Enhancing Coffee Leaf Rust Detection Using DenseNet201: A Comprehensive Analysis of the Mbozi and Public Datasets in Songwe, Tanzania

Adrian Jackob Karia[1]
Juma Said Ally[2]
Stanley Leonard[3]

[1]tajirikaria@gmail.com (+255 757101182)
[2]juma.ally@must.ac.tz
[3]msuya@must.ac.tz

[1,2,3]Mbeya University of Science and Technology, Tanzania

https://doi.org/10.51867/ajernet.6.1.17

ABSTRACT

Coffee Leaf Rust (CLR) is a worldwide devastating fungal disease that threatens coffee production, upsetting economic and farmers' livelihoods. Traditional methods of detecting CLR heavily rely on using machine-learning (ML) models trained through weakly collected datasets and physical inspection which is tedious, time-consuming, and subject to human error. This study explores the performance of the DenseNet201 model using three datasets: Mbozi, Public, and Combined (a merger of Mbozi and Public datasets). Machine Learning Theory guided this research. The study objective is to assess the influence of dataset quality on CLR detection, analyze Mbozi and Public datasets using DenseNet201, and enhance robustness by merging the two datasets. A study on coffee leaf rot (CLR) severity was conducted using systematic sampling techniques. Leaves from multiple coffee farms were collected, representing different levels of infection. The Mbozi dataset, sourced from high-resolution images captured from Tanzania's Songwe coffee plantations, was analyzed for quality under controlled conditions, including environmental factors, image clarity, resolution, labeling consistency, and class balance, based on data completeness, image quality score, visual inspection, and model performance. DenseNet201 was trained and validated on each dataset achieving its highest accuracy with the Mbozi dataset at 98.72% and a validation accuracy of 97.65%, demonstrating the importance of consistent image quality and accurate annotations. In contrast, the public dataset suffered from inconsistencies in resolution and labeling, resulting in a lower training and validation accuracy of 96.86% and 96.42% respectively. The Combined dataset, which integrated the strengths of both datasets, exhibited a stronger generalization with an accuracy of 97.48% and validation accuracy of 97.49%, balancing the need for high-quality images with environmental variability. The study shows improved CLR detection speed and accuracy due to high-quality and consistently labeled images from the Mbozi dataset. It recommends future models integrate regionally relevant and high-resolution datasets for robust performance in real-world agricultural conditions, providing coffee farmers with timely disease intervention tools for better production management and economic stability in coffee-growing regions.

Keywords; Coffee Leaf Rust (CLR) Detection, Dataset Quality, DenseNet201, Image Quality, Machine Learning (ML)

I. INTRODUCTION

Coffee is an important cash crop and one of the most widely planted globally (Kiwelu et al. 2021). Tanzania's economy relies heavily on agriculture, contributing 27% of the Growth Domestic Product (GDP), 4% of export earnings, and 65% of employment in 2020 (Mbwambo et al., 2020). Food crops mainly include maize, cassava, rice, sorghum, millet, and bean, while cash crops include cotton, coffee, tea, cashew nuts, and sisal. Coffee contributed about 0.4% of Tanzania's GDP in 2023 and sustains over 2.4 million people. It is Tanzania's second most exported agricultural crop after tea (Anania & Nade, 2020).

Coffee Leaf Rust (CLR), a pervasive disease caused by the fungus *Hemileia Vastatrix* poses a significant threat to coffee production worldwide, affecting yield, quality, and farmer livelihoods (Soares et al., 2022). Among the common mechanisms used to control pests and diseases is the symptomatic spraying of pesticides (Otieno et al., 2019). This can control the spread of diseases and minimize losses. Disease control must discover and judge the disease type in time and select suitable pesticides for precise treatment. Early detection of CLR is crucial for mitigating its effects, as late-stage infections can drastically reduce crop yields and lead to significant economic losses (Mbwambo et al., 2020) In Tanzania, measures to prevent and control diseases and pests in plants are governed by The Plant Health Act 2020, which plays a major role in improving productivity.

The use of emerging technology for early detection and prevention of the disease is vital for greater quality production (Leandro et al., 2021). Developing efficient ML models for early detection and accurate diagnosis of CLR is important to mitigate its effects (Nawaz et al., 2024).

Observing plant leaves for disease diagnosis requires expertise and continuous monitoring, often failing even for experienced pathologists and agronomists (Kiwelu et al., 2021). Tanzania's government, through the Tanzania Coffee Research Institute (TACRI), introduced CLR-resistant hybrid varieties and distributed fertilizers and pesticides, but farmers still face challenges due to inaccessibility and insufficient extension services (Kiwelu et al., 2021).

Recently, researchers have focused on ML for controlling and monitoring plant diseases, with Convolutional Neural Networks (CNN) improving recognition efficiency and accuracy compared to traditional (Magomba & Ng'atigwa, 2024). Ju et al. (2023) discuss the Identification of Fruit Tree Pests with Deep Learning on embedded drones to achieve accurate pesticide spraying. Since the proposal of LeNet in 1998, CNN has developed continuously upgraded models such as AlexNet in 2012, Google Net in 2014, and ResNet in 2015 (Ruben et al., 2018). Many novel CNN models are also being proposed to apply in the field of plant classification (Bakr et al., 2022). In practice, the diseased coffee image leaves dataset is affected by noise and background clutter which makes it difficult to classify. To address these issues, a quality dataset obtained from Mbozi is analyzed and compared with the Public dataset by using DenseNet201. is needed.

However, high-quality datasets for effective machine-learning models remain challenging (Jepkoech et al., 2021).This study focuses on improving the dataset's quality for CLR disease detection by converging on environmental conditions, image resolution, and metadata integration. The core mission remains to enhance the performance of ML models and their capacity to detect CLR in real-world agricultural settings.

1.1 Statement of the Problem

The effectiveness of ML techniques in agricultural disease detection, especially in identifying coffee leaf rust (CLR), is heavily reliant on the quality and diversity of the image datasets used for model training and validation (Muharromah et al., 2024). CLR, a disease that significantly impacts coffee production, presents specific visual symptoms such as rust-colored spots, yellowing leaves, and premature leaf drop (Cabrera et al., 2024). Capturing these symptoms accurately across various growth stages and environmental settings is essential for building robust ML models capable of reliably diagnosing CLR (Tan et al., 2018). However, a major challenge in CLR detection is the lack of high-quality, comprehensive datasets that represent these symptoms under diverse environmental conditions.

Existing datasets for CLR detection are often limited in resolution, scope, and environmental variability, impeding ML models' ability to generalize effectively to new and unseen conditions, especially when applied outside controlled study settings (Yebasse et al., 2021). In real-world agricultural environments, factors such as lighting variability, background noise, and regional differences in leaf morphology further complicate the detection process, as these elements introduce significant visual variability in images (Hitimana et al., 2023). Models trained on less diverse or lower-quality datasets tend to struggle with accurately identifying CLR symptoms when deployed in dynamic agricultural settings, often yielding inaccurate or inconsistent predictions (Prince et al., 2024).

A study by Kiwelu et al. (2021), pronounces the challenges encountered by farmers in the specific context of the Mbozi coffee plantations in the Songwe Region of Tanzania. The Mbozi region presents unique environmental conditions and crop characteristics that are underrepresented in existing CLR datasets (Otieno et al., 2019). This study seeks to address these data limitations by collecting high-resolution, diverse image datasets from Mbozi's coffee plantations, capturing CLR across various stages and under different environmental factors (Magomba & Ng'atigwa 2024). These newly acquired images are integrated with publicly available CLR datasets to enhance data quality, increase variability, and improve model generalizability.

To maximize detection accuracy and efficiency, this study employs DenseNet201, a convolutional neural network architecture known for its efficient use of parameters and deep feature extraction capabilities (Mengistu et al., 2016). DenseNet201's dense connectivity enables it to capture complex, nuanced features indicative of CLR, making it a promising choice for overcoming challenges associated with variable lighting, leaf texture, and background interference. By leveraging a richer, more diverse dataset and DenseNet201's advanced architecture, this study aims to improve CLR detection accuracy and model robustness, facilitating the deployment of reliable disease detection models that are adaptable to the specific agricultural contexts in the Mbozi's Songwe region, Tanzania and beyond.

1.2 Research Objective

The study objective is to assess the influence of dataset quality on CLR detection, analyze Mbozi and Public datasets using DenseNet201, and enhance robustness by merging the two datasets.

II. LITERATURE REVIEW

2.1 Theoretical Review

2.1.1 Machine Learning Theory

To detect CLR, this study employs ML theoretical framework that combines ML, image processing, and environmental science. The study underlines the importance of high-quality and diverse training data, as well as ML models' capacity to generalize to previously encountered data (Ruben et al., 2018). The goal is to develop a robust, region-specific dataset from Tanzania's Mbozi region, which will improve generalizability and reduce overfitting. Deep CNN Theory is especially important, as CNN architectures such as DenseNet201 may capture fine-grained details and complex patterns linked with CLR (Bakr et al., 2022).

Transfer Learning Theory allows pre-trained models to adapt to specific tasks, even with smaller datasets. DenseNet201, trained on extensive datasets like ImageNet, effectively generalizes with limited labeled agricultural data, especially in agricultural machine-learning applications (Banothu et al., 2024). Pattern Recognition Theory is crucial for CLR detection, as CNNs excel at learning visual and textural patterns that differentiate healthy from diseased leaves, making them ideal for CLR detection.

Image Processing Theory and Ecological and Environmental Theory are used to address challenges in agricultural environments, such as lighting variability and plant morphology differences (Pujiastuti et al., 2023). Data augmentation techniques, such as rotations and brightness adjustments, enhance dataset diversity and performance in complex settings. Understanding the unique environmental factors of the Mbozi region, such as climate, soil type, and agricultural practices, is crucial for adapting the model to Tanzania's specific challenges.

2.2 Empirical Review

In one study by Novtahaning et al. (2022) explores the development and deployment of a mobile application to detect and classify foliar damage on Arabica coffee leaves. The study employs CNN models, specifically EfficientNet-B0, ResNet-152, and VGG-16, to perform the classification task. Although these architectures are known for their high accuracy in image classification, the study's reliance on a relatively small dataset of 1,747 images restricts the model's capacity to generalize effectively to diverse damage patterns and environmental backgrounds. Consequently, the model may struggle to maintain accuracy when applied in varied, real-world farm conditions, where leaf appearances and environmental factors may differ significantly from the controlled dataset (Soares et al., 2022).

In another study, Chavarro et al. (2023) examine the impact of hyperparameters on coffee rust detection models by implementing multiple ML models, including back-propagation neural networks (BPNN), CNNs, and recurrent neural networks (RNNs). This study focuses on optimizing five specific hyperparameters to enhance detection performance. However, the study done by Lyimo et al. (2021) overlooks other influential factors, such as regularization and data augmentation, which are critical for improving the model's resilience and reducing overfitting in diverse environments. The absence of these elements may hinder the model's ability to adapt when faced with data that vary across seasons or coffee plant varieties, limiting its applicability across different farm settings.

Also, the study by Ju et al. (2023), introduces a framework that utilizes vegetation indices derived from UAV (unmanned aerial vehicle) imagery for CLR detection. This method employs algorithms such as artificial neural networks (ANN) and support vector machines (SVM) alongside segmentation techniques to isolate and analyze foliar damage using color spaces and fixed thresholds. While innovative, the framework is limited by its small sample size comprising only 16 plants across four plots which raises questions about its scalability and effectiveness in larger agricultural contexts. The segmentation approach, reliant on fixed thresholds, may also lack flexibility in adjusting to varying levels of disease severity or changes in leaf pigmentation under different lighting conditions.

Further contributions come from a study by Velásquez et al. (2021), in Colombia, combines wireless sensor networks, remote sensing, and deep learning models to detect CLR in the Caturra coffee variety. The ensemble algorithm utilized in this study merges a multi-layer perceptron with three CNN models to enhance diagnostic accuracy. Although this approach offers promising results for regulated farms, its implementation on larger, unregulated farms presents challenges due to limited sensor coverage and increased data variability. The study also highlights the significance of choosing specific hyperparameters, particularly backbone architecture and optimizer types, yet underscores the need for greater interpretability to understand model decision-making processes a critical feature for building trust in automated diagnostics.

In a Brazilian study, Pham et al. (2023), develop an algorithm and mobile application designed to detect CLR and coffee leaf miner damage in coffee plants, utilizing 285 manually captured images. The images, taken in controlled settings with a white background, limit the model's adaptability to actual farm environments where background variability and lighting conditions differ markedly. Although this approach is cost-effective and accessible, it may lack robustness for field application without further enhancements to handle real-world complexities.

Finally, a comprehensive literature review by Gichuru et al. (2021) examined the broader use of image-processing techniques for plant disease detection. These reviews emphasize the role of computer vision in agricultural diagnostics, showcasing advances in segmentation based on crop type and disease location. While these foundational insights have enabled progress in image segmentation, further research and model adaptation are necessary to tailor these techniques specifically to the detection of coffee diseases.

III. METHODOLOGY

The workflow in Figure 1 outlines a structured approach to dataset preparation, model training, and evaluation, focusing on the Mbozi dataset for predictive modeling tasks. The process starts with collecting and integrating the Mbozi dataset with other datasets to enhance model generalizability. Models like DenseNet201 are trained and evaluated on the Mbozi, Public, and Combined datasets, comparing key metrics to identify the best-performing model.

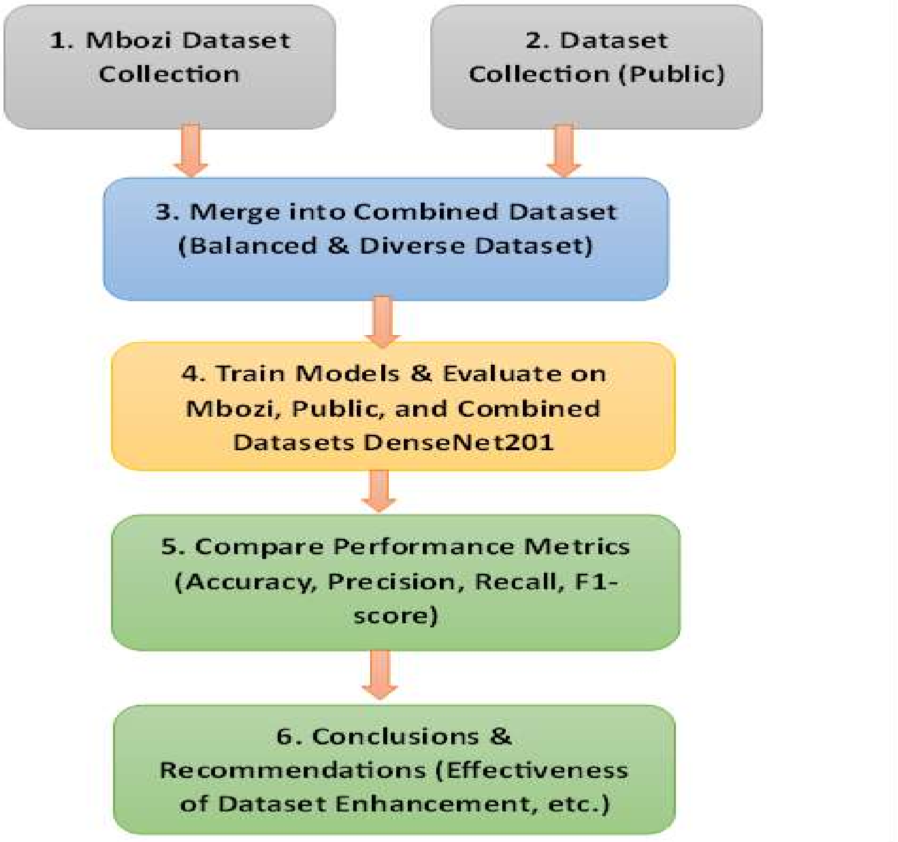


Figure 1
Methodology Flowchart

3.1 Study Area

The study area was captured from Google Earth as shown in *Plate 1*. The red boundary indicates several plots of coffee farms with varying ages and CLR sternness. The area is found near the Tanzania-Zambia highway in the Mbozi district with 1,888.9871419m, an altitude of 1,595.4459385m, the latitude is -9.0909790^0 and the longitude is 32.9554041^0. Mbozi was chosen for its prominence in coffee production and susceptibility to CLR outbreaks, providing a realistic setting for studying disease impacts.

Plate 1
Image of the study area in Mbozi, Songwe, Tanzania
Source: Google Earth Pro (Version 7.3.3), ©2024 (accessed June 21, 2024)

3.2 Sampling Techniques

A study on coffee leaf rot (CLR) severity was conducted using systematic sampling techniques. Leaves from multiple coffee farms were collected, representing different levels of infection. The researchers collaborated with agricultural officers and farmers to identify representative plots within the Mbozi district. 6,018 images were collected, with equal representation across the four CLR severity categories. Sampling was conducted under controlled conditions, using natural daylight between 10:00 AM and 3:00 PM. The study included leaves from trees of varying ages and farm locations, ensuring diversity in the dataset. White backdrops minimized noise and background interference, maintaining consistency in image quality. This approach ensured a balanced dataset with diverse leaf conditions, enhancing the robustness of the ML model.

3.3 Assessment of CLR Disease

Categorizing coffee leaves assessed CLR severity into four distinct groups: healthy, initial, moderate, and severe infection stages visual inspection served as the primary method for identifying and classifying CLR symptoms. Researchers focused on the characteristic yellow-orange rust spots and lesions typically found on the undersides of coffee leaves. This approach enabled accurate classification of CLR progression, ensuring that the dataset reflected the diversity of disease manifestations observed in real-world conditions.

The classification process was carried out in collaboration with agricultural extension officers, leveraging their expertise to verify the accuracy of CLR identification. Each leaf was photographed under controlled conditions, ensuring consistent lighting and background settings to enhance image clarity. The high-quality images allowed the study to capture subtle differences between the four CLR severity categories, providing a robust foundation for ML model development.

By systematically categorizing CLR severity, the study ensured that the dataset encompassed the full spectrum of disease progression. This comprehensive assessment not only improved the dataset's reliability but also enhanced the ML model's ability to generalize across varying levels of CLR infection.

3.4 Data Collection
3.4.1 Data Collection Techniques

Data collection was conducted in June 2024 under controlled environmental conditions to replicate the natural habitat of coffee farms in Mbozi. Key environmental parameters such as atmospheric pressure (86 kPa), temperature (20°C), relative humidity (75%), and light intensity (500 μmol/m²/s) were monitored to ensure consistency. High-resolution images of coffee leaves were captured using a 13-megapixel smartphone camera, with a focus on the underside of leaves where CLR symptoms are most visible.

To maintain image quality, natural daylight between 10:00 AM and 3:00 PM was used, and white backdrops were employed to minimize background noise. Any blurry or poorly lit images were discarded during the review process, ensuring that only high-quality images were included in the dataset. A structured folder system was used to organize images for easy retrieval during preprocessing.

Collaboration with local farmers and agricultural officers facilitated the identification of representative plots and trees, ensuring a diverse and balanced dataset. The systematic approach to data collection minimized variability and noise, providing a consistent dataset for CLR detection. The total of 6,018 images captured reflects the diversity in infection severity and environmental conditions across the Mbozi region.

(a) Healthy leaf *(b) CLR-initial* *(c) CLR-moderate* *(d) CLR-severe*

Plate 2
Illustrates Image Samples of the Diversity of Coffee Leaves for the Mbozi Dataset.

3.4.2 Image Capturing

To ensure uniformity in image quality, all images were captured under controlled lighting conditions using continuous daylight and neutral white backdrops. This setup minimized the influence of external environmental factors, such as inconsistent lighting, which could obscure key characteristics of Coffee Leaf Rust (CLR) on the leaves (Al-Rashidi et al., 2024). The neutral backdrop helped reduce background noise and focused attention on the leaves' features. After capturing, the images were thoroughly reviewed for consistency in resolution and clarity. Any blurry or poorly lit images were discarded from the dataset to maintain high-quality standards and ensure uniformity across all samples.

3.4.3 Dataset Integration and Preprocessing

To create a larger and more diverse training set, the Mbozi dataset was integrated with a publicly available dataset sourced from Kaggle. The public dataset consisted of images of coffee leaves from various Ethiopian coffee plantations, which provided a broader spectrum of leaf conditions and environmental factors (Arathi & Dulhare, 2023). However, inconsistencies in image quality, resolution, and labeling accuracy were evident in the public dataset. These variations posed a challenge to dataset uniformity and model training. To address these issues, a preprocessing pipeline was applied to both datasets. This pipeline standardized image properties, including:

Resolution: Ensuring all images were resized to a consistent resolution.

Size: Uniform dimensions were enforced to prevent variations that could affect model input.

Labeling consistency: Labels from the public dataset were reviewed and corrected where necessary to ensure they aligned with the labels in the Mbozi dataset.

Table 1
Illustrates the Number of Images in each Set from Datasets

Dataset	Number of Images used for	
	Training	Validation
Public	4590	810
Mbozi	5112	902
Combined	9031	1593

3.4.5 Comparing Datasets: Mbozi vs. Public

We compare three datasets: the Mbozi dataset, a publicly available dataset, and a Combined dataset created by merging the two. The Mbozi dataset, collected from Tanzania's coffee plantations, offers high-resolution images captured under controlled lighting conditions, reducing noise that could interfere with ML models' ability to identify CLR symptoms. Public datasets, sourced from various global regions, often suffer from image quality, resolution, and labeling inconsistencies. These differences highlight the importance of high-quality, controlled image capture for CLR detection. The public dataset may introduce biases that could hinder model performance, such as over-representation of certain types of CLR infections or specific environmental conditions. The Mbozi dataset, with its controlled collection methodology, reduces these biases, providing a more reliable foundation for training ML models.

3.5 Model Selection and Training

Model Justification:

DenseNet201 was chosen for this study due to its superior performance in image classification tasks, particularly for datasets with high-resolution images and fine-grained visual details (Tan et al., 2018). In the study of Chavarro et al. (2023), DenseNet201 outperformed other models such as ResNet50 and InceptionV3 in accuracy, demonstrating better feature propagation through its densely connected layers (Hitimana et al., 2023). This dense connectivity was especially beneficial in identifying subtle features of CLR that could be missed by models with simpler architectures (Tan et al., 2018).

Training and Validation:

The Combined dataset (Mbozi + Public) was split into training (85%) and validation (15%). The model was trained independently on each dataset (Mbozi, Public, and Combined) to assess its performance across different data sources.

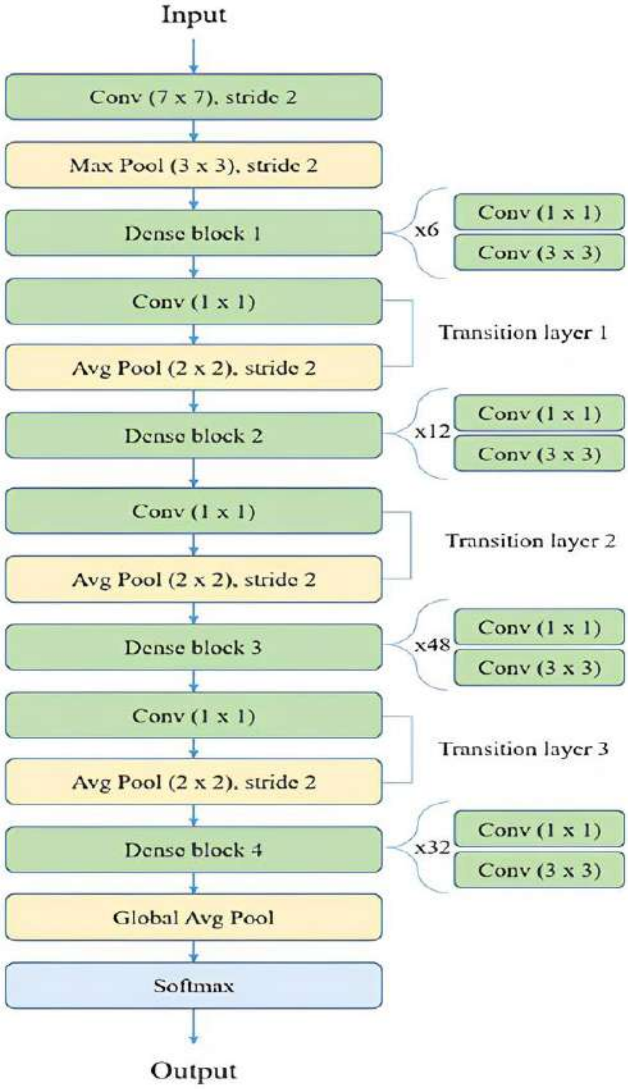


Figure 2

Illustrates the DenseNet201-Architecture (Bakr et al. 2022)

3.6 Tools and Experimental Setup

All experiments were conducted on a personal computer running Windows 11 Pro. The machine had an AMD Ryzen 5 PRO 2500U processor, 24GB RAM, and used the Keras library in Python 3. Hyperparameter tuning was conducted using the grid search method, optimizing parameters such as learning rate, batch size, and dropout rates (Chavarro et al., 2023). The TensorFlow backend was used for deep learning model development. The specification of tools and data obtained is shown in *Table 2.*

Table 2

Illustrate Specifications of the Mbozi Dataset

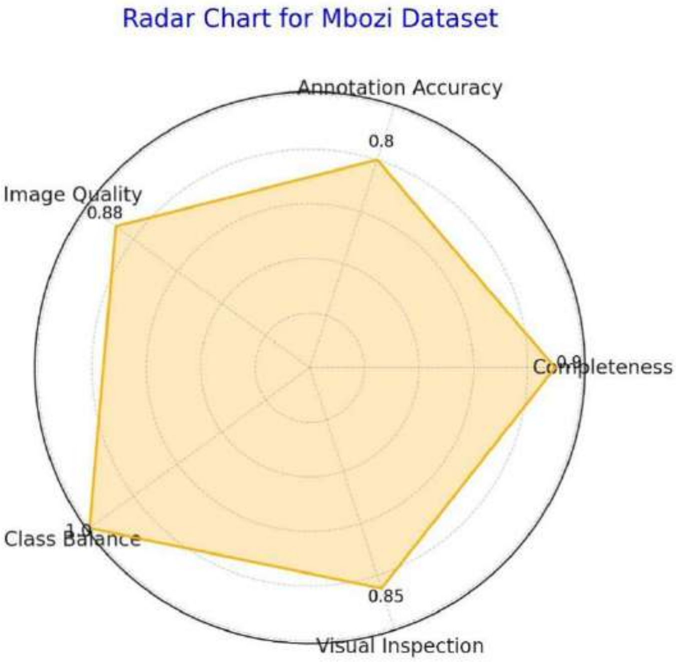
Subject	**Applied Machine Learning**
Particular Domain	Coffee Leaf Disease (CLR) Image detection using computer vision techniques
Kind of data	Image
How data was obtained	The study utilized 13-megapixel Infinix Hot 10 smartphone cameras to capture images of coffee leaves in the field, categorized as healthy or impacted by CLR, with researchers, farmers, plant pathologists, and agricultural extension officers overseeing the data collection procedure.
Data format:	Raw
An explanation of data Collection	Images were acquired in the field over one month. The goal was to create a high-quality collection of coffee leaf disease diagnostics, with a focus on one recognized disease that has a significant impact on production. The disease's name was discovered by glancing at the caption for the coffee image sample.
Where the data source is located	• Town/Region: Mbozi, Songwe as shown in *Figure 2* • Country: Tanzania.

3.7 Dataset Analysis and Evaluation

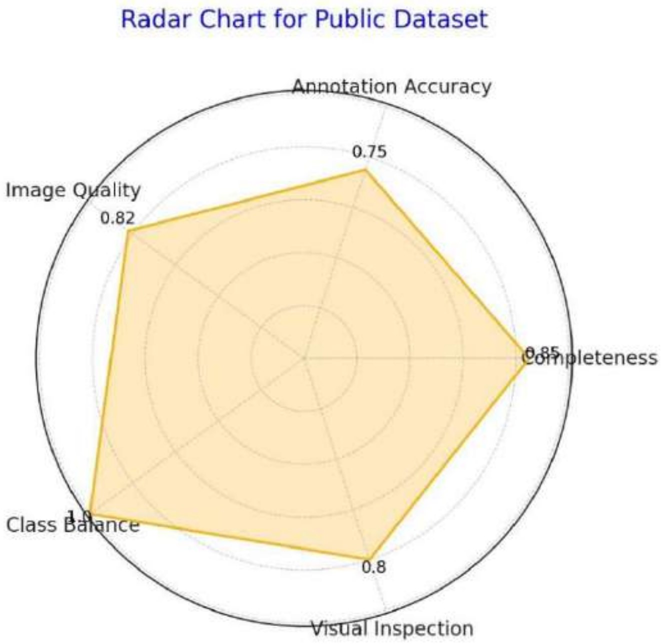
The Mbozi dataset showed superior quality in performance measures like annotation correctness, image quality, class balance, and completeness, while the public dataset had inconsistencies that required preprocessing. The Mbozi dataset, analyzed using the Public and Combined datasets, showed significant performance differences in key metrics, affecting the effectiveness of ML models as depicted in *Figure 3(a)*. The dataset had a completeness score of 0.90, a high image quality score of 0.88, and a perfect class balance score of 1.00, preventing model bias.

The Public dataset shows slightly lower performance in metrics like annotation accuracy, image quality, and class balance, but a higher proportion of medium and low-quality images, potentially affecting model performance, is observed in *Figure 3(b)*.

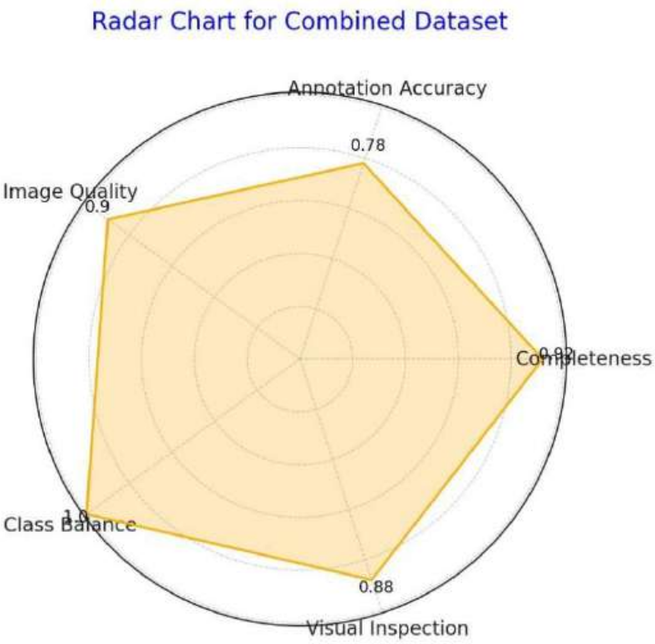
The Combined dataset, which combines Mbozi and Public datasets, performs well in most metrics, with an annotation accuracy of 0.78 and an improved image quality of 0.90 as illustrated in *Figure 3(c)*. The high-quality images from Mbozi offset the lower-quality images from the Public dataset, and the visual inspection score of 0.88 indicates high-quality images. This combination achieves a balance between completeness and image quality, making it a strong candidate for enhancing model robustness and generalization.



(a)



(b)



(c)

Figure 3
Illustrates Radar Charts Overall Summary of all Metrics in Each Dataset

3.7.1. Data Completeness

Data completeness refers to the completeness of a dataset, including images and annotations, which is crucial for effective training and analysis. The Mbozi dataset has higher completeness than the Combined dataset, emphasizing the importance of data integrity when integrating multiple sources to quantitatively assess data completeness, *equation (1)* can be used.

$$Data\ Completeness = \left(\frac{Total\ Number\ of\ Complete\ Samples}{Total\ Number\ of\ Samples}\right) \times 100 \ \ldots\ldots \ \textbf{(1)}$$

The formula in *equation (1)* calculates the percentage of the dataset that is fully annotated and ready for analysis. A dataset with higher completeness ensures that it is fully exploitable for accurate and reliable training of ML models.

3.7.2 Image Quality

Image quality is crucial for ML models, especially in fine-grained feature detection tasks. Mbozi and Combined datasets have high-quality images, enabling better learning of features like distinguishing between healthy and infected coffee leaves. However, the Public dataset has a smaller proportion of high-quality images, potentially affecting model performance due to compression artifacts. Consistent lighting and sharpness are essential for maintaining image quality. Variations in lighting can introduce noise, leading to misclassifications. The Public dataset may require preprocessing steps to improve its quality.

3.6.3 Class Balance

To eliminate bias during training a balanced dataset was used. The imbalanced datasets may overfit the majority class while underperforming the minority class leading to poor generalization (Tan et al., 2018). The Mbozi, Public, and Combined datasets all exhibit similar class distributions, minimizing the risk of model bias. This balanced class representation is crucial for high accuracy, especially in tasks like disease classification, where accurate identification of both healthy and diseased samples is critical. The class balance for each dataset can be calculated using *equation (2)*.

$$\textit{Class Imbalance Ratio} = \frac{\textit{Size of Smallest class}}{\textit{Size of Largest Class}} \quad \ldots\ldots\ldots\ \textbf{(2)}$$

The analysis shows that all datasets have balanced class distributions as shown in *Figure 5*, ensuring fair and unbiased model performance. This balance allows for models to generalize across different leaf conditions, effectively identifying Coffee Leaf Rust (CLR) infections and healthy leaves without bias toward one category. A score close to 100% indicates a well-balanced dataset.

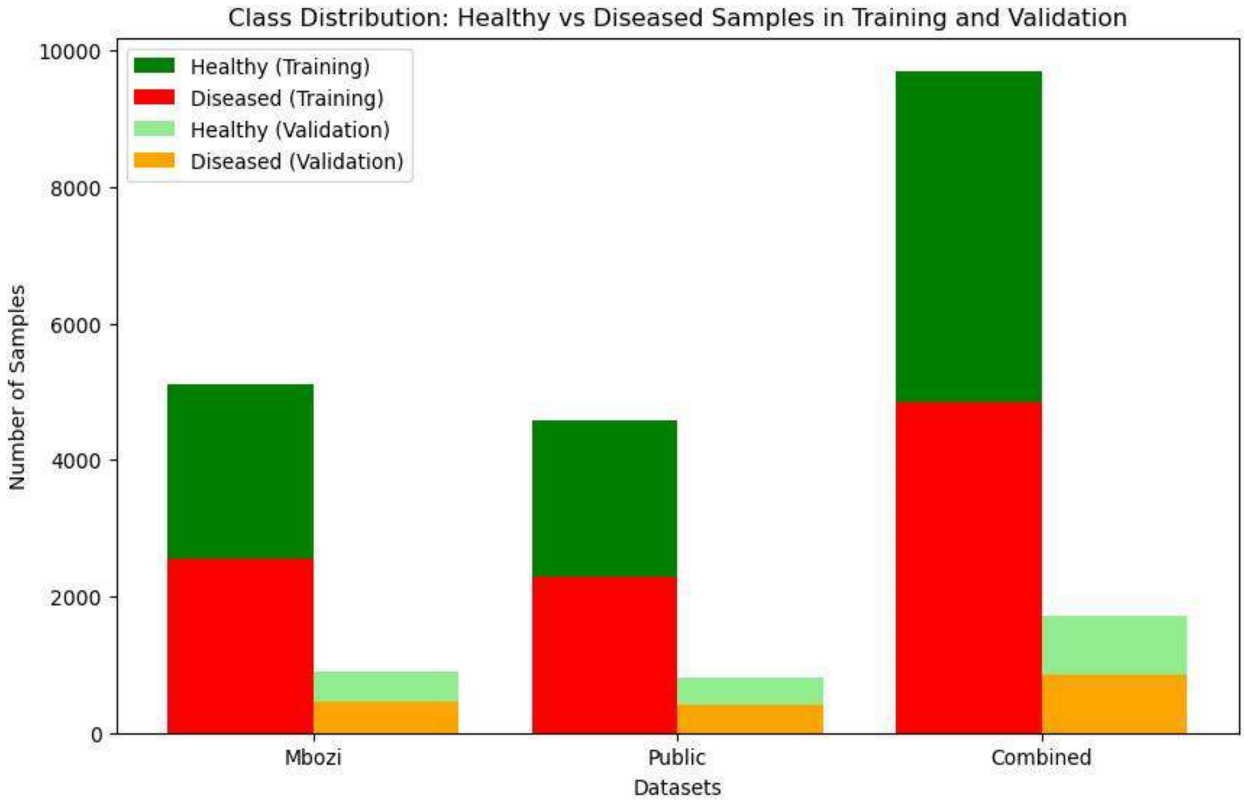


Figure 4
Shows Class Distribution across All Datasets

3.7.4 Annotation Accuracy

Annotation accuracy is crucial for the effective training and evaluation of ML models. The quality of annotations directly impacts the model's ability to learn the correct relationships between features and labels. The Mbozi dataset has high annotation accuracy due to a controlled data collection process, while the Combined dataset also shows high accuracy. However, minor inconsistencies from the Public dataset slightly reduce its overall annotation accuracy. The Public dataset has a significant number of misclassifications and missing annotations, reducing its overall quality. Misclassifications occur when labels are assigned incorrectly, leading to incorrect featurelabel relationships during training. Missing annotations can result in underutilization of the dataset, lowering model performance and generalization ability.

3.7.5 Visual Inspection

Visual inspection is a crucial step in ensuring the consistency of a dataset, preventing inconsistencies that automated metrics might miss. It helps to identify anomalies that may impact the model's learning process. The Mbozi and Combined datasets have high visual quality, maintaining consistent lighting, high resolution, and uniform image characteristics. This helps ML models extract features and classify Coffee Leaf Rust symptoms better. However, the Public dataset has a larger number of poor-quality images, exhibiting issues like inconsistent lighting, resolution degradation, and artifacts. These visual inconsistencies can introduce noise into the model training process, making it harder to focus on key features for disease detection.

3.7.6. Model Performance

The performance of ML models trained on different datasets was influenced by their quality. The Mbozi dataset provided the highest accuracy of 97.90% due to high-quality images, well-labeled annotations, and consistent environmental controls. This allowed the model to effectively learn the distinguishing features of CLR and healthy leaves. The Public dataset, despite its variability and inconsistencies, achieved a slightly lower accuracy of 97.06%. Mislabeled images and poor lighting and resolution likely contributed to this drop. The Combined dataset, which included a wider range of leaf conditions and environmental factors, achieved a performance of 97.55%, balancing the advantages of the Mbozi dataset with the variability introduced by the public dataset. The slight reduction in accuracy compared to the Mbozi dataset can be attributed to the inconsistencies carried over from the public dataset. Overall, the Mbozi dataset provided the highest accuracy due to its image quality, high annotation accuracy, and uniform conditions.

IV. FINDINGS & DISCUSSIONS

4.1 Model Performance and Correlation

In alignment with the study's objective to assess the impact of dataset quality on model accuracy, the DenseNet201 model was trained and evaluated on three distinct datasets: the Mbozi dataset, the Public dataset, and the Combined dataset, as shown in *Figure 6* and values scored in *Table 3*. The results indicate that dataset quality and consistency (independent variables) significantly influence model performance (dependent variable), particularly in the image-based detection of coffee leaf rust (CLR).

The Mbozi dataset achieved the highest training accuracy at 98.72%, with a validation accuracy of 97.65%. These results underscore the role of high-resolution images, consistent lighting conditions, and precise labeling in enhancing model accuracy during both training and validation. The performance spoken by Novtahaning et al. (2022), who found that high-quality images reduce variability and noise, facilitating better feature extraction and learning. The controlled environmental conditions of the Mbozi dataset support the model's generalization capacity, allowing it to identify CLR symptoms with greater accuracy and reliability.

Conversely, the model's performance on the Public dataset was lower, with a training accuracy of 96.86% and a validation accuracy of 96.42%. This discrepancy can be attributed to inconsistencies in the dataset, such as variations in image resolution, lighting, and potential labeling inaccuracies, which introduce noise and hinder learning. The results are mentioned by Chavarro et al. (2023), who noted that data inconsistency leads to overfitting and reduced generalization, particularly in diverse real-world conditions.

When trained on the Combined dataset, which integrates the high-quality images from Mbozi with the diverse environmental contexts of the Public dataset, the model achieved a balanced training accuracy of 97.48% and validation accuracy of 97.49%. This outcome indicates that merging datasets with complementary characteristics can enhance the model's robustness, particularly for applications where generalization to varied environments is essential as suggested by Koutouleas (2023).

4.2 Analysis of Dataset Impact on Model Performance

The study's findings highlight the significant impact of dataset quality on ML performance with the Mbozi dataset, which has high-resolution images, controlled lighting conditions, and precise labeling, consistently outperforms the other three datasets in ML performance. This is due to its high annotation accuracy and minimal noise, allowing the model to achieve a validation accuracy of 97.65%, the highest among the three datasets. As discussed by Ju et al. (2023) datasets with clear, well-labeled images improve model accuracy and reliability by reducing variability.

In contrast, the Public dataset's lower performance (96.42% validation accuracy) can be linked to inconsistencies in image quality and labeling biases, as well as the presence of compression artifacts. These findings reflect those of Velásquez et al. (2021) who emphasized that environmental variability and noise in datasets complicate feature extraction, leading to reduced performance. This underscores the necessity of harmonizing image quality and consistent annotation as part of data preprocessing to enhance the usability of public datasets.

The Combined dataset, which balances the high-quality images from Mbozi with the diversity of the Public dataset, achieved a validation accuracy of 97.49%, illustrating enhanced generalization potential. This approach mentioned by Pham et al. (2023)who suggested that combining datasets from different regions can help mitigate individual dataset limitations, thereby improving model performance in real-world scenarios where environmental variability is inevitable.

4.3 Generalization and Practical Implications

The result of the study emphasizes the importance of dataset quality in achieving high model performance, especially in tasks like disease detection on crops. The Mbozi dataset, with its controlled collection methods and high-quality images, provided the best performance in the training and validation phases. The Combined dataset demonstrated a promising approach to improving model generalization by merging datasets with different characteristics, which could be valuable for real-world deployment. The Public dataset underperformed, but its inclusion in the Combined dataset improved the model's generalization. The findings suggest that data augmentation and preprocessing techniques, such as harmonizing image quality and consistent labeling, could further enhance model performance. The Mbozi dataset provided the highest accuracy, but the Combined dataset offers a balanced approach for real-world deployment.

Table 3

Illustrates training and Validation Accuracy for Each Dataset

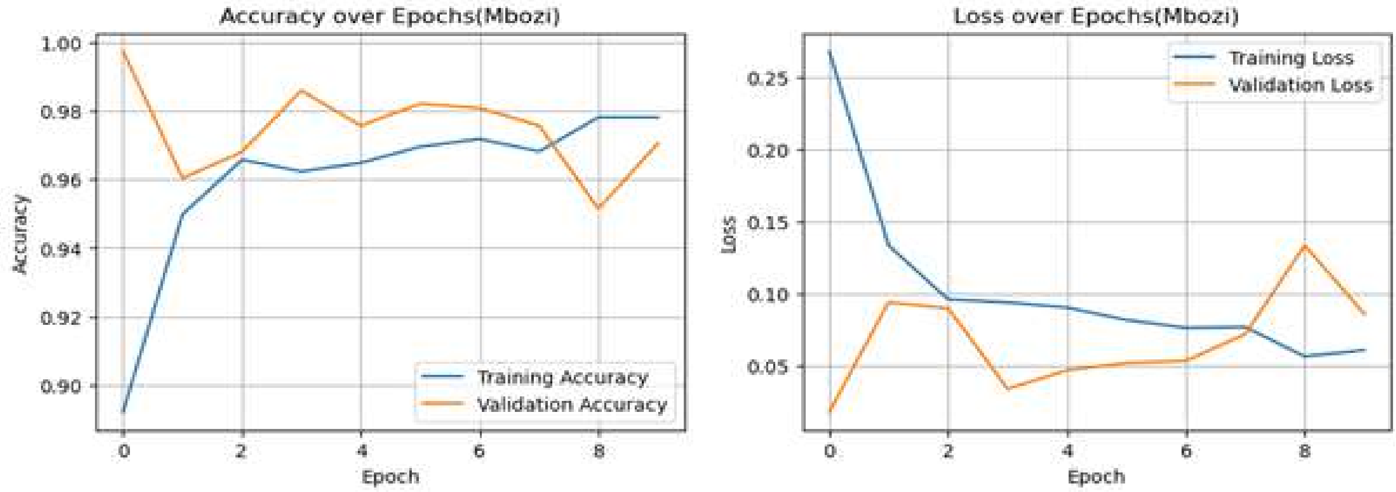
Dataset	Training Accuracy (%)	Validation Accuracy (%)
Mbozi	98.72	97.65
Public	96.86	96.42
Combined	97.48	97.49

4.4 Training and Validation Accuracy

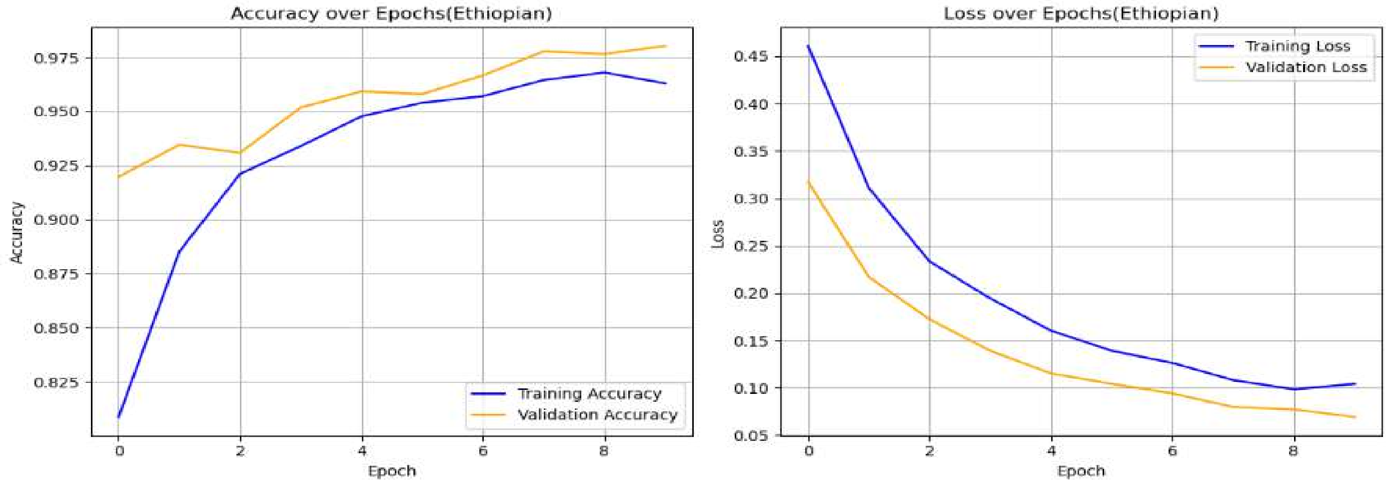
The analysis, as shown in *Figure 5* and *Table 3*, provides a detailed understanding of the DenseNet201 model's performance across different datasets. When trained on the Mbozi dataset *(Figure 5a)*, the model achieved a training accuracy of 98.72% and a validation accuracy of 97.65%. This strong performance reflects the dataset's high-quality images, minimal noise, and accurate labeling, which are essential for efficient feature extraction and learning.

The Public dataset, however, showed a lower training accuracy of 96.86% and the lowest validation accuracy of 96.42% *(Figure 5b*), which indicates that the inconsistent quality of images in the Public dataset negatively affected the model's performance. This finding is consistent with Gichuru et al. (2021) who pointed out that publicly available datasets often require quality control and preprocessing to ensure reliable model training.

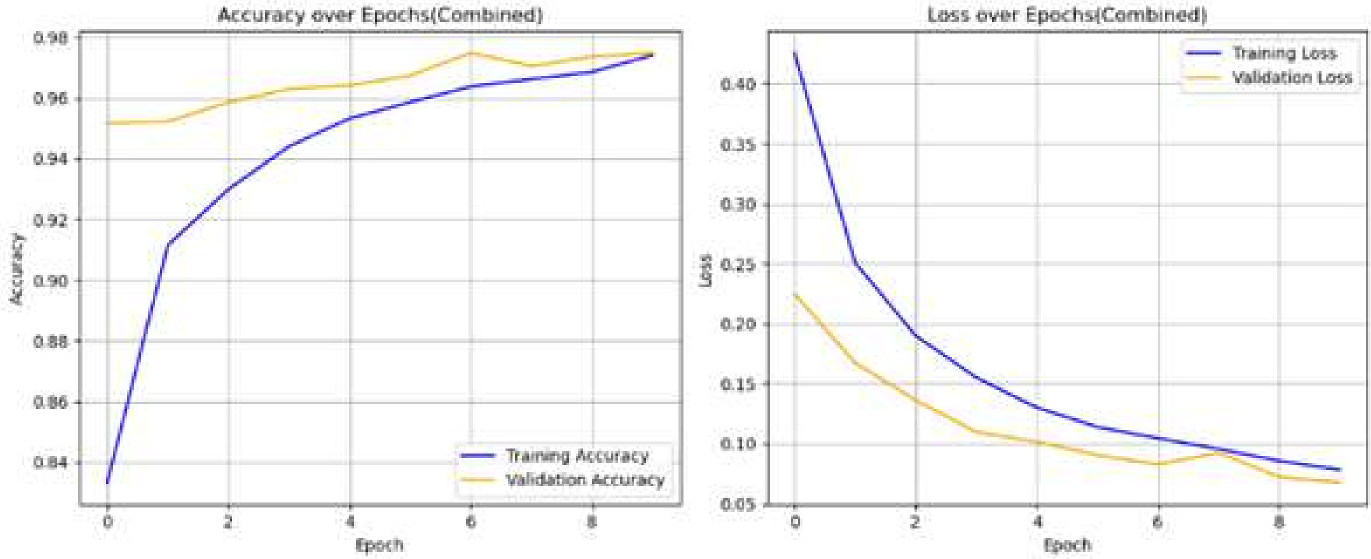
In contrast, the Combined dataset *(Figure 5c)* achieved balanced training and validation accuracies of 97.48% and 97.49%, respectively, suggesting that merging datasets from diverse sources can improve model robustness and adaptability. This result is particularly relevant for applications requiring broad generalization capabilities, as discussed by Jayaprakash and Balamurugan (2021)who highlighted the benefits of mixed datasets for improving model reliability in diverse real-world conditions.



(a)



(b)



(c)

Figure 5
Illustrates model output visualization for training and validation of Mbozi (a), Public(b), and Combined(c) datasets respectively

4.5 Analysis of Dataset Impact on Model Performance

The analysis indicates that dataset quality directly impacts model performance, with the Mbozi dataset outperforming the Public dataset in all performance metrics. The Mbozi dataset's high-resolution images, controlled lighting, and accurate labeling allowed for efficient feature extraction and higher validation accuracy. In contrast, the Public dataset's inconsistent image quality and labeling biases introduced noise, reducing the model's ability to detect CLR accurately. This finding is also highlighted by Yebasse et al. (2021) who observed that datasets with high noise levels and image artifacts negatively impact model generalization.

The Combined dataset demonstrated the potential benefits of merging datasets. While it did not surpass the Mbozi dataset in performance, it achieved near-parity and showed stronger generalization than the Public dataset alone. This suggests that combining high-quality, controlled data with diverse, real-world data can mitigate the limitations of individual datasets, enhancing model adaptability to varying environmental conditions, a finding supported by Prince et al. (2024).

The Mbozi dataset, indicated in *Figure 6(a)*, with its high proportion of good images, is the most reliable for training ML models, resulting in superior model accuracy. In contrast, the public dataset in *Figure 6(b)* suffers from lower image quality, highlighting the need for preprocessing and quality control when using publicly available datasets. The Combined dataset offered an effective compromise, providing near-parity with the Mbozi dataset's accuracy while introducing greater generalizability. The outcome aligns with a study that suggests the integration of high-quality data with diverse conditions can improve model adaptability and robustness in variable environments (Nawaz et al., 2024). Visual inspection *(Figure 6c)* and image quality scores *(Figure 8)* further confirm that high-quality data positively impacts model performance, with the Mbozi dataset supporting better feature extraction and higher accuracy.

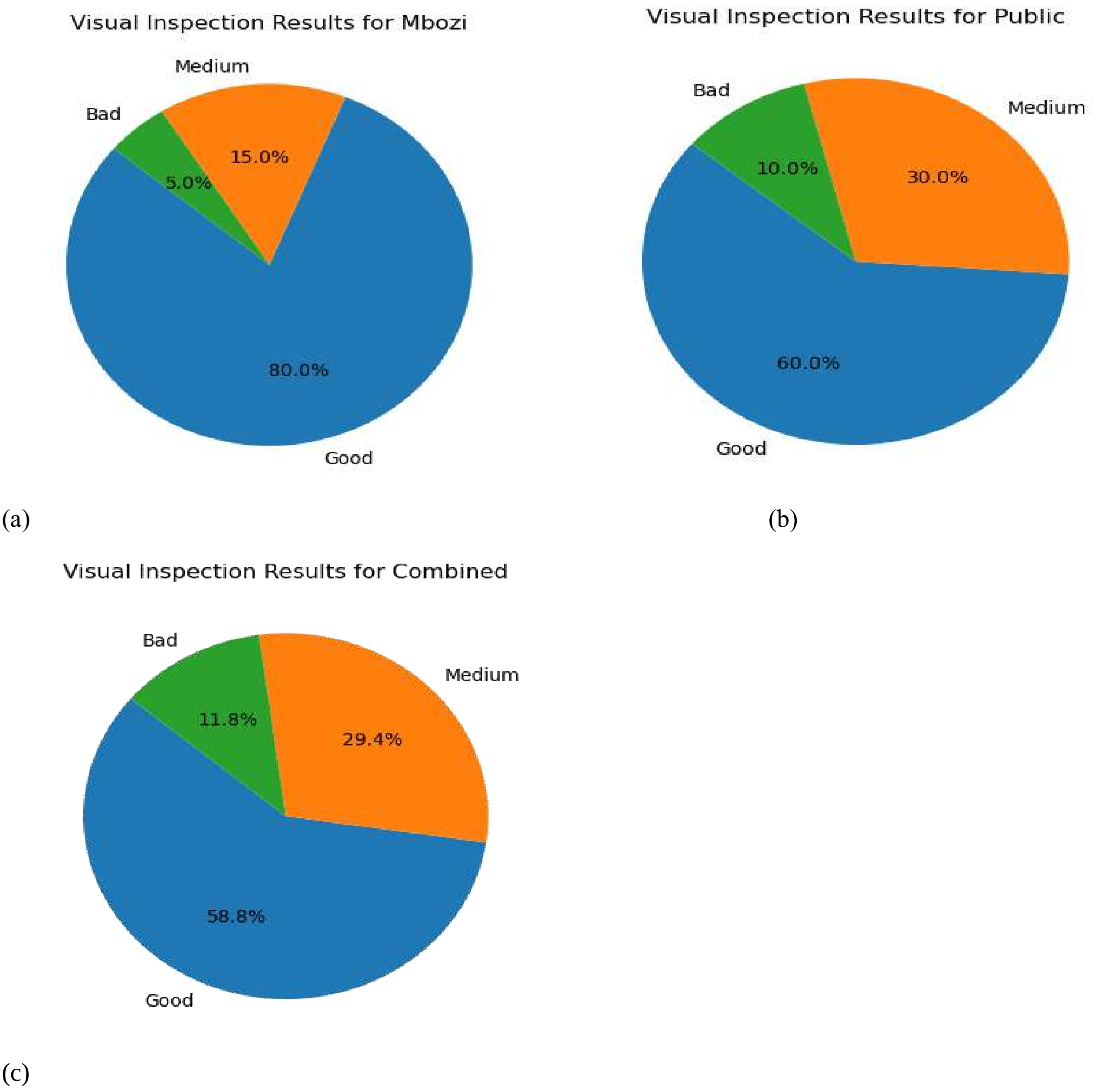


Figure 6
Pie Charts for Visual Inspection Results

The Mbozi dataset, known for its high image quality, is the optimal platform for training ML models. However, despite its lower mean image quality, the Public dataset offers valuable diversity in environmental and image conditions, as shown in *Figure 7*. The Combined dataset merges two datasets, enhancing image quality and diversity. This results in better generalization and robustness to real-world scenarios. The model retains high-quality images while introducing variability, suggesting that merging datasets can improve model robustness. This supports the literature suggesting that combined datasets offer more resilient model performance by balancing high-quality data with diverse environmental conditions (Jepkoech et al., 2021). The image quality score of Mbozi ranges between 80 and 90, while the public dataset has a mean average of nearly 80.

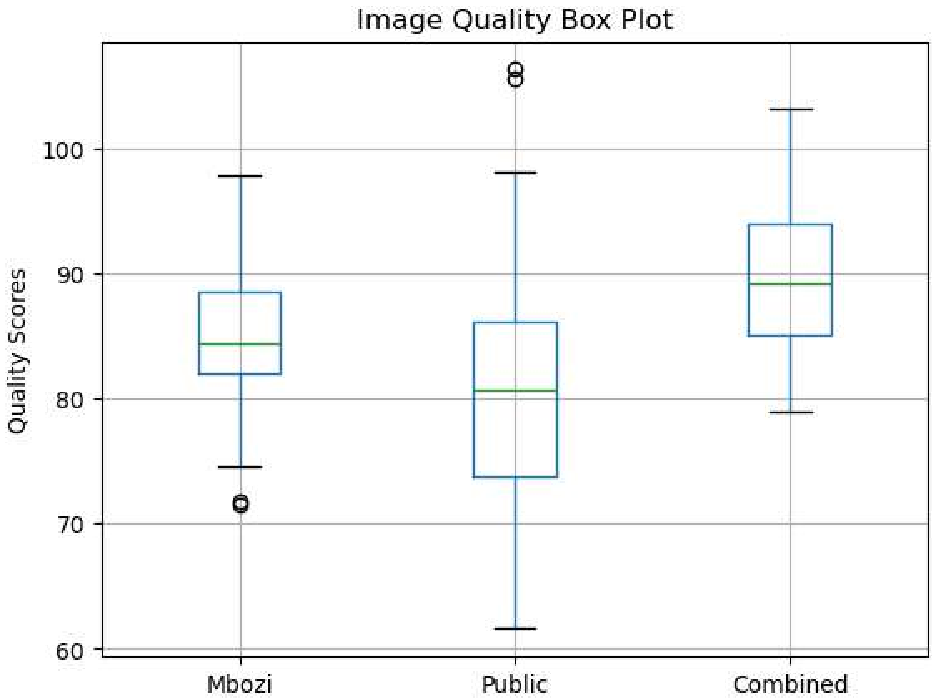
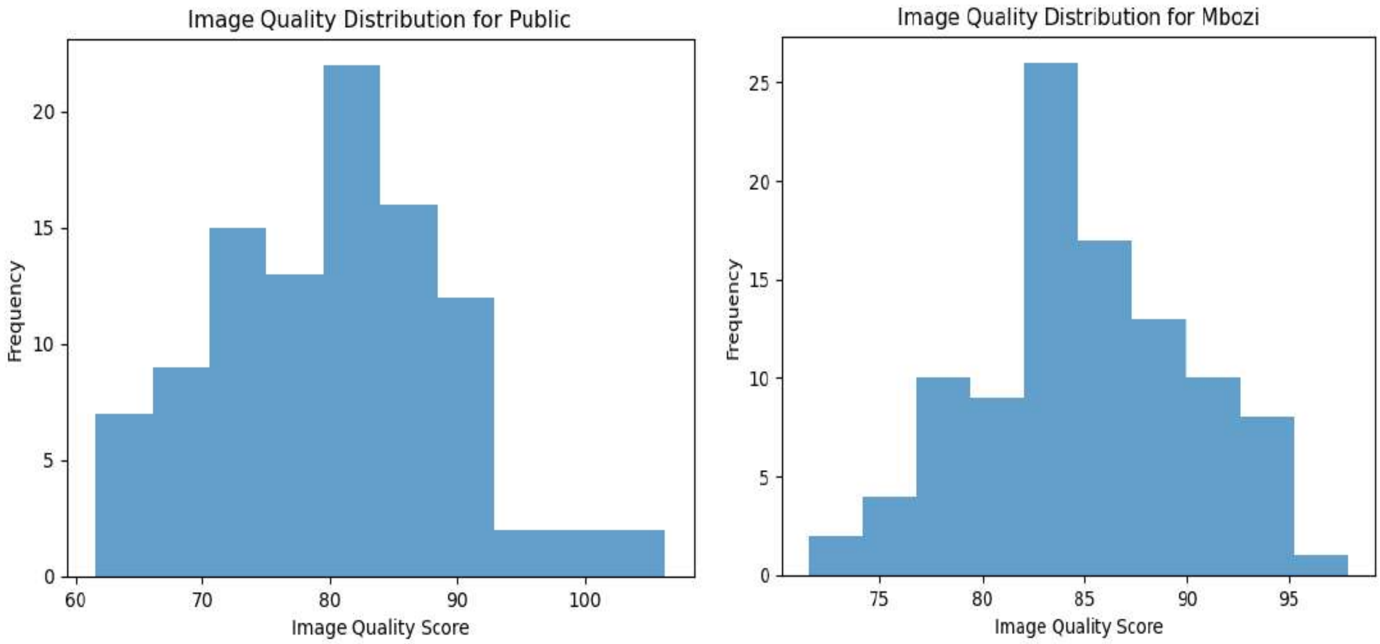


Figure 7
Image Quality Scores for all Datasets

Figure 8 shows the frequency of quality image scores in different datasets. The Mbozi dataset has a high frequency of 80-85, while the public has a frequency of exactly 80. The quality of the Mbozi increases the public's frequency to 85-90.



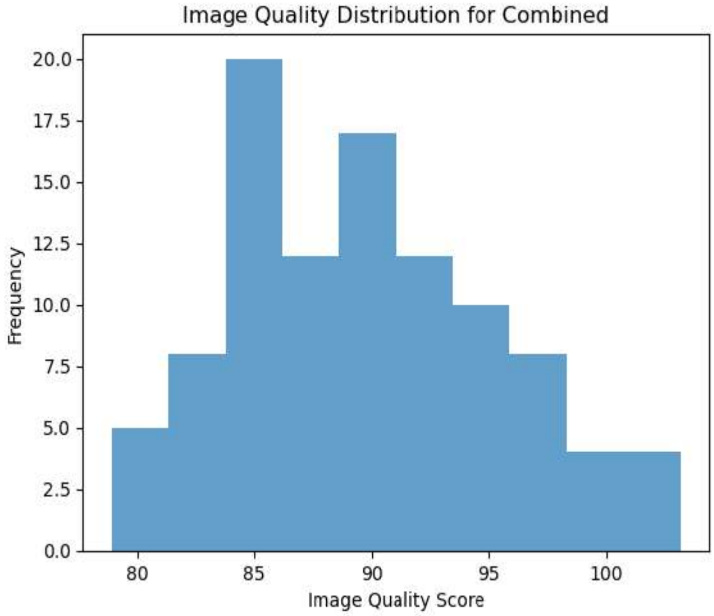


Figure 8
Illustrates Bar Charts for the Frequency Distribution of The Image Quality Scores on all Datasets

4.6 Trade-Offs Between Image Resolution and Model Efficiency

A key insight from this study is the trade-off between image resolution and model efficiency. High-resolution images, such as those in the Mbozi dataset, improved model performance but also increased computational costs due to processing time and memory demands. Conversely, the Public dataset's lower-resolution images required fewer resources, but this came at the expense of model accuracy and generalization. This trade-off has practical implications for scaling models in resource-constrained environments, such as agricultural fields where processing power and storage may be limited.

The results emphasize the importance of high-quality, well-labeled datasets for improving CLR detection performance. The Mbozi dataset showcases the benefits of controlled, high-resolution image collection, leading to more accurate and reliable predictions. Meanwhile, the Combined dataset illustrates that merging datasets from different regions can enhance generalization, though further refinement is needed to optimize performance fully. Moving forward, achieving a balance between image resolution and model efficiency will be critical for developing robust, scalable models for real-world agricultural applications.

V. CONCLUSIONS & RECOMMENDATIONS

5.1 Conclusion

The study emphasizes the significance of dataset quality in determining ML models' performance for CLR detection. The Mbozi dataset, with high-resolution images, minimal noise, and precise labeling, achieved the highest training and validation accuracies with a validation accuracy. The Public dataset showed lower performance due to inconsistencies in image resolution, lighting conditions, and labeling errors. The Combined dataset, which combined the strengths of both datasets, demonstrated stronger generalization across different environmental conditions and leaf types, suggesting that integrating datasets can enhance model robustness.

5.2 Recommendations

Future research should standardize data collection methods for better model performance in CLR detection. This will ensure high image resolution, consistent lighting, and accurate labeling, reducing noise in the dataset. Public datasets should undergo extensive preprocessing to ensure their utility in ML applications. Merging datasets from different regions can improve model robustness, especially in scenarios requiring generalization to varied environmental conditions. Researchers should also consider the trade-off between image resolution and computational efficiency when deploying models in real-world agricultural settings, ensuring accuracy without compromising efficiency. Enhancing dataset quality and accessibility can lead to accurate models for timely disease management, improving crop yield and quality. This supports global coffee production sustainability, necessitating the development of machine-learning solutions for the early detection and mitigation of coffee leaf rust.

REFERENCES

Al-Rashidi, J. F., Al-Enazi, L. A., Al-Mutairi, R. F., Al-Dukhayil, S. Y., Al-Abas, W. A., & Ibrahim, D. M. (2024). Coffee leaf diseases quadruple classifier (CLQC) model using deep learning. In A. Shaikh, A. Alghamdi, Q. Tan, & I. M. M. El Emary (Eds.), *Advances in emerging information and communication technology. ICIEICT 2023.* Signals and Communication Technology. Springer, Cham. https://doi.org/10.1007/978-3-031-53237-5_14

Anania, P., & Nade, P. (2020). *In the quest for semi-industrialized economy: Strategies for agricultural-based industrialization through co-operatives in Tanzania.* Liège (Belgium): CIRIEC International, Université de Liège. https://doi.org/10.25518/ciriec.wp202006

Arathi, B., & Dulhare, U. N. (2023). Classification of cotton leaf diseases using transfer learning-DenseNet-121. In A. B. Reddy, S. Nagini, V. E. Balas, & K. S. Raju (Eds.), *Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems.* Lecture Notes in Networks and Systems, 612. Springer, Singapore. https://doi.org/10.1007/978-981-19-9228-5_33

Bakr, M., Abdel-Gaber, S., Nasr, M., & Hazman, M. (2022). DenseNet based model for plant diseases diagnosis. *European Journal of Electrical Engineering and Computer Science, 6*(5), 1–9. https://doi.org/10.24018/ejece.2022.6.5.458

Banothu, S., Madhavi, K., Madan Kumar, K. M. V., Gajula, R., Mallikarjuna Rao, C., Dixit, S., & Chhetri, A. (2024). Plant disease identification and pesticides recommendation using DenseNet. *Cogent Engineering, 11*(1), 45–49. https://doi.org/10.1080/23311916.2024.2353080

Cabrera, A., Avramidis, K., & Narayanan, S. (2024). Early detection of coffee leaf rust through convolutional neural networks trained on low-resolution images. *arXiv preprint,* arXiv:2407.14737.

Chavarro, A. F., Renza, D., & Ballesteros, D. M. (2023). Influence of hyperparameters in deep learning models for coffee rust detection. *Applied Sciences, 13*(7), 4565. https://doi.org/10.3390/app13074565

Gichuru, E., Alwora, G., Gimase, J., & Kathurima, C. (2021). Coffee leaf rust (Hemileia vastatrix) in Kenya—A review. *Agronomy, 11*(12), 2590. https://doi.org/10.3390/agronomy11122590

Hitimana, E., Sinayobye, O. J., Ufitinema, J. C., Mukamugema, J., Rwibasira, P., Murangira, T., Masabo, E., Chepkwony, L. C., Kamikazi, M. C. A., Uwera, J. A. U., Mvuyekure, S. M., Bajpai, G., & Ngabonziza, J. (2023). An intelligent system-based coffee plant leaf disease recognition using deep learning techniques on Rwandan Arabica dataset. *Technologies, 11*(5), 116. https://doi.org/10.3390/technologies11050116

Jayaprakash, K., & Balamurugan, S. P. (2021). Analysis of plant disease detection and classification models: A computer vision perspective. *Journal of Computational and Theoretical Nanoscience, 17*(12), 5422–5428. https://doi.org/10.1166/jctn.2020.9435

Jepkoech, J., Mugo, D. M., Kenduiywo, B. K., & Too, E. C. (2021). Arabica coffee leaf images dataset for coffee leaf disease detection and classification. *Data in Brief, 36.* https://doi.org/10.1016/j.dib.2021.107142

Ju, C., Chen, C., Li, R., Zhao, Y., Zhong, X., Sun, R., Liu, T., & Sun, C. (2023). Remote sensing monitoring of wheat leaf rust based on UAV multispectral imagery and the BPNN method. *Food and Energy Security, 12,* e477. https://doi.org/10.1002/fes3.477

Kiwelu, L., Damas, P., & Mpenda, Z. (2021). Factors influencing adoption of improved coffee varieties among smallholder farmers in Mbinga and Mbozi districts. *International Journal of Agricultural Economics, 6*(1), 21-32.

Koutouleas, A. (2023). Coffee leaf rust: Wreaking havoc in coffee production areas across the tropics. *Plant Health Cases, 2*(23), 23–28.

Leandro, C. A., de Brito Silva, L., & Faulin, M. (2021). Artificial intelligence for detection and quantification of rust and leaf miner in coffee crop. *arXiv e-prints,* arXiv:2103.

Lyimo, D. A., Lakshmi Narasimhan, V., & Mbero, Z. A. (2021). Sensitivity analysis of coffee leaf rust disease using three deep learning algorithms. *2021 IEEE AFRICON.* https://doi.org/10.1109/AFRICON51333.2021.9571007

Magomba, L., & Ng'atigwa, F. (2024). Assessment of print media coverage of coffee crop farming in Tanzania: A case of Daily News and The Citizen newspapers. *International Journal of Multidisciplinary Research and Growth Evaluation, 2*(4), 275–287.

Mbwambo, S., Maro, G., Monyo, H., & Mosi, E. (2020). Towards expansion of Coffea canephora production in Tanzania: The land suitability perspective. *World Journal of Agricultural Research, 8*(2), 52–56. https://doi.org/10.12691/wjar-8-2-5

Mengistu, A. D., Alemayehu, D. M., & Mengistu, S. G. (2016). Ethiopian coffee plant diseases recognition based on imaging and machine learning techniques. *International Journal of Database Theory and Application, 9*(4), 79–88. https://doi.org/10.14257/ijdta.2016.9.4.07

Montalbo, F. J. P., & Hernandez, A. A. (2020, February). An optimized classification model for *Coffea liberica* disease using deep convolutional neural networks. In *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 213-218).

Muharromah, M. D., Kristiana, A. I., Slamin, S., Dafik, D., Agustin, I. H., & Baihaki, R. I. (2024, July). The analysis of the implementation of convolutional neural network architectures for coffee leaf disease image classification. In *AIP Conference Proceedings* (Vol. 3176, No. 1). AIP Publishing.

Nawaz, M., Nazir, T., Javed, A., Amin, S. T., Jeribi, F., & Tahir, A. (2024). CoffeeNet: A deep learning approach for coffee plant leaves disease recognition. *Expert Systems with Applications*, 237, 121481. doi: https://doi.org/10.1016/j.eswa.2023.121481.

Novtahaning, D., Shah, H. A., & Kang, J.-M. (2022). Deep learning ensemble-based automated and high-performing recognition of coffee leaf disease. *Agriculture*, 12(11), 1909. https://doi.org/10.3390/agriculture12111909.

Otieno, H. M., Alwenge, B. A., & Okumu, O. O. (2019). Coffee production challenges and opportunities in Tanzania: The case of coffee farmers in Iwindi, Msia, and Lwati villages in Mbeya Region. *Asian Journal of Agricultural and Horticultural Research*, 3(2), 1-14. https://doi.org/10.9734/ajahr/2019/v3i229993

Pham, T. C., Nguyen, V. D., Le, C. H., Packianather, M., & Hoang, V.-D. (2023). Artificial intelligence-based solutions for coffee leaf disease classification. *IOP Conference Series: Earth and Environmental Science*, 1278, 012004. https://doi.org/10.1088/1755-1315/1278/1/012004

Prince, R. H., Abdul, A. M., Hasibul, I. P., Shafiun, M., Md Nahiduzzaman, A. K., & Mohamed, A. A. (2024). CSXAI: A lightweight 2D CNN-SVM model for detection and classification of various crop diseases with explainable AI visualization. *Frontiers in Plant Science*, 15(4), 45-49. https://doi.org/10.3389/fpls.2024.1412988

Pujiastuti, E. S., Page, N. R. H., Elisabeth, S. P., Rio, S., Susana, T., & Su, T. (2023). Increasing farmers' knowledge and action in agroforestry-based coffee cultivation. *Journal of Research in Environmental and Earth Sciences*, 9(2), 103-108.

Ruben, R., Catherine, A., Faycal, B., David, M., & Youri, D. (2018). Coffee value chain analysis in the southern highlands. *Wageningen University and Research*.

Soares, A. d. S., Vieira, B. S., Bezerra, T. A., Martins, G. D., & Siquieroli, A. C. S. (2022). Early detection of coffee leaf rust caused by *Hemileia vastatrix* using multispectral images. *Agronomy*, 2(22), 12, 2911. https://doi.org/10.3390/agronomy12122911.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., & Maglogiannis, I. (Eds.). *Artificial Neural Networks and Machine Learning – ICANN 2018*. Lecture Notes in Computer Science, Vol. 11141. Springer, Cham. https://doi.org/10.1007/978-3-030-01424-7_27.

Velásquez, D., Sánchez, A., Sarmiento, S., Velásquez, C., Toro, M., Montoya, E., Trefftz, H., Maiza, M., & Sierra, B. (2021). A cyber-physical data collection system integrating remote sensing and wireless sensor networks for coffee leaf rust diagnosis. *Sensors*, 21(16), 5474. https://doi.org/10.3390/s21165474

Yebasse, M., Shimelis, B., Warku, H., Ko, J., & Cheoi, K. J. (2021). Coffee disease visualization and classification. *Plants*, 10(6), 1257. https://doi.org/10.3390/plants10061257