Banana Leaf Disease Identification Using SqueenzeNet Architecture with Convolutional Block Attention Module

Daniel Wijaya[1], Derry Alamsyah [2]

[1,2] Multi Data Palembang University, Palembang, South Sumatra, Indonesia, 30113
e-mail: *danielwijaya_2226250060@mhs.mdp.ac.id[1], derry@mdp.ac.id [2]*
*Correspondence: danielwijaya_2226250060@mhs.mdp.ac.id[1]

Abstract: *Banana leaf diseases significantly reduce crop productivity and quality, while conventional visual inspection methods are often subjective, time-consuming, and inefficient for large-scale plantations. This study proposes an automated banana leaf disease identification approach using a lightweight Convolutional Neural Network (CNN) based on the SqueezeNet architecture integrated with the Convolutional Block Attention Module (CBAM). The dataset consists of four classes Cordana, Healthy, Pestalotiopsis, and Sigatoka with image augmentation applied to increase data variability. Several experimental scenarios were conducted to evaluate the impact of data augmentation and CBAM integration on model performance. The models were evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results show that SqueezeNet combined with CBAM achieved superior performance compared to the baseline SqueezeNet model, particularly in non-augmented conditions, with an accuracy of 93.75% while maintaining a relatively small number of parameters. Although data augmentation alone led to performance degradation, the inclusion of CBAM mitigated this effect by enhancing spatial and channel-wise feature representation. These findings indicate that the proposed SqueezeNet–CBAM model offers an effective and computationally efficient solution for banana leaf disease identification, with strong potential for real-world agricultural applications.*

Keywords: *banana leaf disease, SqueezeNet, CBAM, convolutional neural network, image classification.*

1. Introduction

Indonesia is recognized as one of the world's megabiodiversity countries with abundant biological resources. One horticultural commodity with high economic value is banana [1]. Bananas are widely cultivated in tropical and subtropical regions and represent the fruit with the highest production level in Indonesia [2]. According to data from the Indonesian Central Bureau of Statistics in 2024, national banana production reached 9.26 million tons [3], highlighting its crucial role in supporting economic growth and national food security [4]. However, banana productivity often declines due to attacks from leaf diseases such as Cordana leaf spot, Pestalotiopsis leaf drop, Sigatoka, and wilt disease [5]. These pathogenic infections can significantly reduce both the quality and quantity of banana yields [6]. Meanwhile, traditional diagnostic methods still rely on visual inspection, which is subjective, requires expert knowledge, and is inefficient for large-scale plantations. This condition emphasizes the need for faster, more accurate, and automated diagnostic methods.

Along with the rapid development of computational technology, digital image processing and deep learning-based approaches have demonstrated great potential in plant disease diagnosis [7]. In particular, Convolutional Neural Networks (CNNs) have become the dominant method due to their ability to automatically extract features through convolutional layers without requiring manual feature engineering [8]. Numerous previous studies have applied CNNs for banana leaf disease classification using various architectures. The VGG19 architecture achieved an accuracy of 82% [9], ResNet-152 reached 61.31% accuracy [10], InceptionV3 obtained 81.05%, while DenseNet demonstrated better performance with an accuracy of 84.73% [11]. However, these models generally have a large number of parameters, leading to high computational costs and making them less suitable for deployment on devices with limited memory and processing capabilities.

To address these limitations, recent studies have explored lightweight CNN architectures such as SqueezeNet, which employs Fire Modules to significantly reduce the number of parameters without sacrificing accuracy. Recent research reported that SqueezeNet achieved an accuracy of 96.25% [12]. Nevertheless, despite its high efficiency, the performance of SqueezeNet can still be further improved by incorporating attention mechanisms that enable the model to focus on important features, thereby enhancing the accuracy and stability of image identification. In order for lightweight CNN models to effectively emphasize disease-indicative regions, an attention mechanism is required. The Convolutional Block Attention Module (CBAM) is one effective approach, as it is capable of highlighting significant features at both the channel attention and spatial attention levels[13], The effectiveness of CBAM in improving CNN performance has been demonstrated in several peer-reviewed studies on image classification tasks, where the integration of CBAM significantly enhanced feature representation and classification accuracy [13]. Similar performance improvements were also reported in academic theses [10]. This substantial improvement indicates that CBAM can strengthen feature representation and increase the model's sensitivity to relevant symptom regions. Therefore, this study proposes the integration of the CBAM module to enhance the accuracy of the employed architecture.

Based on the above discussion, this research aims to address the primary question of whether integrating the SqueezeNet architecture with the CBAM attention mechanism can improve accuracy in banana leaf disease identification. In line with this research question, the objective of this study is to develop and evaluate a SqueezeNet–CBAM-based CNN model, using experimental evaluation tables as the basis for empirically assessing its performance. This study is expected to deliver high identification performance while maintaining a lightweight model size, thereby offering strong potential for implementation as an automated diagnostic system in the agricultural sector.

2. Literature Review
Based on the reviewed studies, it can be observed that most CNN-based approaches for banana leaf disease classification prioritize high accuracy by employing deep and computationally expensive architectures such as VGG[9], ResNet[10], Inception, and DenseNet[11]. Although these models demonstrate competitive performance, their large number of parameters limits their suitability for deployment on resource-constrained environments. Lightweight architectures such as SqueezeNet offer a promising alternative due to their compact model size[14]. however, previous studies mainly focused on accuracy reporting without addressing limitations in feature discrimination, particularly under complex visual variations.

Moreover, while attention mechanisms such as CBAM have been shown to significantly enhance feature representation in deeper CNN architectures, their integration with lightweight networks like SqueezeNet remains insufficiently explored, especially in combination with data augmentation. Existing studies rarely provide a systematic comparison of how augmentation and attention jointly affect model stability and performance. Therefore, a clear research gap exists in evaluating whether the integration of CBAM can compensate for the representational limitations of lightweight architectures while preserving computational efficiency. This study addresses this gap by systematically analyzing the performance of SqueezeNet with and without CBAM under both augmented and non-augmented conditions.

3. Methods
In this study, the methodological design was established in advance to ensure that the research process was more systematic and well focused. The stages of the research methodology are illustrated in Figure 1.
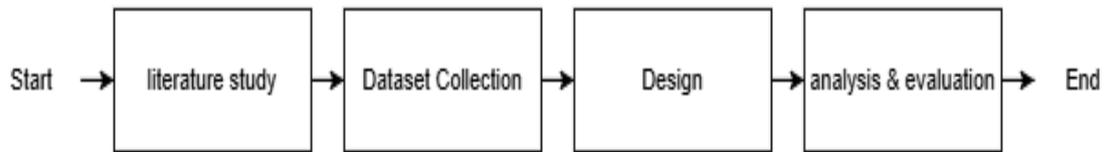
Figure 1. Methodology Flowchart

A. Dataset Collection

In this study, a dataset consisting of banana leaf disease images with four class labels—Cordana, Healthy, Pestalotiopsis, and Sigatoka—was utilized. The dataset comprises a total of 937 original images obtained from Mendeley Data, namely the *Banana Leaf Spot Diseases (BananaLSD) Dataset for Classification of Banana Leaf Diseases Using Machine Learning*. All images were captured using smartphone cameras in banana fields at Bangabandhu Sheikh Mujibur Rahman Agricultural University and surrounding areas in Bangladesh during June 2021. Three different smartphone cameras were used for image acquisition. The images were manually labeled by a plant pathologist to ensure label accuracy. The distribution of the original dataset consists of 162 images of Cordana, 129 images of Healthy leaves, 172 images of Pestalotiopsis, and 472 images of Sigatoka, resulting in an imbalanced class distribution. Data augmentation techniques were subsequently applied to increase the number of training samples and mitigate class imbalance. Figure 2 presents sample images from the banana leaf disease dataset representing the four classes.
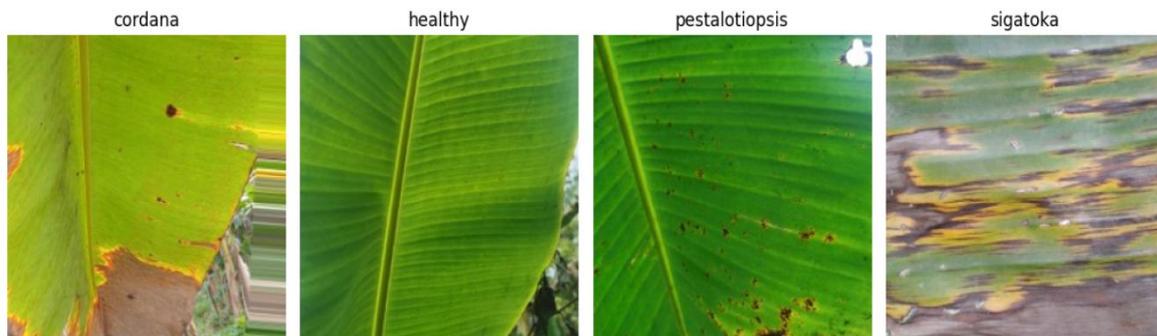

Figure 2. Dataset

B. Design

The system design phase aims to construct and organize the workflow for the development of the banana leaf disease identification model. The overall model development workflow is illustrated in Figure 3.

After the data collection phase, the next stage was data preprocessing, which included data augmentation and image normalization to improve the model's generalization capability and ensure data readiness prior to the training process [15]. As illustrated in Figure 3, data augmentation is an integral part of the preprocessing pipeline, although the figure only presents the overall preprocessing flow.
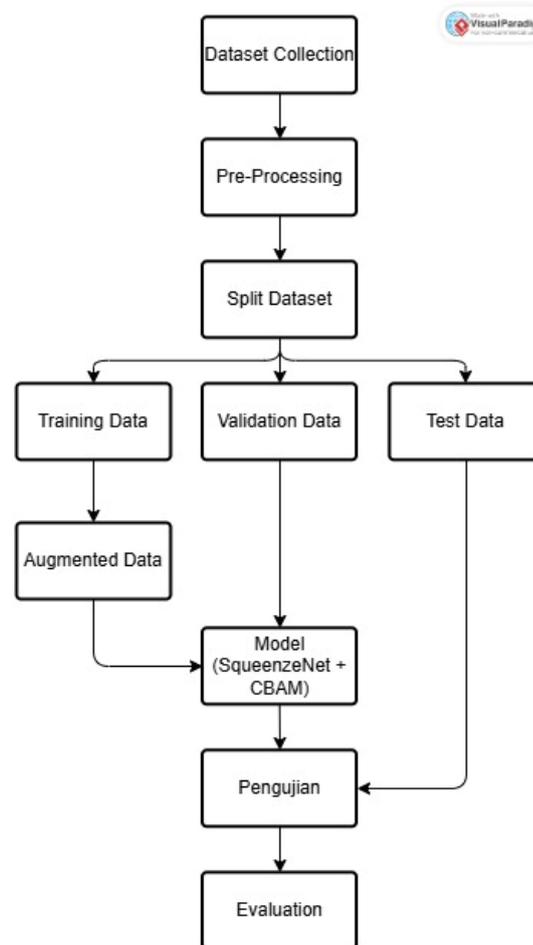
Figure 2. Model design stages

In the initial preprocessing stage, data augmentation was applied to increase the diversity and variability of the training images, enabling the model to be more robust to variations in object shape, position, and orientation. The augmentation techniques employed horizontal flipping; random cropping with a crop ratio of 90% followed by resizing to the original image dimensions; linear contrast adjustment using a scaling factor (α) in the range of 0.7 to 1.3; horizontal shearing with shear factors between −0.2 and 0.2; translation in both horizontal and vertical directions within ±10% of the image dimensions; and rotation with random angles ranging from −30° to +30°.

For each image in the training set, one to three augmentation operations were randomly selected and applied sequentially. Data augmentation was applied exclusively to the training set after the dataset splitting stage as a form of oversampling to balance the number of samples across classes. In this study, the validation and test sets were first constructed by randomly selecting 20 original images per class for each subset, resulting in 80 images for validation and 80 images for testing across four classes.

The remaining images were assigned to the training set and subsequently augmented until the number of training samples reached 500 images per class, including both original and augmented images. Consequently, the final dataset consisted of 2,000 training images, 80 validation images, and 80 test images, yielding a total of 2,160 images across all classes. Subsequently, all images were normalized using the Min–Max normalization method by scaling pixel values to a range between 0 and 1. This normalization step allows the model to process the data more efficiently, stabilizes gradient updates, and prevents the dominance of certain pixel values that could negatively affect the convergence process [16].
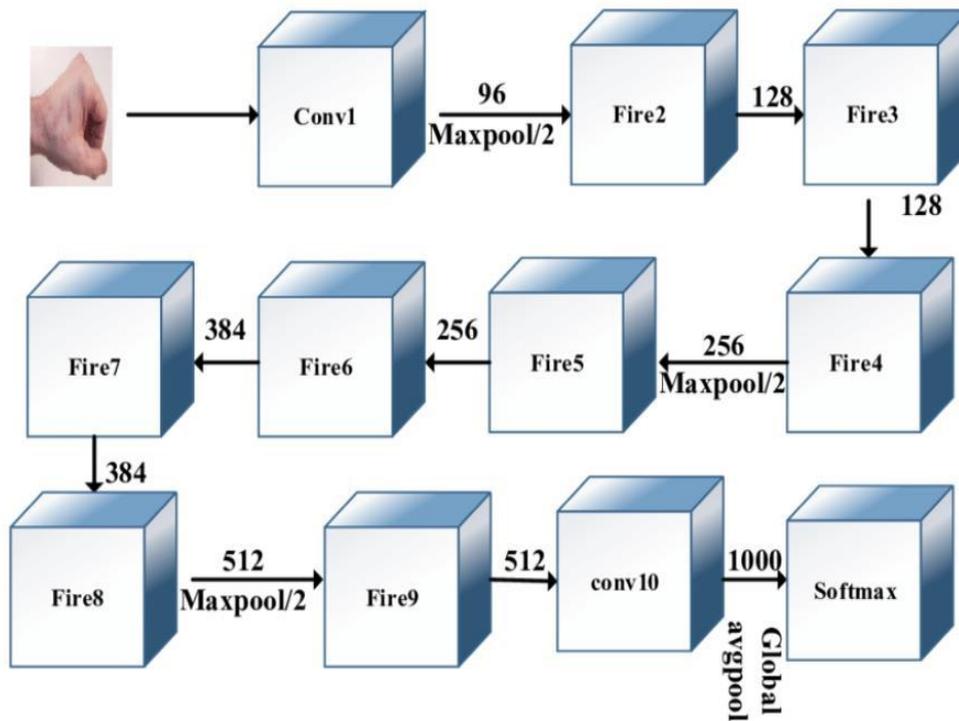
Figure 4. *architecture SqueenzeNet*

After the dataset splitting stage, the process continued with the construction of the baseline architecture using SqueezeNet. SqueezeNet was selected because it is a lightweight CNN architecture capable of achieving high accuracy with a significantly smaller number of parameters compared to conventional models such as VGG or ResNet. The core component of SqueezeNet is the Fire Module, which consists of a squeeze layer using 1×1 convolution and an expand layer that combines 1×1 and 3×3 convolutions [14]. This design enables efficient feature extraction while minimizing parameter growth. As illustrated in Figure 4, the SqueezeNet architecture arranges multiple Fire Modules to reduce model complexity while maintaining reliable classification performance.

To enhance the model's ability to focus on salient features, this study integrates the Convolutional Block Attention Module (CBAM). CBAM operates through a two-stage attention mechanism, namely channel attention and spatial attention. Channel attention aims to identify which feature channels are most relevant to the objects within the image, while spatial attention emphasizes the important regions or locations in leaf images that exhibit disease symptoms. By combining these two attention mechanisms, CBAM is expected to improve the quality of feature representation without introducing a significant computational overhead.

The final stage involves integrating the SqueezeNet architecture with CBAM to form the SqueezeNet–CBAM model. In this design, the CBAM modules are inserted after selected Fire Modules so that attention can be applied during the most critical stages of feature extraction. This integration is expected to improve the model's accuracy in identifying banana leaf disease symptoms while preserving the main advantages of SqueezeNet as a lightweight and computationally efficient architecture. Figure 5 illustrates the combined SqueezeNet–CBAM architecture, depicting the data flow from image input, feature extraction, attention mechanisms, and the final classification stage.
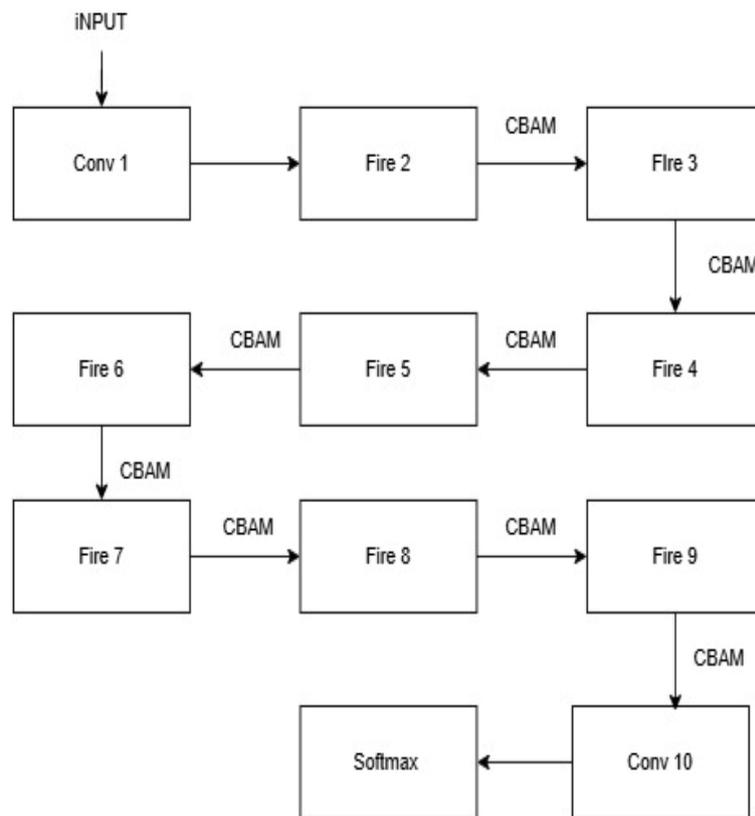
Figure 5. Model *architecture SqueenzeNet with CBAM*

C. Analysis and evaluation

The analysis and evaluation stage was conducted to determine the performance assessment of the proposed banana leaf disease identification model. Model evaluation was carried out using several evaluation metrics, namely accuracy, precision, recall, and F1-score, to provide a comprehensive performance overview in a multiclass classification problem. Accuracy was used to measure the overall correctness of the model's predictions, while precision and recall were employed to assess the model's ability to correctly and completely classify each disease class. The F1-score was used to evaluate the balance between precision and recall simultaneously. A high F1-score indicates that both precision and recall achieve good performance, whereas a low F1-score suggests an imbalance or poor performance of one or both metrics, which may occur when one metric is significantly higher or lower than the other [17].

As an initial benchmark, this study employs the Zero Rule (ZeroR) baseline method, which is a simple classification approach that predicts all test instances based solely on the majority class in the training data without considering any input features. The use of this baseline aims to ensure that the proposed SqueezeNet-CBAM model achieves superior performance compared to the most basic classification approach.

Through this evaluation process, it can be determined to what extent the model is able to effectively learn data characteristics while maintaining its generalization capability on unseen data. The evaluation results are then used as the basis for comparing the effectiveness of data augmentation and the CBAM module within the SqueezeNet architecture. Consequently, this evaluation plays a crucial role in identifying the most optimal model configuration for improving classification performance.

4. Results and Discussion

At this stage, a series of experimental scenarios was conducted to evaluate the performance of the SqueezeNet architecture under several configuration combinations, namely with and without the application of data augmentation and the Convolutional Block Attention Module (CBAM). This evaluation scheme was primarily designed to observe the impact of incorporating the CBAM module on model performance under both non-augmented and augmented data conditions.

All configurations were trained using identical training parameters to ensure a fair comparison. The models were trained for 50 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. All input images were resized to 224×224 pixels, and the training objective was optimized using the Categorical Cross-Entropy loss function. To ensure reproducibility, a fixed random seed of 42 was applied during the training process. All experiments were conducted on a workstation equipped with an Intel Core i5-13450HX processor, 12 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU with 6 GB VRAM. By keeping all training settings and hardware configurations constant, any performance differences observed can be directly attributed to the presence of the CBAM module rather than variations in training configurations. The detailed experimental setups are summarized in Table 1.

Table 1. Test Scheme Table

No	architecture	Augmentation	CBAM
1	SqueenzeNet	No	No
2	SqueenzeNet	No	Yes
3	SqueenzeNet	Yes	No
4	SqueenzeNet	Yes	Yes

A. Model Testing Results

This subsection presents the experimental results obtained from testing the SqueezeNet architecture under four different configuration scenarios, as summarized in Table 2. The evaluation was conducted to analyze the effect of data augmentation and the Convolutional Block Attention Module (CBAM) on model performance. The metrics used for evaluation include accuracy, precision, recall, and F_1-score, where precision, recall, and F1-score are reported using macro average values to ensure equal consideration of all classes.

Table 2. Result testing scheme table

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Number of Parameter
1	SqueenzeNet (No Augmentation, No CBAM)	92,5%	93,48%	92,5%	92.71%	1.231.608
2	SqueenzeNet (No Augmentation, with CBAM)	93,75%	94,32%	93,75%	93,65%	1.339.500
3	SqueenzeNet (Augmentation, No CBAM)	66,25%	63,51%	66,25%	58,98%	1.231.608
4	SqueenzeNet (Augmentation, with CBAM)	70%	55,57%	70%	60,98%	1.339.500

Based on the results presented in Table 2, the baseline SqueezeNet model without data augmentation and CBAM demonstrates a strong initial performance, indicating that the architecture is capable of extracting discriminative features from banana leaf images. Prior to deep learning evaluation, a ZeroR baseline classifier was employed as a reference model, achieving an accuracy of 50.48% by

predicting the majority class. This result serves as a lower-bound benchmark, indicating the minimum performance level expected without feature learning.The incorporation of the CBAM module in the non-augmented configuration further enhances the classification performance across all evaluation metrics. This improvement suggests that the attention mechanism effectively emphasizes informative feature channels and spatial regions, leading to more accurate predictions.

In contrast, the application of data augmentation without CBAM results in a noticeable decline in model performance. This indicates that the additional variability introduced by augmentation increases the complexity of the learning process, which may hinder the model's ability to generalize effectively when attention-based feature refinement is not applied. Nevertheless, when data augmentation is combined with CBAM, the performance improves compared to the augmented model without CBAM, demonstrating that CBAM partially mitigates the adverse effects of augmentation by strengthening feature representation. Furthermore, the parameter analysis reveals that the inclusion of CBAM slightly increases the number of model parameters. However, this increase remains relatively limited and does not compromise the lightweight characteristics of the SqueezeNet architecture. Overall, these results highlight the effectiveness of CBAM in improving classification performance, particularly in configurations without data augmentation, while maintaining computational efficiency.

Table 3. Evaluation Metrics of SqueezeNet–CBAM (No Augmentation)

	Class	Precision (%)	Recall (%)	F1-Score (%)	Support
1	Cordana	86,36	95	90,48	20
2	Healthy	100	100	100	20
3	Pestalotiopsis	90,91	100	95,24	20
4	Sigatoka	100	80	88,89	20
	Average	94,32	93,75	93,65	80

Table 3 presents the per-class evaluation metrics of the SqueezeNet–CBAM model without data augmentation. The results show that the model performs consistently well across all classes, with the Healthy class achieving perfect precision, recall, and F1-score. Sigatoka exhibits a lower recall value, indicating a higher number of false negatives compared to other classes. This per-class analysis confirms that the reported macro-average performance is not dominated by a single class and provides a clearer distribution of classification errors.
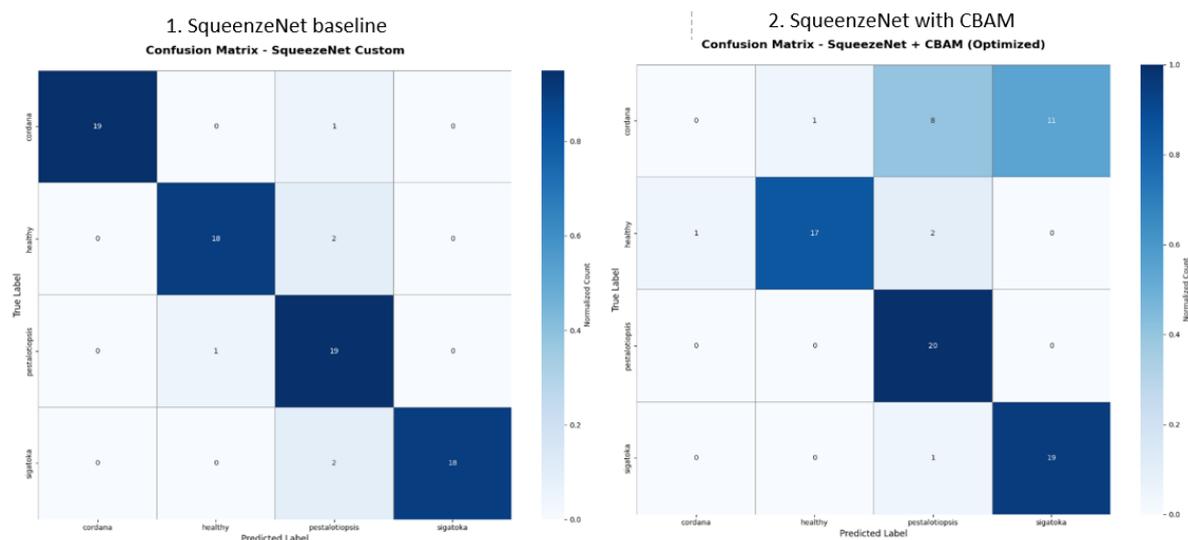

Figure 6. Confusion Matrices of SqueezeNet without Data Augmentation

Figure 6 illustrates the confusion matrices of the SqueezeNet model evaluated without data augmentation under two configurations, namely the baseline model and the model enhanced with the

Convolutional Block Attention Module (CBAM). The baseline configuration exhibits several misclassifications across the banana leaf disease classes, indicating limitations in feature discrimination when attention mechanisms are not applied. In contrast, the CBAM-enhanced model demonstrates a clearer diagonal dominance, reflecting improved classification consistency and more accurate predictions across all classes. These observations suggest that the incorporation of CBAM effectively enhances the model's ability to focus on relevant spatial and channel-wise features, leading to better performance in non-augmented conditions.
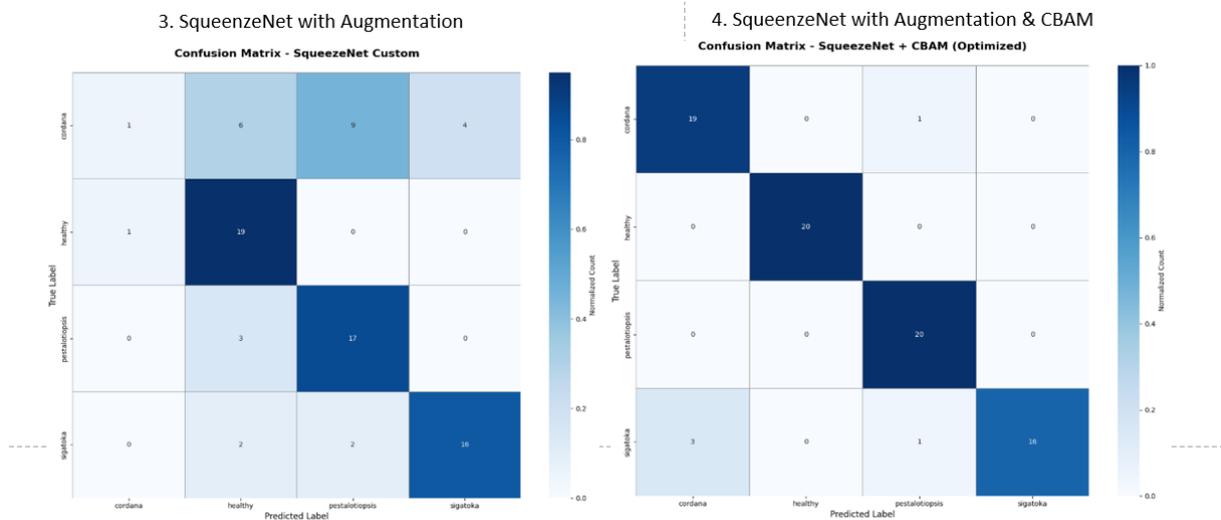

Figure 7. Confusion Matrices of SqueenzeNet with Data Augmentation

Figure 7 presents the confusion matrices of the SqueezeNet model evaluated with data augmentation under two configurations, namely without CBAM and with CBAM. The model trained with data augmentation but without CBAM shows an increased number of misclassifications across several classes, indicating that the augmented variability introduces additional challenges in feature learning. When CBAM is incorporated, the confusion matrix demonstrates improved diagonal concentration, suggesting enhanced classification accuracy and more balanced predictions among classes. Although the performance improvement is not as pronounced as in the non-augmented configuration, the results indicate that CBAM contributes positively to reducing misclassification in augmented data settings.

5. Conclusions

This study evaluated four experimental configurations of the SqueezeNet architecture for banana leaf disease classification by examining the effects of data augmentation and the Convolutional Block Attention Module (CBAM). The highest classification performance was achieved by the SqueezeNet model without data augmentation and with CBAM, reaching an accuracy of 93.75%, which outperformed the baseline SqueezeNet without CBAM accuracy of 92.5%. Conversely, the lowest performance was observed when data augmentation was applied without CBAM, where accuracy dropped significantly to 66.25%, indicating that excessive data variability can negatively affect feature learning. Although the integration of CBAM under augmented conditions improved accuracy to 70%, this result remained inferior to the non-augmented CBAM configuration. Overall, these findings confirm that CBAM consistently enhances classification accuracy while preserving the lightweight characteristics of the SqueezeNet architecture. Future research is recommended to investigate more selective and adaptive data augmentation strategies that are better aligned with disease-specific visual patterns to improve model robustness and generalization.

References

[1] R. Suryani and - Owbel, "Pentingnya Eksplorasi dan Karakterisasi pada Tanaman Pisang sehingga Sumber Daya Genetik Tetap Terjaga," *Agro Bali Agric. J.*, vol. 2, no. 2, pp. 64–76, 2020, doi: 10.37637/ab.v2i2.410.

[2] W. O. S. Sariamanah, A. Munir, and A. Agriansyah, "Karakteristik Morfologi Tanaman Pisang (Musa paradisiaca L.) di Kelurahan Tobimeitake Kecamatan Abeli Kota Kendari," *AMPIBI J. Alumni Pendidik. Biol.*, vol. 1, no. 3, pp. 32–41, 2016, [Online]. Available: http://ojs.uho.ac.id/index.php/ampibi/article/view/5043

[3] B.Statistik, "Produksi Tanaman Buah–Buahan dan Sayuran Tahunan Menurut Jenis Tanaman, 2024," Badan Pusat Statistik. [Online]. Available: https://www.bps.go.id/id/statistics-table/3/WXpSVU5uUTBOSEl5WVhGQmVESTVSVnBSVlhWeVVVUMDkjMw==/produksi-tanaman-buah---buahan-dan-sayuran-tahunan-menurut-jenis-tanaman--2024.html?year=2024

[4] Kasrina dan A. Zulaikha Q, "Pisang Buah (Musa spp): Keragaman dan Etnobotaninya Pada Masyarakat Di Desa Sri Kuncoro Kecamatan Pondok Kelapa Kabupaten Bengkulu Tengah," *Pros. Semirata FMIPA Univ. Lampung*, no. 1995, pp. 33–40, 2013.

[5] N. U. Aliah, Liliek Sulistyowati, and Anton Muhibbudin, "Hubungan Ketebalam Lapisan Epidermis Daun Terhadap Serangan Jamur (Mycosphaerella Musicola) Penyebab Penyakit Bercak Daun Sigatoka Pada Sepuluh Kultivar Pisang," *J. HPT*, vol. 3, no. 1, pp. 35–43, 2015.

[6] R. S. Segall, "Big Data Visualization for Black Sigatoka Disease of Bananas and Pathogen – Host Interactions (PHI) of Other Plants," vol. 13, no. 1, pp. 1–22, 2024, doi: 10.4018/IJARB.361940.

[7] V. S. Dhaka *et al.*, "A survey of deep convolutional neural networks applied for prediction of plant leaf diseases," *Sensors*, vol. 21, no. 14, 2021, doi: 10.3390/s21144749.

[8] A. Ajit and A. C. Layer, "A Review of Convolutional Neural Networks," pp. 1–5, 2020.

[9] D. Pratama, "Jurnal Teknologi Terpadu LEARNING," *J. Teknol. Terpadu*, vol. 8, no. 1, pp. 89–94, 2022.

[10] M. S. L. S. KEMBAREN, "SKRIPSI OLEH : FAKULTAS TEKNIK UNIVERSITAS MEDAN AREA MEDAN SKRIPSI Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana di Fakultas Teknik Universitas Medan Area Oleh : MICHAEL PRAYOGO SIAHAAN," 2024.

[11] Sukri and D. Arisandi, "Jurnal Resti," *Resti*, vol. 1, no. 1, pp. 19–25, 2020.

[12] M. A. B. Bhuiyan, H. M. Abdullah, S. E. Arman, S. Saminur Rahman, and K. Al Mahmud, "BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases," *Smart Agric. Technol.*, vol. 4, no. March, p. 100214, 2023, doi: 10.1016/j.atech.2023.100214.

[13] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11211 LNCS, pp. 3–19, 2018, doi: 10.1007/978-3-030-01234-2_1.

[14] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "50 X FEWER PARAMETERS AND < 0 . 5MB MODEL SIZE," pp. 1–13, 2017.

[15] M. A. Al-Fahrezi, "Pengaruh Augmentasi Data Terhadap Akurasi Pelatihan Model CNN untuk Klasifikasi Jenis Ikan," *JITSI J. Ilm. Teknol. Sist. Inf.*, vol. 6, no. 2, pp. 177–185, 2025, doi: 10.62527/jitsi.6.2.471.

[16] N. Helmawati and E. Utami, "Analysis for Detecting Banana Leaf Disease Using the CNN Method," *JUITA J. Inform.*, vol. 13, no. 1, pp. 29–36, 2025, doi: 10.30595/juita.v13i1.24514.

[17] R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," pp. 79–91, 2020, doi: 10.18653/v1/2020.eval4nlp-1.9.