iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80484

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Attention-Enhanced EfficientNet with Grad-CAM Explainability for Multi-Class Tea Leaf Disease Classification

Akash Nath[1], Shrinjoy Das[2], Dalia Mukherjee[3]
[1, 2, 3]Department of Computer Science and Engineering, Assam University, Silchar, Assam, India

Abstract: Tea is among the most commercially significant crops in northeastern India, yet its leaves are highly prone to a range of fungal, bacterial, and pest-induced diseases that erode both yield and quality each season. While deep learning has shown strong promise for automated plant disease detection, most existing work on tea either overlooks model interpretability or relies on architectures too heavy for practical field use. In this paper, we describe an Attention-Enhanced EfficientNet-B3 model referred to as AE-EffNet that incorporates a Convolutional Block Attention Module (CBAM) after the final feature extraction stage of the backbone. The CBAM helps the network concentrate on disease-relevant spatial regions and feature channels rather than spreading its capacity over background noise. We trained and evaluated the model on a dataset of 5,867 tea leaf images spanning six classes: algal spot, brown blight, gray blight, healthy, helopeltis, and red spot. The images were drawn from a publicly available Kaggle repository and supplemented with field-collected samples from tea gardens in the Silchar region of Assam, India. On the held-out test set of 881 images, AE-EffNet achieved a classification accuracy of 98.98%, with macro-averaged precision, recall, and F1-score of 0.989, 0.990, and 0.990 respectively. Grad-CAM heat maps indicate that the model attends to lesion areas and disease-specific visual patterns rather than irrelevant background features. The framework maintains a compact parameter footprint by adding only a lightweight attention block on top of the already efficient backbone. This makes it a viable framework for deployment on mobile or edge devices in resource-constrained plantation settings.
Keywords: Tea Leaf Disease, Deep Learning, EfficientNet, CBAM, Attention Mechanism, Grad-CAM, Explainable AI, Plant Disease Classification, Precision Agriculture.

I. INTRODUCTION

Tea (Camellia sinensis) occupies a peculiar position in the Indian economy. It is at once a globally traded commodity and a deeply local affair state in Assam, West Bengal, and parts of the southern states owe their economic backbone to the crop. India is the world's second-largest tea producer, and the northeastern state of Assam alone accounts for over half the national output [1]. The livelihoods of millions of workers, from the plucking fields to the auction floors, hinge on the health of the tea bush. And yet, that health is constantly stay under threat. Foliar diseases in plants caused by fungi, algae, and insect pests—brown blight, gray blight, algal leaf spot, helopeltis damage, red spot, they collectively eat into an estimated 10–15% of annual production [2, 3]. In a bad year, the losses are much worse.

The traditional techniques to detect those problems are slow and inefficient. Furthermore the standard approach to detecting these problems is still remarkably low-tech in todays world. Trained workers or agronomists walk through the rows and inspect leaves by eye which consider to be slow. This works to a point, but it has well-known shortcomings like it scales poorly, it depends heavily on individual expertise, and it becomes unreliable when symptoms are subtle or when two diseases look alike in their early stages [4]. There has been a growing push, therefore, to explore automated alternative systems that can look at a photograph of a leaf and tell you weather anything is wrong with it, if anything, in a fraction of a second.

The Deep learning, and specifically Convolutional Neural Networks (CNNs), have made this kind of automation genuinely feasible over the past several years. There are several architectures such as Alexnet, VGG, ResNet, Inception, and DenseNet are been used to classify plant diseases with results that sometimes exceed manual level inspection done by human in controlled conditions [5, 6]. More recently, EfficientNet [7] has attracted attention because of its compound scaling approach, which jointly tunes depth, width, and resolution in a principled way. The result is a family of models that get better accuracy per unit of computation than most competitors a property model that matters enormously when the end goal is running the model on a smartphone camera in the middle of a tea garden.

The implementation of the framework result in high accuracy on a benchmark dataset which is not the same thing as a system that works on real world practices. Most of the published studies shows impressive numbers on test cases where the images are been lab captured, with clean backgrounds, but performance tends to drop noticeably under real world conditions. As on the field for testing it has different lighting, leaf overlap and cluttered surroundings [8]. To implement the frameworks in real world most of the studies achieves best accuracy are simply too large to run on the hardware which are available in a typical plantation.[10]

We designer our approach while keeping those issues in our mind in order to tackle those issues. The backbone which we used is pretrained on ImageNet dataset and we are using it for feature extraction on the top of it we added Convolutional Block Attention Module (CBAM) [11], which learns to emphasize the feature channels and spatial locations, which are responsible for prediction. The CBAM block is looking after disease relevant regions such as spots, the discoloration, lesion margins. Those helps the backbone of the framework to determine which region is responsible for disease. To make the model reasoning visible, we use Gradient-weighted Class Activation Mapping (Grad-CAM) [12], to generate heatmaps which shows which part of the image is responsible for each prediction.

The model is trained and evaluated on a dataset of 5,867 tea leaf images having six classes among which five classes are of diseases and one class is a healthy leaf class, we have discussed further about this on Dataset Description section. The testing of the dataset is done on images chaptured directly from tea gardens around Silchar, Assam specifically from the Rosekandi and Ironmara tea estates and gardens adjacent to Assam University. The testing of the dataset got to achieve around 98.98% classification accuracy with a macro F1-score of 0.990, which compares favourable with previously reported results on tea disease datasets. We called the model as AE-EffNet.

II. LITERATURE REVIEW

A. Deep Learning for Plant Disease Classification

The modern story of deep learning in plant disease detection arguably begins with the work of Mohanty et al. [5], who trained GoogLeNet and AlexNet on the PlantVillage dataset roughly containing 54,000 images across 26 diseases having 14 crop species which reached 99.35% accuracy. But the authors also noticed that during the testing phase of model in the real world the model is struggling to give an accurate result out of the lab conditions. This generalization gap has been a recurring theme ever since. Ferentinos [8] achieves accuracy even higher using VGG and ResNet, but the same domain-shift concern applied. Too et al. [14] ran a thorough head-to-head comparison of DenseNet, VGG, Inception, and ResNet variants for this task and found DenseNet-121 to be the strongest overall model architecture for prediction.

The attention mechanisms which is more recent and arguably more consequential line of work. The core idea is borrowed from human perception, rather than treating every pixel and every feature channel equally. The attention network learn where to look and what to prioritize Woo et al. [11] introduced Convolutional Block Attention Module CBAM, this applies channel attention (deciding which feature maps are important) followed by spatial attention (which decide which pixel locations matter). This two-step process is lightweight it adds negligible parameters but it has been shown to boost accuracy by 2–3% in a range of image classification tasks. In the agricultural domain, Chen et al. [15] applied SE-Net attention to rice disease classification and observed similar improvements. Li et al. [16] replaced the SE module in MobileNetV3 with CBAM for tea disease recognition and achieved 94.80% accuracy, an improvement over the 94.45% baseline without attention.

B. Tea Leaf Disease Detection

Compared to crops like tomato, rice, or wheat, tea has received relatively modest attention in the deep learning literature. Mukhopadhyay et al. [2] were among the first to apply transfer learning with VGG16 to tea leaf disease identification, reporting around 94% accuracy across eight classes. Srivastav et al. [17] built a CNN model trained on tea leaf images and achieved 84% accuracy using the Adam optimizer. Chen et al. [18] proposed AX-RetinaNet, a modified detection network with multi-scale feature fusion and channel attention for identifying tea diseases in natural scenes, reporting an mAP of 86.15%. On the lightweight side, Xia et al. [19] swapped in MobileNeXt as the backbone for YOLOv7 with a dual-layer attention mechanism and obtained 92.1% precision for tea disease classification.

More recently, Shikdar et al. [20] combined DenseNet201, InceptionV3, and EfficientNet-B4 with SE and CBAM attention modules in an ensemble framework. They created a novel seven-class dataset from Bangladeshi tea gardens and tested that dataset on several attention-augmented configurations. Their best ensemble model reached 85.68% accuracy, and they also applied Grad-CAM for visual interpretability. In a different vein, Bhuyan et al. [21] proposed Res4net-CBAM, an attention-based residual network tested

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

on tea leaf disease data, achieving 98.27% accuracy one of the highest figures reported to date. A very recent study by Li et al. [16] proposed WaveLiteNet, integrating wavelet transforms with MobileNetV3 and CBAM, targeting five tea disease categories.

C. Explainability in Agricultural Settings

Trust is a real barrier to adoption in agriculture. Farmers and extension officers are unlikely to accept a system that behaves like a black box. Grad-CAM [12], proposed by Selvaraju et al., has become the dominant technique for peering inside CNN predictions. It works by computing gradients of the target class score with respect to the final convolutional layer's feature maps, producing a heatmap that highlights the most influential image regions. Ghosal et al. [22] applied Grad-CAM to soybean stress phenotyping and found that attention-augmented models produced more localized and pathologically relevant heatmaps than vanilla architectures. Similar observations have been reported in tomato [23] and cereal crop [24] disease classification. The general finding is consistent attention mechanisms make models not only more accurate but also more interpretable, because the attention itself encourages focusing on relevant regions.

III. RESEARCH METHODOLOGY

A. Dataset Description

The dataset used in this study have raw images of 5,867 of tea leaves. The dataset have six classes they are algal spot, gray blight, red spot, helopeltis having one thousand images of each classes whereas for brown blight class have only 867 images and healthy tea leaves class have one thousand images. The primary source is the publicly available Tea Leaf Disease dataset on Kaggle, originally compiled by Datta [13], which was developed from tea gardens in the Sylhet region of Bangladesh. To improve the diversity and regional representativeness of the training data, we also collected supplementary images from tea gardens in the Silchar region of Assam, India specifically from the Rosekandi Tea Estate, Ironmara Tea Garden, and smaller gardens adjacent to Assam University, Silchar. These field-collected images were captured using a standard 12-megapixel smartphone camera under natural and different lighting conditions.

All images are then resized to 224×224 pixels and normalized using the standard ImageNet mean and standard deviation values. The dataset was partitioned into training (70%), validation (15%), and test (15%) subsets using stratified random sampling to maintain approximate class proportions across all splits. Table 1 shows the distribution.

Table 1. Dataset distribution across classes and splits

Class	Total	Train (70%)	Val (15%)	Test (15%)
Algal Spot	1,000	700	150	150
Brown Blight	867	606	130	131
Gray Blight	1,000	700	150	150
Healthy	1,000	700	150	150
Helopeltis	1,000	700	150	150
Red Spot	1,000	700	150	150
Total	5,867	4,106	880	881

It is worth noting that the brown blight class has somewhat fewer samples (867) than the other five classes (1,000 each). While this imbalance is not severe, we address it through class-weighted loss during training, as described in Section 3.4.

B. Data Augmentation

For training our model we used 4,106 images they are splitted across six classes, while training overfitting is a genuine concern, especially for a model with several million parameters. We used Albumentations library [25] to apply a set of on-the-fly augmentations during training With 4,106 training images split across six classes, overfitting is a genuine concern, especially for a model with several million parameters. We used the Albumentations library [25] to apply a set of on-the-fly augmentations during training: random horizontal flips (p=0.5), vertical flips (p=0.2), rotations up to ±15° (p=0.5), random brightness and contrast shifts (p=0.5), slight scale and translation via ShiftScaleRotate (p=0.3), and occasional Gaussian blur with kernel sizes 3–5 (p=0.15).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

These transforms were applied stochastically, so the model effectively sees a different version of each image every epoch. During validation and testing, only resizing and normalization were applied.

C. Proposed Architecture: AE-EffNet

The proposed model uses EfficientNet-B3 [7] as the feature extraction backbone, loaded with ImageNet pre-trained weights through the timm library (version 1.0.26). We use the backbone only to extract features of the images which outputs the feature maps from the final extraction stage. This produces a tensor of shape (batch, 384, H, W)—that is, 384 feature channels at a spatial resolution of 7×7 for 224×224 input images. This is an important detail: we are not using the full EfficientNet-B3 classification head; instead, we extract intermediate features and add our own attention and classification layers on top.

The central architectural contribution is the insertion of a CBAM module [11] immediately after these backbone features. CBAM operates in two sequential stages:

1) Channel Attention: The 384-channel feature map is squeezed spatially through both global average pooling and global max pooling. Both pooled vectors are passed through a shared two-layer MLP with a reduction ratio of 16 (hidden dimension = 384/16 = 24), and the outputs are summed and passed through a sigmoid to produce per-channel attention weights. The feature map is then element-wise multiplied by these weights.

2) Spatial Attention: The channel-refined feature map is compressed along the channel axis by taking the mean and max across all 384 channels, yielding two single-channel maps. These are concatenated and passed through a 7×7 convolution with sigmoid activation to produce a spatial attention mask. This mask is multiplied element-wise with the feature map.

After CBAM, we apply global average pooling to collapse the spatial dimensions, followed by a dropout layer (p=0.3) and a fully connected layer mapping to six output classes. The entire addition amounts to a negligible increase in parameter count over the base EfficientNet-B3 feature extractor.

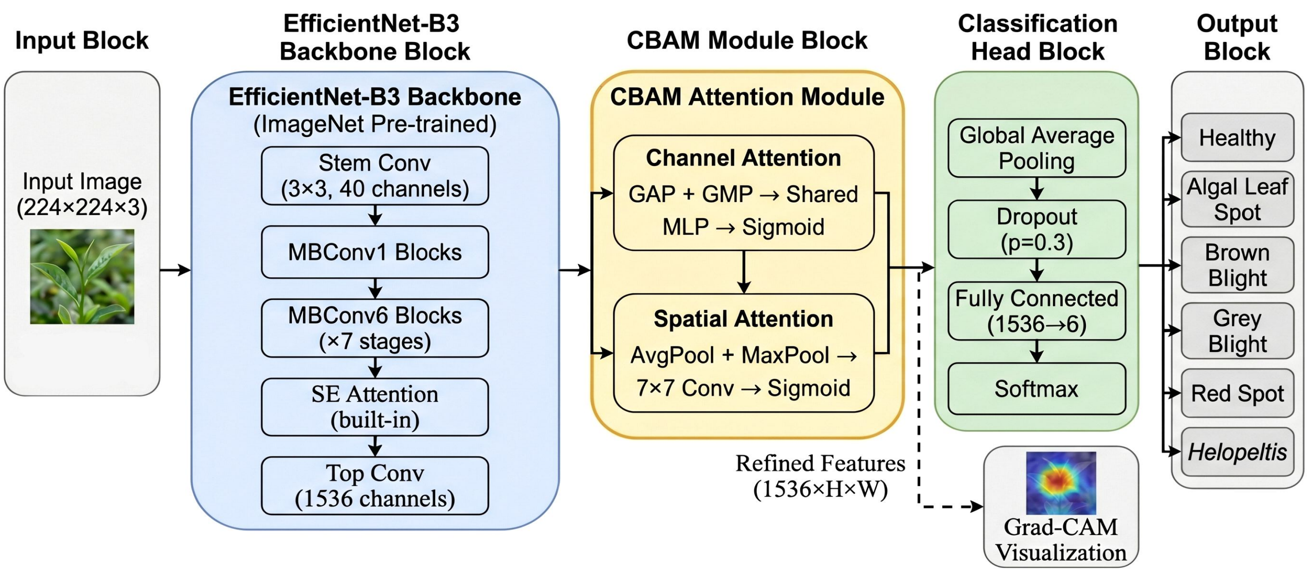


Figure 1: Architecture of the proposed AE-EffNet framework.

D. Training Configuration

We performed all our experiments on a single NVIDIA Tesla T4 GPU (16 GB VRAM) through Google Colaboratory. We used differential learning rates, the backbone parameters were fine-tuned at 1×10^{-5}, while the CBAM module and classification head received 1×10^{-3}. The optimizer which we used was AdamW with weight decay of 1×10^{-4}. We used a cosine annealing schedule with warm restarts (T_0=10, T_mult=2) to cycle the learning rate.

To handle the slight class imbalance (brown blight which has 867 samples), we computed inverse-frequency class weights and applied them to the CrossEntropyLoss. The resulting weights were close to uniform except for brown blight, which received a weight of approximately 1.13 versus ~0.97 for the other classes. Training ran for 50 epochs with a batch size of 32, and we used early stopping with a patience of 10 epochs on validation loss. The best model checkpoint was saved at epoch 43, where the validation loss reached its minimum of 0.0276.

E. Grad-CAM Visualization

We applied Grad-CAM [12] using the pytorch-grad-cam library, targeting the last inverted residual block of the EfficientNet-B3 backbone (specifically model.backbone.blocks[-1][-1]). This layer was chosen because it captures the most semantically rich features while still retaining enough spatial resolution to produce meaningful heatmaps. The resulting grayscale activation maps were overlaid on the original images using the standard jet colormap to produce visual explanations of each prediction.

F. Evaluation Metrics

Model performance on the held-out test set (881 images) was assessed using overall accuracy, macro-averaged precision, macro-averaged recall, and macro-averaged F1-score. Per-class precision, recall, and F1-score were also computed. A confusion matrix was generated to identify specific misclassification patterns between classes.

IV. RESULTS AND DISCUSSION

A. Overall Classification Performance

The proposed AE-EffNet model achieved a test accuracy of 98.98% on the 881-image test set, with macro-averaged precision of 0.989, recall of 0.990, and F1-score of 0.990. These numbers reflect strong and balanced performance across all six disease categories. Table 2 summarizes the overall results.

Table 2. Overall test set performance of AE-EffNet

Metric	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
AE-EffNet	98.98%	0.9893	0.9898	0.9895

The test accuracy approaching 99% is certainly a strong result, though it is important to keep it in context. The dataset, while reasonably large and diverse. It does include a substantial proportion of images with relatively clean backgrounds. Field deployment under truly adverse conditions like rain-soaked leaves, heavy occlusion, extreme lighting might produce lower accuracies. Still, the result is encouraging and compares well with the best figures in the tea disease literature.

B. Per-Class Analysis

Table 3. Per-class performance of AE-EffNet on the test set

Class	Precision	Recall	F1-Score	Support
Algal Spot	1.00	1.00	1.00	150
Brown Blight	0.96	0.99	0.97	131
Gray Blight	0.99	0.96	0.98	150
Healthy	0.99	1.00	0.99	150
Helopeltis	1.00	0.99	1.00	150
Red Spot	1.00	0.99	1.00	150
Macro Average	0.99	0.99	0.99	881

Several things are worth noting here. Algal spot, helopeltis, and red spot are classified essentially perfectly (F1 = 1.00). These three conditions have distinctive visual signatures the greenish-grey concentric rings of algal spot, the characteristic stippling and curling of helopeltis damage, and the prominent reddish discoloration of red spot that make them relatively easy for the model to distinguish. Healthy leaves also score very high (F1 = 0.99), which makes intuitive sense a uniformly green leaf. The unblemished leaf looks very different from any of the disease categories. The two classes where the model is slightly less certain are brown blight having F1 = 0.97 and gray blight having F1 = 0.98. This is not surprising to anyone familiar with these diseases. Both produce irregular necrotic spots on the leaf surface, and in their intermediate stages, the spots can look quite similar. The confusion matrix in Figure 3 confirms that the handful of errors are concentrated along this brown blight–gray blight axis.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

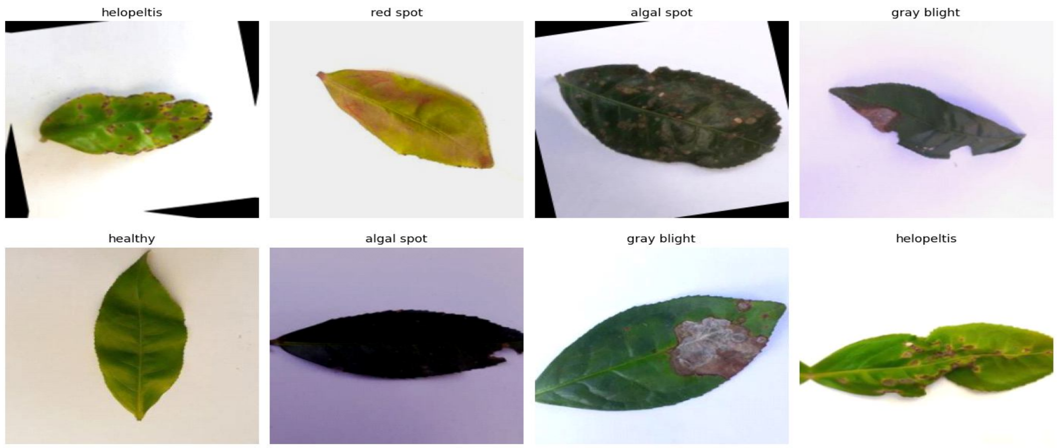


Figure 2: Sample training images from each of the six classes after augmentation, showing the visual diversity within the dataset.

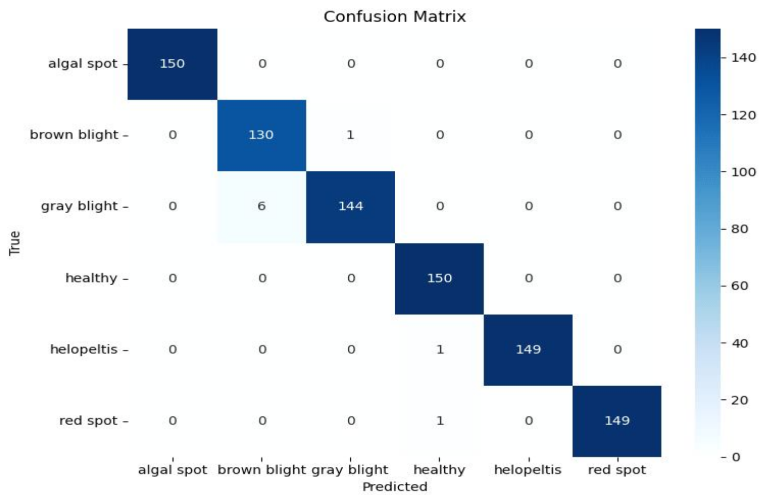


Figure 3: Confusion matrix of AE-EffNet on the test set (881 images).

C. Training Dynamics

Training proceeded smoothly over 50 epochs. The model reaching strong validation performance by around epoch 15 and continuing to refine gradually thereafter. The best validation loss (0.0276) was recorded at epoch 43. Training accuracy at the final epoch was 99.32% with an F1 of 0.993, while validation accuracy hovered around 98.86–98.98% from epoch 30 onwards, indicating good convergence without significant overfitting. The class-weighted loss and the augmentation pipeline both contributed to this stability without them, we would expect the model to overfit much earlier on a dataset of this size.

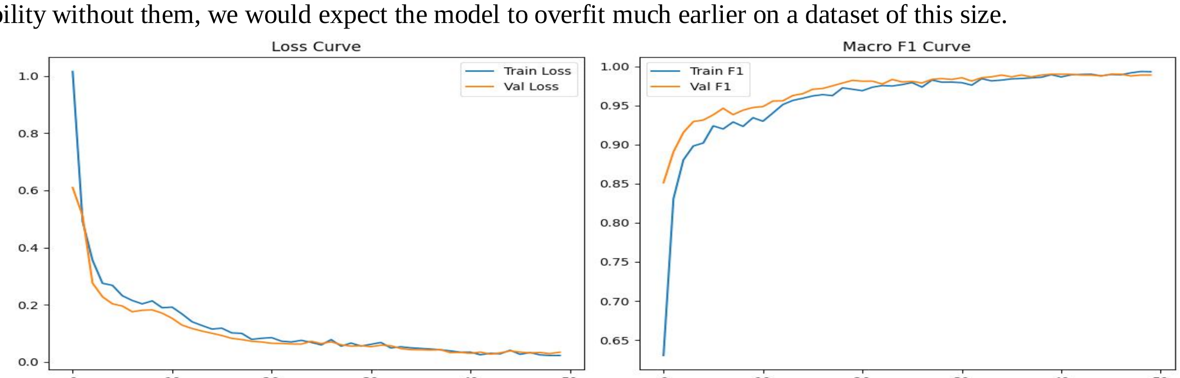


Figure 4: Training and validation curves: (a) loss over 50 epochs, (b) macro F1-score over 50 epochs. The model converges around epoch 30 and the best checkpoint is saved at epoch 43.

D. Grad-CAM Analysis

In order to understand what the model is actually attending to, we generated Grad-CAM heatmaps for a sample of test images from each class which is given in Figure 5. The heatmaps show patterns that align well with what a plant pathologist would focus on. For the disease classes, the model's attention consistently falls on the affected regions the necrotic patches in brown blight, the concentric ring patterns in algal spot, the stippling damage from helopeltis, and the reddish-brown areas in red spot. For gray blight, the activation centres on the whitish-grey lesion margins, which are the primary diagnostic feature for this disease.

For healthy leaves class, The model is distributed more broadly across the leaf surface, which is exactly what we would hope to see as it don't have any specific region to focus on as it has no disease. When there is no localized pathology, the model does not fixate on any particular spot it considers the overall condition of the leaf. This is a reassuring sign that the model has learned genuine disease features which are important for disease detection rather than dataset-specific shortcuts or background artifacts.

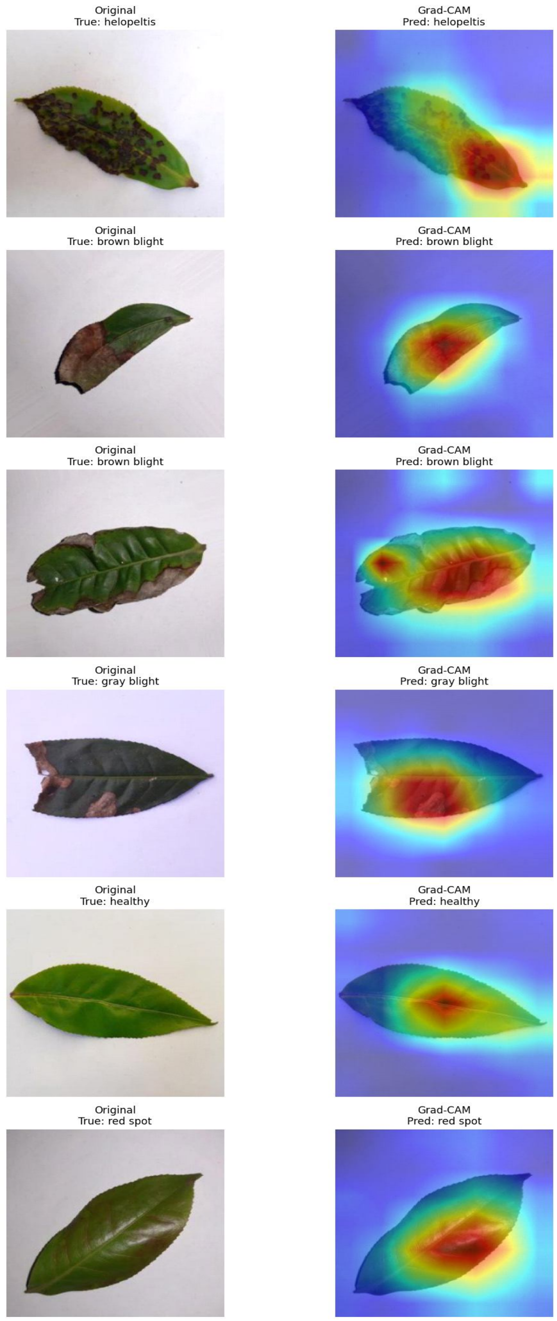


Figure 5: Grad-CAM visualizations demonstrating the interpretability of the proposed AE-EffNet model. The left column displays the original input images with their classes, the right column exhibits the corresponding Grad-CAM overlays.

E. Comparison with Existing Studies

The direct comparison across studies is always complicated due to difference in datasets, classes, split ratios and evaluation protocols. Table 4 shows the result comparison of most relevant published work on tea leaf disease classification with our proposed framework.

Table 4. Comparison with published tea leaf disease classification studies

Study	Method	Classes	Images	Best Metric	Value	XAI
Srivastav et al. [17]	Custom CNN	—	—	Accuracy	84.00%	No
Mukhopadhyay [2]	VGG16	8	2,400	Accuracy	94.00%	No
Chen et al. [18]	AX-RetinaNet	5	3,600	mAP	86.15%	No
Xia et al. [19]	YOLOv7-MobileNeXt	6	4,500	mAP	92.10%	No
Li et al. [16]	MobileNetV3+CBAM	7	Self-built	Accuracy	94.80%	No
Shikdar et al. [20]	Ensemble+Attn	7	5,278	Accuracy	85.68%	Yes
Bhuyan et al. [21]	Res4net-CBAM	—	—	Accuracy	98.27%	No
Proposed (Ours)	AE-EffNet+CBAM	6	5,867	Accuracy	98.98%	Yes

There are few observations are warranted. Our 98.98% accuracy is the highest reported among these studies, narrowly edging the 98.27% of Bhuyan et al. [21]. The two studies use different datasets and have different classes, so the comparison should be treated cautiously. More meaningful, perhaps, is the gap relative to studies using the same general type of attention mechanism (CBAM)—Li et al.'s 94.80% and Shikdar et al.'s 85.68% which suggests that architecture choice and training strategy matter as much as the attention module itself. We are also one of only two studies in this comparison to provide XAI which in our case is Grad-CAM that provide visual explainability.

F. Limitations

The framework is limited to result on specific regions like Assam, Silchar regions. The comparisons in Table 4 are with published results from other studies, which means they should be interpreted as indicative rather than definitive. The dataset, while reasonably diverse, is still predominantly composed of images from a specific geographical region and may not fully represent the variability of tea diseases across all growing areas. The near-99% accuracy should be understood with these caveats in mind.

V. CONCLUSION

In this work, we presented AE-EffNet, an attention-enhanced variant of EfficientNet-B3 designed for multi-class tea leaf disease classification. The model augments the EfficientNet-B3 feature extractor with a CBAM attention module. That enables focused learning of disease-discriminative features. The Grad-CAM is used to make its predictions visually interpretable. The model is trained on 5,867 images having six classes, it achieves 98.98% on test set of the dataset.

In our future work we would like to make a mobile or web application so that people on the field can use this technology to detect diseases and can take steps to protect the plant. We also aim to develop a comprehensive, end-to-end diagnostic pipeline that integrates mobile data collection at the edge with centralized cloud processing. In real-world scenarios, a single tea leaf may suffer from multiple concurrent infections. In our future work we will investigate transitioning the architecture to a multi-label classification framework to detect and untangle overlapping symptomatic features.

REFERENCES

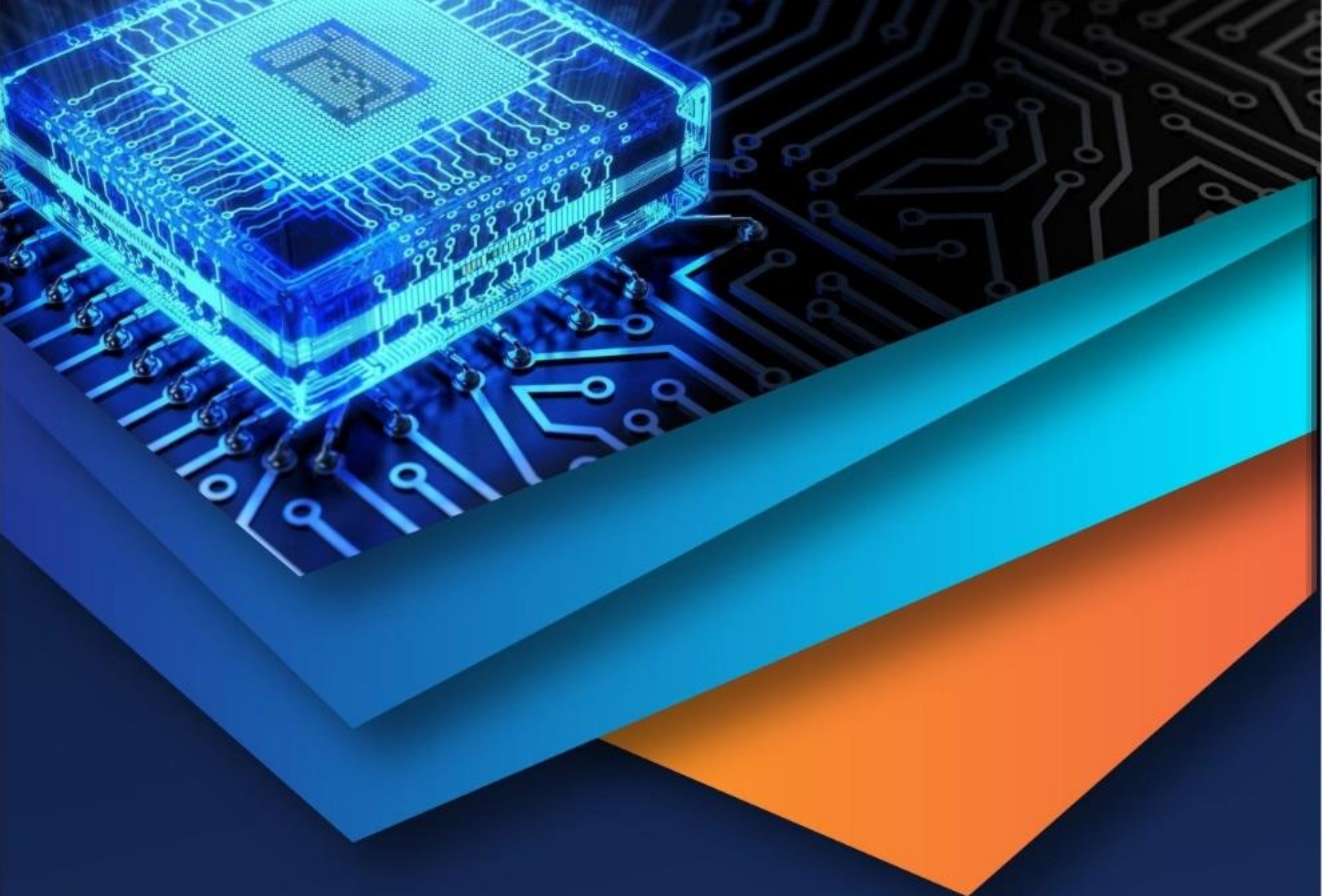
[1] Tea Board of India, "Tea Statistics," Annual Report 2023–2024, Ministry of Commerce and Industry, Government of India, 2024.

[2] S. Mukhopadhyay, M. Paul, R. Pal, and D. De, "Tea leaf disease detection using multi-objective image segmentation," Multimed. Tools Appl., vol. 80, pp. 753–771, 2021.

[3] R. Hazarika, S. Sarmah, and K. K. Sarma, "Tea pest and disease management: challenges and opportunities in Assam," Indian J. Agric. Res., vol. 56, no. 3, pp. 285–294, 2022.

[4] A. Latha, R. S. Raj, and G. Manikandan, "A review on deep learning techniques for plant leaf disease detection," J. Plant Pathol., vol. 103, pp. 441–460, 2021.

[5] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Front. Plant Sci., vol. 7, p. 1419, 2016.
[6] L. Li, S. Zhang, and B. Wang, “Plant disease detection and classification by deep learning: A review,” IEEE Access, vol. 9, pp. 56683–56698, 2021.
[7] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, pp. 6105–6114, 2019.
[8] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Comput. Electron. Agric., vol. 145, pp. 311–318, 2018.
[9] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks,” Int. J. Comput. Vis., vol. 128, no. 2, pp. 336–359, 2020.
[10] A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv:1704.04861, 2017.
[11] S. Woo, J. Park, J. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in Proc. ECCV, pp. 3–19, 2018.
[12] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. ICCV, pp. 618–626, 2017.
[13] S. Datta, “Tea Leaf Disease Dataset,” Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/saikatdatta1994/tea-leaf-disease
[14] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” Comput. Electron. Agric., vol. 161, pp. 272–279, 2019.
[15] J. Chen, D. Zhang, Y. A. Nanehkaran, and D. Li, “Detection of rice plant diseases based on deep transfer learning,” J. Sci. Food Agric., vol. 100, no. 7, pp. 3246–3256, 2020.
[16] Y. Li et al., “Tea leaf disease and insect identification based on improved MobileNetV3,” Front. Plant Sci., vol. 15, p. 1459292, 2024.
[17] S. Srivastav, S. Kumar, and M. Gupta, “Tea leaf disease detection using CNN,” in Proc. Int. Conf. IoT in Social, Mobile, Analytics and Cloud, pp. 210–214, 2022.
[18] Y. Chen et al., “Detection and identification of tea leaf diseases based on AX-RetinaNet,” Sci. Rep., vol. 12, p. 2183, 2022.
[19] Y. Xia et al., “Classification and identification of tea diseases based on improved YOLOv7 model of MobileNeXt,” Sci. Rep., vol. 14, p. 11799, 2024.
[20] O. F. Shikdar et al., “Enhancing tea leaf disease recognition with attention mechanisms and Grad-CAM visualization,” arXiv:2512.17987, 2025.
[21] P. Bhuyan, P. K. Singh, and S. K. Das, “Res4net-CBAM: A deep CNN with CBAM for tea leaf disease diagnosis,” Multimed. Tools Appl., vol. 83, pp. 1–23, 2023.
[22] S. Ghosal et al., “An explainable deep machine vision framework for plant stress phenotyping,” Proc. Natl. Acad. Sci., vol. 115, no. 18, pp. 4613–4618, 2018.
[23] M. Shoaib et al., “Deep learning-based segmentation and classification of leaf images for tomato plant disease,” Front. Plant Sci., vol. 13, p. 1031748, 2022.
[24] A. Mahmood et al., “Deep learning framework using UAV imagery for multi-disease detection in cereal crops,” Sci. Rep., vol. 16, p. 3339, 2026.
[25] A. Buslaev et al., “Albumentations: Fast and flexible image augmentations,” Information, vol. 11, no. 2, p. 125, 2020.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)