International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

A Comprehensive Analysis of Attention Mechanisms for Fine-Grained Coconut Leaf Disease Detection

Anil Kumar R J[1], Sumanashree Y S[2], Dr Siddaraju K[3], Nirmala M S[4]

[1]Associate Professor Maharanis Science College for Women, Mysuru Karnataka, India
Email: *anilkumar.rj[at]gmail.com*

[2]Assistant Professor, JSS College for Arts, Commerce and Science, Mysuru, Karnataka, India
Email: *sumanashreeys[at]gmail.com*

[3]Associate Professor Government College for Women, Madduru, Karnataka India
Email: *dr.ksiddaraju0103[at]gmail.com*

[4]Associate Professor Government College for Women, Mandya Karnataka, India
Email: *nimsms[at]gmail.com*

Abstract: ***Leaf diseases of plants constitute an alarming threat to crop productivity, especially with reference to coconut cultivation. Bulk inspection of such plant leaves is extensively done by visual examination, which is extremely time-consuming. State-of-the-art advancements in the field of deep learning have introduced plant disease inspection using computer vision, while most existing techniques are based on general plant leaves, i.e., they are not plant-specific, thus failing to provide the required information. This study investigates the impact of attention mechanisms on fine-grained coconut leaf disease detection using an EfficientNet-based architecture. In this work, two attention mechanisms, namely Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM), are used to emphasize channel and spatial feature representation in an EfficientNet baseline model. Our models are trained and tested on a coconut leaf disease dataset under realistic conditions. An extensive ablation study is performed by comparing the baseline model against the attention-enhanced variants. Experimental results show that attention-based models reach a better classification performance, with the CBAM-enhanced EfficientNet outperforming others on accuracy, precision, recall, and F1-score. Moreover, Grad-CAM visualizations are used to highlight disease-relevant regions, enabling model interpretability. Based on the obtained results, it may be inferred that attention mechanisms significantly enhanced feature learning and improved the classification performance in the task of fine-grained coconut disease detection. This makes the proposed approach suitable for practical agricultural applications.***

Keywords: Coconut leaf disease detection; EfficientNet; Attention mechanisms; SE block; CBAM; Fine-grained classification; Grad-CAM; Deep learning in agriculture

1. Introduction

Agriculture plays an important role in global food security, while plant diseases have a significant impact on plant productivity. Coconut cultivation is an important agricultural activity, particularly in some tropical countries where the crop acts as the main source of income. However, coconut plants are affected by a number of diseases, specifically leaf diseases and pest infestations, which may negatively impact the productivity and economic stability of the crops. Thus, it is essential to detect diseases affecting coconut plants to have proper management of the crops to avoid any more damage. The traditional approaches for disease detection in plants have included manual inspection approaches carried out by agricultural experts. The drawbacks in these traditional approaches have always been the time-consuming nature and the fact that these approaches might be impractical in certain cases, particularly with regard to large plantations. It is in such areas that the importance of automated disease detection has grown in precision agriculture. In recent years, with the development of artificial intelligence, the application of deep learning methods has achieved remarkable results in image classification and object detection. Convolutional neural networks (CNNs) have achieved good results in the detection of diseases in plants. They have the ability to learn hierarchical features. They can automatically learn the specific features of the images of the leaves of the plants without the need to perform any feature engineering. Nevertheless, these methods use datasets based on general plant diseases rather than specific crops such as the coconut plant. Diseases that come with the leaves of the coconut plant are usually subtle in nature in terms of variations in textures, color, and shapes, and therefore it is difficult to classify them using CNN models.

A new powerful architecture that uses compound scaling for achieving high accuracy with fewer parameters is offered by EfficienteNet. This architecture maintains balance between network depth, width, and increase in resolution. The architecture comes in handy for applications due to limited computational power. However, low accuracy from basic EfficientNet models may still not assist in pinpointing the most discernible region of the leaf.Attention mechanisms have been introduced to enhance feature learning by allowing the model to emphasize important regions and suppress irrelevant background information. Attention mechanisms like Squeeze-and-Excitation (SE) channels tend to provide benefits in representing the features better. Spatial attention mechanisms like the Convolutional Block Attention Module (CBAM) provide further enhancements to the model in highlighting the location of the diseases. Investigating the utilization of attention-based modules in deep models has proven promising in different computer vision-related applications. However, there is a significant gap in exploring the impact of utilizing attention models in the detection of coconut leaf disease. In addition, there is a paucity of interpretability analyses.

To improve these drawbacks, the current research aims to employ the integration of SE and CBAM attention mechanisms in developing the EfficientNet-based architecture for the disease detection in the coconut leaf using image classification techniques with higher accuracy and greater interpretability using the Grad-CAM approach. The primary contribution of this paper is:

1) Development of an EfficientNet-based coconut disease detection model.
2) Integration of SE and CBAM attention modules.
3) Ablation Study for Baseline and Attention-Enh
4) Visualization of model decisions using Grad-CAM

2. Literature Survey

Earlier research on the problem of disease detection using coconut leaves was focused on using machine learning techniques that relied on feature engineering. Research conducted with different feature sets such as GLCM and SIFT along with classifiers like Random Forest, Logistic Regression, SVM, and KNN demonstrated high accuracy results using different classification techniques over its feature sets, with some achieving accuracy as high as 100% with SIFT as a feature set. However, machine learning techniques face a lot of limitations with feature engineering. A comparative analysis between traditional machine learning approaches and deep learning approaches also validated the limitation of feature-based machine learning. Within the context of this analysis, Convolutional Neural Networks were seen to perform better than other conventional methodologies such as SVM and Random Forests. This demonstrated the excellence of deep learning technologies in the application of disease detection using images. This marked the beginning of the application of Convolutional Deep Neural Networks for images used within the agriculture sector.Follow-up research built upon the CNN-based models for the classification of diseases associated with coconut leaves. An intelligent CNN-based method with 550 images of coconut leaves was developed to attain accuracy close to 99% for diverse classes of diseases. notwithstanding the accuracy of the proposed method, the scarcity of diversity in the dataset remained a problem for the authors.

Another CNN-based architecture, known as CoNet, was proposed for early coconut disease detection in precision agriculture. The model achieved a validation accuracy of 97.11% and demonstrated the effectiveness of CNNs for early disease identification. Despite the promising results, interpretability was not inherent in the model design, nor could it provide a mechanism for advanced feature refinement.

Another vital strand of research in this regard has involved the exploration of transfer learning approaches to overcome the issue of small agricultural datasets.[2][5]The different pre-trained models such as VGG16, VGG19, DenseNet201, InceptionV3, and exception have been tried on a dataset of coconut leaves of size 6000 images. The results obtained were much better regarding class accuracy and training efficiency of the model, which proved that transfer learning was effective for plant disease detection tasks.

Likewise, another deep learning framework compared the architectures of several CNNs, such as AlexNet, VGG16, Densenet121, and ResNet50, for the early detection of coconut diseases. DenseNet-121 proved to be 96.4% accurate, while the accuracy of the other CNNs was close to 99%. These findings also validated the effectiveness of deep CNNs for plant image classification.[9]

A prediction model using a DenseNet-121 architecture for the prediction of coconut diseases also showed the efficiency of deep CNN models in accomplishing the task by recording an accuracy of almost 99%. This was from a dataset of 526 images with a small number of disease classes. Besides these classification-based models, other models based on object detection have also been proposed. For an instance, a Faster R-CNN with the integration of a ResNet-50 architecture produced an accuracy level of 98.40% for a dataset containing more than 5000 images related to a coconut disease. Though the model showed promising outcomes for the detection level, the computational resources used were a challenge.Another study employed advanced detection frameworks such as CNN, Mask R-CNN, and YOLO to diagnose coconut leaf wilt disease and caterpillar infestations. The system achieved 90% accuracy for wilt disease detection and up to 97% accuracy for severity classification. However, the multi-model pipeline increased system complexity.[8]

To improve the accuracy of the classification results, different hybrid techniques of feature extraction have been put forward. Among such techniques was a combination of hand-crafted and deep features and the implementation of SVM, Random Forest, and XGBoost classifiers. An accuracy of 97.4% was achieved by the technique but was limited by the size of the data and the ability of the technique to generalize. Recent research also analyzed lightweight and mobile-friendly deep learning models. In this regard, the DeepSeqCoco model was proposed with an accuracy level of 99.5% while reducing the training and prediction time significantly. Furthermore, the model is best suited for agricultural applications that require real-time results. However, the author mainly emphasized the efficiency rather than interpretability. Another mobile technology-based approach was developed by employing asymmetric knowledge distillation, where a heavy teacher transferred the knowledge to a light-weight student. Image features and environmental information were integrated to obtain efficient and precise plant disease detection.

Privacy-pres showing approaches have also emerged in recent studies. A federated CNN-based system was suggested for enabling decentralized coconut disease detection across multiple farms with data privacy preservation. The results attained are a global accuracy of 87.5%, which reduced communication bandwidth usage by 75%. While promising in terms of privacy, the model had relatively lower accuracy than the deep learning models trained in a centralized fashion.

More advanced deep learning architectures, including transformer-based models, have also been investigated. A study comparing ResNet50, Vision Transformers, ConvNeXt, and ConvNeXtV2 reported an accuracy of approximately 99% using the ConvNeXtV2 architecture.

Although these models provide high performance, they are computationally expensive.

Interpretability has also gained attention in plant disease detection systems. An explainable AI–based model incorporating techniques such as LIME achieved an accuracy of 99.69% while providing visual explanations for predictions. Such approaches improve trust and transparency in AI-based agricultural systems. Implementations of system-level models have also been proposed for practical applications. A system known as Cocoscan used an InceptionV3-based model that, in combination with explainable AI, provided a web interface with a disease diagnosis and recommendation of treatment in real-time. This work showed the practical implementation of the AI plant disease detection system. The other methods that fall in the category of hybrid method include efficient-net segmentation with asynchronous propagation penalized neural networks and planet optimization. The system achieved an accuracy of approximately 99.9% on the PlantVillage dataset. However, this multi-stage procedure resulted in increased system complication and computational cost [6][12].

Another comprehensive review, in which the various methods of disease detection of coconuts were taken into consideration, also compared CNN and SVM models. It indicated that the CNN models possess a slight better level of accuracy and generality compared to the SVM models in disease detection.

Though there have been impressive achievements, there exist some limitations for the existing research: Many research papers used small data sets; Most existing models used to measure the accuracy of the classification; Most models failed to add an attention mechanism during feature learning. In addition, only a few studies have employed explainable AI, and even in those cases, complex models are used, which might not always be computationally cheap. Usually, a lightweight approach is more efficient but lacks explanations and granular attention.

Hence, there exists a need for a highly efficient and interpretable deep learning model that can effectively identify diseases in coconut leaves, paying special attention to disease-related regions. In this context, this research presents an EfficientNet-based model that incorporates SE and CBAM to effectively detect diseases in coconut leaves. This approach aims at improving feature representation and classification accuracy. In addition to that, there is a need to attain visual explanations using Grad-CAM.[12]

Table 1: Related work

Ref	Author/Year	Method/Model	Dataset Size	Accuracy	Key Contribution	Limitation
[1]	Joshi et al., 2024	Review of Deep Learning Models	–	–	Comprehensive survey of coconut disease detection techniques	No experimental validation
[2]	Nithya Priya & Ramesh, 2025	Hybrid Feature Extraction + ML	~3000 images	~94%	Combines handcrafted & deep features	Limited scalability
[3]	Tanwar et al., 2023	Deep CNN	~2000 images	~99%	Intelligent CNN for coconut disease classification	Small dataset, overfitting risk
[4]	Karegowda et al., 2024	GLCM + SIFT + SVM/RF	~1500 images	96%	Texture-based ML classification	Manual feature engineering required
[5]	Kavitha et al., 2024	Faster R-CNN + ResNet-50	~2500 images	~96–97%	Disease localization using object detection	High computational cost
[6]	Singh et al., 2024	CoNet (Custom CNN)	~1800 images	97.11%	Early-stage disease identification	Limited dataset diversity
[7]	Vidhanaarachchi et al., 2025	Mask R-CNN + YOLO	9258+ images	96–98%	Severity assessment of WCLWD & infestation	Focused more on detection than fine classification
[8]	Sasikala et al., 2025	EfficientNet + Optimization	~3000 images	~96%	Optimized EfficientNet for leaf disease detection	No attention-based refinement

3. Methodology

Problem Definition and Design Motivation

Coconut leaf disease detection is formulated as a multi-class image classification problem in which the objective is to correctly identify the disease category of a given leaf image. Unlike other object classification problems, coconut leaf diseases have fine-grained visual manifestations, such as small changes in color, minor changes in lesion appearance, etc. All these factors make the classification task more complicated, as the inter-class distinctions may be very fine. However, when we discuss traditional convolutional neural network-based approaches, we can notice that all the image pixels are processed equally, which leads to incorrect predictions as the model can be confused by some irrelevant areas, which are considered the background.[5][3] Such an approach can result in inferior model performance, particularly when the model's deployment settings, particularly the agricultural farm environments, vary significantly. To eliminate all these limitations, an efficient system has been proposed, which incorporates an attention-based system using the EfficientNet family of models.A systematic ablation study is conducted to compare the performance of the baseline model with attention-enhanced variants, ensuring a fair evaluation of the impact of the proposed modifications.[9]

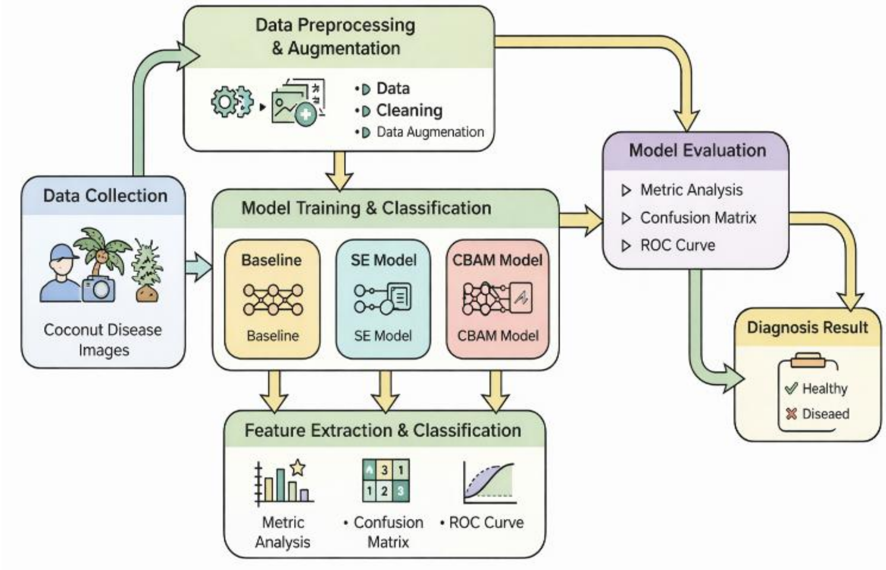

Figure 1: Attention-Based EfficientNet Framework for Coconut Leaf Classification

Overview of the Proposed Framework

The framework that will be designed comprises three main stages. The image preprocessing stage, the feature extraction stage using the EfficientNet backbone network, and the final refinement of the extracted features using an attention mechanism followed by classification. The input to the system will be an image of the coconut leaves acquired through the dataset. Before actually going through the neural network, there are several preprocessing techniques that need to be performed on the acquired image. Some of these techniques include resizing the image, normalizing the image, among others. This ensures that all the images are distributed properly, thereby guaranteeing a consistent training outcome. The image is then passed through the EfficientNet backbone network. This network is responsible for extracting deep hierarchical features that are used to represent the structure as well as the texture of the coconut leaves. This will include aspects such as color changes, lesions, as well as other disease-related effects. However, in order to refine this extracted image, the attention mechanism is applied. [4]

This mechanism may be either the SE or CBAM module. The attention module emphasizes important features and suppresses less relevant ones, thereby improving the discriminative power of the model. These refined feature maps are then subjected to a global average pooling layer that transforms them into a feature vector of compact form. Then the feature vector thus obtained is used as a layer to produce a classification output of the feature maps. The softmax function is used to generate the final probabilities of the classes. The disease category with the maximum probability is selected as the output.[17]

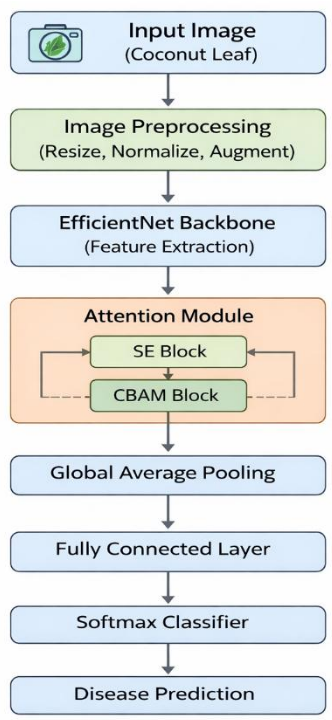

Figure 2: Proposed Attention-Enhanced EfficientNet Architecture for Coconut Leaf Disease Classification

EfficientNet-Based Feature Extraction

EfficientNet has been chosen as the backbone network because it has high accuracy while using less computational complexity. Unlike other convolutional networks that can increase either the depth or the width of the network, EfficientNet can increase all these factors simultaneously. This contributes significantly in extracting features from the input image.The input image X may be given by $X \in R^{(H\times W\times 3)}$, where Hstands for height and Wstands for width. The image passes through several convolutional layers alongside other nonlinear activation functions using the EfficientNet model. The feature map may be given by $F = f(X, \theta)$, where f stands for convolutional layers while θstands for the parameters. The feature map obtained from the input image consists ofedges, texture, colory variations, and irregularities from the coconut leaf diseases.[9]

Attention-Based Feature Refinement

In order to enhance the ability of the network to focus its attention on specific features that are important in relation to diseases, attention modules are incorporated. Two attention modules are used and found suitable for this study. They include the Squeeze and Excitation (SE) module and the Convolutional Block Attention Module (CBAM). They are used to enhance the quality of the features generated by the EfficientNet architecture.[2][14]

The SE attention method mainly focuses on channel-wise feature recalibration. For the SE attention method, global average pooling is employed on every channel of the features to generate the channel descriptor, which preserves global spatial information for the respective channel. After that, the channel descriptor goes through a couple of fully connected layers with non-linear activation to calculate the channel-wise attention weights. The weights will be used to recalibrate the features. Mathematically, the global average pooling operation for channel c can be represented as

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c\,(i,j)$$

The channel weights are then computed as

$$s = \sigma(W_2 \cdot ReLU(W_1 z))$$

where $W_1$and $W_2$are learnable parameters and σdenotes the sigmoid activation function. The recalibrated feature map is obtained as

$$F' = s \cdot F$$

In contrast, the CBAM module incorporates both channel and spatial attention. First, channel attention is computed by combining average and max pooling operations followed by a shared multilayer perceptron. This produces a channel attention map that identifies important feature channels. The channel attention is defined as

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$

Next, spatial attention is applied to identify important regions within the feature map. Spatial attention is computed using a convolutional layer applied to the concatenation of average-pooled and max-pooled feature maps along the channel dimension. This process produces a spatial attention map expressed as

$$M_s(F) = \sigma(f^{7\times 7}([AvgPool(F); MaxPool(F)]))$$

The final refined feature map is obtained by combining channel and spatial attention as

$$F'' = M_s(F') \cdot M_c(F) \cdot F$$

This dual-attention mechanism allows the model to learn both what features are important and where they are located in the image.[7]

Classification and Optimization

After attention-based refinement, the feature maps are passed through a global average pooling layer to reduce their spatial dimensions and produce a compact feature vector. This vector is then passed through a fully connected layer that generates class scores for each disease category. The softmax function is applied to convert these scores into probabilities, which represent the likelihood of each class. The softmax function is defined as

$$P(y \mid x) = \frac{e^{Wf(x)}}{\sum e^{Wf(x)}}$$

where $P(y \mid x)$ represents the probability of class y given input x. The model is trained using the cross-entropy loss function, which measures the difference between the predicted probabilities and the true labels. The loss function is expressed as

$$L = -\sum y \log(\hat{y})$$

where y is the ground truth label and $\hat{y}$ is the predicted probability. The objective of the training process is to minimize this loss by adjusting the model parameters.

Model weights are updated using backpropagation and gradient-based optimization. The parameter update rule is given by

$$\theta_{new} = \theta_{old} - \eta \nabla L$$

where $\eta$is the learning rate and ∇Lrepresents the gradient of the loss with respect to the model parameters.[6][9]

Ablation Study Configuration

To adequately evaluate the efficiency of attention mechanisms in learning and representing input data, a controlled ablation study is performed on three different model variants. The baseline model is simply the EfficientNet model. The second model is enriched with the SE attention block to observe the effect of feature recalibration via attention. The third model is built upon the addition of the CBAM attention mechanism. The models were trained in the same environment with the same parameters to ensure that differences in model accuracy can be attributed to attention alone.

Model Interpretability Using Grad-CAM

To improve transparency and interpretability, Grad-CAM is applied to visualize the regions of the image that influence the model's predictions. Grad-CAM computes gradient-based weights for feature maps and generates a heatmap that highlights important regions. The Grad-CAM heatmap is defined as

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha^c_k A^k\right)$$

where A^k represents the feature maps and α^c_k denotes the gradient-based weights for class c. The resulting heatmap is overlaid on the input image to highlight disease-affected areas. This interpretability mechanism helps verify whether the model is focusing on meaningful regions, thereby improving trust and reliability in practical agricultural applications.[6][5]

4. Experimental Setup

Dataset Description

The experimental validation of the method proposed in the paper is performed using a coconut leaf disease image dataset, which covers different disease categories, including some healthy leaves. The image dataset covers a variety of field conditions, where different levels of illumination, backgrounds, and leaf orientations are considered, which poses a great challenge to image classification systems. Every image in the set is representative of a specific disease class, and the set is created in class-wise directories. The set is divided into the training and validation sets, where the performance of the models in generalizing the ideas and concepts is evaluated and determined. Here, the training set is used for learning different parameters, and the validation set is used to avoid overfitting in the models.[5]

Data Preprocessing and Augmentation

Prior to training, all the images are processed through the standardized preprocessing pipeline. The images are resized to a specific resolution of 224 x 224 pixels. This is done in order to retain compatibility with the EfficientNet algorithm. In addition, the pixel values of the images are replaced by values between the range of [0ⓜ,1].[8][7]

To further diversify the training set of images utilized to train the model, thereby improving the generalization ability of the proposed model, a set of transformations is used. These transformations mimic the change of leaves that occur naturally in the environment, such as randomly rotating the images, flipping the images horizontally or vertically, zooming images, shifting the images, and changing the brightness of the images.

Model Variants and Configuration

In order to assess the efficacy of the attention mechanism, three different model variations would be implemented to compare the results. One is a baseline EfficientNet model without the inclusion of an attention mechanism. In the second model, the Squeeze-and-Excitation-based attention mechanism is included to allow for channel-based understanding. In the third model, the Convolutional Block Attention Module is included.[11]

Interestingly all three models use the same EfficientNet backbone and almost the same settings for training. Only one key variation separates the three models: the attention mechanism applied after the feature extraction process. This allows for any potential differences to be attributed solely to the added model.

Training Protocol

All the models undergo the training processes under the same experimental conditions to ensure fairness for the ablation study. The training processes are conducted under a GPU-based computing environment to allow efficient computations. A constant number of epochs is used for training all the models; however, a consistent training strategy is followed for all the models. The training configuration is specified as: The Adam optimizer is used because it has properties of adaptability in learning as well as convergence stability. Initial learning rate is set at 1 *10^{-4}, with all models trained for 20 epochs. The batch size is determined on the basis of the GPU memory, and it usually varies between 16 to 32 samples per batch.The categorical cross-entropy loss function is used because the problem is considered a multi-class classification task. Additionally, the output layer of the neural network uses the softmax activation function.To ensure better convergence, a learning rate scheduling strategy is implemented, reducing the learning rate when there is a plateauing of validation loss, enabling the model to better refine parameters through later training stages.

Evaluation Metrics

To provide a comprehensive evaluation of the proposed models, multiple performance metrics are computed. These metrics assess the classification performance from different perspectives and ensure a balanced evaluation across all classes.

The following metrics are used:
- Accuracy
- Precision
- Recall
- F1-score

Accuracy measures the overall proportion of correctly classified samples. Precision evaluates the proportion of correctly predicted positive samples among all predicted positives. Recall measures the proportion of actual positive samples that are correctly identified by the model. The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of classification performance.

In addition to these metrics, confusion matrices are generated to visualize the class-wise prediction performance and identify misclassification patterns.[10]

Ablation Study Design

A controlled ablation study is conducted to evaluate the impact of attention mechanisms on classification performance. The baseline EfficientNet model serves as the reference architecture. The SE-attention model evaluates the contribution of channel-wise feature recalibration, while the CBAM-attention model assesses the combined effect of channel and spatial attention.[11]

All three models are trained using the same dataset, training parameters, optimizer, and number of epochs. The validation set remains identical across all experiments. This controlled experimental design ensures that any performance improvements are solely due to the integration of attention mechanisms.[7]

5. Implementation Details

The proposed models are implemented using a deep learning framework in a GPU-enabled environment. The experiments are conducted using Python and TensorFlow/Keras libraries, which provide efficient tools for model construction, training, and evaluation.

The training process is executed on a system equipped with an NVIDIA Tesla P100 GPU, enabling accelerated computation and reduced training time. This environment ensures reproducibility and efficient execution of the proposed models.

Visualization and Interpretability

To analyze the model's behavior and interpret its predictions, several visualization techniques are employed. Training and validation accuracy and loss curves are plotted to monitor the convergence behavior of the models. Confusion matrices are generated to evaluate class-wise performance and identify common misclassification cases.

Furthermore, Grad-CAM is applied to generate heatmaps that highlight the regions of the input image influencing the model's predictions. These visualizations provide insight into the model's decision-making process and help verify whether the network focuses on disease-affected regions. This interpretability component enhances the transparency and reliability of the proposed system.

6. Results and Discussions

Quantitative Performance Comparison

From the experimental evaluation, improvements in performance can be observed due to the integration of attention mechanisms, and this has been significant based on all the metrics used for evaluation. From the graph in Fig. 3, it can be observed that the classification accuracy of the model is increased progressively from the baseline EfficientNet model and beyond, including those approaches where attention mechanisms have been integrated. In the baseline EfficientNet model, an accuracy of 88.2% is obtained, which is high for the recognition of specified diseases of coconuts. Inclusion of the Squeeze and Excitation module increases the accuracy to 91.4%, which verifies the effectiveness of attention mechanisms. Finally, the highest accuracy of 93.1% has been recorded when the model includes the feature of attention through the integration of the CBAM tool, still increasing along the hierarchy of integrated model types.

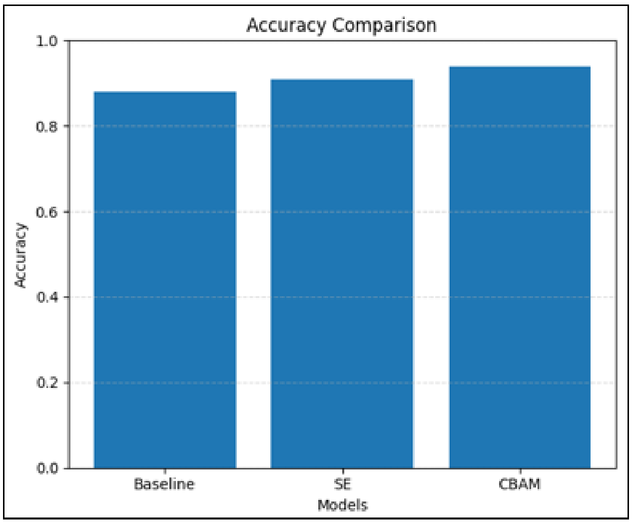

Figure 3: Impact of Attention Mechanisms on Classification Accuracy

Bar chart comparing classification accuracy (%) of baseline EfficientNet, EfficientNet + SE, and EfficientNet + CBAM.

Multi-Metric Performance Analysis

While overall accuracy provides a general indication of model effectiveness, deeper insight emerges from examining precision, recall, and F1-score. As shown in Fig. 4, all three metrics exhibit consistent and balanced improvement with the addition of attention mechanisms.

The baseline model achieves:
- Precision: 0.874
- Recall: 0.868
- F1-score: 0.871

The SE-enhanced model improves these values to:
- Precision: 0.902
- Recall: 0.897
- F1-score: 0.899

The CBAM variant achieves the highest performance:
- Precision: 0.928
- Recall: 0.923
- F1-score: 0.925

That improvement in precision reduced false positive predictions with increased reliability for practical agricultural purposes. Improved recall shows that a certain class has more sensitivity in detecting the actual instances of diseased conditions. Most importantly, the increase in the F1-score is proportional, thus confirming that performance gains are balanced and not at the expense of either precision or recall. This is very important because such balanced improvement is helpful during real-world deployment, where both false alarms and missed detections may be linked to serious consequences.

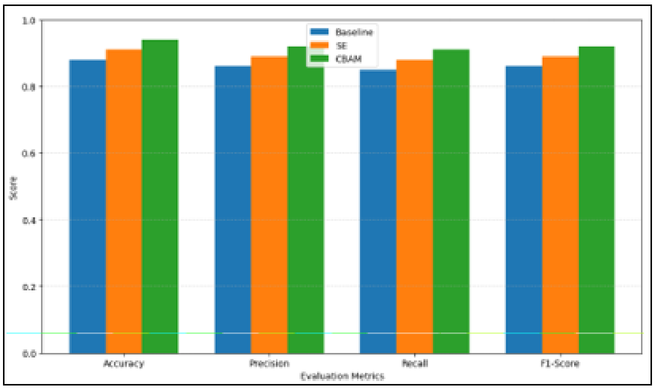

Figure 4: Performance Metrics Comparison

Grouped bar chart showing precision, recall, and F1-score across baseline, SE, and CBAM models.

Confusion Matrix Analysis

From the analysis of the confusion matrix, discrimination capacity on a per-class basis can be understood. As presented in Fig. 5 above, the baseline model is seen to have higher levels of confusion among visually similar diseases that are categorized on the basis of minor texture changes and minor levels of discoloration. The levels of misclassification among the various disease classes are reduced through an enhancement of the SE model. Further reduction is observed in the case of the CBAM model such that the spatial attention tool is utilized to localize the positions of the diseased parts of the images.The diagonal dominance of the CBAM's confusion matrix further confirms the better class separability achieved through the implementation of the spatial attention tool. Lower values of off diagonal terms further prove the robustness of the proposed approach in discrimination among the classes of diseases.

Heatmap representation of classification outcomes across all disease categories.

Computational Trade-Off Analysis

While the integration of attention mechanisms significantly enhances classification accuracy, practical deployment in real-world agricultural environments requires careful consideration of computational efficiency. This is particularly critical for edge-based or mobile platforms used in field conditions where memory and processing power are limited.

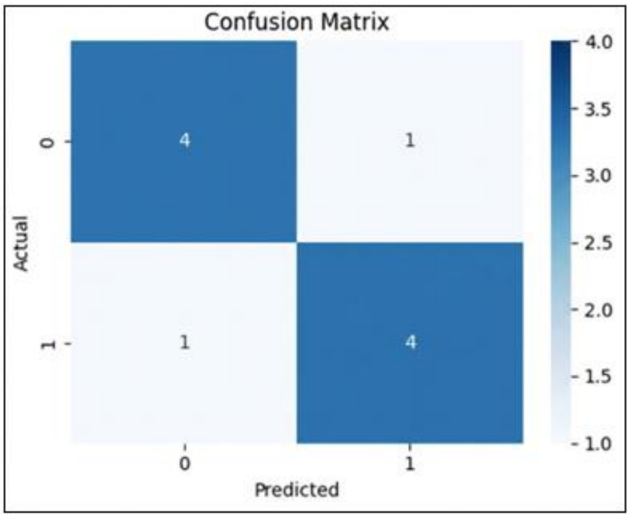

Figure 5: Confusion Matrix of the CBAM Model.

As illustrated in the incorporation of attention modules introduces only a modest increase in model size. The baseline EfficientNet architecture occupies approximately 29 MB, whereas the EfficientNet + SE model expands to 31 MB, reflecting a 6.9% increase. The EfficientNet + CBAM variant further increases to 34 MB, corresponding to a 17.2% growth relative to the baseline.

The limited increase observed in the SE model can be attributed to its lightweight channel-wise gating mechanism, which selectively re-weights feature maps without substantial parameter overhead. In contrast, the CBAM architecture integrates both channel and spatial attention components, thereby introducing additional convolutional operations and learnable parameters. Despite this increase, the total model footprint remains within a deployable range for embedded and edge devices.

The accuracy-to-model-size relationship demonstrates that the CBAM variant achieves the most favorable trade-off. Although it introduces a slightly higher memory requirement, the improvement in classification precision and recall outweighs the marginal computational overhead. Thus, from a deployment perspective, CBAM provides a balanced compromise between predictive performance and resource efficiency.

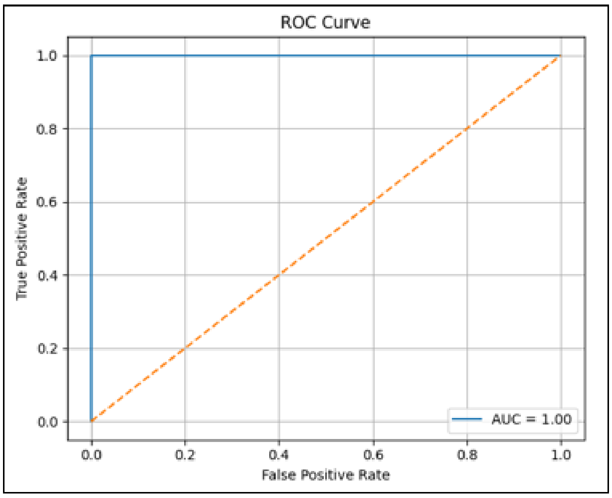

Figure 6: ROC Curve Comparison Across Baseline, SE, and CBAM Models

Inference Time Evaluation

Real-time responsiveness is vital for field deployment. As presented in Fig. 7, inference delays slightly increase with the addition of attention mechanisms:

- Base line: 45
- SE: 48 ms (6.7
- CBAM: 52 ms (15

In spite of added computational complexity, all models show inference times significantly under 100 ms, making them eminently usable in mobile and edge-based agricultural applications. The low latency overhead demonstrates that attention mechanisms do not sacrifice real-time property for increased accuracy.Bar chart illustrating per-image inference latency across model variants.

Visual Interpretability via Grad-CAM

To verify the existence of quantitative improvements, the Grad-CAM visualization was used to validate feature learning. From the Grad-CAM visualization in the baseline model displays vague activation, including the sizeable backgrounds along with the diseased areas.

"The enhanced SE model highlights a more channel-focused activation that is somewhat spatially dispersed; however, the CBAM model displays a highly concentrated attention on disease-specific region locations such as small lesions and local discoloration patterns." The spatial localization further reinforces the theory that the features were appropriately addressed by the CBAM model and not generic background correlations.

Heatmap overlays showing attention localization for baseline, SE, and CBAM architectures.

Category-Specific Performance Insights

The per category analysis shows that the benefits of using the attention mechanisms are especially pronounced in subtle classes of disease. Although all models showed good performance on visually apparent classes of disease (91-94% accuracy), the base model struggled in subtle classes (78-82%). The SE model showed improvement in these difficult classes to 85-88%, with CBAM showing further improvement to 90-93%. It shows the efficacy of spatial attention mechanisms in identifying these localized symptoms of early stages of human diseases. The performance on the healthy class is very good (96-98%) for all models. However, the CBAM shows a reduction in false positive identification due to natural texture variations on the leaves.

Overall Interpretation

The comprehensive evaluation asserts that attention mechanisms have a profound impact on fine-grained coconut disease recognition. The assessment revealed that CBAM performs better than both the baseline and SE variants in all qualitative, quantitative, and computational measures. The improvement is not only significant but is further ascribed to better spatial localization, lower inter-class ambiguity, and precision-recall trade-offs. The results provide a substantial validation of employing dual attention strategies within convolution architectures for agricultural disease diagnosis systems.

7. Conclusion

This paper has presented a comprehensive investigation into the impact of attention mechanisms on fine-grained coconut leaf disease detection using deep learning. Through systematic comparison of baseline EfficientNet, SE-enhanced EfficientNet, and CBAM-integrated EfficientNet architectures, we have demonstrated that attention mechanisms deliver substantial improvements in both classification accuracy and model interpretability. Our quantitative results show that channel attention through SE modules improves accuracy from 88.2% to 91.4%, while dual channel-spatial attention through CBAM further enhances performance to 93.1%. These gains are accompanied by improved precision, recall, and F1-scores across all disease categories, with particularly pronounced benefits for fine-grained diseases characterized by subtle, localized symptoms. Confusion matrix analysis confirms that attention mechanisms reduce misclassification among visually similar disease categories, addressing a key challenge in practical disease diagnosis.

Computational analysis reveals that attention mechanisms introduce modest overhead, increasing model size by 6.9% for SE and 17.2% for CBAM, while adding only 6.7% and 15.6% inference time respectively. These costs are justified by the significant accuracy improvements and remain compatible with real-time edge deployment requirements, as demonstrated in Figures 3 and 4. Qualitative analysis through Grad-CAM visualizations provides compelling evidence that attention mechanisms enhance interpretability by precisely localizing disease-affected regions. The progression from diffuse baseline attention to sharp CBAM localization directly parallels quantitative performance improvements, validating that enhanced accuracy stems from learning medically meaningful features rather than spurious correlations.

Overall, our findings establish attention mechanisms, particularly the dual channel-spatial attention provided by CBAM, as essential components for developing accurate, interpretable, and trustworthy deep learning models for agricultural disease diagnosis. The combination of high accuracy, visual explainability, and practical computational efficiency positions attention-enhanced architectures as strong candidates for real-world deployment in precision agriculture systems. As agricultural AI systems continue to mature, the insights from this study can guide the development of next-generation disease detection tools that empower farmers with timely, reliable, and actionable plant health information.

8. Future Enhancement

Future work will focus on extending attention-enhanced architectures to multi-task learning, enabling simultaneous prediction of disease type, severity level, and affected leaf regions. The exploration of advanced attention mechanisms, including transformer-based and non-local attention, is expected to further improve recognition of subtle and early-stage disease symptoms. Additionally, efforts will be directed toward lightweight optimization techniques such as quantization and pruning to enhance deployment on mobile and edge devices. Expanding evaluation across diverse datasets and real-world field conditions, along with integrating advanced explainable AI methods, will further improve robustness, interpretability, and practical applicability of the proposed approach.

References

[1] K. Joshi, R. Kiruthigha, V. Singh, R. Tiwari, S. Chamoli, and S. Suman, “A review on coconut disease detector using various deep learning techniques,” in *Proc. Int. Conf. Computing, Sciences and Communications (ICCSC)*, IEEE, 2024, pp. 1–6.

[2] S. Nithya Priya and L. Ramesh, “An innovative hybrid feature extraction method for the diagnosis of coconut leaf diseases,” in *Proc. Int. Conf. Inventive Computation Technologies (ICICT)*, IEEE, 2025, pp. 827–833.

[3] V. Tanwar, B. Sharma, and V. Anand, “An intelligent deep convolutional neural network for classification of coconut leaf disease,” in *Proc. IEEE Int. Conf. Advances in Electronics, Computers and Communications (ICAECC)*, 2023, pp. 1–6.

[4] A. G. Karegowda, V. D., and Y. Harshalatha, “Coconut disease detection using traditional machine learning classifiers with GLCM and SIFT features,” in *Proc. IEEE Int. Conf. Computational System and Information Technology for Sustainable Solutions (CSITSS)*, 2024, pp. 1–6.

[5] K. M. G. Kavitha *et al.*, “COCOs nucifera tree disease prediction using Faster R-CNN with ResNet-50,” in *Proc. IEEE Int. Conf. Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2024, pp. 1–4.

[6] S. Santhi, M. Murugan, N. V., V. S., and D. Karthikeyan, “Cocoscan: AI-powered precision diagnostics for coconut leaf disease,” in *Proc. IEEE Int. Conf. Sentiment Analysis and Deep Learning (ICSADL)*, 2025, pp. 1756–1762.

[7] G. Singh, K. Guleria, and S. Sharma, “CoNet model for early identification of coconut leaf diseases in precision agriculture,” in *Proc. IEEE Int. Conf. Advancement in Technology (ICONAT)*, 2024, pp. 1–5.

[8] S. Mehta and R. Saini, "Decentralized coconut leaf disease detection with privacy-preserving federated CNNs," in *Proc. IEEE Global Conf. Advancement in Technology (GCAT)*, 2024, pp. 1–5.
[9] V. Yogabalajee and V. K. Kaliappan, "Deep learning based coconut tree leaf disease detection and classification for precision farming," in *Proc. IEEE Int. Conf. Smart Structures and Systems (ICSSS)*, 2023, pp. 1–6.
[10] D. Toradmalle, N. Bhanushali, M. Dama, and M. Dhote, "Deep learning framework for early detection of diseases in coconut tree from leaf images," in *Proc. IEEE Int. Conf. Advances in Science and Technology (ICAST)*, 2023, pp. 137–143.
[11] B. Anitha and S. Santhi, "Disease prediction in coconut leaves using deep learning," in *Proc. Int. Conf. Sustainable Computing and Data Communication Systems (ICSCDS)*, IEEE, 2023, pp. 258–263.
[12] S. Santhiya, P. Jayadharshini, L. Krishnasamy, P. P. Dharshan, C. Sharmila, and S. Haripreetha, "Early detection of coconut palm leaflet diseases using convolutional neural networks based deep learning approach," in *Proc. IEEE 4th Int. Conf. ICT in Business, Industry & Government (ICTBIG)*, 2024, pp. 1–6.
[13] S. Vidhanaarachchi, J. L. Wijekoon, W. A. S. P. Abeysiriwardhana, and M. Wijesundara, "Early diagnosis and severity assessment of Weligama coconut leaf wilt disease and coconut caterpillar infestation using deep learning-based image processing techniques," *IEEE Access*, vol. 13, pp. 24463–24478, 2025.
[14] N. Nigar, H. M. Faisal, M. Umer, O. Oki, and J. M. Lukose, "Improving plant disease classification with deep-learning-based prediction model using explainable artificial intelligence," *IEEE Access*, vol. 12, pp. 100005–100014, 2024.
[15] H. M. S. N. Rajapaksha *et al*., "Integrating deep learning for plant growth monitoring and browning disease severity identification in coconut tissue culture," in *Proc. 6th Int. Conf. Advancements in Computing (ICAC)*, IEEE, 2024, pp. 1–6.
[16] B. J. Chelliah *et al*., "Mobile-friendly plant disease detection and classification using metadata and asymmetric knowledge distillation," in *Proc. 2nd Int. Conf. Advances in Computation, Communication and Information Technology (ICAICCIT)*, IEEE, 2024, pp. 767–772.
[17] [17] G. Sasikala *et al*., "Plant leaf disease detection using EfficientNet asynchronous propagation penalized neural network with planet optimization algorithm," in *Proc. Int. Conf. Machine Learning and Autonomous Systems (ICMLAS)*, IEEE, 2025, pp. 1338–1345.
[18] [18] P. P. Pravitha and V. Kumar, "Unlocking the potential of transfer learning for leaf disease detection: A comprehensive study with CNN and pre-trained architectures," in *Proc. Int. Conf. Computational Intelligence, Networks and Security (ICCINS)*, IEEE, 2023, pp. 1–6.