

# Modification of C-Grabcut for Segmentation and Classification of Coffee Leaf Diseases in Complex Backgrounds

Anastia Ivanabilla Novanti, Agus Harjoko\*

Dept. of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia

**Abstract**—Visual changes, including spots, discoloration, and deformation characterize coffee leaf diseases. In real-world image data, complex backgrounds present challenges for classification using deep learning models. Irrelevant objects, such as soil, other leaves, and miscellaneous items, can hinder the model's ability to accurately recognize disease patterns. Furthermore, the absence of effective segmentation techniques has resulted in low accuracy in previous studies. This work aims to address these limitations by enhancing the performance of the MobileNet-V2 model for coffee leaf disease classification. We applied a modified C-Grabcut segmentation technique to improve the isolation of diseased areas from complex backgrounds. The results demonstrate a significant performance improvement, achieving an Intersection over Union (IoU) of 0.8369 and an accuracy of 94.83%. These findings suggest that the modified MobileNet-V2 model, combined with the improved C-Grabcut segmentation, offers robust performance for in-field coffee leaf disease classification, striking a better balance between effectiveness and accuracy compared to previous studies.

**Keywords**—Image segmentation; in-field image; mobilenet-v2; coffee leaf diseases; background complexity

## I. INTRODUCTION

Coffee is an important agricultural commodity with significant economic value. In 2020, the coffee industry was valued at USD 102 billion [1] and is projected to grow at a compound annual growth rate (CAGR) of 4.28% through 2026, supporting approximately 125 million jobs [2] worldwide. Maintaining the health of coffee plants is crucial for ensuring both quality and productivity.

One of the main challenges in coffee cultivation is the occurrence of leaf diseases, which are often caused by pathogens such as fungi, bacteria, and viruses [3]. These diseases exhibit visual symptoms on leaves, including spots, discoloration, and deformation. Early and accurate detection of these symptoms is essential to control disease spread and enhance crop yield.

In recent years, deep learning models have gained popularity for automating plant disease detection. Among these, Convolutional Neural Networks (CNNs) are particularly effective for image classification tasks. Previous studies have explored CNN models like MobileNet-V2 for classifying coffee leaf diseases. However, in-field images often contain complex backgrounds, including soil, other leaves, and environmental artifacts, which introduce noise and decrease model performance. Without effective image segmentation

techniques, deep learning models struggle to differentiate disease-affected areas from irrelevant objects. Some studies report accuracy drops as low as 34% when classifying multiple disease types [4]. This highlights the need for an approach that combines segmentation and classification to enhance model robustness in in-field agricultural settings.

To address these limitations, this study introduces an enhanced MobileNet-V2 model that incorporates C-Grabcut segmentation technique. The research aims to:

- 1) Develop an image segmentation approach that effectively isolates disease-relevant features from complex backgrounds using modified C-Grabcut.
- 2) Improve the accuracy of coffee leaf disease detection through transfer learning with MobileNet-V2.
- 3) Optimize hyperparameters and augmentation techniques to enhance the generalization capability of the model for in-field classification tasks.

This study contributes to agricultural image processing by integrating segmentation and classification techniques for automated plant disease recognition. The findings provide insights into optimizing deep learning models for precision agriculture, enabling early disease detection and intervention.

The rest of this paper is organized as follows. Section II presents a literature review on existing methods for coffee leaf disease classification and segmentation. Section III describes the research methodology, including data collection, preprocessing, segmentation, model training, and evaluation metrics. Section IV discusses the experimental results and performance comparisons. Finally, Section V concludes the study and suggests future research directions.

## II. LITERATURE REVIEW

The classification of coffee leaf diseases has become a major focus in agricultural research, especially because of its impact on crop yield and quality. Developing a strong classification model using deep learning that performs effectively in field conditions presents unique challenges, such as managing complex backgrounds in field images. Many existing studies focus primarily on classification using deep learning models but do not incorporate effective segmentation techniques to isolate disease features from irrelevant background elements. Table I provides an overview of several studies on the classification of coffee leaf diseases.

\*Corresponding Author.

TABLE I. RELATED STUDIES

| Methods                          | Data                 | Preprocessing               | No. of class | Accuracy   |
|----------------------------------|----------------------|-----------------------------|--------------|--|
| MobileNet-V2 [4]                 | Real condition image | Augmentation                | 2, 3, 6      | 99.93% ( 2 classes),<br>34% (3 classes),<br>16% (6 classess) |
| Extreme Learning Machine ELM [5] | Controlled image     | Segmentation                | 3            | 99.09%   |
| Inception v3 [6]                 | Controlled image     | Augmentation                | 5            | 97.61%   |
| VGG16 [7]                        | Controlled image     | Augmentation                | 4            | 97.20%   |
| EfficientNet-B0 [8]              | Real condition image | Augmentation                | 6            | 91%  |
| ResNet-50 [9]                    | Controlled image     | Augmentation                | 2            | 99%  |
| ResNet-50 [10]                   | Real condition image | Segmentation & Augmentation | 2, 6         | 92% (2 classes),<br>88.98% (6 classes)                       |

Despite the advances in coffee leaf disease classification, several challenges remain unaddressed, particularly in in-field conditions. Many existing models rely solely on data augmentation for performance enhancement but lack proper segmentation techniques, leading to suboptimal classification in complex environments.

For instance, MobileNet-V2 achieved an accuracy of 99.93% for binary classification but significantly dropped to 34% and 16% for three and six-class classification tasks, respectively, in in-field conditions [4]. This highlights the model's difficulty in distinguishing diseased areas from background noise such as soil and other foliage, reducing overall accuracy. Other studies employing models like Extreme Learning Machine (ELM), Inception v3, VGG16, and EfficientNet-B0 have reported high accuracy (above 90%) in controlled environments but have struggled to generalize to in-field settings [5], [6], [7], [8].

A study using ResNet-50 with segmentation and augmentation demonstrated enhanced accuracy (92%) in in-field images [10]. This underscores the importance of integrating segmentation techniques to improve classification robustness. However, existing segmentation approaches, such as Grabcut, have shown limited effectiveness in isolating disease-affected areas from complex backgrounds.

The C-Grabcut algorithm, originally developed for detecting apple leaf diseases, improves upon the traditional Grabcut method by incorporating contour detection to more accurately isolate areas [11] affected by the disease. This approach effectively reduces background noise, allowing models to concentrate on relevant features, thus enhancing classification accuracy while lowering computational demands. However, C-Grabcut has not yet been widely explored for coffee leaf disease classification, leaving a gap in its application to agricultural disease detection under in-field conditions.

To bridge these gaps, this study proposes an improved MobileNet-V2 model incorporating modified C-Grabcut segmentation to enhance coffee leaf disease classification under in-field conditions. The proposed approach aims to improve segmentation accuracy by modifying C-Grabcut to better isolate diseased areas from background elements, reducing noise interference from soil, other leaves, and environmental artifacts. Additionally, this study integrates segmentation, augmentation, and transfer learning techniques to enhance the model's ability to recognize disease patterns more effectively, particularly in complex agricultural environments. By balancing computational efficiency and classification accuracy, this approach ensures that the model remains lightweight and practical for real-world agricultural applications. Through the combination of segmentation with deep learning, this study provides a more effective and scalable solution for coffee leaf disease detection, addressing the key limitations identified in previous research.

### III. RESEARCH METHODS

This research aims to improve the accuracy and robustness of coffee leaf disease classification under real-world agricultural conditions by employing a MobileNet-V2 model combined with a modified C-Grabcut segmentation technique. The methods are presented in Fig. 1.

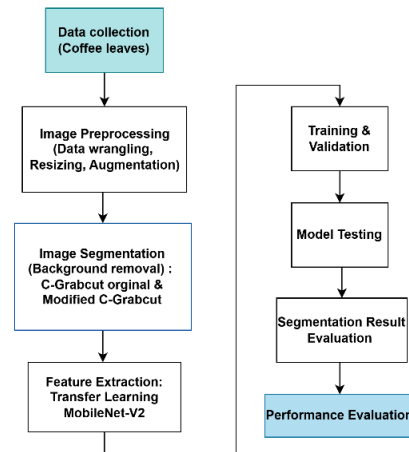


Fig. 1. Research method.

#### A. Data Collection




The dataset used in this study comprises images of coffee leaves collected from a public dataset [12] containing three classes: healthy, rust, and spot disease. Each image represents realistic field conditions, including complex backgrounds with noise elements such as soil, other leaves, and environmental artifacts. Data wrangling is performed to label each image according to its class and remove duplicates, ensuring data quality and preventing model bias. The dataset is then split into training (80%), validation (10%), and test (10%) subsets to facilitate model training and performance assessment.

#### B. Image Preprocessing

Data preprocessing in this study involved several steps to prepare the coffee leaf images for analysis. First, the images were organized into class-specific folders through labeling to

ensure proper categorization. Duplicate images were identified and removed using hash-based techniques to maintain data integrity. After segmentation, the images were resized to  $224 \times 224$  pixels to meet the input requirements of the MobileNet-V2 model. Lastly, data augmentation techniques were employed to generate diverse dataset variations, including rotation, blurring, noise addition, and contrast adjustment. This approach enhances the model's robustness and generalization capabilities [13]. The augmented dataset helps the model recognize disease features across various field conditions. Table II presents the dataset distributions after augmentation.

TABLE II. DATASET DISTRIBUTIONS AFTER PREPROCESSING

| Classes   | Data Distributions |            |      | Preview  |
|-----------|--------------------|------------|------|--|
|           | Train              | Validation | Test |  |
| Healthy   | 1600               | 200        | 200  |   |
| Leaf Rust | 1600               | 200        | 200  |   |
| Leaf Spot | 1600               | 200        | 200  |  |
| Total     | 4800               | 600        | 600  |  |

### C. Image Segmentation

To effectively isolate diseased areas of leaves and minimize background noise, the modified C-Grabcut algorithm is applied to each image. This enhanced version of the traditional Grabcut algorithm includes contour detection, which allows for more accurate differentiation between diseased leaf areas and surrounding elements, such as soil and other foliage. The modifications made to the original C-Grabcut involve adjustments to key functions and parameter settings, resulting in improved segmentation accuracy while retaining essential leaf and disease features. A step-by-step illustration of the modified C-Grabcut process is presented in Fig. 2, and the procedure is outlined as follows:

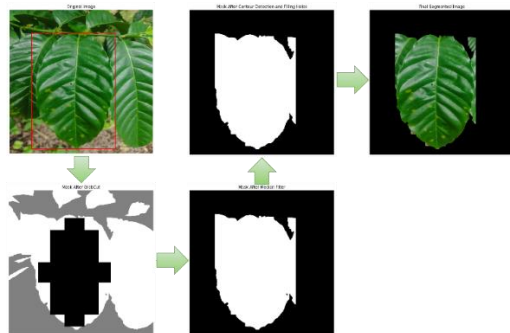


Fig. 2. Foreground segmentation with modified C-Grabcut algorithm.

1) *Initialization and modified mask function*: Segmentation begins by defining an initial bounding box around the leaf, focusing the algorithm on the relevant area. To enhance foreground detection, two markers are added within this bounding box: a foreground box and intersecting vertical-horizontal lines. The width and height of the bounding box are calculated to determine the leaf's orientation, guiding the accurate placement of the foreground markers. The modified mask function is visualized in Fig. 3.

2) *Foreground box and vertical-horizontal lines*: The foreground box is assigned a value of 1 in the mask, marking it as a definite foreground. A vertical and horizontal line intersecting at the bounding box center creates a cross ("+"), extending 90% of the box's width and height with a thickness of 90 pixels. Pixels within this cross are set to 1, reinforcing the foreground, while areas outside the box and cross are set to 2, marking probable background. This ensures that key leaf features, such as lesions, are preserved during segmentation, unlike in the original C-Grabcut.

3) *Bounding box limitation*: To address a common issue where irrelevant background features remain outside the bounding box, the modified mask is restricted to the bounding box area only. This ensures the mask applies solely within the bounding box, eliminating non-relevant features outside it.

4) *Median filtering*: After applying the bounding box limitation, a median filter with a  $3 \times 3$  kernel size is used on the mask. This step smooths the mask by reducing noise and softening edges. The smaller kernel size provides a gentle smoothing effect that preserves critical details of the leaf, such as disease features, while effectively eliminating isolated noise. The median filter is particularly effective in maintaining the shape and texture of small lesions, which are essential for accurate disease identification.

5) *Contour functions*: Contour detection is performed to refine the leaf's boundary within the bounding box. This step identifies the edges of the segmented leaf and adjusts the mask accordingly, ensuring that it accurately captures the leaf's shape and any disease-specific features along its edges. Contour detection is particularly effective in preventing background elements, which may share similar color properties, from being incorrectly included in the foreground. Contour detection significantly enhances the overall segmentation accuracy by maintaining clear and precise edges.

6) *Bitwise operation*: Bitwise operations are used to isolate the segmented leaf from the background. Specifically, a bitwise AND operation is performed between the mask and the original image. This operation retains only the foreground (the segmented leaf) while setting the background pixels to zero. As a result, any remaining background noise within the bounding box is eliminated, producing a clean and focused image of the leaf that is ready for disease classification.

### D. Feature Extraction Using Transfer Learning

This study employs a pre-trained MobileNet-V2 model, which has been trained on the ImageNet dataset, for feature extraction. Transfer learning enables a model to be trained and

fine-tuned for a specific task, then adapted to a related task [14]. Transfer learning is utilized by taking weights from this pre-trained model, which has already learned general features from ImageNet [7]. ImageNet is a large dataset widely used for training deep learning models, particularly Convolutional Neural Networks (CNNs). It consists of approximately 1.2 million images organized into 1,000 categories [15]. By leveraging prior knowledge, this approach enhances the model's performance and efficiency in a new context [14].

In this model, the bottom layers of MobileNet-V2 are frozen to preserve these general features, while the deeper layers are modified to learn features specific to the task at hand. These layers can be trained or fine-tuned to enhance model performance [16]. The top layers, or classifier, are adapted by adding three fully connected layers, one batch normalization layer, and two dropout layers to reduce the risk of overfitting. ReLU activation functions are also utilized. These additional layers improve the model's ability to process the extracted features effectively, allowing it to classify them into the three target classes.

### E. Training and Validation

The training and validation process starts with loading the respective datasets. Training is carried out on the training set, while validation is performed using the validation set. The initial hyperparameter settings, presented in Table III, are applied consistently throughout the initial experiment to maintain baseline conditions.

TABLE III. INITIAL HYPERPARAMETER TUNING

| Hyperparameter    | Value                           |
|-------------------|---------------------------------|
| Number of classes | 3                               |
| Pre-trained Model | MobileNet-V2                    |
| Trainable Layers  | Only final classification layer |
| Optimizer         | SGD                             |
| Loss function     | Cross-Entropy                   |
| Batch size        | 64                              |
| Learning rate     | 0.001                           |
| Patience          | 5                               |
| Epochs            | 25                              |

TABLE IV. EXPERIMENTAL SETUP PHASE 1

| Scenario | Segmentation Techniques         |
|----------|---------------------------------|
| 1.1      | No Segmentation                 |
| 1.2      | GrabCut Segmentation            |
| 1.3      | C-Grabcut Segmentation          |
| 1.4      | Modified C-Grabcut Segmentation |

### F. Model Testing

The testing process consists of four experimental phases, evaluating the impact of segmentation, trainable layers, hyperparameter tuning, and background complexity on the model's classification performance.

1) *Phase 1: Data segmentation setup*: The first experiment evaluates the impact of different data segmentation techniques on the model's performance in recognizing images. The best-performing setup from this phase is selected for the next experiment phase (Table IV).

2) *Phase 2: Trainable layer experiment*: In the second phase, the number of trainable layers in the model is adjusted to assess how different layer configurations affect performance under transfer learning. The optimal configuration from this experiment is used in the final experiment phase (Table V).

TABLE V. EXPERIMENTAL SETUP PHASE 2

| Scenario | Number of trainable layer        |
|----------|----------------------------------|
| 2.1      | -                                |
| 2.2      | Last 5% of layers are trainable  |
| 2.3      | Last 20% of layers are trainable |
| 2.4      | Last 50% of layers are trainable |

3) *Phase 3: Hyperparameter tuning*: To optimize the model's performance, the final phase involves fine-tuning hyperparameters, including batch size, learning rate, optimizer, and the number of epochs. Multiple combinations are tested, and the model configuration yielding the highest performance is selected as the final model, saved for testing (Table VI).

TABLE VI. EXPERIMENTAL SETUP PHASE 3

| Hyperparameter | Value         |
|----------------|---------------|
| Optimizer      | SGD, Adam     |
| Epoch          | 25, 50, 75    |
| Learning rate  | 0.001, 0.0001 |

4) *Additional experiment: The effect of background complexity on model performance*: An additional experiment was conducted to evaluate the influence of background complexity on model accuracy by testing two types of images: natural background images, as used in the main study, and plain background images, obtained from a public dataset, Roboflow [20]. This experiment aimed to determine whether a simplified background could improve classification performance compared to natural backgrounds. Both datasets underwent the same preprocessing steps, except that segmentation was not applied to the plain-background images, ensuring a fair comparison. The best hyperparameters from previous experiments were utilized for both datasets, allowing for an objective assessment of performance differences. The results of this experiment provide valuable insight into the extent to which background complexity affects classification accuracy and whether segmentation techniques remain essential when dealing with plain-background images.

### G. Segmentation Result Evaluation

Segmentation result evaluation aims to assess the outcomes of both the original C-Grabcut and the modified C-Grabcut

using predefined evaluation metrics. The evaluation process consists of two types: quantitative and qualitative.

Quantitative evaluation provides objective measurements utilizing metrics such as Intersection over Union (IoU), Dice Coefficient, Pixel Accuracy and Precision. These metrics enable a measurable comparison of the performance of the two methods. In contrast, qualitative evaluation involves visual observation to ensure that the segmentation results meet specific visual standards, such as boundary clarity and consistency in the target area. The combination of these evaluation methods offers a comprehensive assessment of segmentation quality.

IoU measures the agreement between the region predicted by the segmentation model and the ground truth region [17]. The Dice Coefficient measures the similarity between a segmentation model's predicted region and the ground truth [17]; higher Dice Coefficient values indicate better model performance. Pixel Accuracy, also known as the Rand Index, defines the number of correct predictions (both positive and negative) relative to the total number of predictions [18]. The formulas for the quantitative evaluations are presented in Table VII.

TABLE VII. THE QUANTITATIVE SEGMENTATION EVALUATION

| Evaluation       | Formula  |
|------------------|--|
| IoU              | $IoU = \frac{TP}{TP + FP + FN}$                  |
| Dice Coefficient | $Dice = \frac{2 * TP}{(2 * TP + FP + FN)}$       |
| Pixel Accuracy   | $Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$ |
| Precision        | $Precision = \frac{TP}{TP + FN}$                 |

where TP (True Positive) represents pixels correctly classified as part of the class, TN (True Negative) refers to pixels correctly predicted as background, FP (False Positive) denotes pixels incorrectly classified as part of the class, and FN (False Negative) indicates pixels that were not classified as part of the class [19].

Two main elements are required to compute these metrics: the data mask (ground truth) and the prediction mask. The data mask represents annotated ground truth areas (e.g., leaf objects) and is manually created using the Roboflow platform. It generates XML files for each image, which are then converted into binary images in .png format. The prediction mask is derived from the segmentation results of the original and modified C-grabcut methods applied to the test dataset.

#### H. Performance Evaluation

To assess the effectiveness of the proposed model in identifying feature patterns across different disease categories and evaluating classification accuracy for each class, the model's performance is measured using key evaluation metrics derived from the confusion matrix. These metrics include accuracy, precision, recall, and F1-score, as summarized in Table VIII. The best model's performance is evaluated on the test set using these metrics thoroughly analyze the model's

performance for each class. A confusion matrix is also created to visualize the classification results across the different classes.

TABLE VIII. PERFORMANCE EVALUATION METRIC

| Metric          | Formula                                     |
|-----------------|---|
| Accuracy (ACC)  | $\frac{TN + TP}{TP + FP + TN + FN}$         |
| Precision (PRE) | $\frac{TP}{TP + FP}$                        |
| Recall (TPR)    | $\frac{TP}{TP + FN}$                        |
| F1-Score (F1)   | $2 \times \frac{PRE \times REC}{PRE + REC}$ |

## IV. RESULTS

### A. Segmentation Evaluation Results

Table IX presents the quantitative evaluation results of the GrabCut, C-Grabcut Original, and Modified C-Grabcut methods, assessed using four key metrics: Intersection over Union (IoU), Dice Coefficient, Pixel Accuracy, and Precision.

TABLE IX. QUANTITATIVE EVALUATION RESULT OF SEGMENTATION TECHNIQUES

| Technique          | IoU    | Dice Coefficient | Pixel Accuracy | Precision |
|--------------------|--------|------------------|----------------|-----------|
| GrabCut            | 0,6821 | 0,7934           | 0,8445         | 0,7019    |
| C-Grabcut Original | 0,683  | 0,7941           | 0,8446         | 0,7018    |
| Modified C-Grabcut | 0,8369 | 0,9091           | 0,9344         | 0,8402    |

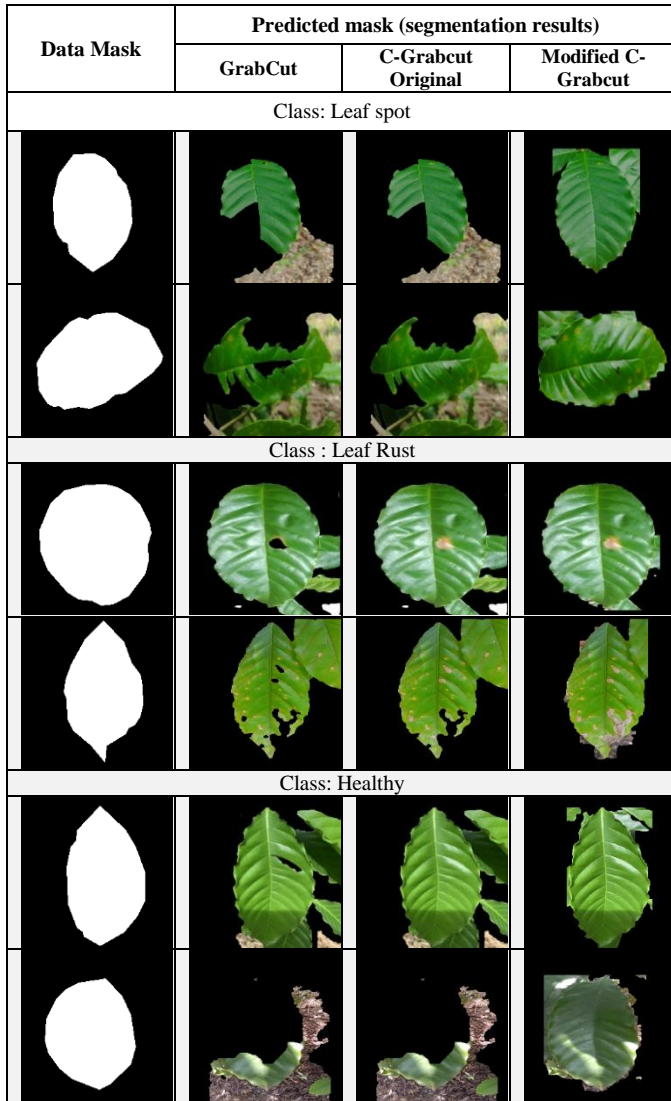
These results indicate that C-Grabcut Original does not show significant improvement compared to GrabCut, as reflected in the minimal differences in all four metrics. However, Modified C-Grabcut outperforms both methods, demonstrating higher segmentation accuracy and precision.

Table X presents the visual results of the GrabCut, C-Grabcut Original, and Modified C-Grabcut. The results indicate that C-Grabcut Original performs better than traditional GrabCut, as it is able to retain lesion features more effectively. In contrast, GrabCut often removes critical disease lesions, leading to loss of essential features for classification. However, C-Grabcut Original still has some segmentation inaccuracies, particularly in areas where the leaf color or texture closely resembles the background, causing parts of the leaf to be mistakenly removed.

In contrast, Modified C-Grabcut demonstrates significant improvements over both GrabCut and C-Grabcut Original. The segmentation results show that Modified C-Grabcut effectively retains lesion structures, reducing the likelihood of misclassification between disease spots and the background. Compared to GrabCut, which often erases crucial lesion areas, and C-Grabcut Original, which still exhibits some errors in boundary detection, Modified C-Grabcut achieves better object preservation and noise reduction. The enhanced contour detection and optimized bounding box adjustments in Modified C-Grabcut allow for sharper, more defined segmentation while minimizing the loss of lesion information.



TABLE X. VISUAL RESULT OF SEGMENTATION TECHNIQUES



B. Performance Evaluation

1) Phase 1: Data segmentation setup: The models trained in the first experiment were used for testing. The performance metrics of phase 1, including precision, recall, F1-score, and accuracy for each class, are presented in Table XI.

The Modified C-Grabcut approach achieved the highest accuracy (88.33%), outperforming traditional Grabcut and C-Grabcut. The results indicate that segmentation improves disease detection, especially for leaf spot classification.

2) Phase 2: Trainable layer experiment: Table XII presents the testing result for experiments utilizing different trainable layers in transfer learning. Scenario 2.4, which allowed only the last 50% of the layers to be trainable, resulted in the best accuracy (92.5%), suggesting that fine-tuning a larger portion of MobileNet-V2 enhances feature extraction for coffee leaf disease classification.

3) Phase 3: Hyperparameter tuning: Table XIII presents the result from the hyperparameter tuning experiments.

TABLE XI. THE RESULT OF PHASE 1

| Segmentation Setup          | Class     | Precision | Recall | F1-Score | Acc           |
|-----------------------------|-----------|-----------|--------|----------|---------------|
| No Segmentation (Sc 1.1)    | Healthy   | 0,84      | 0,92   | 0,88     | 0,8433        |
|                             | Leaf Spot | 0,79      | 0,82   | 0,8      |               |
|                             | Leaf Rust | 0,91      | 0,8    | 0,85     |               |
| GrabCut (Sc 1.2)            | Healthy   | 0,83      | 0,93   | 0,88     | 0,8450        |
|                             | Leaf Spot | 0,82      | 0,77   | 0,79     |               |
|                             | Leaf Rust | 0,89      | 0,83   | 0,86     |               |
| C-Grabcut Original (Sc 1.3) | Healthy   | 0,81      | 0,94   | 0,87     | 0,8517        |
|                             | Leaf Spot | 0,85      | 0,77   | 0,81     |               |
|                             | Leaf Rust | 0,9       | 0,85   | 0,88     |               |
| Modified C-Grabcut (Sc 1.4) | Healthy   | 0,86      | 0,95   | 0,9      | <b>0,8833</b> |
|                             | Leaf Spot | 0,88      | 0,83   | 0,86     |               |
|                             | Leaf Rust | 0,89      | 0,83   | 0,86     |               |

TABLE XII. THE RESULT OF PHASE 2

| Number of trainable layer | Class     | Precision | Recall | F1-Score | Acc    |
|---------------------------|-----------|-----------|--------|----------|--------|
| - (Sc 2.1)                | Healthy   | 0,82      | 0,95   | 0,88     | 0,8600 |
|                           | Leaf Spot | 0,87      | 0,78   | 0,82     |        |
|                           | Leaf Rust | 0,90      | 0,84   | 0,87     |        |
| Last 5% layers (Sc 2.2)   | Healthy   | 0,88      | 0,96   | 0,92     | 0,8967 |
|                           | Leaf Spot | 0,89      | 0,84   | 0,87     |        |
|                           | Leaf Rust | 0,93      | 0,89   | 0,91     |        |
| Last 20% layers (Sc 2.3)  | Healthy   | 0,89      | 0,96   | 0,93     | 0,8933 |
|                           | Leaf Spot | 0,86      | 0,86   | 0,86     |        |
|                           | Leaf Rust | 0,93      | 0,86   | 0,89     |        |
| Last 50% layers (Sc 2.4)  | Healthy   | 0,91      | 0,95   | 0,93     | 0,925  |
|                           | Leaf Spot | 0,92      | 0,91   | 0,91     |        |
|                           | Leaf Rust | 0,94      | 0,92   | 0,93     |        |

TABLE XIII. THE RESULT OF PHASE 3

| Scenarios         | Hyperparameter |               |            | Accuracy      |
|-------------------|----------------|---------------|------------|---------------|
|                   | Optimizer      | Learning rate | Batch size |               |
| <b>Epoch : 25</b> |                |               |            |               |
| 3.1               | SGD            | 0,001         | 32         | 0,925         |
| 3.2               | SGD            | 0,001         | 64         | 0,9367        |
| <b>3.3</b>        | <b>SGD</b>     | <b>0,001</b>  | <b>128</b> | <b>0,94</b>   |
| 3.4               | SGD            | 0,0001        | 32         | 0,9233        |
| 3.5               | SGD            | 0,0001        | 64         | 0,8917        |
| 3.6               | SGD            | 0,0001        | 128        | 0,88          |
| 3.7               | Adam           | 0,001         | 32         | 0,9283        |
| 3.8               | Adam           | 0,001         | 64         | 0,9333        |
| 3.9               | Adam           | 0,001         | 128        | 0,9333        |
| 3.10              | Adam           | 0,0001        | 32         | 0,9317        |
| <b>3.11</b>       | <b>Adam</b>    | <b>0,0001</b> | <b>64</b>  | <b>0,9483</b> |
| 3.12              | Adam           | 0,0001        | 128        | 0,93          |
| <b>Epoch : 50</b> |                |               |            |               |
| 3.13              | SGD            | 0,0001        | 64         | 0,8983        |

|                   |     |        |     |      |
|-------------------|-----|--------|-----|------|
| 3.14              | SGD | 0,0001 | 128 | 0,9  |
| <b>Epoch : 75</b> |     |        |     |      |
| 3.15              | SGD | 0,0001 | 128 | 0,91 |

Overall, the result indicate that the Adam optimizer performed better, particularly with a learning rate of 0.0001. When comparing learning rates, a lower learning rate of 0.0001 was found to be more effective with the Adam optimizer, while the Stochastic Gradient Descent (SGD) optimizer showed slightly better performance with a higher learning rate of 0.001. The optimal configuration was identified as Adam with a learning rate of 0.0001 and a batch size of 128 which resulted in a training accuracy of 99% and an F1 score of 0.9049, achieving the highest validation accuracy across all tested configurations.

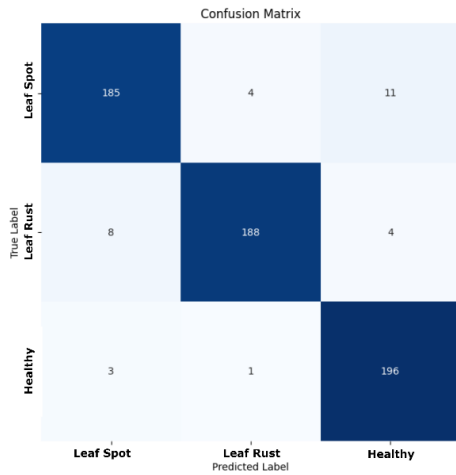


Fig. 3. Confusion Matrix for the best performance model.

The confusion matrix shown in Fig. 3 indicates that the classification results are primarily concentrated along the diagonal. This pattern suggests that the model generates more True Positives than False Negatives and False Positives. As a result, the model demonstrates high accuracy in correctly predicting the class of each image.

In the optimal scenario of the third experiment, the model performs well in recognizing all classes, surpassing the results of both the first and second experiments.

4) *Additional experiment: The Effect of Background Complexity on Model Performance.*

The testing results of additional experiment are presented in Table XIV.

TABLE XIV. TESTING RESULT OF ADDITIONAL EXPERIMENT

| Scenario | Image Type         | Testing Accuracy |
|----------|--------------------|------------------|
| 1.1      | Complex background | 0,8433           |
| 3.11     |                    | 0,9483           |
| 1.1      | Plain Background   | 0,965            |
| 3.11     |                    | 0,9983           |

The result indicates that background complexity significantly affects classification performance. When tested with natural background images, the model in Scenario 1.1 achieved 77.5% validation accuracy and 84.33% test accuracy. After hyperparameter optimization in Scenario 3.11, validation accuracy improved to 90.17%, and test accuracy increased to 94.83%, demonstrating that optimized hyperparameters enhance the model’s ability to handle complex backgrounds.

In contrast, models trained with plain background images exhibited higher performance across all metrics. In Scenario 1.1, the validation accuracy reached 94.17%, and test accuracy was 96.5%. After applying the best hyperparameters in Scenario 3.11, the model achieved 99.50% validation accuracy and 99.83% test accuracy, indicating that simplified backgrounds facilitate more effective feature extraction.

V. DISCUSSION

A. *Advantages of Modified C-Grabcut*

The superior performance of Modified C-Grabcut over traditional GrabCut and C-Grabcut Original is attributed to a series of refinements that enhance segmentation accuracy, particularly in complex backgrounds and varying lighting conditions.

One key improvement is the addition of a ‘+’ marker in the initial mask, which ensures that the leaf’s edges and disease lesions remain intact, preventing accidental removal. This enhancement is particularly beneficial for objects that share similar colors with the background, maintaining their structure more effectively.

Increasing the number of GrabCut iterations from 5 to 10 allows the model to refine the segmentation mask, resulting in sharper contours and fewer errors caused by noise or slight color differences. Additionally, reducing the median filter kernel size from 5 to 3 helps retain fine lesion details, preventing excessive blurring that could lead to information loss.

Through these modifications, Modified C-Grabcut significantly improves segmentation quality, effectively isolating the disease-affected areas while minimizing background interference. The results confirm that this approach enhances feature extraction for classification, making it a more reliable and efficient segmentation technique for coffee leaf disease detection.

B. *Analysis of Experiment Results*

The results confirm that segmentation plays a crucial role in enhancing classification performance, particularly in in-field conditions with complex backgrounds. In the first experiment, the model achieved an initial accuracy of 88.33%, establishing a baseline performance before applying further optimizations. The introduction of Modified C-Grabcut segmentation significantly improved disease feature extraction by isolating lesions from background noise, leading to more stable classification performance compared to models without segmentation. These findings validate that effective segmentation enhances model robustness by reducing misclassification due to background interference.

Further improvements were observed when 50% of MobileNet-V2 layers were fine-tuned, resulting in an accuracy of 92.5%. This indicates that selective layer tuning enhances feature extraction, allowing the model to capture disease-specific patterns more effectively. These findings are consistent with previous studies, where freezing too many layers reduced adaptability, while excessive fine-tuning led to overfitting and decreased generalization ability [14].

Hyperparameter tuning also played a crucial role in optimizing model performance. The Adam optimizer with a learning rate of 0.0001 and batch size of 128 achieved the highest accuracy at 94.83%, outperforming SGD. Adam's adaptive optimization strategy contributed to faster convergence and better classification robustness, reinforcing the importance of fine-tuning hyperparameters for deep learning-based plant disease detection [21], [22].

The trend of accuracy improvement across the three experiments is illustrated in Fig. 4, showing a consistent upward trajectory as various optimizations were applied. As depicted, the model initially achieved 88.33% accuracy in Experiment 1, which increased to 92.5% in Experiment 2 after trainable layer optimization, and finally reached 94.83% in Experiment 3 following hyperparameter tuning. This trend confirms that a structured approach to segmentation, transfer learning, and hyperparameter tuning leads to significant improvements in classification accuracy.

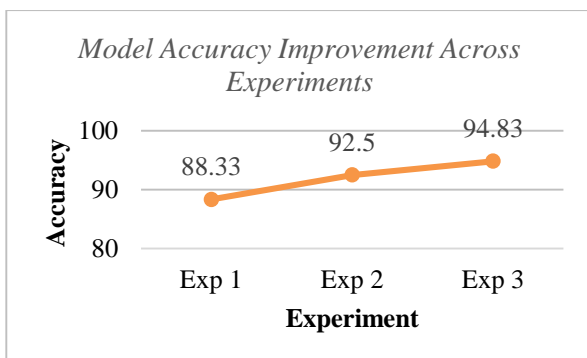


Fig. 4. Trend of model accuracy improvement across experiments.

### C. The Effect of Background Complexity on Model Performance

The additional experiment highlights the significant role of background complexity in deep learning-based plant disease classification. The results indicate that models trained with plain background images achieved higher accuracy across all evaluation metrics, confirming that background noise in natural images negatively affects classification performance. The increase in test accuracy from 84.33% to 94.83% for natural background images after hyperparameter tuning suggests that model optimization helps mitigate background interference but does not fully eliminate its impact.

The superior accuracy observed in plain background images (99.83%) indicates that a simplified background enables the model to focus on key object features without distractions. Conversely, natural background images introduce additional challenges, including color variations, shadows, and overlapping objects, which can lead to misclassification errors.

These findings align with prior computer vision research, which has shown that complex backgrounds hinder feature extraction and reduce model performance.

Despite the improved accuracy with plain background images, real-world agricultural settings rarely provide such controlled conditions. In practical applications, coffee leaves are surrounded by other foliage, exposed to uneven lighting, and subject to various environmental factors. As a result, models trained exclusively on plain background datasets may struggle to generalize effectively in in-field conditions, where background complexity is unavoidable.

To address these challenges, segmentation remains a critical preprocessing step. By isolating the primary object, Modified C-Grabcut significantly reduces background interference, allowing the model to extract more relevant disease features. The results reinforce the importance of integrating segmentation techniques into deep learning workflows, ensuring more reliable classification performance in diverse and uncontrolled environments.

## VI. CONCLUSION

This study investigated the impact of Modified C-Grabcut segmentation and model optimization on coffee leaf disease classification in in-field conditions. The research aimed to enhance classification accuracy by addressing the challenges posed by complex backgrounds in agricultural images.

The results confirm that effective segmentation significantly improves classification performance. The Modified C-Grabcut technique outperformed GrabCut and C-Grabcut Original, achieving an IoU of 0.8369, Dice Coefficient of 0.9091, and test accuracy of 94.83%. These findings validate that better contour detection and refined boundary constraints help isolate disease-relevant features, reducing misclassification due to background noise.

Further improvements were observed through model optimization techniques, particularly in trainable layer selection and hyperparameter tuning. Fine-tuning 50% of MobileNet-V2 layers resulted in an accuracy increase to 92.5%, while the Adam optimizer (learning rate 0.0001, batch size 128) achieved the highest accuracy of 94.83%. Additionally, experiments on background complexity demonstrated that models trained with plain background images performed better (99.83% accuracy) than those with natural backgrounds, confirming that background noise negatively impacts feature extraction.

In summary, this research demonstrates that segmentation-based preprocessing is crucial for improving deep learning-based plant disease classification, especially in real-world agricultural applications. The findings contribute to precision agriculture and automated disease detection by offering a robust segmentation-enhanced classification approach.

## REFERENCES

- [1] S. Bermudez, V. Voora, and C. Larrea, "Coffee prices and sustainability SUSTAINABLE COMMODITIES MARKETPLACE SERIES," 2022.
- [2] M. Intelligence, "Coffee Market Report - Industry Analysis, Size & Forecast (2025 - 2030)," 2021.
- [3] W. Cheppy et al., Hama dan Penyakit Tanaman. 2021.



- [4] Y. AUFAR and T. P. Kaloka, "Robusta coffee leaf diseases detection based on MobileNetV2 model," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, pp. 6675–6683, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6675-6683.
- [5] G. L. Manso, H. Knidel, R. A. Krohling, and J. A. Ventura, "A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust," Mar. 2019, [Online]. Available: <http://arxiv.org/abs/1904.00742>
- [6] M. Kumar, P. Gupta, P. Madhav, and Sachin, "Disease Detection in Coffee Plants Using Convolutional Neural Network," in *Proceedings of the 5th International Conference on Communication and Electronics Systems (ICCES 2020)*, 2020, pp. 755–760.
- [7] F. J. P. Montalbo and A. A. Hernandez, "Classifying barako coffee leaf diseases using deep convolutional models," *International Journal of Advances in Intelligent Informatics*, vol. 6, no. 2, pp. 197–209, Jul. 2020, doi: 10.26555/ijain.v6i2.495.
- [8] S. A. Sabrina and W. F. Al Maki, "Klasifikasi Penyakit pada Tanaman Kopi Robusta Berdasarkan Citra Daun Menggunakan Convolutional Neural Network," in *e-Proceeding of Engineering*, 2022, pp. 1919–1927.
- [9] A. Fatchurrachman and D. Udjulawa, "Identifikasi Penyakit Pada Tanaman Kopi Berdasarkan Citra Daun Menggunakan Metode Convolution Neural Network," *Jurnal Algoritme*, vol. 3, no. 2, pp. 151–159, 2023, doi: 10.35957/algoritme.xxxx.
- [10] Suprihanto, I. Awaludin, M. Fadhill, and M. Anhdika Zaini Zulfikor, "Analisis Kinerja ResNet-50 dalam Klasifikasi Penyakit pada Daun Kopi Robusta," *JURNAL INFORMATIKA*, vol. 9, no. 2, 2022, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [11] S. Lian, L. Guan, J. Pei, G. Zeng, and M. Li, "Identification of apple leaf diseases using C-Grabcut algorithm and improved transfer learning base on low shot learning," *Multimed Tools Appl*, vol. 83, no. 9, pp. 27411–27433, Mar. 2024, doi: 10.1007/s11042-023-16602-4.
- [12] Anonim, "Deteksi Penyakit Daun Kopi Robusta Dataset," Nov. 2023, Roboflow. Accessed: Jul. 07, 2024. [Online]. Available: <https://universe.roboflow.com/tugas-akhir-70fw5/deteksi-penyakit-daun-kopi-robusta>
- [13] K. Kusriani et al., "Data augmentation for automated pest classification in Mango farms," *Comput Electron Agric*, vol. 179, Dec. 2020, doi: 10.1016/j.compag.2020.105842.
- [14] N. Ayni, M. Pauzi, S. Mastura Mustaza, N. Zainal, and M. Faiz Bukhori, "Transfer Learning-based Weed Classification and Detection for Precision Agriculture," *IJACSA (International Journal of Advanced Computer Science and Applications)*, vol. 15, no. 6, 2024, [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
- [16] L. T. Duong, T. B. Tran, N. H. Le, V. M. Ngo, and P. T. Nguyen, "Automatic detection of weeds: synergy between EfficientNet and transfer learning to enhance the prediction accuracy," *Soft comput*, vol. 28, no. 6, pp. 5029–5044, Mar. 2024, doi: 10.1007/s00500-023-09212-7.
- [17] M. Ijaz, N. Tariq, and A. Malik, "Performance Evaluation of the U-Net Model for Medical Image Segmentation Using Dice Coefficient, IOU, and Loss Metrics," *Hist Med*, vol. 10, no. 2, Sep. 2024, doi: 10.48047/HM.10.2.2024.1314-1324.
- [18] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med Imaging*, vol. 15, no. 1, Aug. 2015, doi: 10.1186/s12880-015-0068-x.
- [19] L. Yu, Z. Li, M. Xu, Y. Gao, J. Luo, and J. Zhang, "Distribution-aware Margin Calibration for Semantic Segmentation in Images," Dec. 2021, doi: 10.1007/s11263-021-01533-0.
- [20] Anonim, "Coffee Leaf Computer Vision Project," Jul. 2024, Roboflow. Accessed: Mar. 04, 2025. [Online]. Available: <https://universe.roboflow.com/tugas-akhir-adf4p/coffee-leaf>
- [21] M. Lavanya and R. Parameswari, "A Multiple Linear Regressions Model for Crop Prediction with Adam Optimizer and Neural Network Mlraonn," *IJACSA (International Journal of Advanced Computer Science and Applications)*, vol. 11, no. 4, 2020, [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [22] M. Sandler, A. Howard, M. Zhu, and A. Zhmoginov, "Sandler\_MobileNetV2\_Inverted\_Residuals\_CVPR\_2018\_paper.pdf," *ArXiv*, pp. 4510–4520, 2018.