SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Attention Augmented Deep Learning Model for Enhanced Feature Extraction in Cacao Disease Recognition

Robet[1)*], Johanes Terang Kita Perangin Angin[2)], Tarq Hilmar Siregar[3)]
[1)3)]Department of Informatics, STMIK Time, Medan, Indonesia
[2)]Department of Information Systems, STMIK Time, Medan, Indonesia
[1)*]robet@stmik-time.ac.id, [2)]time.johanes@gmail.com, [3)]tarqhilmarsiregar@gmail.com

Submitted :Aug 21, 2025 | **Accepted** : Sep 12, 2025 | **Published** : Oct 2, 2025

Abstract: Accurate cacao disease recognition is critical for safeguarding yields and reducing losses. Prior cacao studies primarily rely on handcrafted descriptors (eg, Color Histogram, LBP, GLCM) or standard CNN/transfer-learning pipelines, often limited to ≤ 3 classes and a single plant organ; explicit channel-spatial attention and comprehensive multiclass evaluation remain uncommon. To the best of our knowledge, no prior work integrates Squeeze-and-Excitation (SE) and the Convolutional Block Attention Module (CBAM) on a ResNeXt50 backbone for six-class cacao disease classification, accompanied by a standardized ablation study and t-SNE-based interpretability. We propose a six-class classifier (five diseases + healthy) built on ResNeXt-50 enhanced with SE (channel recalibration) and CBAM (channel-spatial emphasis) to highlight lesion-relevant cues. The dataset comprises labeled leaf and pod images from public sources collected under field-like conditions; preprocessing includes resizing to 224x224, normalization, and augmentation (flips, small rotations, color jitter, random resized crops). Trained with Adam and early stopping, ResNeXt50+SE+CBAM attains 97% test accuracy and 0.97 macro-F1, surpassing a ResNeXt50 baseline of 94% and 0.95 and SE-only/CBAM-only variants. Confusion matrix and t-SNE analyses show fewer mix-ups among visual classes and clearer separability, while the ablation validates complementary benefits of SE and CBAM. On a desktop-hosted, web-based setup, batch-1 inference at 224x224 is 7.46 ms/image (134 FPS), demonstrating real-time capability. The findings support deployment as browser-based decision-support tools for farmers and integration into continuous field-monitoring systems.
Keywords: Cacao Disease; ResNeXt50; SE; CBAM; Attention Mechanisms

INTRODUCTION

Cacao plays a strategic role in Indonesia's plantation sector, contributing substantially to farmers' livelihoods and regional economic growth. In recent years, its productivity has shown a notable decline, primarily caused by plant pest organisms, including insect infestations and various diseases affecting leaves, stems, and pods. This reduction impacts both the quantity and quality of beans, with potential losses reaching up to 70% of total production (Hamida, 2024; Indotama, 2022). Such circumstances emphasize the necessity for accurate, user-friendly, and field-adaptable early detection technologies.

Advancements in Computer Vision and Deep Learning (DL) have paved the way for the development of automated plant disease identification systems. Research on crops other than cacao indicates that Convolutional Neural Network (CNN) can surpass conventional machine learning techniques such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), as demonstrated in leaf disease detection (Nikith et al., 2022) and in the classification of three maize leaf diseases (Jeswani et al., 2023). For cacao, reported approaches range from classical feature extraction methods, such as Color Histogram (CH), Local Binary Pattern (LBP), and Gray Level Co-occurrence Matrix (GLCM) combined with Artificial Neural Network (ANN) or KNN (Baculio & Barbosa, 2022; Clarence et al., 2025), to DL based and transfer learning strategies for assessing fruit maturity, classifying infection severity, and detecting specific diseases (Bueno et al., 2020; Brosas et al., 2020; Hortinela & Tupas, 2022; Godmalin et al., 2022, 2023; Cagadas et al., 2024; Vera et al., 2024; Jesse et al., 2024). Popular backbone architectures, including VGG, ResNet, EfficientNet, MobileNet, and DarkNet, have been explored, along with state-of-the-art detection-based models and hybrid CNN-Transformer designs (Sing Soh et al., 2024; Moore &

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Modupe, 2025; Harvyanti et al., 2023; Rola et al., 2024; Miracle, 2024; Mamadou et al., 2023; Kouassi et al., 2025).

Despite encouraging progress in DL for cacao disease classification, most studies remain constrained to binary or no more than three class settings, focus on a single plant part (leaf, stem, or pod), and rarely examine the combined use of channel and spatial attention within modern CNN backbones. Meanwhile, recent agricultural-vision works (2024–2025) increasingly evaluate attention modules, ECA, SE, and CBAM to amplify lesion-relevant cues and suppress background clutter under challenging field photometry, generally reporting robustness gains on leaf/fruit imagery (Duhan et al., 2024; Yang et al., 2024; Karthikeyan et al., 2025). However, for cacao specifically, research that (i) addresses realistic multiclass scenarios reflecting diverse field symptoms, (ii) integrates channel and spatial attention on contemporary backbones, and (iii) conducts standardised ablation alongside t-SNE–based feature interpretability remains limited. Recent cacao-focused surveys and studies have noted that DL is growing but remains fragmented in terms of task scope and dataset coverage (Alvarado et al., 2025; Vera et al., 2024). To the best of our knowledge, no prior work has integrated SE and CBAM on a ResNeXt50 backbone for six-class cacao disease classification, utilizing standardized ablation and t-SNE analysis.

This paper proposes a ResNeXt50 model enhanced with Squeeze-and-Excitation (SE) and the Convolutional Block Attention Module (CBAM) to increase sensitivity to both subtle and dispersed lesion patterns. This study (1) evaluates a six-class setting covering five cacao diseases and a healthy class that reflects diverse field symptoms, (2) quantifies the contribution of each attention component through a controlled ablation (SE-only, CBAM-only, SE+CBAM), and (3) examines feature interpretability via t-SNE. The findings are expected to broaden the classification scope for cacao diseases, provide empirical evidence on the effectiveness of combining channel and spatial attention in ResNeXt50, and supply error analyses that support a practical early-detection system.

LITERATURE REVIEW

Advances in Computer Vision and Deep Learning (DL) have driven their widespread adoption across various sectors, ranging from infrastructure to industrial safety (Robet et al., 2022, 2024, 2025). In agriculture, particularly for non-cacao commodities, Convolutional Neural Networks (CNNs) have consistently outperformed conventional Machine Learning methods, such as Support Vector Machines (SVMs) and K-Nearest Neighbors (KNNs). For instance, CNNs achieved 96% accuracy in classifying eight leaf diseases and proved effective in classifying three maize leaf diseases (Nikith et al., 2022; Jeswani et al., 2023). Similar findings have been reported for other tropical commodities such as rice (Mahadevan et al., 2024) and oil palm (Yarak et al., 2021), reinforcing the strength of end-to-end hierarchical representation in capturing complex lesion patterns compared to manual feature engineering. The achievements have encouraged the increasing adoption of CNNs and other DL-based approaches in plant disease detection, including cacao.

In cacao, reported approaches range from classical feature engineering to CNN and transfer learning-based methods, with research subjects encompassing beans, pods, and leaves. Techniques such as Color Histogram (CH) and Local Binary Pattern (LBP) combined with Artificial Neural Network (ANN) have been shown to outperform SVM and Logistic Regression in distinguishing healthy from unhealthy pods, while Gray Level Co-occurrence Matrix (GLCM) followed by KNN has proven effective for pod disease detection (Baculio & Barbosa, 2022; Clarence et al., 2025). In the post-harvest domain, the combination of image processing and Adaptive Neuro Fuzzy Inference System (ANFIS) achieved 99.7% accuracy for bean quality grading, whereas CNN-based detection of physical bean defects reported 90.67% accuracy (Brosas et al., 2020; Hortinela & Tupas, 2022). These classical feature approaches are advantageous in terms of interpretability and computational efficiency, although their performance often degrades under variations in spatial patterns, delicate textures, and inconsistent lighting. Such limitations have driven the shift toward DL-based approaches capable of learning directly from image data.

Recent developments indicate the dominance of DL and transfer learning methods in various cacao-related tasks, similar to trends observed in other horticultural commodities such as mango (Varma et al., 2025) and banana (Bhuiyan et al., 2023). CNN-based classification of cacao pod maturity has been reported to achieve high accuracy, while transfer learning-based classification of pod conditions and infection severity also consistently exceeds 90% (Bueno et al., 2020; Godmalin et al., 2022, 2023). To achieve such performance, a variety of backbone architectures have been explored, including VGG, ResNet, EfficientNet, MobileNet, and DarkNet for specific diseases like “Monilia” and “Black Pod” (Cagadas et al., 2024; Harvyanti et al., 2023; Vera et al., 2024).

Recent explorations have led to head-to-head comparisons, for example, ResNet50, VGG16, and a Vision Transformer (ViT), with ResNet50 achieving the best performance on three-class cacao disease classification (Jesse et al., 2024). This outcome is consistent with ViT’s greater reliance on large-scale pretraining and tailored optimization, whereas convolutional backbones tend to be more data-efficient in medium-sized settings, such as 224×224.

In addition to large-scale models, lightweight architectures such as MobileNetV2 paired with classical classifiers have performed well in specific scenarios (Mamadou et al., 2023), and hybrid CNN-Transformer

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

approaches, which fuse convolutional inductive bias with global self-attention, show strong potential for diseases such as "Swollen Shoot", offering richer multi-scale representation and a more favorable accuracy-efficiency trade-off than either pure CNNs or pure Vits (Kouassi et al., 2025).

Despite these advancements, several limitations remain prominent. Most studies restrict classification to binary or no more than three class schemes and focus on a single plant part, thus failing to capture the diversity of symptoms encountered in the field fully. Explicit attention mechanisms, particularly the simultaneous combination of channel attention and spatial attention, remain rarely explored, with many studies relying on baseline CNNs lacking modules that could enhance sensitivity to subtle and dispersed lesions (Sing Soh et al., 2024; Moore & Modupe, 2025). In the context of cacao disease research, a truly comprehensive multiclass evaluation, covering precision, recall, and per-class F1-score, macro/micro averaging, and confusion matrix analysis, has not been consistently implemented; most studies only report the overall accuracy of basic aggregate metrics. Interpretability aspects are often overlooked, despite techniques like t-distributed stochastic neighbor embedding (t-SNE) revealing feature distribution patterns and potential inter-class misclassifications (Bueno et al., 2020; Rola et al., 2024; Miracle, 2024).

Table 1. Summary of Cacao Disease Classification Studies with Datasets, Methods, Accuracy, and Limitations

Authors	Year	Datasets	Methods	Accuracy	Limitations
(Brosas et al., 2020)	2020	Private images 200 samples cacao beans (3 classes: moldy, slaty, defect types	Image processing + ANFIS (Grading), CNN (Defect Detection)	99.71% (grading), 72.4%–98.7% (defect)	Small dataset (N=200); reports accuracy only (no precision/recall/F1 or confusion matrix)
(Baculio & Barbosa, 2022)	2022	Private images (cacao pods): 217 healthy and 201 unhealthy (2 classes)	CH, LBP + ANN	98.3%	Binary scope; classical features; lighting variations impact
(Hortinela & Tupas, 2022)	2022	Private images 500 samples cacao beans, each class 100 images (5 classes: good, broken, clumped, flat, moldy)	Preprocessing using Gaussian Blur, Otsu Thresholding, masking, and VGG-16.	Overall accuracy 90.67% from 75 samples (68/75)	Small dataset size and a single location, Uniform class distribution (100 per class)
(Godmalin et al., 2022;2023)	2022, 2023	Private images (3 classes: healthy, black pod disease, pest attack)	CNN "Lightweight and Transfer learning	94% (2022), 91% (2023)	The dataset and protocol details are incomplete. There is no comparison with ResNet50 and VGG16 (or MobileNetV2 for the lightweight version) in the same protocol.
(Harvyanti et al., 2023)	2023	Private images (cacao leaf) 2 classes: 343 healthy and 327 Vascular Streak Dieback (VSD.	Comparison of four CNN (transfer learning) architectures: AlexNet (97.78%), SqueezeNet (97.50%), DarkNet-19 (98.61%), and a Modified-CNN (94.17%)	DarkNet-19: 98.61%,	Binary task (VSD only) and strong dependence on augmentation (from 670 to 1,200 images). No report on model size/latency/compute cost for practical deployment.
(Jesse et al., 2024)	2024	Public Kara AgroAI Cocoa dataset (leaf) 3 classes: Anthracnose (5,162), CSSVD (7,292), Healthy (5,249)	Transfer learning VGG16, ResNet50, and Vision Transformer (ViT).	ResNet50: 94%, VGG16: 92%; ViT: 80%.	3 classes classification, and the hyperparameter table shows different learning rates (VGG16 1e-2 vs. ResNet50/ViT 1e-3) and markedly different weight decay (ResNet50 1e-4 vs. ViT 0.03)
(Cagadas et al., 2024)	2024	Public dataset (cacao leaf ~2,000 images) 4 classes: Healthy, CSSVD, VSD, Witches' Broom	Pretrained VGG19	Overall accuracy: 88.75%	Deployment metrics absent: no latency/model size measured for practical use; no comparison to lighter backbones.
(Rola et al., 2024)	2024	2,000 cacao pod images collected (healthy and unhealthy)	Handcrafted color, shape, and texture features (Haralick) were evaluated with classical learners (Naïve Bayes, Decision Stump, Random Forest, Hoeffding Tree, Multilayer NN) and a CNN.	CNN: 99% testing accuracy	Binary task only, and 150×150 crops may remove relevant morphological/contextual cues.
(Kouassi et al., 2025)	2025	3500 cacao pod images (1,200 healthy and 2,300 infected pods)	Hybrid CNN–ViT + SVM & LightGBM):	99.24%	Binary (healthy vs. infected pods focused on Swollen Shoot)
Ours	2025	Leaf & Pod (6 classes: CSSVD, Anthracnose, Black Pod Rot, Monilia, Pod Borer, Healthy)	ResNeXt50 + SE + CBAM	97%	Reliance on a limited public dataset lacking geographical representativeness.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The synthesis of this literature underscores the need for research that addresses more realistic, large-scale multiclass classification scenarios, adopts metric-rich evaluation protocols, and integrates attention mechanisms capable of highlighting relevant lesion features. In this context, a ResNeXt50 architecture enhanced with Squeeze-and-Excitation (SE) and the Convolutional Block Attention Module (CBAM) is considered to hold strong potential for overcoming the challenges of detecting subtle and dispersed lesion patterns, while improving feature discrimination for six-class cacao classification.

METHOD

The overall experimental workflow is illustrated in Figure 1. The process begins with cacao dataset preparation, followed by experimental setup and preprocessing, including resizing, normalization, and augmentation. The dataset was split into training (70%), validation (15%), and testing (15%) subsets. Four model configurations were trained, namely baseline ResNeXt50, ResNeXt50 with Squeeze-and-Excitation (SE), ResNeXt50 with Convolutional Block Attention Module (CBAM), and the proposed ResNeXt50 with SE and CBAM. Finally, the trained models were evaluated on the test set using accuracy, precision, recall, and F1-score as evaluation metrics.

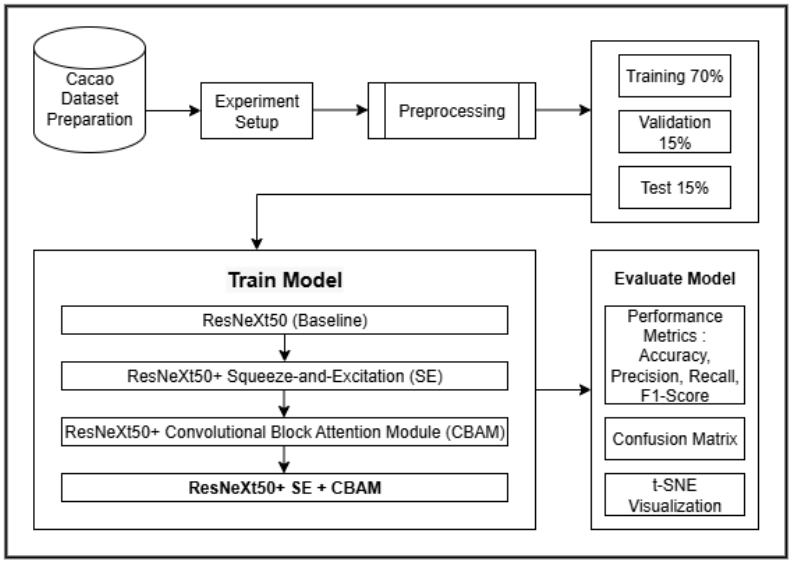

Fig. 1 Research Stages

Dataset Preparation

The dataset used in this study consists of images of cacao leaves and pods, collected from open sources, including Kaggle and other publicly available online repositories. All images were manually curated to ensure quality and consistent labelling across six target classes :

1. Cacao Swollen Shoot Virus Disease (CSSVD): a viral disease affecting the cacao vascular system, characterised by leaf mosaic and chlorosis, as well as swelling of petioles and stems.
2. Anthracnose: a fungal disease that affects both leaves and pods, producing necrotic brown-to-black lesions often surrounded by yellow halos.
3. Black Pod Rot: caused by Phytopthora, this disease primarily affects cacao pods, beginning as small brown spots that rapidly expand into complete black rot.
4. Healthy: leaf and pod samples showing no symptoms of disease or pest damage, serving as the control class.
5. Monilia: a fungal disease caused by Moniliophthora roreri, characterised by white, frosty-like patches on the surface of pods.
6. Pod Borer: a pest infestation caused by Conopomorpha Cramerella, leading to premature pod yellowing, shrivelling, and poorly developed beans.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

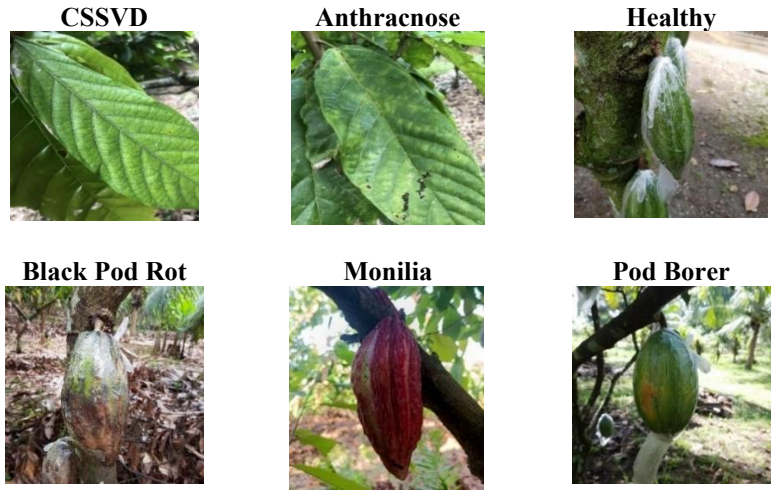

Fig. 2 Samples of the Dataset

Table 2. Distribution of Cacao Dataset Images

Class	Number of Images
CSSVD	826
Anthracnose	755
Black Pod Rot	943
Healthy	1000
Monilia	796
Pod Borer	254
Total	**4574**

Experiment Setup

The hardware and software configurations employed in this study are summarized in Table 3.

Table 3. Experimental Setup

Hardware/Software	Description
Google Colab	Software Specification
Jupyter Notebook	Virtual Environment for Python Code
Google Drive	Dataset Storage
Pytorch 2.4, Scikit-Learn, Matplotlib, Seaborn, Numpy, Flask	Open-source software for training and testing the images

Data Preprocessing

All training images were preprocessed through a set of transformations designed to standardize input size and enhance data diversity. Each image was randomly resized and cropped to 224 × 224 pixels with a scale range of 0.5 to 1.0, ensuring variations in zoom levels. To introduce geometric diversity, random horizontal flipping and random rotations up to 45° were applied, along with a random affine shear (10°) to simulate mild deformations. Furthermore, color jitter was used to adjust brightness and contrast, improving robustness against lighting variations. After augmentation, the images were converted into tensors and normalized using the mean and standard deviation of the ImageNet dataset (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]) to match the pre-trained ResNeXt-50 backbone. For validation and testing, images were resized to 256 pixels on the shorter side**,** center-cropped to 224 × 224, converted to tensors, and normalized with the same ImageNet statistics. This ensured consistency during evaluation while avoiding data augmentation that could bias performance measurement.

Train Model

The dataset was divided into three subsets with a ratio of 70% for training, 15% for validation, and 15% for testing. The training set was used to optimize model parameters, the validation set was employed for manual hyperparameter tuning and early stopping, and the independent test set was reserved for final evaluation. The

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

models were trained using multi-class cross-entropy loss, and the Adam optimizer with an initial learning rate and weight decay was set to 1×10^{-4}. Training was performed with a batch size of 32 for up to 50 epochs. In comparison, an early stopping criterion with a patience of 5 epochs was applied to terminate training once the validation loss stopped improving.

The backbone network was based on ResNeXt50 pretrained on ImageNet, which was further extended with attention mechanisms. The Squeeze-and-Excitation (SE) block was used for channel-wise recalibration via a bottleneck of ratio r 16; a global average pooling (GAP) vector of length C = 2048 is passed through a two-layer MLP C → C/r → C, followed by a sigmoid gate that rescales each channel. The Convolutional Block Attention Module (CBAM) was applied to refine channel and spatial features jointly before global pooling. In its channel stage, CBAM aggregates features with both global average and global max pooling, feeds them to a shared MLP with the same reduction ratio r=16, sums the outputs, and applies a sigmoid gate. In its spatial stage, CBAM concatenates the per-pixel average and max maps across channels and applies a 7x7 convolution followed by a sigmoid to produce a spatial attention mask. The attended feature maps are finally aggregated by global average pooling and passed to a fully connected layer (2048 → 6) and a softmax to predict the six cacao classes.

Pseudocode : SE+CBAM Using In ResNeXt50 Backbone

```
Input: image x
F ← ResNeXt50(x)                    # feature maps
# SE
s ← σ(W2 δ(W1 GAP(F)))
F ← F ⊙ s
# CBAM (channel then spatial)
m_c ← σ(MLP(GAP(F)) + MLP(GMP(F)))
F ← F ⊙ m_c
m_s ← σ(fk×k([AvgPool_c(F); MaxPool_c(F)]))
F ← F ⊙ m_s
# Classification
z ← GAP(F); y ← W_cls z; p ← softmax(y)
return p
```

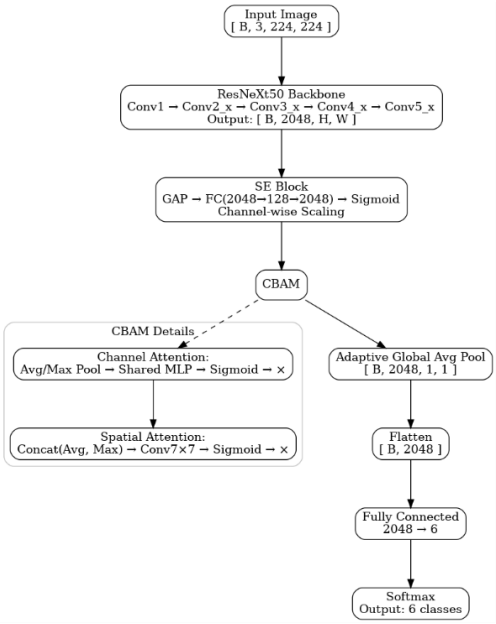

Fig. 3 Architecture of Proposed Method ResNeXt50+SE+CBAM

Evaluation Model

The evaluation was conducted on a held-out test set (15%) that was never used during training or validation. During training, performance was monitored on the validation set (15%) to enable early stopping based on validation loss, thereby preventing overfitting. The best-performing model on the validation set was selected and subsequently evaluated on the test set to obtain the final metrics. We employed standard multi-class classification metrics to assess model performance comprehensively:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Accuracy: the proportion of correctly classified samples out of the total.

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}i=yi)}{N} \quad (1)$$

Precision(per class): the fraction of correct predictions within a given class.

$$\text{Precision } \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

Recall: the fraction of actual class samples correctly identified.

$$\text{Recall } \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

F1-score: harmonic mean of Precision and Recall.

$$\text{F1} = 2 * \frac{\text{Precision*Recall}}{\text{Precision+Recall}} \quad (4)$$

Macro-average F1 (Macro-F1): the unweighted average of F1 across all classes, reflecting balanced performance.

$$\text{Macro} - \text{F1} = \frac{1}{\text{K}} \sum_{\text{k=1}}^{\text{K}} \text{F1}_{\text{k}} \quad (5)$$

Weighted-average F1 (weighted-F1): the class-size-weighted average F1, accounting for class imbalance

$$\text{Weighted} - \text{F1} = \frac{\sum_{\text{k=1}}^{\text{K}} \text{n}_{\text{k}} * \text{F1}_{\text{k}}}{\sum_{\text{k=1}}^{\text{K}} \text{n}_{\text{k}}} \quad (6)$$

Predictions were obtained via the argmax of the softmax probability distribution. Additionally, a confusion matrix was generated to analyze class-level errors, while t-SNE visualizations were used to assess feature separability qualitatively.

RESULT

Training and Validation Performance

Figures 4 and 5 depict the training and validation accuracy and loss curves, respectively, for all four model configurations: ResNeXt50 (baseline), ResNeXt50+SE, ResNeXt50+CBAM, and ResNeXt50+SE+CBAM. These learning curves provide insight into the convergence speed, generalization behavior, and stability of each model.

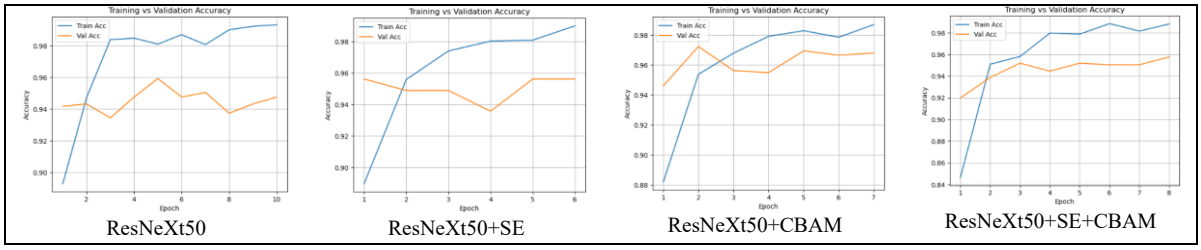

Fig. 4 Training and Validation Accuracy

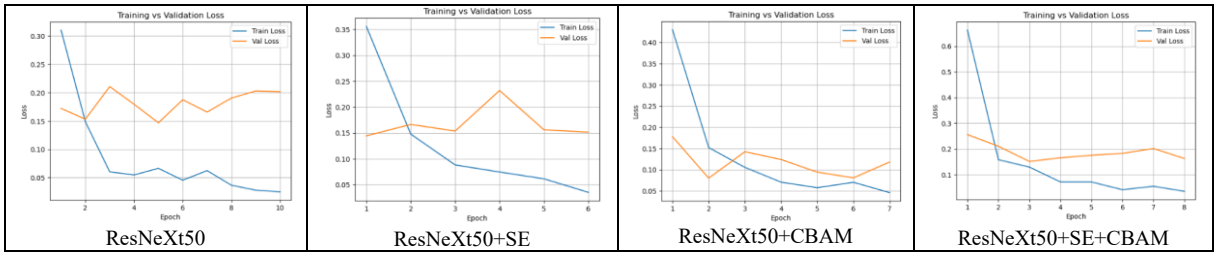

Fig. 5 Training and Validation Loss

The baseline ResNeXt50 converged rapidly, achieving a best validation accuracy of 95.9% at epoch 5, with a minimum validation loss of 0.1469. However, validation loss exhibited minor oscillations after epoch 5, indicating

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

mild overfitting. The addition of SE slightly improved training stability, producing a smoother accuracy trajectory and achieving a best validation accuracy of 95.6% with a minimum validation loss of 0.14442 at epoch 1.

The CBAM variant outperformed both the baseline and SE, reaching the highest validation accuracy of 97.2% at epoch two and the lowest validation loss of 0.0802 early in training, indicating fast convergence and stable optimization. In contrast, the SE+CBAM configuration achieved the best validation accuracy of 95.8% at epoch 8 with a validation loss of 0.1627, which was lower than the baseline but not superior to CBAM in the validation phase.

Table 4. Summary of the Best Validation Accuracy and Loss for Each Model Configuration

Model	Best Val Accuracy	Epoch (Acc)	Min Val Loss	Epoch (Loss)
ResNeXt50	95.9%	5	0.1469	5
ResNeXt50+SE	95.6%	1,5,6	0.1442	1
ResNeXt50+CBAM	97.2%	2	0.0802	2,6
ResNeXt50+SE+CBAM	95.8%	8	0.1627	8

Test Set Performance

The models were further evaluated on the unseen test set. Table 4 summarizes the classification metrics, including class-wise precision, F1-score, overall accuracy, and macro averages. The proposed ResNeXt50+SE+CBAM achieved the best test accuracy of 97% and a macro-F1 of 0.97, surpassing the baseline (94%) and single-attention variants (SE: 95%, CBAM: 96%). Notably, SE+CBAM yielded consistently high per-class F1-scores (≥0.95) across all six classes. Both monilia and pod borer classes reached perfect or near-perfect F1(1.00), while CSSVD recall increased to 0.98, reflecting the model's improved sensitivity to subtle lesion features.

Table 5. Test Set Performance of Each Model Configuration Across Six Cacao Disease Classes

Model	Accuracy	Macro-F1	Weighted-F1	CSSVD F1	Anthracnose F1	Black Pod Rot F1	Healthy F1	Monilia F1	Pod Borer F1
ResNeXt50	0.94	0.95	0.94	0.91	0.93	0.93	0.94	0.98	0.98
ResNeXt50+SE	0.95	0.95	0.95	0.93	0.92	**0.96**	**0.97**	0.98	0.95
ResNeXt50+CBAM	0.96	0.96	0.96	0.94	0.94	**0.96**	0.96	0.98	**1.00**
ResNeXt50+SE+CBAM	**0.97**	**0.97**	**0.97**	**0.97**	**0.96**	0.95	0.96	**1.00**	**1.00**

Ablation Study

Table 5 summarizes the ablation study results for four model configurations: ResNeXt50 (baseline), ResNeXt50+SE. ResNeXt50+CBAM, and ResNeXt50+SE+CBAM. Compared with the baseline (accuracy: 94%, macro-F1: 95%), adding SE improved accuracy to 95% while maintaining the same macro-F1, whereas adding CBAM resulted in 96% accuracy and 96% macro-F1. The combination of SE+CBAM achieved the best results, with 97% accuracy and 97% macro-F1, showing the largest improvements for the CSSVD (+0.006 F1) and anthracnose (+0.03 F1) classes. These findings suggest a synergistic effect between channel re-weighting from SE and spatial attention from CBAM, leading to enhanced inter-class separability.

Table 6. Summary of Test Set Accuracy and Macro-F1 for All Model Configurations in the Ablation Study.

Model	Accuracy	Macro-F1	Inference Time (ms)
ResNeXt50	94%	95%	7.32 ms/img (136 FPS)
ResNeXt50+SE	95%	95%	7.33 ms/img (136 FPS)
ResNeXt50+CBAM	96%	96%	7.35 ms/img (136 FPS)
ResNeXt50+SE+CBAM	97%	97%	7.46 ms/img (134 FPS)

Error Analysis

To better understand the misclassification, confusion matrices were generated for all model variants. For the baseline model, the most frequent errors occurred between CSSVD and anthracnose, as well as between black pod rot and monilia, which share visual similarities in lesion texture and color under specific lighting conditions. The incorporation of SE reduced misclassification in CSSVD, suggesting that channel re-weighting improved sensitivity to subtle spectral differences. CBAM further decreased confusion between black pod rot and monilia by enhancing the spatial localization of lesion areas. The SE+CBAM combination minimized cross-class errors to only a few isolated cases, demonstrating the complementary nature of the two attention mechanisms.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

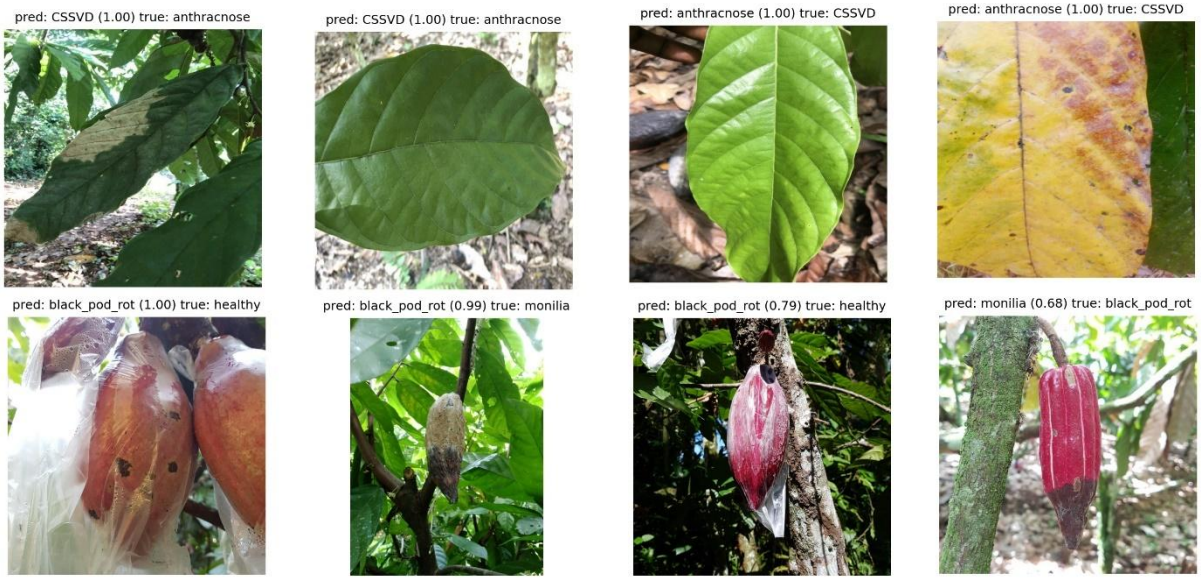

Fig. 6 Examples of Misclassified Cacao Diseases with Visual Similarity Analysis

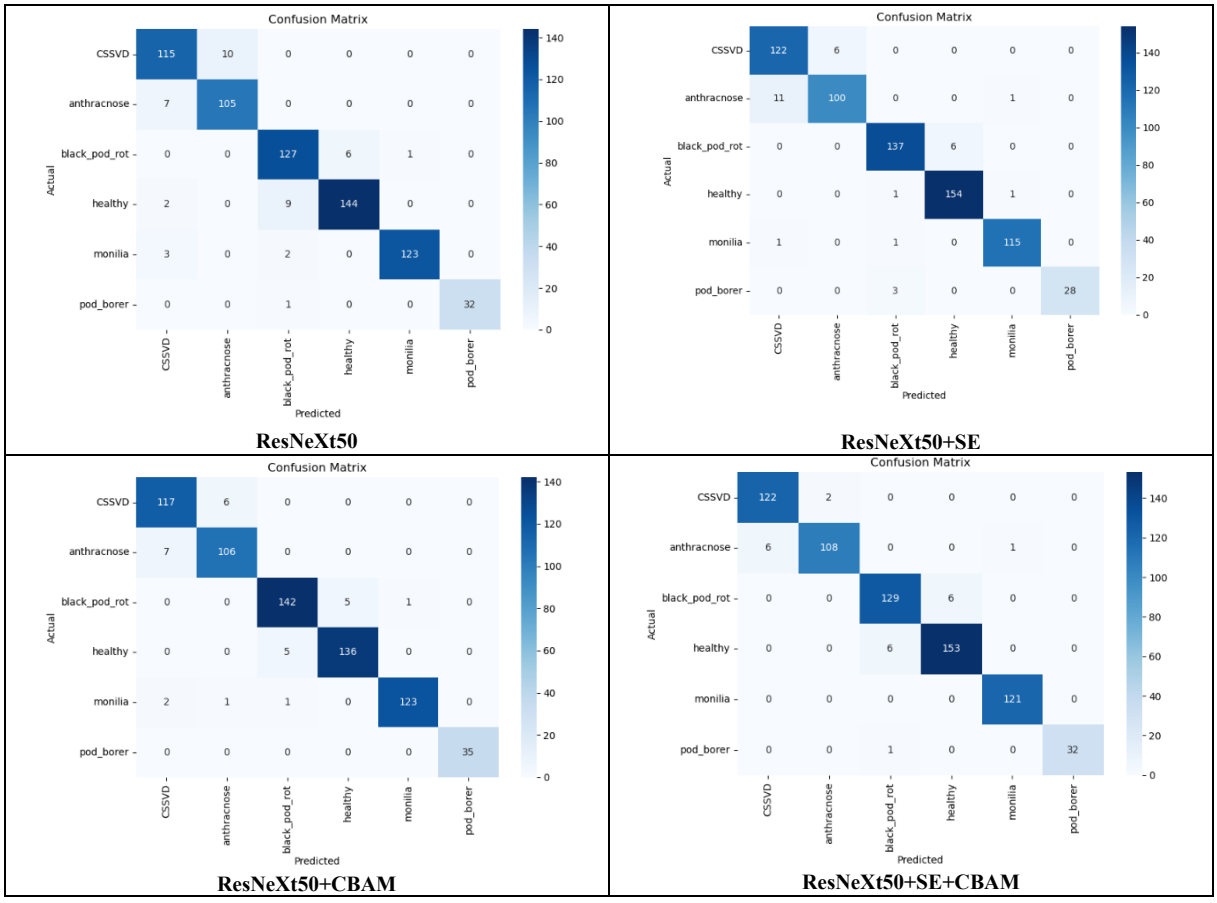

Fig. 7 Confusion Matrix of Four Models

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

t-SNE Visualization

To qualitatively assess the separability of learned feature representations, t-distributed stochastic neighbor embedding (t-SNE) was applied to the penultimate-layer feature vectors extracted from the test set. Each point corresponds to an image, colored according to its ground-truth class.

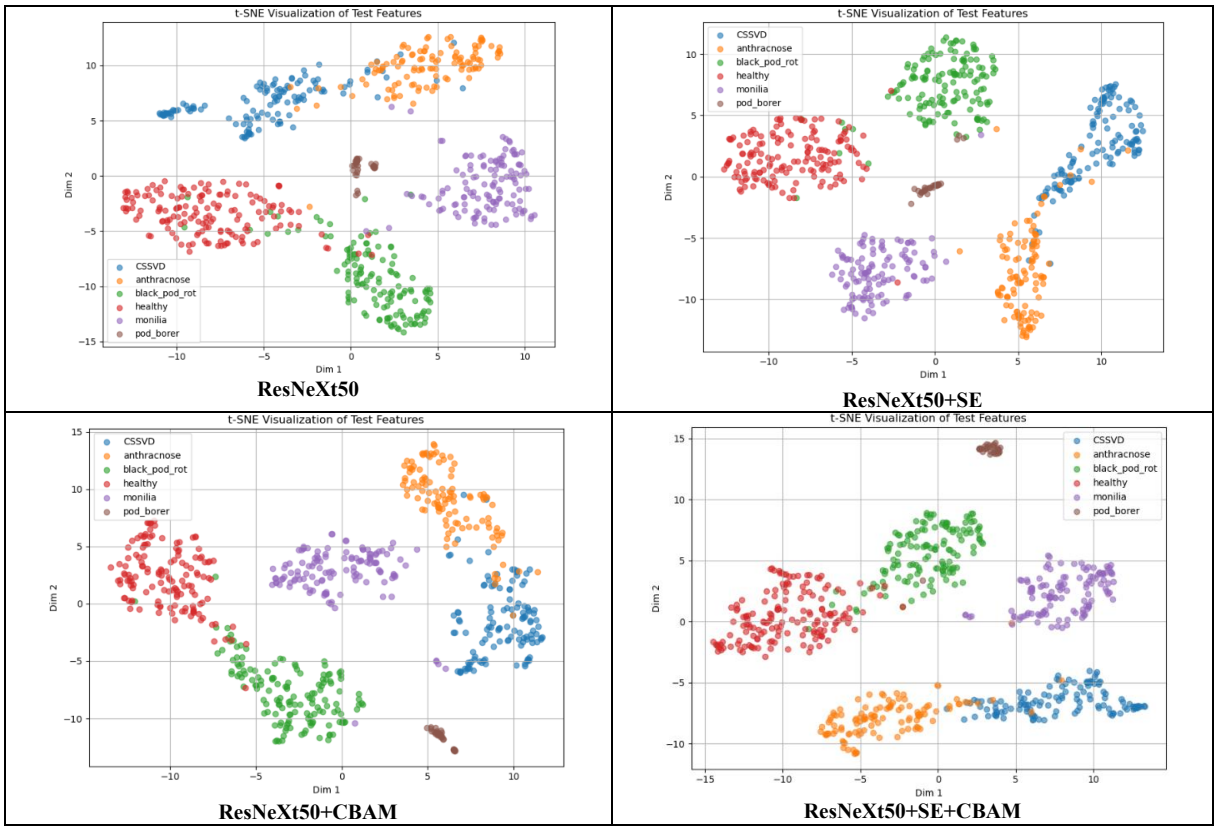

Fig. 8 t-SNE Distribution

In the baseline model, class clusters were visible but exhibited overlaps, particularly between CSSVD and anthracnose. The addition of SE yielded more compact and well-seperated clusters, indicating improved discriminative power. CBAM further enhanced class boundaries by focusing on spatially relevant regions. The SE+CBAM model produced the most apparent separation, with minimal inter-class overlap, confirming that attention integration improves the quality of feature embeddings and, consequently, classification performance.

User Interface Implementation

In addition to the quantitative evaluation, the proposed classification model was also deployed into a simple user interface, as shown in Figure 8.

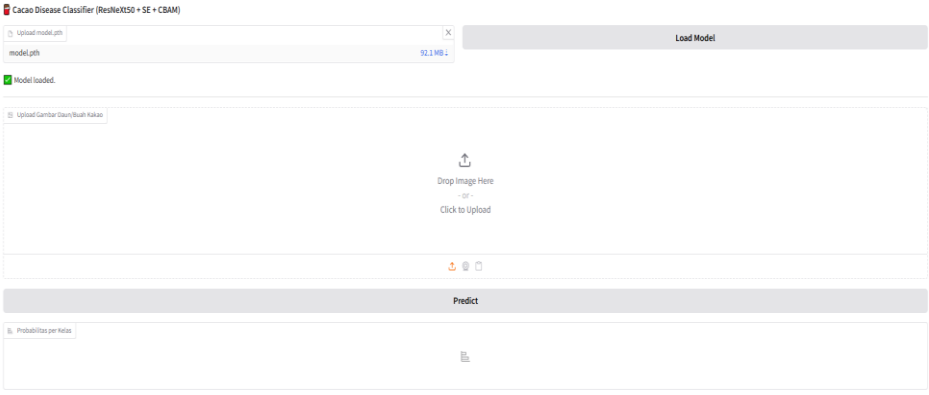

Fig. 9 Web User Interface

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The web interface comprises a model loader, an image upload/drag-and-drop input, a Predict button, and a results panel showing the top predicted class with its confidence and the per-class probability distribution across six classes. As shown in Figure 9, the interface displays the top prediction with its confidence and the six-class probability distribution for each test image.

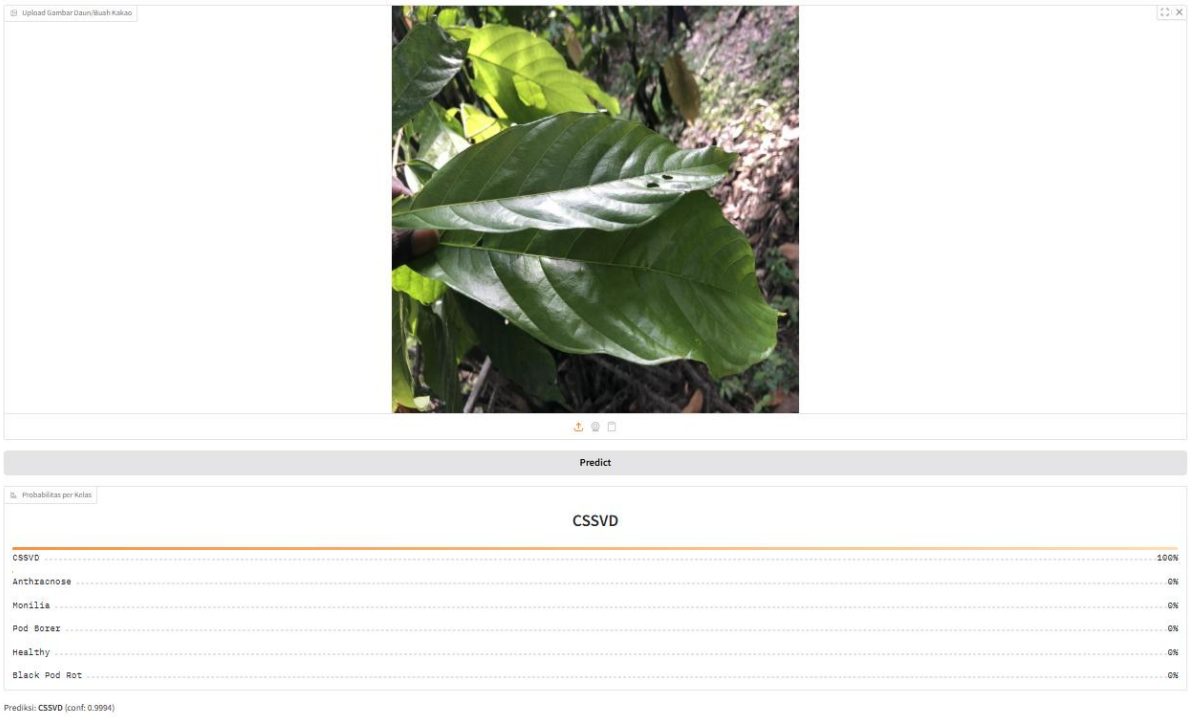

Fig. 10 Web User Interface Example Prediction

The example in Figure 9 illustrates the end-to-end flow, from image upload to probability outputs, supporting the system's usability.

DISCUSSIONS

The experimental findings indicate that integrating attention mechanisms into a high-capacity backbone such as ResNeXt50 significantly enhances classification performance for cacao healthy and disease recognition. The proposed ResNeXt50+SE+CBAM model achieved the highest accuracy (97%) and macro-F1 score (0.97), outperforming both the baseline and single-attention variants. This improvement is consistent with previous studies reporting that attention modules can refine feature representations by selectively amplifying relevant information while suppressing noise.

From the ablation study, SE primarily improved channel-level feature discrimination, allowing the network to weight the contribution of different feature channels adaptively. CBAM, on the other hand, enhanced both channel and spatial attention, enabling better localization of lesion regions even in cluttered backgrounds. Their combination provided complementary benefits, leading to more balanced predictions across all six classes, as reflected in the confusion matrix and per-class metrics. Notably, misclassifications between visually similar diseases such as CSSVD and anthracnose were reduced in the proposed model, supporting the hypothesis that spatial attention plays a critical role in differentiating fine-grained lesion patterns.

t-SNE visualizations further revealed that the feature embeddings produced by ResNeXt50+SE+CBAM formed more compact and separable clusters compared to the baseline. This indicates that the model not only achieves higher predictive accuracy but also learns a more discriminative representation of disease features.

In practical terms, the model can be integrated into a web-based workflow hosted on a desktop and accessed locally by farmers via computers or laptops. Compared with transformer-based models such as ViT and Swin, which excel at modeling global context but typically require large-scale pretraining and incur higher memory/latency budgets (Touvron et al., 2021; Liu et al., 2021), the proposed attention-CNN offers a favorable accuracy efficiency trade-off for medium-scale, field-image datasets and interactive, real-time desktop use (Yang et al., 2024). This positioning sharpens the novelty: dual attention (SE+CBAM) atop ResNeXt50 delivers practice-grade accuracy with measured real-time operation.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

CONCLUSION

This study proposes a cacao classification framework that utilizes an enhanced ResNeXt50 backbone, integrated with Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) techniques. Evaluated on a six-class dataset (five diseases + one healthy class), the ResNeXt50+SE+CBAM architecture achieved **97%** test accuracy and a macro-F1 score of **0.97**, outperforming the baseline ResNeXt50 and its single-attention variants. In a desktop-hosted, web-based setup, batch-1 inference at a 224x224 resolution achieved **7.46 ms/image (134 FPS)**. The ablation study confirmed SE improved channel-level feature discrimination, while CBAM enhanced spatial localization of lesion regions. Error analysis and t-SNE visualization demonstrated well-separated feature clusters with minimal misclassifications. Limitations include reliance on a limited public dataset lacking geographical representativeness and the absence of real-time testing on mobile devices. Future work should focus on: dataset expansion with geographically diverse samples; integration with lightweight CNN architectures (e.g., MobileNetV3); field testing on handheld devices; development of farmer-friendly mobile applications; and implementation of domain adaptation techniques for cross-site generalization. The architecture offers an effective balance between accuracy and computational efficiency, making it promising for agricultural disease monitoring systems, though practical deployment requires further optimization for resource-constrained environments.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the Ministry of Higher Education, Science, and Technology under the 2025 Research Grant Program.

REFERENCES

Alvarado, J., Restrepo-Arias, J. F., Velásquez, D., & Maiza, M. (2025). Disease Detection on Cocoa Crops Based on Computer-Vision Techniques: A Systematic Literature Review. *Agriculture (Switzerland)*, *15*(10), 1–27. https://doi.org/10.3390/agriculture15101032

Baculio, N. G., & Barbosa, J. B. (2022). An Objective Classification Approach of Cacao Pods using Local Binary Pattern Features and Artificial Neural Network Architecture (ANN). *Indian Journal of Science and Technology*, *15*(11), 495–504. https://doi.org/10.17485/ijst/v15i11.60

Bhuiyan, M. A. B., Abdullah, H. M., Arman, S. E., Saminur Rahman, S., & Al Mahmud, K. (2023). BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases. *Smart Agricultural Technology*, *4*(March), 100214. https://doi.org/10.1016/j.atech.2023.100214

Brosas, D. G., Villafuerte, R. S., & Obediencia, D. C. (2020). Adaptive Neuro-Fuzzy Approach for Cacao Bean Grading Classification Process. *Proceeding - 2020 3rd International Conference on Vocational Education and Electrical Engineering: Strengthening the Framework of Society 5.0 through Innovations in Education, Electrical, Engineering and Informatics Engineering, ICVEE 2020*. https://doi.org/10.1109/ICVEE50212.2020.9243281

Bueno, G. E., Valenzuela, K. A., & Arboleda, E. R. (2020). Maturity classification of cacao through spectrogram and convolutional neural network. *Jurnal Teknologi Dan Sistem Komputer*, *8*(3), 228–233. https://doi.org/10.14710/jtsiskom.2020.13733

Cagadas, D. O., Labajan, R. A., Cagadas, D. O., April, R., Cacao, L. L., Classification, D., Image, U., Cagadas, D. O., & Labajan, R. A. A. (2024). Leaf-Based Cacao Diseases Classification Using Image Processing To cite this version : HAL Id : hal-04483097 Leaf-Based Cacao Diseases Classification Using Image. *HAL Open Science*.

Clarence, E., Diego, S. S., Rodrin, S. G. C., & Arboleda, E. R. (2025). *Multi-Feature Visual Analysis And K-Nearest*. 28–34.

Duhan, S., Gulia, P., Gill, N. S., Shukla, P. K., Khan, S. B., Almusharraf, A., & Alkhaldi, N. (2024). Investigating attention mechanisms for plant disease identification in challenging environments. *Heliyon*, *10*(9), e29802. https://doi.org/10.1016/j.heliyon.2024.e29802

Godmalin, R. A., Aliac, C. J., & Feliscuzo, L. (2022). Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm. *4th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2022*. https://doi.org/10.1109/IICAIET55139.2022.9936817

Godmalin, R. A., Aliac, C. J., & Feliscuzo, L. (2023). Cacao Pod Infection Level Classification Using Transfer Learning. *2023 IEEE Open Conference of Electrical, Electronic and Information Sciences, EStream 2023 - Proceedings*. https://doi.org/10.1109/eStream59056.2023.10135062

Hamida, H. (2024). *10 Negara Produsen Kakao Terbesar Di Dunia*. Tempo.Co. https://www.tempo.co/ekonomi/10-negara-produsen-kakao-terbesar-di-dunia-39695

Harvyanti, A. F. M., Baihaki, R. I., Dafik, Ridlo, Z. R., & Agustin, I. H. (2023). *Application of Convolutional Neural Network for Identifying Cocoa Leaf Disease* (Vol. 2). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-174-6_21

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15249
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Hortinela, C. C., & Tupas, K. J. R. (2022). Classification of Cacao Beans Based on their External Physical Features Using Convolutional Neural Network. *2022 IEEE Region 10 Symposium, TENSYMP 2022*. https://doi.org/10.1109/TENSYMP54529.2022.9864337

Indotama, P. F. (2022). *Indonesia Masuk Daftar 7 Negara Penghasil Kakao Terbesar di Dunia*. Freyabadi.Com. https://www.freyabadi.com/id/blog/indonesia-masuk-daftar-7-negara-penghasil-kakao-terbesar-di-dunia

Jesse, A., Douha, N. Y.-R., & Lenka, P. (2024). Image Classification for CSSVD Detection in Cacao Plants. *ArXiv*, 1–5. http://arxiv.org/abs/2405.04535

Jeswani, J., Ahmed, A. S., Zaid, K., & Fernandes, R. (2023). *Leaf Disease Detection Using Deep Learning*. Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-136-4_87

Karthikeyan, S., Charan, R., Narayanan, S., & Jani Anbarasi, L. (2025). Enhanced plant disease classification with attention-based convolutional neural network using squeeze and excitation mechanism. *Frontiers in Artificial Intelligence*, *8*(August). https://doi.org/10.3389/frai.2025.1640549

Kouassi, K. S., Diarra, M., Edi, K. H., & Jean-Claude, K. B. (2025). Detection of cocoa pod diseases using a hybrid feature extractor combining CNN and vision transformer with dual classifier. *Edelweiss Applied Science and Technology*, *9*(1), 668–681. https://doi.org/10.55214/25768484.V9I1.4209

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *Proceedings of the IEEE International Conference on Computer Vision*, 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

Mahadevan, K., Punitha, A., & Suresh, J. (2024). Automatic recognition of Rice Plant leaf diseases detection using deep neural network with improved threshold neural network. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *8*(February), 100534. https://doi.org/10.1016/j.prime.2024.100534

Mamadou, D., Ayikpa, K. J., Ballo, A. B., & Kouassi, B. M. (2023). Cocoa Pods Diseases Detection by MobileNet Confluence and Classification Algorithms. *International Journal of Advanced Computer Science and Applications*, *14*(9), 344–352. https://doi.org/10.14569/IJACSA.2023.0140937

Miracle, A. (2024). Enhancing Cocoa Crop Resilience in Ghana: The Application of Convolutional Neural Networks for Early Detection of Disease and Pest Infestations. *Qeios*, 1–14. https://doi.org/10.32388/dps5zh

Moore, S. E., & Modupe, A. (2025). Cacao Plant Disease Detection and Classification. *Research Square*, *January*. https://doi.org/10.21203/rs.3.rs-5763786/v1

Nikith, B. V., Keerthan, N. K. S., Praneeth, M. S., & Amrita, D. T. (2022). Leaf Disease Detection and Classification. *Procedia Computer Science*, *218*, 291–300. https://doi.org/10.1016/j.procs.2023.01.011

Robet, Juliandy, C., Andi, Hendri, Hendrik, J., & Tarigan, F. A. (2022). Image Road Surface Classification Based on GLCM Feature Using LGBM Classifier. *IOP Conference Series: Earth and Environmental Science*, *1083*(1). https://doi.org/10.1088/1755-1315/1083/1/012006

Robet, R., Terang, J., Perangin, K., & Pribadi, O. (2024). Implementation of Deep Learning Model for Classification of Household Trash Image. *Sinkron*, *8*(October), 2575–2583.

Robet, R., Terang, J., Perangin, K., & Wijaya, E. (2025). Improving Resnet Model in Safety Gear Classification using Finest Optimizer. *Journal of Artificial Intelligence and Engineering Applications*, *4*(2).

Rola, J. B., Barrera, J. J. A., Calhoun, M. V., Maaghop, J. F. O., Unajan, M. C., Boncalon, J. M., Sebios, E. T., & Espinosa, J. S. (2024). Convolutional Neural Network Model for Cacao Phytophthora Palmivora Disease Recognition. *International Journal of Advanced Computer Science and Applications*, *15*(8), 986–990. https://doi.org/10.14569/IJACSA.2024.0150897

Sing Soh, K., Gubin Moung, E., John Julius Danker, K., Dargham, J. A., & Farzamnia, A. (2024). Cocoa Diseases Classification using Deep Learning Algorithm. *ITM Web of Conferences*, *63*, 01014. https://doi.org/10.1051/itmconf/20246301014

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. *Proceedings of Machine Learning Research*, *139*, 10347–10357.

Varma, T., Mate, P., Azeem, N. A., Sharma, S., & Singh, B. (2025). Automatic mango leaf disease detection using different transfer learning models. *Multimedia Tools and Applications*, *84*(11), 9185–9218. https://doi.org/10.1007/s11042-024-19265-x

Vera, D. B., Oviedo, B., Casanova, W. C., & Zambrano-Vega, C. (2024). Deep Learning-Based Computational Model for Disease Identification in Cocoa Pods (Theobroma cacao L.). *ArXiv*, 1–16. https://doi.org/10.1109/PIC62406.2024.10892748

Yang, W., Yuan, Y., Zhang, D., Zheng, L., & Nie, F. (2024). An Effective Image Classification Method for Plant Diseases with Improved Channel Attention Mechanism aECAnet Based on Deep Learning. *Symmetry*, *16*(4). https://doi.org/10.3390/sym16040451

Yarak, K., Witayangkurn, A., Kritiyutanont, K., Arunplod, C., & Shibasaki, R. (2021). Oil palm tree detection and health classification on high-resolution imagery using deep learning. *Agriculture (Switzerland)*, *11*(2), 1–17. https://doi.org/10.3390/agriculture11020183

*name of corresponding author