Implementation of an EfficientNet-B4 Model Architecture with a Convolutional Block Attention Module (CBAM) for Betel Leaf Disease Classification

Rima Tri Wahyunigrum
Department of Informatics Engineering, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia
rimatriwahyuningrum@trunojoyo.ac.id (corresponding author)

Indah Agustien Siradjuddin
Department of Informatics Engineering, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia
indah.siradjuddin@trunojoyo.ac.id

Triasmi Dwi Farawati
Department of Informatics Engineering, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia
triasmidwii08@gmail.com

Irmalia Suryani Faradisa
Department of Electrical Engineering, Institut Teknologi Nasional, Malang, Indonesia
irmalia_suryani_faradisa@lecturer.itn.ac.id

Achmad Bauravindah
Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia
baurav99@gmail.com

Deshinta Arrova Dewi
Center for Data Science and Sustainable Technologies, INTI International University, Nilai, Malaysia
deshinta.ad@newinti.edu.my

Received: 3 May 2025 | Revised: 28 June 2025, 8 July 2025, and 17 July 2025 | Accepted: 19 July 2025

 DOI: https://doi.org/10.48084/etasr.11900

ABSTRACT

Betel leaf farming plays a significant role in the agricultural economy of Southeast Asia, particularly in Indonesia, supporting the cultural practices and rural livelihoods. However, the sector faces challenges from diseases caused by fungal, bacterial, and viral pathogens, leading to significant yield losses. Traditional leaf disease detection methods are limited in accuracy and timeliness, necessitating innovative solutions. This study presents an advanced approach leveraging EfficientNet-B4, enhanced with the Convolutional Block Attention Module (CBAM), for betel leaf disease detection. A localized dataset of 4,000 high-resolution images of betel leaves, categorized into four classes, was used to ensure relevance to the Indonesian agriculture. The CBAM-enhanced model demonstrated superior performance in identifying disease-specific patterns, achieving an average accuracy of 95.6%, compared to the 90% with the base model. Using metrics, such as precision, recall, F1-score, and ROC-AUC, the proposed methodology highlights its robustness and reliability. The study's findings underscore the importance of

localized datasets and attention mechanisms in improving the disease classification accuracy. Practical implications include the potential for real-time deployment of the model on mobile platforms, enabling early detection and intervention. This approach promises to enhance the sustainability of betel leaf farming by minimizing the crop losses, reducing chemical usage, and supporting the farmers' economic well-being.

Keywords-classification; betel leaf; EfficientNet-B4; Convolutional Block Attention Module (CBAM); process innovation

I. INTRODUCTION

Betel leaf farming plays a crucial role in the agricultural economy of Southeast Asia, especially in Indonesia [1]. It contributes significantly to the rural livelihoods and cultural traditions, such as ceremonies and medicinal use [2]. In several regions, such as West Bengal and Bangladesh, betel leaf cultivation supports the economic growth, with reports of improved living standards among the farming communities [3, 4]. In Indonesia, the traditional practice of chewing betel leaves (known as nyirih) remains prevalent and holds cultural significance [2-5]. Despite its cultural and economic value, betel leaf farming faces critical challenges due to diseases caused by fungal, bacterial, and viral pathogens. These diseases, including foot rot (Phytophthora parasitica), collar rot (Sclerotium rolfsii), and bacterial leaf spot (Xanthomonas campestris pv. betlicola), can cause yield losses of up to 100% [6]. In humid conditions, common diseases, such as leaf spot and anthracnose spread rapidly, leading to discoloration, lesions, and defoliation that degrade both the quality and quantity [6, 7]. Visual inspections remain the primary detection method, yet they are prone to error and ineffective during the early disease stages. While chemical treatments exist, they pose environmental risks and can lead to pathogen resistance [6, 8].

Advancements in Artificial Intelligence (AI) and Machine Learning (ML) present promising solutions for disease detection in agriculture. Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown high accuracy in classifying the plant diseases. For instance, DenseNet201 has achieved up to 98.77% accuracy in betel leaf disease detection [9, 10]. Extensive datasets have been leveraged to train models capable of identifying the diseases across various crops [1]. These image-based disease detection systems powered by deep learning models also utilize large datasets, such as the 54,306 images of 14 crop species analyzed in [11].

Models integrating attention mechanisms have further improved the performance by focusing on disease-relevant features in complex environments. The Spatial Feature-Enhanced Attention Neural Network achieved a Top-1 accuracy of 85.29% in pest recognition tasks [12], highlighting the importance of fine-grained feature extraction. These models enhance the early detection, enabling timely interventions and minimizing the crop loss.

Despite the advancements, the adoption of AI in betel leaf farming in Indonesia remains limited, primarily due to the lack of localized datasets, computational inefficiencies, and the absence of farmer-friendly tools. Local datasets have proven essential for improving the model accuracy, as shown by the utilization of region-specific data, such as 1,189 images from a betel farm in India [9, 10] and the CropDP-181 dataset with over 123,000 diverse images [13]. In Indonesia, most models rely on generic datasets, reducing their effectiveness in recognizing region-specific disease symptoms.

Another major challenge is the ability of models to extract detailed features from leaf images. The variability in training data can lead to significant drops in performance, particularly in underrepresented classes [13, 14]. Attention mechanisms, such as the CBAM, have been proposed to address this issue by enhancing the model's ability to focus on relevant image regions [15].

For example, combining CBAM with EfficientNet-V2 has improved the classification accuracy in histopathology and plant disease datasets, reaching up to 98.96% in some cases [16]. This demonstrates how integrating advanced attention mechanisms into deep learning architectures can significantly enhance the performance of disease detection systems for specific agricultural tasks. EfficientNet-B4, a state-of-the-art model, leverages compound scaling to balance the depth, width, and resolution, achieving high performance with reduced computational requirements [16, 17]. When integrated with CBAM, the model demonstrates superior performance in extracting fine-grained features, achieving competitive results across agricultural tasks. EfficientNet-B4 has shown faster inference speeds compared to traditional models, making it suitable for real-time deployment on mobile platforms [18].

This study proposes the integration of EfficientNet-B4 and CBAM using a localized dataset specialized in Indonesian betel leaf farming. This combination aims to overcome the limitations in existing detection systems by improving accuracy, scalability, and regional relevance. The approach supports the early disease detection, helping farmers reduce losses, minimize chemical usage, and ensure crop sustainability.

The integration of advanced AI architectures with localized datasets presents a promising solution to the challenges faced by betel leaf farmers in Indonesia. By addressing issues related to data specificity, feature extraction, and deployment constraints, this study contributes to the development of effective and scalable tools for disease detection. Ultimately, the proposed approach supports the long-term sustainability of betel leaf farming by minimizing the yield losses, reducing the dependency on chemical treatments, and improving farmers' livelihoods.

II. MATERIALS AND METHODS

A. Dataset

The dataset used in this research consists of 4,000 high-resolution images of betel leaves, categorized into four distinct classes: healthy green betel leaves, green betel leaves with anthracnose, green betel leaves with bacterial leaf spot, and healthy red betel leaves, with 1,000 images per class. These images were captured using a Samsung Galaxy A52s smartphone equipped with a 64 MP ISOCELL GW3 sensor (0.8μm pixel size, f/1.8 aperture, autofocus, and Optical Image Stabilization). Each image has a native resolution of 9248 × 6936 pixels and was photographed at approximately 15 cm distance under varying natural lighting conditions between 09:00 and 17:00 WIB, across several days. To ensure variability and robustness, each of the 25 sampled leaves per class was captured in 40 different poses, resulting in diverse perspectives in terms of angle, lighting, and background.

To promote transparency and reproducibility, the entire dataset has been made publicly available via Kaggle. The dataset includes original images in JPEG format as well as preprocessed .npy files. Additionally, a CSV file is provided that details the image paths, class labels, one-hot encodings, and 5-fold cross-validation splits used during training and evaluation. This structured organization supports the easy integration into the deep learning pipelines and ensures that future research can directly build upon this work. Sample images from each class are illustrated in Figure 1.

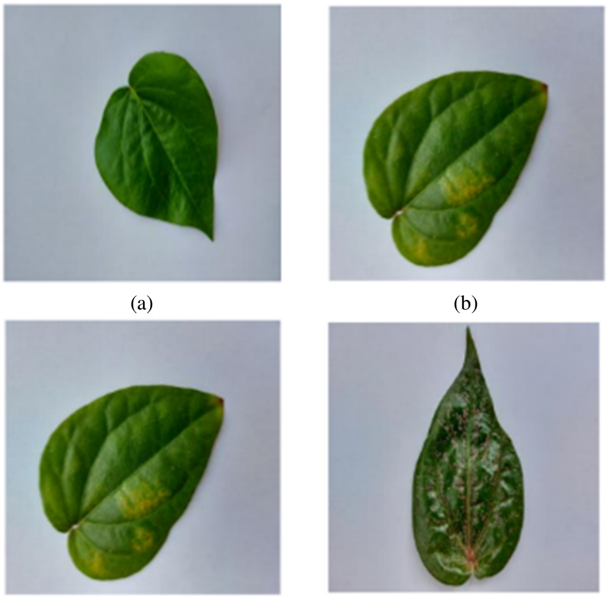
Original images were stored in JPEG format with 8-bit depth per channel, resulting in RGB images with pixel intensity values ranging from 0 to 255 (uint8). During preprocessing, the images were resized to 224×224 pixels and normalized to a (0,1) range using TensorFlow/Keras preprocessing utilities before being input into the EfficientNet-B4-CBAM model.

To ensure the authenticity and relevance of the data, the dataset was reviewed and validated by the Agriculture and Horticulture Office of Bangkalan Regency. This validation was formalized through letter number 521/233/433.110/2023, issued on February 24, 2023. The dataset preparation process emphasized the quality and classification accuracy to enable a reliable analysis.

B. *Research Workflow*

The research workflow, as illustrated in Figure 2, outlines the methodological steps undertaken in this study. The process begins with data preprocessing, where input images are resized to 224×224 pixels and normalized to a (0,1) range. These operations are performed using the TensorFlow and Keras preprocessing utilities to ensure consistency and compatibility with the neural network architecture. Subsequently, the dataset was randomly split into training and testing subsets, with 80% allocated for training and 20% for testing, maintaining diversity and class balance.

In the training phase, the EfficientNet-B4 model was employed in combination with the CBAM. The model was initialized using pretrained ImageNet weights and fine-tuned on the localized dataset. The implementation utilized TensorFlow and Keras libraries, with GPU acceleration to handle the computational load.



(a) (b) (c) (d)

Fig. 1. Examples of betel leaf conditions used in this study: (a) healthy green leaf, (b) green leaf with anthracnose, (c) green leaf with bacterial leaf spot, (d) healthy red leaf.

For model evaluation, a 5-fold cross-validation strategy was adopted, as shown in Figure 2. The dataset was partitioned into five subsets, with each subset used as a validation set once, while the remaining served for training. This approach provided a robust and unbiased estimate of the model's generalization capability. The performance was evaluated using key metrics, including accuracy, precision, recall, and F1-score. After training, the model was tested on previously unseen data, and its predictions were compared against ground truth labels to compute the final evaluation metrics, as illustrated in the "Predicted Class" and "Evaluation" blocks in Figure 2.

C. *EfficientNet Architecture*

EfficientNet is a state-of-the-art CNN architecture designed to optimize both the accuracy and computational efficiency, making it suitable for resource-constrained platforms, such as mobile devices [17]. The architecture builds upon the Mobile Inverted Bottleneck Convolution (MBConv) module, initially introduced in MobileNetV2 [19]. The core of EfficientNet lies in its compound scaling method, which uniformly scales the network's depth, width, and resolution based on a fixed set of scaling coefficients. This allows the model to achieve better performance compared to conventional scaling strategies that only adjust a single dimension [17]. The MBConv block, as portrayed in Figure 3, is a key building block of the EfficientNet architecture. It consists of two 1×1 pointwise convolutions, a depth-wise convolution, and a Squeeze-and-Excitation (SE) attention module. These components collectively reduce the computational cost while preserving the model's ability to extract critical features. The SE module further enhances the performance by allowing the network to recalibrate the channel-wise feature responses adaptively [19, 20].

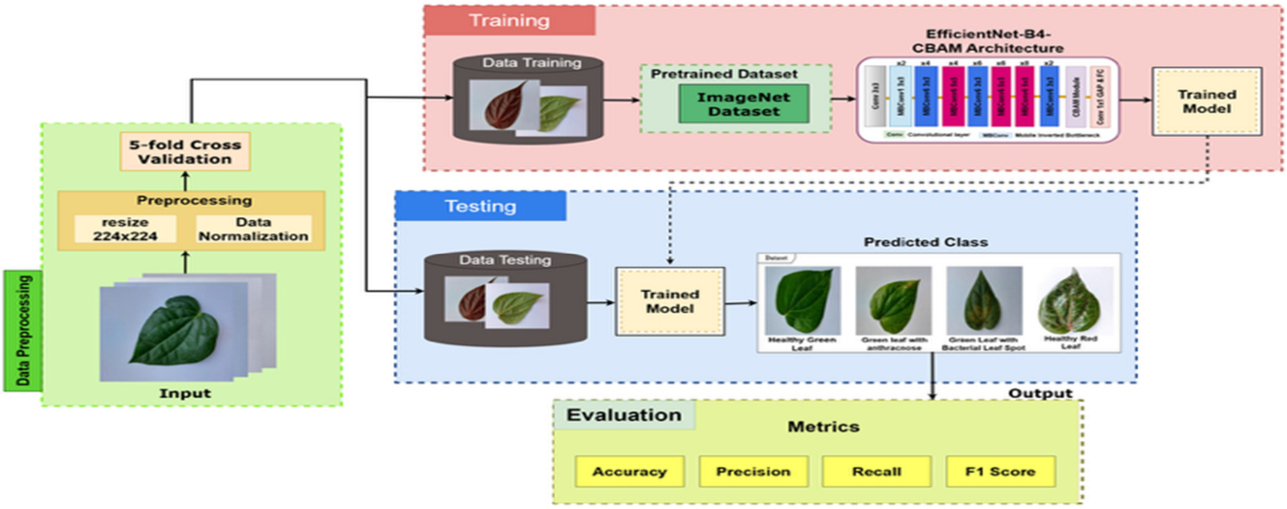


Fig. 2. Research workflow illustrating the steps of preprocessing, training, validation, and evaluation.

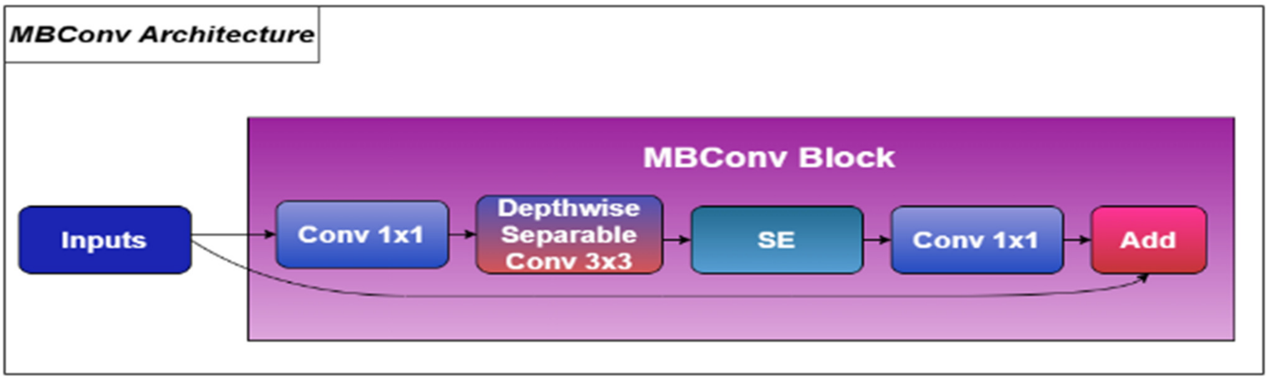


Fig. 3. Structure of the MBConv block used in EfficientNet.

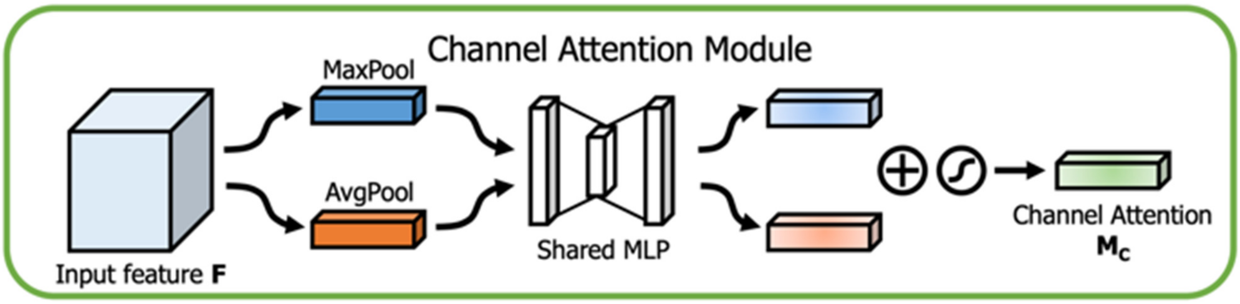


Fig. 4. Channel Attention Module (CAM) structure.

D. Convolutional Block Attention Module

The CBAM is a lightweight and effective attention mechanism designed to refine the feature representations in CNNs by focusing on the most informative parts of the input data [15]. CBAM sequentially applies attention in both the channel and spatial dimensions to enhance the discriminative power of the features. The CBAM module comprises two sub-modules: the CAM and the Spatial Attention Module (SAM).

1) Channel Attention Module

The CAM emphasizes the important features along the channel axis using global average pooling and max pooling, followed by a shared Multi-Layer Perceptron (MLP). The two pooling outputs are combined and passed through a sigmoid activation function to produce the channel attention map M_c, which is then multiplied element-wise with the input feature map F, as described by:

$$M_c F =$$

$$\sigma\left(MLP\left(MaxPool(F)\right)\right) + \sigma\left(MLP\left(AvgPool(F)\right)\right) \quad (1)$$

$$F' = M_c(F) \times F \quad (2)$$

2) Spatial Attention Module

The SAM emphasizes the key spatial regions of the feature map by applying average pooling and max pooling along the channel axis. These are concatenated and passed through a convolutional layer followed by a sigmoid activation to generate the spatial attention map M_s. The refined feature map F'' is computed as:

$$M_s(F') = \sigma(f_{7x7}[MaxPool(F'); AvgPool(F')]) \quad (3)$$

$$F'' = M_s(F') \times F' \quad (4)$$

The combination of CAM and SAM in CBAM enhances the model's focus on disease-specific regions in leaf images. While this dual-attention mechanism improves the classification performance, it also increases the computational burden and memory usage, which may limit its deployment on resource-constrained devices [16].

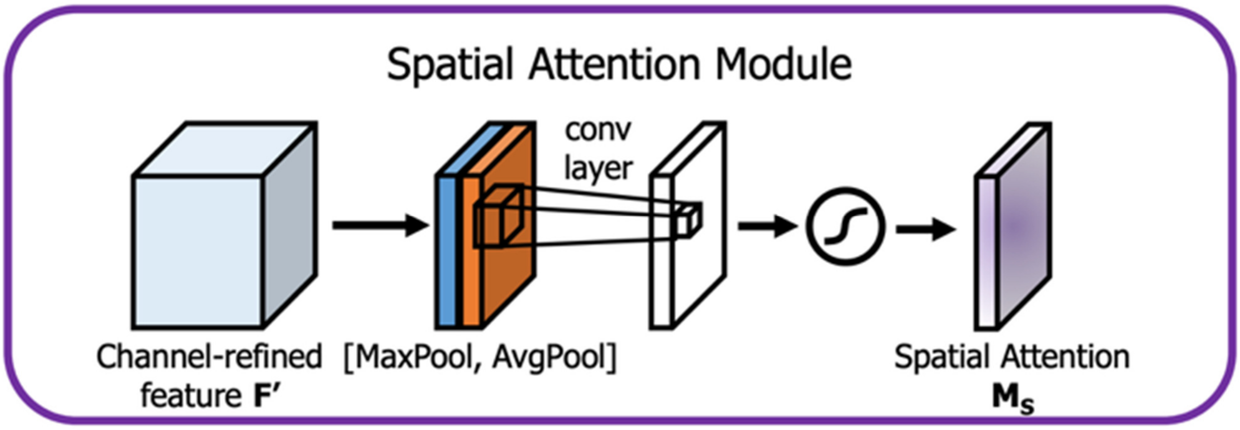


Fig. 5. SAM structure.

E. K-Fold Cross Validation

K-Fold Cross Validation is a robust evaluation technique used to assess the generalization capability of an ML model. In this method, the dataset is randomly partitioned into *K* equally sized subsets or "folds." During the validation process, each fold is systematically used as the validation set, while the remaining *K–1* folds are utilized for training. This procedure is repeated *K* times, with each fold serving as the validation set exactly once. This technique ensures that every observation in the dataset is utilized for both training and validation, thus providing a more comprehensive and unbiased evaluation of the model performance. Moreover, it helps to minimize the overfitting and yields reliable statistical estimates of the performance metrics.

In this study, a 5-fold cross-validation approach ($K = 5$) was implemented, dividing the entire dataset into five equal parts. Each fold underwent the training-validation cycle as described, and the final evaluation metrics, namely accuracy, precision, recall, and F1-score, were computed as the average of the results obtained across all folds.

F. Performance Evaluation

The performance evaluation of the proposed methodology is crucial to validate its effectiveness and reliability. This section outlines the metrics and strategies used to assess the model, with a focus on the applied methods and tools.

To quantify the model performance, three primary metrics were employed; accuracy, precision, and recall. These metrics collectively provide a comprehensive overview of the model's predictive quality, each highlighting different evaluation perspectives [21-23].

1) Accuracy

Accuracy reflects the overall proportion of the correctly classified samples and is given by:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (5)$$

where TP = True Positives, TN = True Negatives, FP= False Positives, and FN = False Negatives.

Although accuracy gives a general performance measure, it may be misleading in the context of imbalanced datasets.

2) Precision

Precision assesses the proportion of the TP predictions among all positive predictions:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

High precision indicates fewer FP, which is particularly important in scenarios where false alarms must be minimized.

3) Recall

Also known as sensitivity, recall evaluates the model's ability to identify the actual positive instances:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (7)$$

High recall is critical in applications where missing positive cases has serious consequences.

4) F1-score

The F1-score represents the harmonic mean of precision and recall, balancing both metrics:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

This metric is especially relevant when dealing with class imbalance, as it considers both the FP and FN.

5) Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)

The ROC curve illustrates the trade-off between the true positive rate (Recall) and the False Positive Rate (FPR), with AUC summarizing the model's discriminatory ability, where FPR is computed from:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (9)$$

A higher AUC indicates a better ability to distinguish between classes across various threshold settings. To evaluate the proposed approach, EfficientNet-B4 and its CBAM-enhanced counterpart were trained and tested using a curated dataset of 4,000 annotated betel leaf images. The dataset was preprocessed and split using 5-fold cross-validation to ensure the model reliability. The CBAM module enhanced the feature extraction by focusing on disease-specific spatial and channel patterns. The performance metrics including accuracy, precision, recall, F1-score, and AUC demonstrated the effectiveness of the proposed approach in accurately identifying the betel leaf diseases. These results provide a strong foundation for further comparison and real-world implementation.

III. RESULTS AND DISCUSSION

The experimental results indicate a notable improvement in the detection of the betel leaf diseases using the EfficientNet-B4 model integrated with the CBAM. The performance metrics, summarized in Tables I and II, demonstrate a consistent enhancement across all validation folds when CBAM is incorporated.

A. Performance Metrics

Table I presents the classification results for the base EfficientNet-B4 model. The model achieves an average accuracy of approximately 90%, with the highest accuracy (93.12%) observed in Fold 1. Other performance metrics, such as precision, recall, and F1-score also remain consistent across folds, ranging around 90%.

TABLE I. TESTING RESULTS OF EFFICIENTNET-B4

Fold	Accuracy (%)	Loss (%)	Precision (%)	Recall (%)	F1-score (%)
1	93.12	0.258	93.72	93.12	93.21
2	88.37	0.514	89.59	88.37	88.60
3	90.25	0.318	90.82	90.25	90.38
4	89.12	0.511	89.78	89.12	89.26
5	92.37	0.240	92.88	92.37	92.51

Table II displays the results after the integration of CBAM into the base model. The CBAM-enhanced model achieves a significantly better performance, with a peak accuracy of 97.25% in Fold 1 and an average accuracy exceeding 95%. The loss values are also substantially reduced, with a minimum loss of 0.101 in Fold 1, indicating improved generalization and training stability.

TABLE II. TEST RESULTS OF EFFICIENTNET-B4 WITH CBAM

Fold	Accuracy (%)	Loss (%)	Precision (%)	Recall (%)	F1-score (%)
1	97.25	0.101	97.34	97.25	95.11
2	93.00	0.294	93.32	93.00	93.86
3	95.75	0.174	95.84	95.75	91.03
4	96.00	0.230	96.17	96.00	95.42
5	95.00	0.372	95.60	95.00	95.37

The implementation of CBAM significantly enhances both the recall and F1-score, with an average close to 96%. For instance, in Fold 4, the model achieved a recall of 96.00% and an F1-score of 95.42%, indicating a reliable classification performance, especially in identifying the disease-affected samples. The base EfficientNet-B4 model, on the other hand, demonstrated lower recall and F1-scores, with averages around 90%, confirming its limited capability in handling complex visual patterns.

B. Visualization and Confusion Matrix Analysis

The baseline EfficientNet-B4 model achieved a notable performance across all five folds, with the peak accuracy reaching 93.12% in Fold 1 and an overall average accuracy of approximately 90%. Nevertheless, the integration of the CBAM significantly improved the performance, as depicted in Table II, with the enhanced model attaining a maximum accuracy of 97.25% in Fold 1 and an average accuracy surpassing 95%. This enhancement demonstrates the effectiveness of CBAM in refining the feature representations by focusing on spatial and channel-wise informative regions [19].

In addition, the CBAM-enhanced model exhibited a markedly reduced loss percentage, with the lowest recorded loss being 0.101% in Fold 1. This result implies enhanced model generalization and stability during training. The superiority of the proposed model is also evident in the recall and F1-score metrics, with both averaging around 96%, indicating a robust ability to correctly classify the disease-affected leaves while maintaining high precision. For example, in Fold 4, the model achieved a recall of 96.00% and an F1-score of 95.42%, indicating a balanced trade-off between the TP identification and FN minimization.

Conversely, the baseline EfficientNet-B4 model demonstrated variability in recall and F1-score, with averages near 90%, highlighting its relative limitations in complex classification scenarios.

The confusion matrix in Figure 6 provides a detailed view of the CBAM-enhanced model's classification capability across four target classes: Green Health, Green Anthracnose, Green Bacterial Leaf-spot, and Red Health.

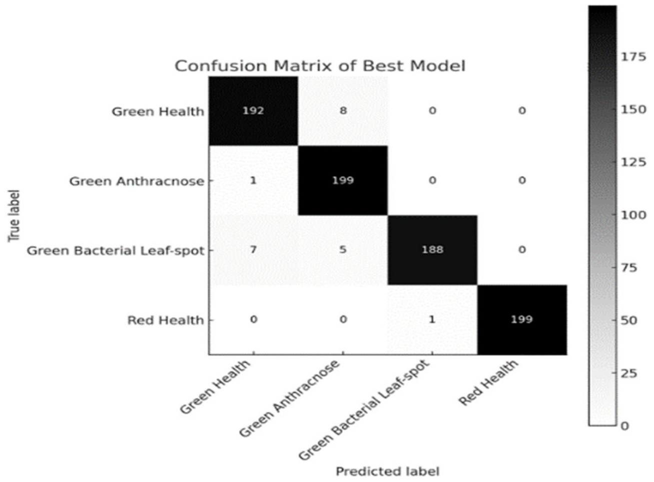


Fig. 6. Confusion matrix of the best-performing model.

The model demonstrated near-perfect performance, particularly in the Red Health category, where it achieved 100% precision and 0% FN. However, minor misclassifications were observed in the Green Bacterial Leaf-spot class, with seven samples being erroneously classified as FN. This misclassification likely results from overlapping visual features between the bacterial-infected and healthy green leaves under varying imaging conditions. Future enhancements could include extended dataset augmentation or advanced attention mechanisms to improve the class separability.

As shown in Table III, the Green Anthracnose class exhibited a recall of 0.995 and an F1-score of 0.966, though the precision value of 0.939 indicates occasional FP, possibly due to the visual similarity with healthy leaves. The Green Health class maintained a balanced performance across all metrics, each recorded at 0.960, with most errors occurring in the classification overlap with Green Anthracnose. Meanwhile, the Green Bacterial Leaf-spot class had the lowest recall (0.940), reflecting its subtle symptom patterns and highlighting the challenge of differentiating the bacterial infections from the natural leaf textures. The ROC curves in Figure 7 further validate the robustness of the CBAM-enhanced model. All classes yielded AUC values greater than 0.96, with the Red Health class achieving an AUC of 0.998. This outcome confirms the model's strong discriminative ability across varying classification thresholds and underscores its practical applicability in real-world agricultural scenarios, where accurate disease detection is critical for effective crop management.

TABLE III. TEST RESULTS OF THE EFFICIENTNET-B4-CBAM MODEL

Class	Accuracy	Precision	Recall	F1-score	ROC AUC
Green Health	0.960	0.960	0.960	0.960	0.973
Green Anthracnose	0.995	0.939	0.995	0.966	0.987
Green Bacterial Leaf-Spot	0.940	0.995	0.940	0.967	0.969
Red Health	0.995	1.000	0.995	0.997	0.998
Green Health	0.960	0.960	0.960	0.960	0.973

The ROC curves in Figure 7 further validate the robustness of the CBAM-enhanced model. All classes yielded AUC values greater than 0.96, with the Red Health class achieving an AUC of 0.998.

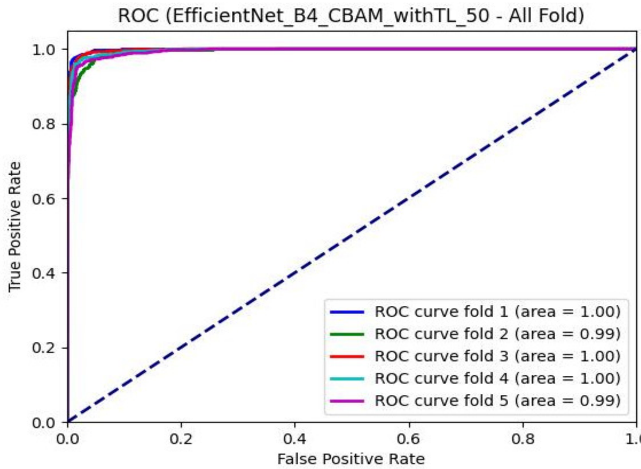


Fig. 7. ROC curve of the best model.

This outcome confirms the model's strong discriminative ability across varying classification thresholds and underscores its practical applicability in real-world agricultural scenarios, where accurate disease detection is critical for effective crop management.

C. Model Comparative Analysis

While the primary focus of this study is on the EfficientNet-B4 and its CBAM-enhanced variant for betel leaf disease detection, the study also trained and evaluated other state-of-the-art deep learning models—including XceptionResNet52 [25], InceptionResNetV2 [26], and DenseNet201 [27]—using the same localized dataset. The results provide a broader context, highlighting the advancements introduced by the EfficientNet-B4-CBAM model in terms of the classification accuracy and feature extraction. By training these models on the created dataset, the study ensures a fair comparison and reinforces the effectiveness of integrating attention mechanisms for specialized agricultural tasks. Table IV presents a comparative overview of the classification accuracy of the aforementioned models. Notably, the CBAM-enhanced EfficientNet-B4 consistently outperforms other architectures, achieving an accuracy of 97.25%. This superior performance underscores the importance of incorporating attention mechanisms to refine the feature extraction, particularly in challenging image classification tasks involving overlapping and subtle disease features.

The graphical representation in Figure 8 further highlights these results, showcasing the significant leap in the performance brought by the CBAM integration. XceptionResNet52, despite its depth, demonstrates limitations in extracting complex visual patterns characteristic of leaf disease symptoms. InceptionResNetV2 and DenseNet201 offer moderate improvements, yet still fall short compared to the EfficientNet-based models.

TABLE IV. ACCURACY COMPARISON OF DEEP LEARNING MODELS

Model	Accuracy
XceptionResNet52	76.6
InceptionResnet50V2	86.0
DenseNet201	82.5
EfficientNet-B4	93.12
EfficientNet-B4-CBAM	97.25

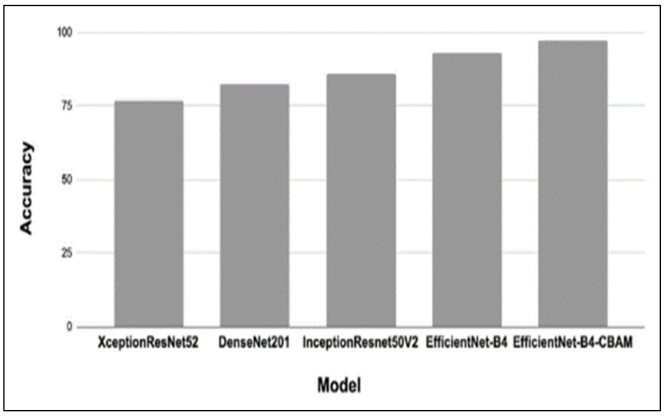


Fig. 8. Comparison of accuracy for various deep learning models.

EfficientNet-B4, with its compound scaling strategy, already exhibits high accuracy due to its ability to balance the model depth, width, and input resolution. The incorporation of CBAM into this architecture enhances its discriminative capability further by focusing on disease-relevant spatial and channel information. This leads to reduced misclassification rates and a better handling of visually similar disease categories, such as anthracnose and bacterial leaf spot. The practical value of EfficientNet-B4-CBAM lies in its ability to facilitate a precise and timely identification of the plant diseases, which is crucial for sustainable agriculture. Its high classification accuracy, combined with computational efficiency, makes it suitable for real-time deployment on portable devices, offering a scalable solution for modern digital farming practices.

D. Key Observations and Practical Implications

The integration of CBAM into EfficientNet-B4 effectively addresses the core challenges in image-based disease classification, including inconsistent lighting, background noise, and inter-class similarities. Moreover, data augmentation and preprocessing contributed significantly to the model's ability to generalize across variations in angle, scale, and illumination, which are critical factors for deployment in field conditions. These characteristics are essential for real-world applicability, especially in smallholder farming contexts. The implications of the model's performance are multifaceted. The accurate early detection minimizes the crop loss and reduces the reliance on chemical treatments, directly benefiting farmers' productivity and economic outcomes [22]. The CBAM-enhanced EfficientNet-B4, with its high accuracy and low inference time, offers a feasible solution for integration into mobile or edge-based platforms, facilitating the on-site diagnostics without requiring high-end computational infrastructure.

IV. CONCLUSION

This study presents an effective deep learning-based approach for the classification of betel leaf diseases using a hybrid architecture that integrates EfficientNet-B4 with the Convolutional Block Attention Module (CBAM). Leveraging a localized dataset comprising 4,000 high-resolution images from Indonesian betel leaf farms, the proposed model achieved a classification accuracy exceeding 95%, significantly outperforming the base EfficientNet-B4 and other state-of-the-art models, like DenseNet201 and InceptionResNetV2.

By embedding CBAM into EfficientNet-B4, the model demonstrated enhanced capability in spatial and channel feature refinement, resulting in improved precision, recall, F1-scores, and reduced classification errors. The evaluation through 5-fold cross-validation, confusion matrices, and Receiver Operating Characteristic (ROC)-Area Under the Curve (AUC) analysis further confirmed the robustness and generalizability of the model. It showed strong performance across all disease categories, particularly excelling in classifying Red Health and Green Anthracnose leaves.

The study emphasizes the importance of using localized datasets that reflect region-specific disease characteristics, which substantially contributes to the model accuracy and relevance. Additionally, it addresses key challenges, such as feature extraction complexity and variability in environmental imaging conditions. The findings also suggest practical implications for deploying the model in real-time scenarios, especially in low-power mobile or edge devices, potentially empowering farmers with on-site disease diagnostics.

In conclusion, the EfficientNet-B4-CBAM model provides a scalable, accurate, and computationally efficient solution for betel leaf disease detection. Its deployment can revolutionize the disease management in agriculture by enabling early detection, minimizing crop loss, and improving the economic resilience of betel leaf farmers in Indonesia. The success of this model sets a precedent for future AI-driven agricultural systems, underscoring the transformative potential of combining attention mechanisms with optimized deep learning architectures in real-world farming applications.

ACKNOWLEDGEMENT

Gratitude is extended to the Department of Agriculture in Bangkalan for validating the betel leaf dataset for the experimental method in this research. Special thanks are also conveyed to the Research and Community Service Institute (LPPM) of Trunojoyo Madura University for providing funding for the National Collaborative Research under contract number 5895/UN46.4.1/PT.01.03/2023.

DATASET AVAILABILITY

The entire dataset is publicly available at: https://www.kaggle.com/datasets/achmadbauravindah/betel-leaf-disease-classification/data

REFERENCES

[1] A. I. Pratama, H. Fitriawan, M. L. Aulia, D. A. Prastyo, and R. Ridwan, "Pemberdayaan Masyarakat Dalam Betel Leaf Empowerment Hub Program PT Pertamina Patra Niaga Shafthi," *Focus : Jurnal Pekerjaan*

Sosial, vol. 7, no. 1, pp. 80–89, Aug. 2024, https://doi.org/10.24198/focus.v7i1.54124.

[2] Raidi, "Nyirih, Indonesia Betel Chewing Tradition Deep in Symbolic and Ritual Meaning," *Indonesia Sentinel*, Apr. 2025, https://indonesiasentinel.com/nyirih-indonesia-betel-chewing-tradition-deep-in-symbolic-and-ritual-meaning/.

[3] E. J. Mahfuza, Md. S. Ahamed, and Md. F. Hassan, "Impact of Betel Leaf Farming on Livelihood and Income Generation in Some Selected Areas of the Rajshahi District of Bangladesh," *South Asian Journal of Social Studies and Economics*, vol. 20, no. 3, pp. 250–259, Oct. 2023, https://doi.org/10.9734/sajsse/2023/v20i3728.

[4] A. Samanta, "Present Socio Economic Condition Of Betel Leaf Farmer: A Case Study Of Alipurduar District In West Bengal," *International Journal of Research and Analytical Reviews*, vol. 9, no. 1, pp. 526–535, Ja 2022.

[5] I. M. Y. Pramana, "Chewing Betel in Bali: An Ancient Tradition Faces Modern Times," *Bali Tourism Journal*, vol. 7, no. 2, pp. 31–34, Jun. 2023, https://doi.org/10.36675/btj.v7i2.92.

[6] A. Banerjj and R. Ghosh, "Fungal diseases of Betel vine (Piper betle L.) in India and their management: An overview," *Journal of Mycopathological research*, vol. 62, no. 1, pp. 21–29, Mar. 2024, https://doi.org/10.57023/jmycr.62.1.2024.021.

[7] S. L. Prasanna, *Diseases of Betelvine*, Hyderabad, Telangana, India: Acharya N. G. Ranga Agricultural University, 2022.

[8] N. M. D. Nayaka *et al.*, "Piper betle (L): Recent Review of Antibacterial and Antifungal Properties, Safety Profiles, and Commercial Applications," *Molecules*, vol. 26, no. 8, Apr. 2021, Art. no. 2321, https://doi.org/10.3390/molecules26082321.

[9] S. Kusuma and K. R. Jothi, "Early betel leaf disease detection using vision transformer and deep learning algorithms," *International Journal of Information Technology*, vol. 16, no. 1, pp. 169–180, Jan. 2024, https://doi.org/10.1007/s41870-023-01647-3.

[10] B. Prashanthi, A. V. Praveen Krishna, and Ch. Mallikarjuna Rao, "A Comparative Study of Fine-Tuning Deep Learning Models for Leaf Disease Identification and Classification," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19661–19669, Feb. 2025, https://doi.org/10.48084/etasr.9017.

[11] A. S. Devi, S. Albert, and A. Raj, "Optimized Betel Leaf Disease Detection Using Improved CNN Model for Precision Agriculture," *Journal of Electrical Systems*, vol. 20, no. 10, pp. 20–30, 2024.

[12] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, Sep. 2016, https://doi.org/10.3389/fpls.2016.01419.

[13] J. Kong, H. Wang, C. Yang, X. Jin, M. Zuo, and X. Zhang, "A Spatial Feature-Enhanced Attention Neural Network with High-Order Pooling Representation for Application in Pest and Disease Recognition," *Agriculture*, vol. 12, no. 4, Mar. 2022, Art. no. 500, https://doi.org/10.3390/agriculture12040500.

[14] V. Bali and D. Syal, *AI for Disease Detection from Medical Images*. Punjab, India: GreyB, 2020.

[15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2018, pp. 3–19.

[16] N. Sengodan, "Breast Cancer Histopathology Classification using CBAM-EfficientNetV2 with Transfer Learning." arXiv, May 13, 2025, https://doi.org/10.48550/arXiv.2410.22392.

[17] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36 th International Conference on Machine Learning*, Long Beach, California, 2019, vol. 97, pp. 6105–6114.

[18] *EfficientNet-B4*, Qualcomm AI Hub, 2025, [Online]. Available: https://aihub.qualcomm.com/models/efficientnet_b4.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, https://doi.org/10.1109/cvpr.2018.00745.

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 4510–4520, https://doi.org/10.1109/cvpr.2018.00474.

[21] H. Dalianis, *Clinical Text Mining: Secondary Use of Electronic Patient Records*, 1st ed. Cham: Springer International Publishing, 2018.

[22] S. S. Putro, N. Ansori, M. Fuad, E. M. S. Rochman, Y. P. Asmara, and A. Rachmad, "Corn Leaf Disease Classification Using Convolutional Neural Network Based on MobileNetV2 with RMSProp Optimization," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 2, pp. 465–474, Feb. 2025, https://doi.org/10.18280/mmep.120211.

[23] O. Cordón, P. Kazienko, and B. Trawiński, "Special Issue on Hybrid and Ensemble Methods in Machine Learning," *New Generation Computing*, vol. 29, no. 3, pp. 241–244, Jul. 2011, https://doi.org/10.1007/s00354-011-0300-3.

[24] Y. Yang, Z. Cai, S. Qiu, and P. Xu, "Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image," *PLOS ONE*, vol. 19, no. 3, Mar. 2024, Art. no. e0299265, https://doi.org/10.1371/journal.pone.0299265.

[25] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 1800–1807, https://doi.org/10.1109/cvpr.2017.195.

[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, https://doi.org/10.1609/aaai.v31i1.11231.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, https://doi.org/10.1109/cvpr.2017.243.