizing yield losses and preserving the economic viability of coffee production.

This manuscript aims to contribute to the field of agricultural technology by demonstrating the potential of deep learning, specifically CNNs, in combating the Coffee Leaf Rust disease. The proposed methodology, findings, and insights from this research can serve as a foundation for the development of scalable and cost-effective solutions to tackle this and other plant diseases, ultimately benefiting coffee farmers and the global coffee industry as a whole.

2. METHODOLOGY

This section describes the architecture and training method of the proposed CNN model, as well as the experimental setup and dataset preparation. The suggested workflow for plant leaf disease detection begins with dataset preparation and ends with model prediction. TensorFlow 2.9.1, numpy Version 1.19.2, matplotlib Version 3.5.2, and OpenCV Version 4.5.5 libraries are utilized for dataset preparation and CNN model building respectively.

2.1 Dataset Collection

A diverse dataset, comprising high-resolution images of coffee leaves affected by Coffee Leaf Rust (CLR) disease as well as images of healthy coffee leaves, was obtained from the RoCole image dataset (Parraga-Alava, 2019) [13]. This dataset encompasses 1560 images of robusta coffee leaves, featuring noticeable mites and spots which indicate the presence of coffee leaf rust in the case of infections, and images without such abnormalities for healthy instances. In order to establish an image-based recognition system, every image underwent a thorough examination for CLR symptoms. Only images devoid of any deformities were considered as healthy. Images were selected based on the presence of visible rust pustules for diseased cases and healthy leaves for the unaffected ones. The dataset was then annotated with appropriate labels denoting the presence or absence of CLR disease, categorizing them as diseased or healthy, respectively. Figure 1 showcases sample images of both rust-affected and healthy coffee leaves from this dataset.

2.2 Data Preprocessing

To maintain consistent input dimensions, all images were resized to 224x224 pixels. To augment the dataset and increase the number of images in each class, various data augmentation techniques were applied. These techniques serve to expand the dataset size and mitigate overfitting during the model training process by incorporating augmented images into the training set. Augmentation methods included rotation, flipping, zooming, and rescaling, which were implemented with specific parameters. These parameters included a width shift range of 0.2, height shift range of 0.2, shear range of 0.2, zoom range of 0.2, and horizontal flip. The dataset was then divided into an 80:20 train-test split, meaning 80% of the dataset was utilized to train the proposed model, while the remaining 20% was reserved for testing the model's performance. Detailed information about the dataset can be found in Table 1.

Table 1. Details of the proposed dataset

	Leaf rust	**Healthy**	**Total**
Train	422	671	1093
Test	104	168	272
Total	526	839	**1365**

2.3 Feature extraction

Deep learning techniques were used for feature extraction to extract deep characteristics automatically from the acquired images. This helps to classify the given images into predefined classes (CLR, and healthy).

Feature extraction using the proposed CNN model

Feature extraction is a crucial step in object recognition, necessitating features that are distinct enough to distinguish between different object classes, while also maintaining consistent characteristics within the same class. It serves as a dimensionality reduction process, essential for efficient pattern recognition and machine learning in image analysis. In this study, a feature extraction method based on deep learning using Convolutional Neural Networks (CNN) was employed and compared to process CLR images for reliable image classification. The proposed classifiers effectively discern between diseased and healthy coffee plant leaves.

CNN, also known as ConvNet, is a potent deep learning technique inspired by biological models that mirror how humans perceive images through various layers of processing. CNN captures spatial and temporal dependencies in an image by applying filters at different layers. Feature extraction with CNN condenses the image, requiring fewer computations while retaining the necessary features for accurate prediction. The CNN model is structured with a layered architecture, comprising a convolutional layer, ReLU layer, pooling layer, dropout layers, and fully connected layers.

To extract deep features from the images, a hierarchy of feature maps is constructed by consecutively applying learnable filters to the input image. The initial convolutional layer captures low-level features like edges, corners, texture, and lines, while the subsequent layers extract high-level features based on more complex characteristics to identify objects and structures within the image. Consequently, to achieve the optimal CNN model, three convolutional layers were utilized to attain the highest accuracy in the experiment.

2.4 Model Training

After the CNN network architecture was used to extract features from the input images, the CNN model was trained using a set of labelled training images. Then the classification process categorizes the data into the desired category using the retrieved features.

2.5 Softmax

In this study, CNN models were used as input data for the SoftMax classifier to determine the probability of the expected label for coffee diseases. The Softmax classifier was employed for recognition purposes, aiming to ascertain the likelihood that the input belongs to a specific class. It produces values in the range of 0 to 1, where the sum of all probabilities equals one (Ho and Wookey, 2020) [3]. The notable advantage of using SoftMax lies in its ability to easily define the output probabilities range, and it is also efficient in terms of training speed and prediction. Furthermore, it accepts the output from the last fully connected layer and is employed for classifying coffee images into specific classes (CLR or healthy).

The CNN model was initialized with pre-trained weights derived from the selected architecture. The training data was fed into the model, and the model parameters were optimized using a suitable optimization algorithm such as Adam. The training process involved iterating over the training set for a predetermined number of epochs, while adjusting the learning rate to minimize the classification loss.

2.6 Hyperparameter settings

Hyperparameters are settings outside of the DL algorithm whose value is determined before the training begins. There is no universally accepted method for determining the appropriate hyperparameters for a given situation. As a result, numerous tests are carried out in order to pick appropriate hyperparameters. Table 2 summarizes the hyperparameters employed throughout our model training. The selected hyper-parameters for the model are described in the sections A-E.

A. Optimization algorithms

To minimize the error rate, the proposed model is trained using the Adam optimization technique, and the weights are updated via backpropagation of error. In DL research, Adam is by far the most popular and commonly utilized optimization method. It adjusts the model’s weight and modifies parameters to minimize the loss function. Adam calculates an adaptive learning rate for each parameter, scaling the learning rate with squared gradients and a moving average of the gradient.

B. Learning rate

Because backpropagation was used to train the proposed model, a learning rate is required during weight updates. Its purpose is to limit the amount of weight that is updated during backpropagation. The most difficult component of our experiment was determining the appropriate learning rate. In our experiment, we discovered that training at a low learning rate takes longer than training at a high rate. However, when we supply a lower value, the model outperforms one with a higher learning rate. Therefore, a learning rate of 0.001 was set for all of the experiments determine by simulations.

C. Loss function

The loss function, sometimes called a cost function, defines how well a model achieves the specified goal. The activation functions employed in the model's output layer (final fully connected layer) and the type of problem to be solved all influenced the loss function choosen (whether regression or classification). In the proposed model, SoftMax is employed as an activation function in the last completely linked layer to determine the class label because a classification challenge, specifically categorical classification, is being addressed. For our model, Categorical Cross-Entropy (CCE) loss was employed as a loss function. Despite the existence of other loss functions such as Binary Cross-Entropy (BCE) and Mean Squared Error (MSE), Categorical Cross-Entropy is the recommended loss function for problems with more than two classes.

D. Number of epochs

The number of epochs is the number of iterations the complete dataset goes through while testing the model. During training, we discovered that using too little or too large epochs causes the model to have a large gap between the training and validation errors. With an epoch of hundred (100), the model becomes optimal after many experiments.

E. Batch size

Batch size is used to calculate how many inputs can be sent to the network at once. Since the computer can not be fed all of the data in a single period since it's too complex, so the input is broken into smaller groups. In the model training, it is better to reduce the machine's computing time. In our experiment, a batch size of 32 was employed for training the model.

Table 2. Summary of hyper parameters used for model training

Parameter	Values
Epoch	100
Batch Size	32
Activation Function	softMax
Loss Function	Categorical Cross-Entropy
Optimization Algorithm	Adam
Learning Rate	0.001

2.7 Model Evaluation

Evaluated the trained model on the testing set to assess its performance. Evaluation metrics such as accuracy, precision, recall, and F1-score to measure the model's effectiveness in detecting CLR disease were estimated. Confusion matrix was generated to analyze the model's performance in terms of true positives, true negatives, false positives, and false negatives.

Accuracy is the ratio of the number of correctly classified samples divided by the total number of samples, whereas the error rate is the proportion of samples identified incorrectly. When observations for each class are not at variance, accuracy can be considered as a good metric. The equation to calculate the accuracy is expressed mathematically as:

$$\textbf{Accuracy}\ (\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100$$

Precision is the number of true positives divided by the number of true positives + the number of false positives. It provides an impression of how well a classifier correctly classifies each class. Precision can be expressed mathematically as:

$$\textbf{Precision}\ (\%) = \frac{TP}{TP + FP} \times 100$$

Recall expresses the ability to find all relevant instances in a dataset. Recall presents the performance of the model to avoid false negatives. Recall can be expressed mathematically as:

$$\textbf{Recall}\ (\%) = \frac{TP}{TP + FN} \times 100$$

F1 Score is a performance metric for classification and is calculated as the harmonic mean of precision and recall. The F1 score is commonly used to measure performance of binary classification, but extensions to multi-class classifications exist.

$$\textbf{F1 Score}\ (\%) = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

where,

True positive (TP)- No. of samples correctly identified as Healthy

False positive (FP)- No. of samples incorrectly identified as Healthy

True Negative (TN)- No. of samples correctly identified as diseased

False Negative (FN)- No. of samples incorrectly identified as diseased

3. RESULTS AND DISCUSSION

All of the details of the experiment including the outcomes of each experiment and the discussions of these results are described in this section. The experimental results are presented in the form of figures and tables.

3.1 Experimental setting

Python was used as the programming language and anaconda Jupyter notebook as a tool for writing the code. For the experiment, the following hyperparameters were used; loss function was categorical cross-entropy, Adam optimizer was used as the optimization function with a learning rate of 0.001, with 100 epochs and a batch size of 32. The designed model was tested on HLBS Technologies for Tomorrow, Windows 11 Pro, 64-bit operating system, x64-based processor. The system embeds 12th Gen Intel(R) Core (TM) i7-12700, 2.10 GHz, with 16 GB RAM.

3.2 Experimental results

In this study, we assessed the accuracy of the CNN classifier in identifying CLR disease from images. Sample images from the dataset depicted CLR conditions, featuring distinct rust pustules on the lower leaf surface, corresponding to chlorosis symptoms on the upper surface. In contrast, healthy leaves showed no visual signs or symptoms (Fig. 1). Additionally, we subjected sample images from the dataset to data augmentation techniques such as flipping, rotation, and zooming, among others, as shown in Fig. 2.

For the model training, we adjusted several hyperparameters: 100 epochs, a batch size of 32, SoftMax activation function, Categorical Cross-Entropy loss function, Adam optimization algorithm, and a learning rate of 0.001 (Table 2). Fig. 3A illustrates the progressive decrease in training and validation set losses with successive iterations. As the model learns, it effectively minimizes the loss values, demonstrating a convergence towards the global minima. The training and validation losses reached their minimum values around the 40th epoch, after which the reduction in loss values was marginal, eventually stabilizing at 0.2 by the 100th epoch. Fig. 3B displays the upward trend in accuracy for both the training and validation sets with successive iterations. The highest accuracy was attained around the 40th epoch, beyond which a notable disparity in accuracy between the training and validation curves was observed. The training curve peaked at 90.00% accuracy by the 100th epoch, while the validation curve reached its peak of 90.00% at the 50th epoch and consistently maintained an accuracy exceeding 90.00% in subsequent iterations. This graph underscores that our model achieved an accuracy of at least 90% in most iterations, culminating in an overall classification accuracy of 98.89%.

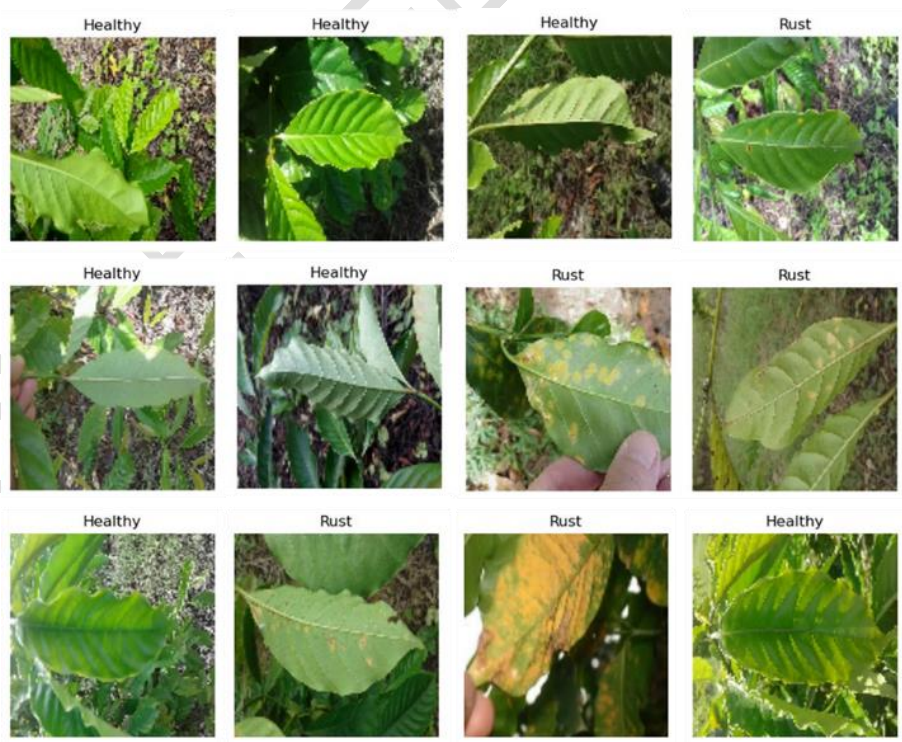

Fig. 1: Sample images of coffee healthy and rust affected leaves from the dataset

Fig.2: Sample images from the data set after image augmentation

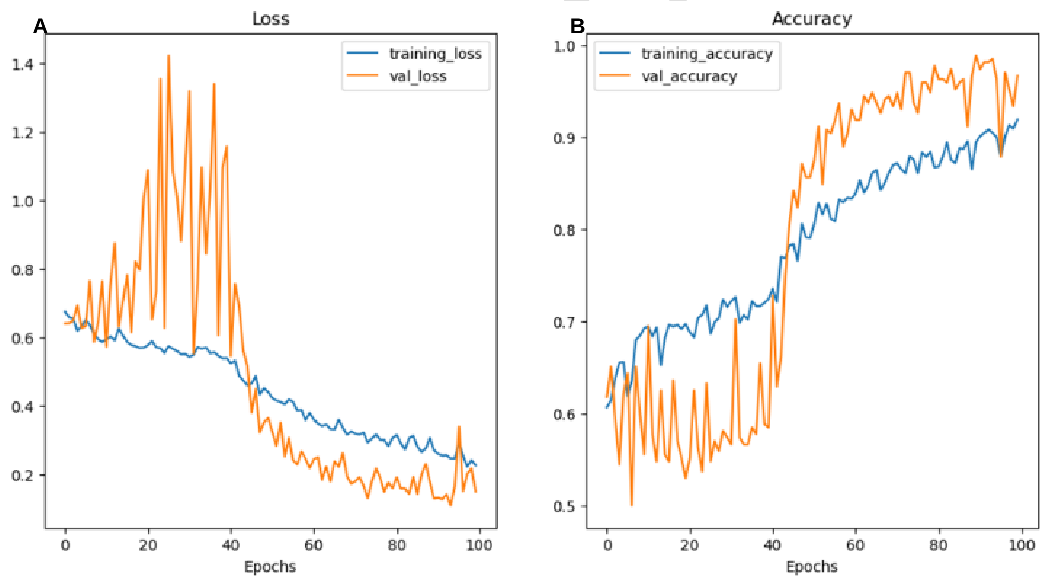

Fig. 3: Model accuracy and loss curves

Prediction made using test set with developed trained model also showed 98.89 per cent accuracy by predicting CLR correctly as CLR and coffee healthy class images correctly as coffee healthy (Fig. 4).

Fig.4: Prediction accuracy of the model

Confusion matrix is a specific table that makes it easy to see if the model is mislabeling one class as another. The performance of the model can be visualized using the confusion matrix. Fig. 5 depict the confusion matrix comparing the true class against the predicted class in the split test set of images for the CLR disease. The calculated value describes the classification rate for individual classes. In the matrix higher colour density signifies higher accuracy for the individual classes. A total 272 images which were in test set, consisting of 104 CLR diseased images and 168 coffee healthy images were subjected to prediction using pre-trained model. The trained model correctly predicted 102 images as CLR disease and 167 images as coffee healthy images and only 1 coffee healthy images was predicted as rust and 2 CLR images were predicted as healthy images by the trained CNN model. The evaluation metrics such as accuracy, precision, recall and F1 score of developed model on recognition of CLR disease were determined using confusion matrix. Table 3 depicts the evaluation metrices for the developed coffee CLR disease detection using CNN. It is evident from Table 3 that developed model shows an Accuracy-98.89 %, Precision-99.00 %, Recall-98.07 %, and F1 Score-98.55 %.

The table 4 showcases the results of various studies focused on the classification of Coffee Leaf Rust (CLR) and other coffee leaf diseases using Convolutional Neural Networks (CNNs), along with their respective accuracy percentages. The current study, utilizing CNN for CLR classification, achieved an impressive accuracy of 98.89%. This indicates a high degree of success in accurately identifying CLR, suggesting that the CNN model used in this study is highly effective for this task. Similarly, Paulos and Woldeyohannis (2022) [14] and Novtahaning *et al.* (2022) [12] both employed CNNs for the classification of coffee leaf diseases, achieving high accuracies of 98.50% and 97.31% respectively. These results suggest that CNN models are highly suitable for accurately distinguishing various coffee leaf diseases. Comparatively, Marcos *et al.* (2019) [8] also utilized a CNN for CLR classification, achieving an accuracy of 95.00%. While this accuracy is slightly lower than the current study, it still demonstrates a high level of effectiveness in identifying CLR.

Table 3. Model evaluation metrices

Evaluation metrics	Per cent
Accuracy	98.89
Precision	99.00
Recall	98.07
F1 Score	98.55

Table 4. Comparison of present and two previously published classification algorithms

Reference	Method	Description of the study	Accuracy (%)
Current study	CNN	CLR detection	98.89
Paulos and Woldeyohannis, 2022	CNN	Coffee leaf diseases detection	98.50
Novtahaning et al., 2022	CNN	Coffee leaf diseases detection	97.31
Marcos et al., 2019	CNN	CLR detection	95.00

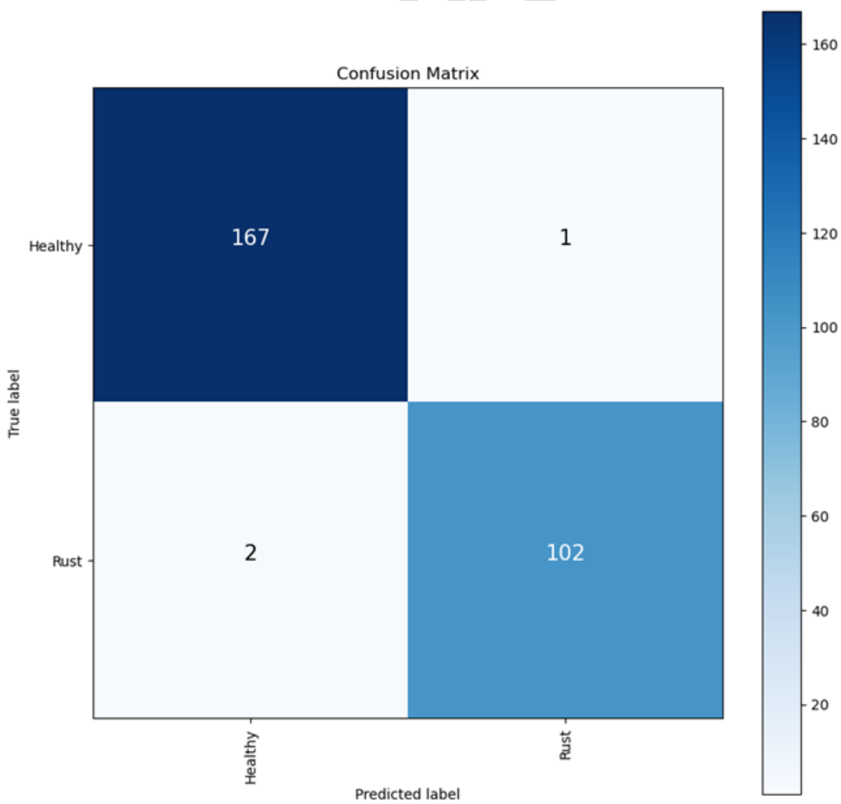

Fig. 5: Confusion matrix for the developed model

4. CONCLUSION

In conclusion, the studies presented collectively demonstrate the remarkable potential of CNNs in coffee leaf disease classification. Future research could delve into further refining these models, exploring transfer learning techniques, and potentially integrating additional data sources for even more accurate and robust classifications.

REFERENCES

1. Dutta L, Rana AK. Disease Detection Using Transfer Learning In Coffee Plants. In Proceedings of the 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 1-3 October 2021; 1-4.
2. Esgario JG, de Castro PB, Tassis LM, Krohling RA. An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning. Information Processing in Agriculture 2022;9: 38-47.
3. Ho Y, Wookey S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. IEEE Access, 2020; 8: 4806-4813.
4. Kouadio L, Deo RC, Byrareddy V, Adamowski JF, Mushtaq S, Nguyen VP. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. Computers and Electronics in Agriculture, 2018; 155: 324-338.
5. Kranz J. Measuring plant disease. In Experimental Techniques in Plant Disease Epidemiology; Springer: Berlin/Heidelberg, Germany, 1988; 35-50.
6. Kushalappa AC, Akutsu SH, Oseguera GM, Chaves, et al. Equations for Predicting the Rate of Coffee Rust Development Based on Net Survival Ratio for Monocyclic Process of *Hemileia Vastatrix* [*Coffea Arabica*]. *Fitopatologia Brasileira* (*Brazil),* 1984.
7. Lee SH, Goeau H, Bonnet P, Joly A. New perspectives on plant disease characterization based on deep learning. Computers and Electronics in Agriculture, 2020; 170: 105220.
8. Marcos AP, Silva NL, Rodovalho, Backes AR. Coffee Leaf Rust Detection Using Convolutional Neural Network. 2019 XV Workshop de Visão Computacional (WVC), São Bernardo do Campo, Brazil, 2019; 38-42. doi: 10.1109/WVC.2019.8876931.
9. Mendieta M, Neff C, Lingerfelt D, Beam C, George A, Rogers S, Ravindran A, Tabkhi H. A Novel Application/Infrastructure Co-design Approach for Real-time Edge Video Analytics. In Proceedings of the 2019 Southeast Con, Huntsville, AL,USA, 11-14 April 2019; 1-7.
10. Montalbo FJ Hernandez A. Classifying Barako coffee leaf diseases using deep convolutional models. International Journal of Advances in Intelligent Informatics, 2020; 6: 197-209.
11. Montalbo FJP, Hernandez AA. An Optimized Classification Model for *Coffea Liberica* Disease using Deep Convolutional Neural Networks. In Proceedings of the 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, 28-29 February 2020; 213-218.
12. Novtahaning, Damar, Shah, Hasnain Ali and Kang, Jae-Mo. Deep Learning Ensemble-Based Automated and High-Performing Recognition of Coffee Leaf Disease. Agriculture, 2022; 12 (11): 1-16.
13. Parraga-Alava J, Cusme K, Loor A, Santander E. RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition. Data Brief, 2019; 25:104414.
14. Paulos EB, Woldeyohannis MM. Detection and Classification of Coffee Leaf Disease using Deep Learning. 2022 International Conference on Information and

Communication Technology for Development for Africa (ICT4DA), Bahir Dar, Ethiopia, 2022;1-6. doi: 10.1109/ICT4DA56482.2022.9971300.
15. Sabrina SA, Maki WFA. Klasifikasi Penyakit Pada Tanaman Kopi Robusta Berdasarkan Citra Daun Menggunakan Convolutional Neural Network. eProc Eng 2022; 3: 1919.
16. Shah HA, Saeed F, Yun, S.; Park, J.H.; Paul, A.; Kang, J.M. A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Fine tuned EfficientNet. IEEE Access 2022; 10: 65426-65438.
17. Suhartono, Derwin, Aditya W, Lestari M, Yasin M. Expert System in Detecting Coffee Plant Diseases. International Journal of Electrical Energy, 2013; 1 (3): 156-162.
18. Voora V, Bermúdez S, Larrea C, Balino S. Global Market Report: Coffee. Winnipeg, Manitoba, Canada 2019.