

Revisiting Attention Mechanisms for Tea Leaf Disease Classification: An Experimental Study of EfficientNetV2-B0 with CBAM

Dewa Setyo Nugroho¹, Christy Atika Sari², Eko Hari Rachmawanto³, Ajib Susanto⁴

Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia

nugrohodewa05@gmail.com¹, christy.atika.sari@dsn.dinus.ac.id², eko.hari@dsn.dinus.ac.id³, ajib.susanto@dsn.dinus.ac.id⁴

Article Info

Article history:

Received 2026-09-02

Revised 2026-10-09

Accepted 2026-10-17

Keywords:

CBAM,
EfficientNetV2-B0,
Grad-CAM,
Tea Leaf Disease,
Transfer Learning.

ABSTRACT

Plant leaf diseases can reduce crop quality and productivity, making accurate disease identification important for supporting plant management. This study evaluates the performance of EfficientNetV2-B0 with and without the Convolutional Block Attention Module (CBAM) for tea leaf disease classification. The Tea Sickness Dataset was examined and preprocessed to address class imbalance. After preprocessing, six classes were retained with 90 images per class, resulting in 540 images. The dataset was divided into training, validation, and testing sets using a stratified split of 70%, 10%, and 20%, respectively. Two model configurations were evaluated under the same training and testing conditions. EfficientNetV2-B0 achieved a test accuracy of 92.59%, macro precision of 93.85%, macro recall of 92.59%, and macro F1-score of 92.61%. In comparison, EfficientNetV2-B0 + CBAM achieved 89.81% accuracy, 90.93% macro precision, 89.81% macro recall, and 89.68% macro F1-score. The results show that adding CBAM did not improve classification performance on the evaluated dataset. Furthermore, CBAM increased the number of model parameters without improving test performance. These findings indicate that the effectiveness of an attention mechanism depends on the characteristics of the dataset and model architecture. Under the experimental conditions of this study, EfficientNetV2-B0 without CBAM provided the best classification performance.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Tea (*Camellia sinensis*) is one of the agricultural commodities with significant economic value, making the health of tea plants and the quality of their production important to maintain. One of the problems that can affect tea plant productivity is the occurrence of leaf diseases. Leaf diseases can reduce crop quality and production yield, making rapid and accurate disease identification important for supporting plant management. Research on tea leaf disease classification has shown that the development of *deep learning* methods provides opportunities for automatic image-based disease identification [1], [2], [3].

The development of computer vision and deep learning has encouraged the use of image-based methods for automatic plant disease identification. One of the widely used approaches is Convolutional Neural Network (CNN), as it can

learn visual representations directly from images and can therefore be used to recognize disease characteristics such as color changes, spots, textures, and patterns on leaf surfaces. Various studies have applied deep learning approaches to leaf disease classification across different plant species and have demonstrated good classification performance [4], [5], [6], [7], [8], [9].

The development of deep learning models has also led to the use of lightweight architectures through a transfer learning approach. This approach allows models to utilize weights learned from large-scale datasets and adapt them to classification tasks on specific datasets. The application of transfer learning to leaf disease classification has been implemented using MobileNet and has demonstrated good performance in recognizing several leaf disease classes [10]. In tea leaf disease classification, Li et al. (2024) also applied fine-tuning and transfer learning strategies to MobileNetV3,

while Lien et al. (2025) used EfficientNet-B0 for tea leaf disease identification. Other studies have also shown that lightweight models with fewer parameters can provide a balance between classification performance and computational efficiency [11], [12].

Although CNN architectures have achieved good results in plant disease classification, the ability of models to distinguish symptoms with similar visual characteristics remains a challenge. Attention mechanisms have therefore been widely used to help models assign greater weight to feature information considered important. In plant disease classification, focusing on channel features and spatial locations can help models highlight characteristics related to disease symptoms [13], [14]. In tea leaf diseases, complex background conditions and similarities in disease characteristics also pose challenges for recognition, leading several studies to employ attention mechanisms to improve feature representation [14], [15], [16], [17].

Although various studies have shown that the use of lightweight deep learning architectures and attention mechanisms can provide good results in plant disease classification, their results still vary depending on the architecture, dataset, and model configuration used. Li and Zhao (2025) used ECA with ResNet50 for tea disease identification, while Tang et al. (2025) integrated CBAM into a tea leaf recognition model. Nigam et al. (2025) also combined CBAM with EfficientNet-B0 to estimate disease severity. The differences in architecture and classification objectives among these studies indicate that the effect of attention mechanisms needs to be evaluated on specific architectures and datasets before it can be concluded that they improve performance [14], [15], [20].

Based on previous studies, attention mechanisms have been applied to various architectures for plant disease classification, including EfficientNet-B0, ResNet50, and models specifically developed for tea leaf diseases. However, these studies used different base architectures, datasets, and classification objectives, so their results cannot yet be used to determine whether adding CBAM to EfficientNetV2-B0 will improve performance on the tea leaf disease dataset used in this study [14], [15], [20], [3]. Therefore, an experimental evaluation is needed to compare EfficientNetV2-B0 without CBAM and EfficientNetV2-B0 with CBAM using the same dataset and testing configuration.

Based on the identified problem and research gap, this study aims to classify tea leaf diseases using the EfficientNetV2-B0 architecture and evaluate the effect of adding the Convolutional Block Attention Module (CBAM) on classification performance. Two model configurations are compared using the same dataset, data split, and training parameters so that differences in performance can be observed in a more controlled manner. In addition to quantitative evaluation using accuracy, precision, recall, and F1-score, this study also uses a confusion matrix and Grad-CAM to help analyze classification error patterns and provide visual interpretations of model predictions. The main contribution of

this study is a controlled experimental comparison of EfficientNetV2-B0 with and without CBAM for tea leaf disease classification, providing empirical evidence on whether the addition of CBAM improves classification performance and model interpretability under the same experimental conditions [2], [21], [22].

II. METHODS

A. Dataset and Preparation

This study uses the Tea Sickness Dataset as the main dataset for tea leaf disease classification. The initial stage involved examining the dataset structure, the number of images in each class, and the data distribution before the training process. Dataset inspection and management are important steps in plant disease classification because an imbalance in the number of images between classes can affect the learning process and cause the model to become more dominant in learning classes with a larger number of samples. Dataset imbalance has also been identified in previous leaf disease classification studies, indicating that data management and augmentation strategies are needed to increase sample diversity and reduce bias toward certain classes [18]. In tea leaf disease studies, dataset management is also important because differences in the number and conditions of images can affect the training process and the model's generalization ability [3].

The initial dataset used in this study consisted of eight classes, namely *algal leaf*, *Anthraco*se, *bird eye spot*, *brown blight*, *gray light*, *healthy*, *red leaf spot*, and *white spot*. The number of images in each class was 113, 100, 99, 113, 98, 74, 144, and 142 images, respectively. The differences in the number of images indicate that the data distribution was not balanced, so an adjustment stage was required before the data were used for training. Data imbalance is also a common problem in leaf disease classification because classes with fewer samples may provide more limited feature representations than majority classes [18], [1].

Based on the results of the dataset examination, the *Anthraco*se and *white spot* classes were not used in the final experiment. The balancing process was then performed by setting the number of images to 90 for each class. Classes with more than 90 images underwent *undersampling*, while the *healthy* class, which had 74 images, was increased through augmentation-based *oversampling* until it reached 90 images. Thus, the final dataset consisted of six classes, namely *algal leaf*, *bird eye spot*, *brown blight*, *gray light*, *healthy*, and *red leaf spot*, with a total of 540 images. The use of a combination of reducing samples from majority classes and increasing data variation in classes with fewer samples is one of the strategies used to reduce the effect of dataset imbalance in leaf disease classification [18].

After the balancing process, the dataset was divided into training, validation, and test data using a stratified split with proportions of 70% for training, 10% for validation, and 20% for test. Stratified splitting was used to maintain the proportion of each class in each subset. The resulting split

consisted of 372 training images, 60 validation images, and 108 test images. The separation of training, validation, and test data is an approach used in plant disease classification studies to measure the model's ability to perform on data that are not used during the learning process [1], [5].

After the dataset was split, all images were resized to 224×224 pixels to ensure uniform input dimensions before being used in the model. For the training data, data augmentation was applied, including rotation of up to 20° , horizontal flipping, zooming of up to 15%, and brightness adjustment within the range of 0.8–1.2. Augmentation was applied only to the training data, while the validation and test data were used without additional transformations. The use of rotation, flipping, zooming, and brightness adjustment has been adopted in previous leaf disease classification studies to increase image variation and help models handle changes in visual conditions [18], [19].

B. Model Architecture

This study uses EfficientNetV2-B0 as the main architecture for tea leaf disease classification. The selection of this architecture is based on the use of the EfficientNet family in previous studies on plant disease and tea leaf disease classification. Lien et al. (2025) used EfficientNet-B0 as one of the models for tea leaf disease identification, while Chowdhury et al. (2026) used the EfficientNetV2 architecture for tea leaf disease classification. The results of these studies indicate that the EfficientNet family can be used to learn visual characteristics of leaf diseases and is one of the suitable approaches for plant image classification tasks [23], [3].

EfficientNetV2-B0 was used with a transfer learning approach using ImageNet pretrained weights. This approach allows the model to utilize previously learned feature representations so that the learning process can be adapted to the characteristics of the leaf disease dataset used in this study. The application of transfer learning in leaf disease classification has been used in previous studies. Li et al. (2024) applied a fine-tuning and transfer learning strategy to MobileNetV3 for tea leaf disease identification, while Alhammad et al. (2024) used a pretrained model for leaf disease classification and integrated Grad-CAM as part of the interpretability analysis. These studies indicate that the use of pretrained models can be an appropriate approach when a model needs to be adapted to a specific plant disease dataset [1], [21].

After the feature extraction process, the feature output from EfficientNetV2-B0 is summarized using Global Average Pooling (GAP) before being passed to the classification layer. GAP is used to calculate the average value across the spatial dimensions of each feature map, resulting in a more compact feature representation for the classification process. The global pooling approach has also been used in the development of plant disease classification models to reduce feature dimensions before entering the classification layer [1], [3]. In this study, the GAP output is then passed to a Dropout layer with a rate of 0.3 and a Dense layer with 256 neurons.

This structure follows the architectural configuration used in the experiment of this study, as shown in Equation (1).

$$zk = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{i,j,k} \quad (1)$$

After Global Average Pooling, the feature vector is passed to a Dense layer with 256 neurons and a ReLU activation function. The ReLU function is used to provide nonlinearity to the network, allowing the model to learn more complex feature representations. The use of ReLU in plant disease classification architectures has also been applied in lightweight deep learning models [24]. Subsequently, the output of the Dense layer is passed to the output layer with a softmax activation function to generate the probability of each class. The use of a probability-based output layer has also been applied in leaf disease classification using deep learning architectures [10]. In this study, the output layer consists of six neurons corresponding to the number of classes in the dataset, as shown in Equation (2).

$$ReLU(x) = \max(0, x) \quad (2)$$

The output of the Dense layer is then passed to the output layer with a softmax activation function to generate probabilities for each class. The softmax function is used in multiclass classification to convert the network output values into probabilities whose sum equals one. The softmax equation can be written as shown in Equation (3).

$$P(y_i | x) = \frac{e^{z_i}}{\sum_{j=1}^c e^{z_j}} \quad (3)$$

To improve the model's ability to emphasize features relevant to disease characteristics, this study adds the Convolutional Block Attention Module (CBAM) to EfficientNetV2-B0. CBAM is an attention mechanism that processes features through two stages, namely channel attention and spatial attention. The use of attention mechanisms in plant disease classification has been applied in previous studies, including EfficientNet-based models and tea leaf disease models [20], [14], [15]. In this study, CBAM is placed after the top activation output of EfficientNetV2-B0 and before Global Average Pooling. The CBAM process is performed sequentially through channel attention and spatial attention. The channel attention and spatial attention processes can be expressed sequentially as shown in Equation (4).

$$\begin{aligned} F' &= Mc(F) \otimes F \\ F'' &= Ms(F') \otimes F' \end{aligned} \quad (4)$$

During the training process, the model uses categorical cross-entropy as the loss function to measure the difference between the actual labels and the output probabilities of the model in multiclass classification. The use of categorical cross-entropy in plant disease classification has also been

applied in deep learning-based studies with probability outputs for each class [24], [1]. In this study, the loss function is used for both EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM together with the Adam optimizer and a learning rate of 0.0001, as shown in Equation (5).

$$L = - \sum_{i=1}^c y_i \log(\hat{y}_i) \quad (5)$$

C. Training Configuration

The training process was conducted using two model configurations, namely EfficientNetV2-B0 as the baseline and EfficientNetV2-B0 + CBAM as the model with the attention mechanism. Both models used an input image size of 224×224 pixels and a batch size of 8. The training configurations were kept the same for both models so that the performance comparison could be conducted in a more controlled manner and differences in the results could be attributed to the architectural changes rather than differences in training settings. The use of consistent training configurations has also been applied in plant disease classification studies when comparing several deep learning architectures [19], [5].

During the training process, image augmentation was applied to the training data to increase the variety of samples received by the model. The augmentation techniques used included rotation of up to 20° , horizontal flipping, zooming of up to 15%, and brightness adjustment within the range of 0.8–1.2. The validation and test data were not given any additional augmentation, so the evaluation was performed using images that did not undergo transformation during the testing process. The use of augmentation on training data to increase image diversity and help address data limitations has also been applied in leaf disease classification studies [18], [19].

During the training process, the initial weights of EfficientNetV2-B0 were initialized using ImageNet pretrained weights, while the base model was frozen so that parameter updates were focused on the added classification layers. The use of transfer learning by utilizing pretrained models and making adjustments to the required parts of the model has been applied in plant disease and tea leaf disease classification studies [1], [21]. Subsequently, the model was compiled using the Adam optimizer with a learning rate of 0.0001 and categorical cross-entropy as the loss function. The optimizer and learning rate configurations were applied identically to EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM so that the comparison between the two models was conducted under equivalent training conditions.

The training process was conducted for a maximum of 100 epochs, with validation accuracy used as the main metric to monitor model performance during training. The best model was selected based on the highest validation accuracy starting from the 10th epoch, and the model weights from the best epoch were saved for use in the testing stage. The use of validation data as the basis for model selection is an approach

used in plant disease classification studies to help obtain models with better generalization ability [5], [19]. In this study, the procedure was applied equally to EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM so that both models were compared under equivalent training conditions.

D. Model Evaluation

Model evaluation was conducted using test data that were not used during the training process or in selecting the best model. The test data consisted of 108 images from six classes, with 18 images in each class. The evaluation was performed after the best model from each configuration was obtained based on validation accuracy. The use of a test set separated from the training and validation data is intended to measure the model's generalization ability on data that were not used during the learning process [5], [24]. In this study, the same test set was used to evaluate EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM so that the comparison between the two models was conducted under the same testing conditions.

Model performance was evaluated using accuracy, precision, recall, and F1-score to determine the overall classification ability as well as the performance for each class. The use of multiple metrics provides a more comprehensive evaluation because accuracy indicates the proportion of correct predictions overall, while precision, recall, and F1-score are used to assess the accuracy and ability of the model to recognize each class. These metrics have also been used in leaf and plant disease classification studies to provide a more comprehensive evaluation of model performance [24], [1]. In this study, precision, recall, and F1-score were calculated using macro averaging so that each class contributed equally to the overall evaluation score.

Accuracy is calculated as the ratio of the number of correct predictions to the total number of test data, as shown in Equation (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision is calculated as the ratio of the number of correct positive predictions to the total number of positive predictions, as shown in Equation (7).

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall indicates the model's ability to correctly identify actual positive data, as shown in Equation (8).

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

F1-score is the harmonic mean of precision and recall, as shown in Equation (9).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

TP (*True Positive*) represents the number of positive data correctly predicted as positive, TN (*True Negative*) represents the number of negative data correctly predicted as negative, FP (*False Positive*) represents the number of negative data incorrectly predicted as positive, while FN (*False Negative*) represents the number of positive data incorrectly predicted as negative.

The precision, recall, and F1-score values in this study were also calculated using macro averaging, so that each class contributed equally to the overall evaluation. The notebook implementation uses `average='macro'` to obtain these three metrics for both the baseline model and the CBAM model.

In addition to the quantitative metrics, a confusion matrix was used to show the prediction patterns for each class. The confusion matrix compares the actual classes with the predicted classes, allowing correct predictions to be identified on the diagonal of the matrix and classification errors to be identified outside the diagonal. The use of a confusion matrix has also been applied in leaf disease classification studies to determine the distribution of model errors across classes and evaluate the model's ability on data that were not used during the learning process [25]. In this study, the confusion matrix was used on the test set consisting of six classes to analyze the classification error patterns of EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM.

In addition to the confusion matrix, this study uses Gradient-weighted Class Activation Mapping (Grad-CAM) to provide a visual interpretation of the model's predictions. Grad-CAM is used to show the areas in an image that contribute to the model's decision, helping to understand which parts of the image receive attention during classification. The use of Grad-CAM as an interpretability approach has been applied in tea leaf disease classification and other plant disease studies to examine the relationship between image regions and model predictions [2], [21], [22]. In this study, Grad-CAM is used to compare the attention patterns between EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM on one *bird eye spot* image from the test set.

III. RESULTS AND DISCUSSION

A. Results of Dataset Preparation and Balancing

The initial examination showed that the *Tea Sickness Dataset* consisted of eight classes, namely *algal leaf*, *Anthraco*se, *bird eye spot*, *brown blight*, *gray light*, *healthy*, *red leaf spot*, and *white spot*. The numbers of images in each class were 113, 100, 99, 113, 98, 74, 144, and 142 images, respectively. The distribution indicates an imbalance in the number of samples across classes, particularly between the

healthy class and the classes with a larger number of images. This condition needs to be considered because data imbalance can cause the model to learn more characteristics from majority classes and provide less optimal representations for classes with fewer samples. Data imbalance is also one of the challenges in leaf disease classification that encourages the use of augmentation strategies and sample distribution management before the training process [18].

In the preparation stage, the *Anthraco*se and *white spot* classes were excluded from the final experiment. Subsequently, the number of images in the five classes that had more than 90 images was reduced using *undersampling*, while the *healthy* class, which had 74 images, was increased by 16 images through augmentation so that each class contained 90 images. Thus, the final dataset consisted of six classes with a total of 540 images. Balancing the class distribution by reducing samples from majority classes and adding variations to classes with fewer samples was used to reduce the effect of data imbalance on the model learning process [18].

The balanced dataset was then divided using a stratified split into training, test, and validation data with a ratio of 70:20:10. The resulting split consisted of 372 training images, 108 test images, and 60 validation images, with each class containing 62 training images, 18 test images, and 10 validation images, as shown in Table I. Stratified data splitting was used to maintain the proportion of each class in each subset so that the training and evaluation processes could be conducted in a more controlled manner. The approach of separating data into training, validation, and test sets has also been used in plant disease classification studies to obtain more measurable model evaluation results [1], [5].

TABLE I
DATASET DISTRIBUTION AFTER BALANCING

Class	Training	Validation	Test	Total
<i>Algal leaf</i>	62	10	18	90
<i>Bird eye spot</i>	62	10	18	90
<i>Brown blight</i>	62	10	18	90
<i>Gray light</i>	62	10	18	90
<i>Healthy</i>	62	10	18	90
<i>Red leaf spot</i>	62	10	18	90
Total	372	60	108	540

After the dataset splitting process, data augmentation was applied to the training data to increase the variety of images used during the learning process. The applied augmentation included rotation of up to 20°, horizontal flipping, zooming of up to 15%, and brightness adjustment within the range of 0.8–1.2. Meanwhile, the validation and test data were not subjected to additional augmentation. This process allowed the model to obtain more diverse image variations without modifying the data used for evaluation. The use of augmentation on training data has also been reported to help increase data diversity and generalization ability in leaf disease classification [18], [19].

B. Training Results of EfficientNetV2-B0

The EfficientNetV2-B0 model without CBAM was used as the baseline model to determine its classification performance before adding the attention mechanism. The training process was conducted for a maximum of 100 epochs, with validation accuracy used as the basis for selecting the best model. At the 10th epoch, the validation accuracy reached 80.00%, then increased to 88.33% at the 18th epoch and 91.67% at the 25th epoch. The highest validation accuracy was obtained at the 71st epoch, reaching 95.00%, so the model at that epoch was selected as the best model for the testing stage. The use of validation accuracy as a basis for model selection is consistent with plant disease classification studies that use validation data to monitor the learning process and evaluate the model's generalization ability [5], [19].

After the best model was selected based on validation accuracy, testing was conducted using 108 images from the test set. The EfficientNetV2-B0 model obtained a test loss of 0.2134 and a test accuracy of 92.59%. These results indicate that the model was able to correctly classify most of the images in the test data. The test accuracy obtained shows that EfficientNetV2-B0 can be used as a reasonably strong baseline model for leaf disease classification on the dataset used in this study. The use of EfficientNet models in plant and leaf disease classification has also demonstrated good classification performance in previous studies [23], [3].

Based on the classification report in Table II, the performance of the EfficientNetV2-B0 model varied across classes. The healthy class achieved the best results, with precision, recall, and F1-score each reaching 1.00. The bird eye spot and gray light classes also showed high performance, with an F1-score of 0.97 for each class. Meanwhile, the algal leaf class achieved a precision of 1.00 but a recall of only 0.72, whereas the brown blight class obtained a precision of 0.74 and a recall of 0.94. The differences in the values across classes indicate that the model's ability to recognize the visual characteristics of each disease was not yet fully consistent. The analysis of precision, recall, and F1-score in plant disease classification has also been used to examine model performance in greater detail for each class [24], [20].

The results show that EfficientNetV2-B0 was able to achieve reasonably good classification performance without the addition of CBAM. However, the performance across classes was not yet fully consistent. The most noticeable difference was observed in the algal leaf class, which had a precision of 1.00 but a recall of only 0.72. This means that when the model predicted an image as algal leaf, the prediction tended to be correct, but some algal leaf images were still predicted as other classes. In contrast, brown blight had a recall of 0.94 but a precision of 0.74, indicating that the model was relatively capable of identifying brown blight images, but some images predicted as brown blight actually belonged to other classes. These results show that the overall accuracy value does not fully represent the same level of model performance across all classes, so per-class metric analysis is needed to examine the characteristics of classification errors in greater detail [20], [24].

TABLE II
CLASSIFICATION RESULT OF EFFICIENTNETV2-B0

Class	Precision	Recall	F1-score
<i>Algal leaf</i>	1.00	0.72	0.84
<i>Bird eye spot</i>	1.00	0.94	0.97
<i>Brown blight</i>	0.74	0.94	0.83
<i>Gray light</i>	0.95	1.00	0.97
<i>Healthy</i>	1.00	1.00	1.00
<i>Red leaf spot</i>	0.94	0.94	0.94
<i>Macro average</i>	0.94	0.93	0.93
Accuracy	0.93	-	-

The evaluation results on the test set show that EfficientNetV2-B0 without CBAM achieved good classification performance, with a test accuracy of 92.59% and a macro F1-score of 0.93. However, differences in recognition ability were still observed across classes, particularly for algal leaf and brown blight. This condition indicates that the baseline model was able to learn the visual characteristics of most classes but still encountered difficulties with certain classes that had more difficult-to-distinguish visual characteristics. These findings provide a basis for comparing the performance of the baseline model with EfficientNetV2-B0 with CBAM in the following section. Per-class evaluation metrics such as precision, recall, and F1-score have also been used in plant disease studies to examine model performance in greater detail for each class [24].

C. Training Results of EfficientNetV2-B0 + CBAM

The second model used the EfficientNetV2-B0 architecture with the addition of the Convolutional Block Attention Module (CBAM). All dataset and training parameter configurations were kept the same as those of the baseline model, so that differences in the results could be used to examine the effect of adding CBAM. In this model, CBAM was placed after the top activation output of EfficientNetV2-B0 and consisted of channel attention and spatial attention before the features were passed to Global Average Pooling and the classification layer. The use of the CBAM mechanism in EfficientNet models has been reported to help the model focus on relevant features in disease images, although its effect may vary depending on the characteristics of the dataset and model configuration [20], [15].

The training of EfficientNetV2-B0 + CBAM was conducted for a maximum of 100 epochs, with validation accuracy used as the basis for monitoring the best model. Based on the training results, the validation accuracy increased from 45.00% in the first epoch to 85.00% at the 10th epoch, then reached 90.00% at the 29th epoch and 91.67% at the 31st epoch. The highest validation accuracy was obtained at the 37th epoch, reaching 93.33%, so the model at that epoch was selected as the best model for the testing stage. After the 37th epoch, the validation accuracy fluctuated and did not exceed this value again.

After the best model was obtained at the 37th epoch, testing was conducted using 108 images from the test set. The EfficientNetV2-B0 + CBAM model obtained a test loss of 0.26697 and a test accuracy of 89.81%. This value was lower

than that of the EfficientNetV2-B0 model without CBAM, which obtained a test accuracy of 92.59%. Therefore, under the dataset and experimental configuration used in this study, the addition of CBAM did not improve classification accuracy. These results indicate that adding an attention mechanism does not automatically improve performance when applied to different architectures and datasets. This finding differs from the results reported by Nigam et al. (2025), who reported improved performance after CBAM was added to EfficientNet-B0. Therefore, the effectiveness of CBAM needs to be considered based on the characteristics of the dataset and the model configuration used.

Based on the classification report, the performance of EfficientNetV2-B0 + CBAM varied across classes. The healthy class achieved the highest results, with precision, recall, and F1-score each reaching 1.00. The red leaf spot class also showed high performance, with an F1-score of 0.97. Meanwhile, the algal leaf class had a precision of 0.93, a recall of 0.72, and an F1-score of 0.81, while the bird eye spot class had a recall of 0.78. These results show that although the model with CBAM was still able to recognize several classes well, its classification ability was not evenly distributed across all classes. The use of precision, recall, and F1-score to examine the performance of each class has also been applied in plant disease classification studies [1], [20].

TABLE III
CLASSIFICATION RESULT OF EFFICIENTNETV2-B0 + CBAM

Class	Precision	Recall	F1-score
Algal leaf	0.93	0.72	0.81
Bird eye spot	1.00	0.78	0.88
Brown blight	0.76	0.89	0.82
Gray light	0.82	1.00	0.90

TABLE IV
PERFORMANCE COMPARISON OF THE TWO MODELS

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score
EfficientNetV2-B0	92.59%	93.85%	92.59%	92.61%
EfficientNetV2-B0 + CBAM	89.81%	90.93%	89.81%	89.68%
Change	-2.78%	-2.93%	-2.78%	-2.93%

Based on Table IV, EfficientNetV2-B0 without CBAM achieved higher performance than EfficientNetV2-B0 + CBAM across all evaluation metrics. The test accuracy of EfficientNetV2-B0 reached 92.59%, while the model with CBAM achieved 89.81%, resulting in a decrease of 2.78 percentage points. A decrease was also observed in macro precision, macro recall, and macro F1-score. Macro precision decreased from 93.85% to 90.93%, macro recall from 92.59% to 89.81%, and macro F1-score from 92.61% to 89.68%. These results indicate that under the dataset and experimental configuration used in this study, the addition of CBAM did not provide an overall improvement in classification performance.

The results of this study show a different pattern from the study by Nigam et al. (2025), which reported that adding CBAM to EfficientNet-B0 increased the test accuracy from

Healthy	1.00	1.00	1.00
Red leaf spot	0.95	1.00	0.97
Macro average	0.91	0.90	0.90

Based on Table III, the healthy and red leaf spot classes showed the best classification performance after the addition of CBAM, with a recall of 1.00 for each class. In contrast, the algal leaf class still had a recall of 0.72, while the bird eye spot class obtained a recall of 0.78. This condition indicates that CBAM was not able to improve the classification errors in several classes that had also experienced difficulties in the baseline model. These results show that the addition of an attention mechanism does not always result in improvement for every class, as its effect may depend on the feature characteristics and visual patterns present in the dataset. This finding is consistent with studies showing that model performance with CBAM can differ across datasets and architectural configurations [13], [20].

D. Comparison of EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM

The performance comparison was conducted using the same test set, consisting of 108 images from six classes. The evaluation was performed on the best model from each configuration using accuracy, precision, recall, and F1-score. The use of multiple evaluation metrics is necessary to provide a more comprehensive overview of model performance because the overall accuracy value does not necessarily indicate balanced performance across all classes. The use of multiple evaluation metrics has also been applied in plant disease classification studies to compare model performance and examine classification ability for each class [1], [20].

92.19% to 96.68%. That study was conducted on a different dataset and task, namely wheat rust disease severity estimation, so its results cannot be directly equated with those of this study. Another study also showed that model performance is strongly influenced by the selection of architecture, dataset, and experimental configuration [8]. In this study, the addition of CBAM instead decreased the test accuracy from 92.59% to 89.81%. This difference indicates that the effect of the attention mechanism strongly depends on the characteristics of the dataset, the base architecture, and the experimental configuration used.

In addition to the differences in classification performance, the addition of CBAM also increased the model complexity. EfficientNetV2-B0 without CBAM had 6,248,790 parameters, with 329,478 trainable parameters, whereas EfficientNetV2-B0 + CBAM had 6,659,928 parameters, with 740,616 trainable parameters. Thus, the addition of CBAM

resulted in an increase of approximately 411,138 parameters, but this increase in complexity was not accompanied by improved performance on the test data. This condition indicates that the addition of an attention module needs to be considered not only based on its potential to improve accuracy, but also based on architectural efficiency and the number of parameters. Consideration of model size and computational efficiency is also important in the development of lightweight models for plant disease classification [11], [9].

E. Analysis of Confusion Matrix

The confusion matrix analysis was conducted to determine the classification error patterns for each class in the EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM models. The confusion matrix was generated based on a comparison between the actual labels and the predicted results on the test set. The test data consisted of six classes with 18 images for each class, resulting in 108 images being evaluated. The confusion matrix analysis was used as a complement to the evaluation metrics to examine the distribution of correct and incorrect predictions for each class. The use of confusion matrices and classification metrics has also been applied in deep learning studies for leaf diseases to evaluate model performance in greater detail [20], [1], [7].

Based on the confusion matrix in Figure 1, the EfficientNetV2-B0 model correctly classified all images in the gray light and healthy classes. For the bird eye spot class, 17 out of 18 images were correctly classified, while one image was incorrectly classified. The red leaf spot class also obtained 17 correct predictions out of 18 images. Meanwhile, classification errors were relatively more noticeable in the algal leaf class, which had a recall of 0.72, and the brown blight class, which had a precision of 0.74. This pattern shows that the baseline model was able to distinguish several classes well but still encountered difficulties with certain classes that had more difficult-to-distinguish visual characteristics.

Based on the confusion matrix in Figure 2, the EfficientNetV2-B0 + CBAM model correctly classified all images in the healthy and red leaf spot classes. However, several classes still showed classification errors. The bird eye spot class had a recall of 0.78, while the algal leaf class still had a recall of 0.72. For the brown blight class, the recall reached 0.89. These results show that after CBAM was added, several classes still experienced difficulties during classification and prediction errors were not completely reduced. This condition indicates that the attention mechanism does not always improve all classification error patterns across every class.

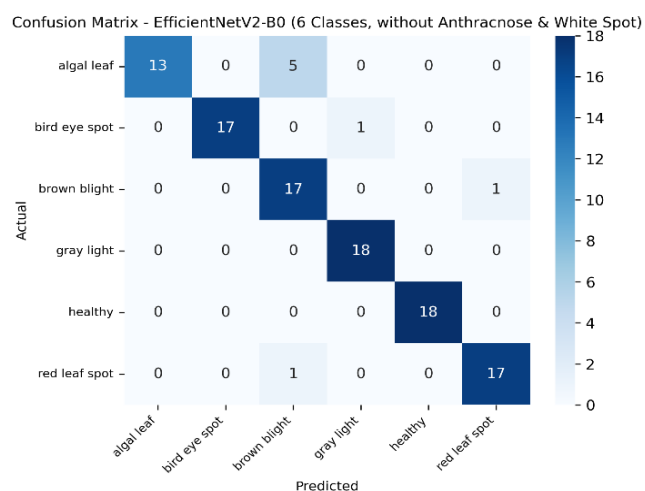


Figure 1. Confusion Matrix of EfficientNetV2-B0

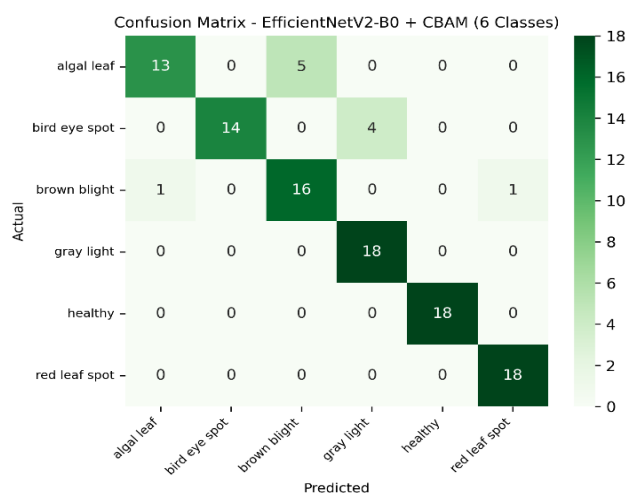


Figure 2. Confusion Matrix of EfficientNetV2-B0 + CBAM

The comparison of Figures 1 and 2 shows that the classification error patterns of the two models were not completely the same. In the EfficientNetV2-B0 model, the bird eye spot class had a recall of 0.94, whereas after the addition of CBAM, its value decreased to 0.78. For the algal leaf class, the recall remained at 0.72 in both models, indicating that the addition of CBAM did not improve this class. Meanwhile, brown blight had a recall of 0.94 in the baseline model and 0.89 in the model with CBAM. These changes show that the addition of CBAM was not able to consistently reduce classification errors in the classes that had previously experienced difficulties. This result is consistent with the overall performance comparison, which shows that EfficientNetV2-B0 without CBAM achieved better results than the model with CBAM.

F. Analysis of Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) analysis was used as an additional analysis to examine the image areas that received attention from the model during

prediction. In this study, Grad-CAM was applied to one *bird eye spot* image from the test set to compare the attention patterns of EfficientNetV2-B0 as the baseline model and EfficientNetV2-B0 + CBAM. For the baseline model, Grad-CAM used the output of the *top_conv* layer, whereas the

model with CBAM used the output of the *spatial_attention* layer. The use of Grad-CAM to provide visual interpretation of model decisions has also been applied in tea leaf disease classification and other plant disease classification studies [2], [21], [22].

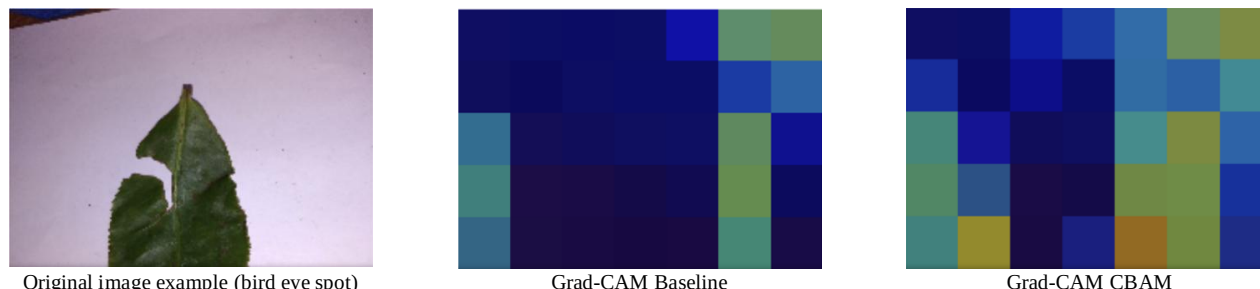


Figure 3. Grad-CAM Comparison of EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM

Based on the Grad-CAM visualization in Figure 3, there were differences in activation patterns between EfficientNetV2-B0 and EfficientNetV2-B0 + CBAM on the bird eye spot image. Both models responded to different areas of the leaf in terms of activation intensity and location, indicating differences in the feature patterns used when the models generated their predictions. In the model with CBAM, the activation pattern was derived from the output of the spatial attention layer, whereas in the baseline model, it was derived from the *top_conv* layer. These differences can provide additional information for understanding how each model configuration utilizes visual features in leaf images. The Grad-CAM results in this study were used to support the interpretation of the model evaluation results and not as a direct measure for determining which model was better. The visualization was used to help understand the areas receiving the model's attention, which were then compared with the accuracy, precision, recall, F1-score, and confusion matrix results.

IV. CONCLUSION

Based on the results of the tea leaf disease classification study using EfficientNetV2-B0 with and without the addition of the Convolutional Block Attention Module (CBAM), the dataset examination and balancing process resulted in six final classes with 90 images per class, resulting in a total of 540 images. The dataset was then divided using a stratified split into 372 training images, 60 validation images, and 108 test images. The testing results showed that EfficientNetV2-B0 without CBAM achieved the best performance, with a test accuracy of 92.59%, macro precision of 93.85%, macro recall of 92.59%, and macro F1-score of 92.61%. Meanwhile, EfficientNetV2-B0 + CBAM achieved a test accuracy of 89.81%, macro precision of 90.93%, macro recall of 89.81%, and macro F1-score of 89.68%. The comparison of the two models showed that the addition of CBAM did not improve classification performance on the dataset used. Test accuracy decreased by 2.78 percentage points, from 92.59% to 89.81%.

In addition, the number of parameters increased from 6,248,790 in EfficientNetV2-B0 to 6,659,928 in EfficientNetV2-B0 + CBAM. Under the dataset and experimental configuration used in this study, EfficientNetV2-B0 without CBAM was the more effective model compared with EfficientNetV2-B0 + CBAM.

REFERENCES

- [1] Y. Li *et al.*, "Tea leaf disease and insect identification based on improved MobileNetV3," *Front. Plant Sci.*, vol. 15, 2024, doi: 10.3389/fpls.2024.1459292.
- [2] O. Ozturk, B. Sarica, and D. Z. Seker, "Interpretable and Robust Ensemble Deep Learning Framework for Tea Leaf Disease Classification," *Horticulturae*, vol. 11, no. 4, Apr. 2025, doi: 10.3390/horticulturae11040437.
- [3] S. H. Chowdhury, M. S. Arman, M. B. H. Siam, M. R. Khan, A. Hasan, and P. A. M. Fahim, "A curated dataset and lightweight deep learning framework for tea leaf disease classification," *PLoS One*, vol. 21, no. 5 May, May 2026, doi: 10.1371/journal.pone.0349210.
- [4] N. Ullah *et al.*, "An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model," *Front. Plant Sci.*, vol. 14, 2023, doi: 10.3389/fpls.2023.1212747.
- [5] M. S. Krishna, P. Machado, R. I. Otuka, S. W. Yahaya, F. Neves dos Santos, and I. K. Ihianle, "Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach," *J (Basel)*, vol. 8, no. 1, p. 4, Jan. 2025, doi: 10.3390/j8010004.
- [6] S. Murugesan, J. Chinnadurai, S. Srinivasan, S. K. Mathivanan, R. R. Chandan, and U. Moorthy, "Robust multiclass classification of crop leaf diseases using hybrid deep learning and Grad-CAM interpretability," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-14847-7.
- [7] C. Subbarayudu and M. Kubendiran, "An automated hybrid deep learning framework for paddy leaf disease identification and classification," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-08071-6.
- [8] B. Naresh kumar and S. Sakthivel, "Rice leaf disease classification using a fusion vision approach," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-87800-3.
- [9] G. Sambasivam, G. Prabu kanna, M. S. Chauhan, P. Raja, and Y. Kumar, "A hybrid deep learning model approach for automated detection and classification of cassava leaf diseases," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-90646-4.
- [10] E. Elfatimi, R. Eryigit, and L. Elfatimi, "Beans Leaf Diseases Classification Using MobileNet Models," *IEEE Access*, vol. 10, pp. 9471–9482, 2022, doi: 10.1109/ACCESS.2022.3142817.

- [11] M. Asghar *et al.*, "A lightweight hybrid model for scalable and robust plant leaf disease classification," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-08788-4.
- [12] A. Srivastava *et al.*, "Deep convolutional models for robust multi-crop disease recognition in real-world conditions," *Sci. Rep.*, May 2026, doi: 10.1038/s41598-026-52369-y.
- [13] Y. Wang *et al.*, "Classification of Plant Leaf Disease Recognition Based on Self-Supervised Learning," *Agronomy*, vol. 14, no. 3, Mar. 2024, doi: 10.3390/agronomy14030500.
- [14] L. Li and Y. Zhao, "Tea disease identification based on ECA attention mechanism ResNet50 network," *Front. Plant Sci.*, vol. 16, 2025, doi: 10.3389/fpls.2025.1489655.
- [15] X. Tang, L. Tang, J. Li, and X. Guo, "Enhancing multilevel tea leaf recognition based on improved YOLOv8n," *Front. Plant Sci.*, vol. 16, 2025, doi: 10.3389/fpls.2025.1540670.
- [16] J. Ai, Y. Li, S. Gao, R. Hu, and W. Che, "Tea Disease Detection Method Based on Improved YOLOv8 in Complex Background," *Sensors*, vol. 25, no. 13, Jul. 2025, doi: 10.3390/s25134129.
- [17] J. Song, Y. Zhang, S. Lin, H. Han, and X. Yu, "TLDDM: An Enhanced Tea Leaf Pest and Disease Detection Model Based on YOLOv8," *Agronomy*, vol. 15, no. 3, Mar. 2025, doi: 10.3390/agronomy15030727.
- [18] B. Min, T. Kim, D. Shin, and D. Shin, "Data Augmentation Method for Plant Leaf Disease Recognition," *Applied Sciences (Switzerland)*, vol. 13, no. 3, Feb. 2023, doi: 10.3390/app13031465.
- [19] J. Rahaman, P. Paul, A. Chowdhury, and M. Quamruzzaman, "A customized MobileNetV3Large-based deep learning framework for plant disease detection," *Discover Artificial Intelligence*, vol. 6, no. 1, Dec. 2026, doi: 10.1007/s44163-025-00733-8.
- [20] S. Nigam, R. Jain, V. K. Singh, A. K. Singh, and H. Krishna, "Automated severity level estimation of wheat rust using an EfficientNet-CBAM hybrid model," *Front. Plant Sci.*, vol. 16, 2025, doi: 10.3389/fpls.2025.1540642.
- [21] S. M. Alhammad, D. S. Khafaga, W. M. El-hady, F. M. Samy, and K. M. Hosny, "Deep learning and explainable AI for classification of potato leaf diseases," *Front. Artif. Intell.*, vol. 7, 2024, doi: 10.3389/frai.2024.1449329.
- [22] E. Udoh, M. A. Alaoui Mhamdi, and M. Allili, "An Efficient Attention-Enhanced MobileNetV2 Framework for Plant Disease Detection on Resource-Constrained Devices," *Electronics (Basel)*, vol. 15, no. 15, p. 3343, Jul. 2026, doi: 10.3390/electronics15153343.
- [23] B. Y. Lien and C. C. Lai, "A Lightweight Deep Learning Model for Tea Leaf Disease Identification," *Mach. Learn. Knowl. Extr.*, vol. 7, no. 4, Dec. 2025, doi: 10.3390/make7040123.
- [24] U. V. Nnamdi and V. Abolghasemi, "Optimised MobileNet for very lightweight and accurate plant leaf disease detection," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-27393-z.
- [25] K. Joshi *et al.*, "Classification of cotton leaf disease using YOLOv8 based k-fold cross validation deep learning method for precision agriculture," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-13147-4.