

Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation

Moïse K. Kasambya^{1, 2, ✉}, Kiro M. Lauraine²

1. Computer Science, University of the Philippines Diliman, Quezon, PHL

2. Information Systems, Université de l'Assomption au Congo, Butembo, COD

Received: March 31, 2026 | Review began: April 02, 2026 | Review ended: June 08, 2026 | Published: June 10, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Banana cultivation is significantly affected by fungal and foliar diseases, leading to substantial yield losses worldwide. Early and accurate detection is therefore essential for sustainable crop management. This study presents a comparative evaluation of three convolutional neural network architectures - MobileNet, ResNet50, and EfficientNetB0 - for automated classification of banana leaf diseases using the BananaLSD dataset. The dataset includes four classes: Cordana leaf spot, *Pestalotiopsis* leaf spot, Black Sigatoka, and healthy leaves. To enhance robustness, Focal Loss and a Convolutional Block Attention Module are integrated into an improved MobileNet variant. Model interpretability is achieved using Gradient-weighted Class Activation Mapping (Grad-CAM), while robustness is evaluated through stratified 5-fold cross-validation. Statistical significance is assessed using McNemar's test and Cochran's Q test. Results show that ResNet50 achieves the highest test accuracy (99.24%) and macro-area under the curve (0.9997), while MobileNet achieves comparable performance (99.08% accuracy; macro-area under the curve = 0.9993) with significantly lower computational cost. Cross-validation confirms the stability of MobileNet (0.9915 ± 0.0066). Statistical analysis reveals significant overall differences among models ($Q = 29.65$, $p < 0.001$) but no significant difference between MobileNet and ResNet50 ($p = 0.625$). Grad-CAM visualizations confirm biologically meaningful lesion localization. These findings highlight MobileNet as an efficient and reliable solution for real-time banana disease detection.

Categories: AI applications, Algorithm Analysis, Computer Vision

Keywords: banana leaf disease, mobilenet, resnet50, efficientb0, grad-cam, explainable ai, focal loss

Introduction

Banana (*Musa* spp.) is among the world's most economically important crops, providing a primary food source and livelihood for hundreds of millions of people across tropical and subtropical regions. However, banana cultivation is severely threatened by a range of foliar diseases, including *Cordana* leaf spot (caused by *Cordana musae*), *Pestalotiopsis* leaf spot, and Black Sigatoka (caused by *Mycosphaerella fijiensis*), which can cause yield losses of 30% to 100% if left undetected [1]. Strobl and Mohan [2] documented the global spread of Black Sigatoka driven by climate change, while Esguerra et al. [3] provided a comprehensive overview of the Sigatoka leaf spot complex and its management implications.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI https://doi.org/10.7759/s44389-026-00114-x

Recent advances in deep learning, particularly in convolutional neural networks (CNNs), have demonstrated remarkable performance in plant disease detection tasks [4-6]. Architectures such as MobileNet [7], ResNet50 [8], and EfficientNetB0 [9] offer different trade-offs between accuracy, computational efficiency, and deployability in resource-constrained field environments. For banana leaf disease specifically, Bhuiyan et al. [10] developed BananaSqueezeNet, a lightweight CNN for diagnosing three banana diseases. Lakshmi Narayanan et al. [11] proposed a hybrid CNN architecture, achieving competitive accuracy on banana disease data. Despite these advances, several critical limitations remain.

First, most prior work treats CNN models as black boxes, providing accuracy metrics without explaining which visual features drive predictions [12]. In high-stakes domains such as plant disease management, this lack of transparency undermines user trust. Second, performance claims are frequently reported without rigorous statistical testing [12,13], making it difficult to determine whether observed differences between models are meaningful or merely due to random variation [14]. Third, lightweight models - essential for mobile deployment - are rarely compared systematically against deeper architectures under identical experimental conditions.

This study addresses these limitations through three principal contributions:

- (1) A systematic comparative evaluation of MobileNet, ResNet50, and EfficientNetB0 under identical training protocols on the BananaLSD dataset, using a two-phase fine-tuning strategy with class-balanced training.
- (2) Integration of Focal Loss and Convolutional Block Attention Module (CBAM) attention into an enhanced MobileNet variant to improve robustness and feature localization under class imbalance.
- (3) Rigorous explainability analysis via Gradient-weighted Class Activation Mapping (Grad-CAM) and statistical validation using stratified 5-fold cross validation, McNemar's pairwise test, and Cochran's Q global test.

Related work

Deep Learning for Plant Disease Classification

The seminal work of Mohanty et al. [4] demonstrated that CNNs trained on the PlantVillage dataset could achieve over 99% accuracy under controlled conditions, establishing deep learning as a viable approach for automated plant disease diagnosis. Ferentinos [5] confirmed this potential across a wider range of plant-pathogen combinations. Ramcharan et al. [15] highlighted the practical challenges of real-world agricultural imagery, including variable lighting, occlusion, and image quality for cassava disease detection. Recent systematic reviews have synthesized the rapid progress in this field: Kumar et al. [16] analyzed machine and deep learning approaches for plant leaf disease classification across dozens of studies, while Demilie [17] provided comparative performance benchmarks. Dong et al. [18] proposed PlantPAD, a large-scale platform for image phenomics analysis enabling standardized disease benchmarking.

Transfer Learning Architectures in Agriculture

Transfer learning with pre-trained CNN backbones has become the dominant approach for plant disease classification, compensating for limited agricultural labeled data [19]. He et al. [8] introduced ResNet50 with residual connections, enabling the training of very deep networks. Howard et al. [7] proposed MobileNet with depthwise separable convolutions, reducing computational cost while maintaining competitive accuracy. Tan and Le [20] introduced EfficientNetB0 through compound scaling. Sandler et al. [21] extended MobileNet to MobileNetV2 with inverted residuals, while Tan and Le [20] further developed EfficientNetV2 for faster training. For plant disease tasks, Atila et al. [22] demonstrated EfficientNet's strong performance across 38 disease classes with 99.91% accuracy. Ashwinkumar et al. [23] achieved competitive results using an optimized MobileNet approach. Chen et al. [24] integrated attention mechanisms into MobileNetV2 for crop disease identification. Fan et al. [25] proposed feature fusion strategies for leaf image disease identification using transfer learning.

Banana Leaf Disease Classification

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

For banana leaf disease specifically, Bhuiyan et al. [10] developed BananaSqueezeNet, a very fast, lightweight CNN for diagnosing three banana diseases, and introduced the BananaLSD dataset used in this study [26]. Narayanan et al. [11] proposed a hybrid CNN architecture for banana plant disease classification. Jimenez et al. [27] compared EfficientNetB0, ResNet50, and VGG19 on a banana leaf dataset from Ecuador. Thiagarajan et al. [28] analyzed banana plant health using multiple machine learning techniques. Rajalakshmi et al. [29] developed a novel deep CNN for early banana leaf disease detection. Mathew et al. [30] classified three banana leaf spot diseases using pre-trained CNNs. Unal [31] conducted a comparative study of deep learning architectures for banana leaf spot diagnosis.

Explainable AI and Grad-CAM in Agriculture

Selvaraju et al. [32] introduced Grad-CAM, which has since become a standard technique for generating visual explanations of CNN decisions in agricultural contexts. Quach et al. [33] applied Grad-CAM to smart agricultural applications, demonstrating improved transparency and farmer trust. Rajbongshi et al. [34] provided a comprehensive review of Explainable Artificial Intelligence (XAI) for sustainable agriculture, synthesising 145 studies employing Grad-CAM, Local Interpretable Model-agnostic Explanations, and SHapley Additive exPlanations techniques. Kondaveeti and Simhadri [35] provided both qualitative and quantitative XAI evaluation for rice leaf disease detection. Pal et al. [36] developed a lightweight and explainable CNN using Grad-CAM and Grad-CAM++ for plant disease diagnosis. Bhandari et al. [37] introduced BotanicX-AI using EfficientNetB5 with Grad-CAM for tomato leaf disease. Shawon et al. [38] reviewed federated learning and XAI integration for plant disease detection, highlighting privacy-preserving interpretable approaches. Oad et al. [39] demonstrated ensemble learning combined with XAI for plant leaf disease detection via IEEE Access.

Attention Mechanisms and Loss Functions

Woo et al. [40] proposed CBAM, which applies sequential channel and spatial attention with negligible computational overhead and integrates seamlessly into any CNN backbone. Hu et al. [41] introduced Squeeze-and-Excitation Networks (SENet) as a channel attention precursor. Agac and Durmaz Incel [42] applied CBAM to DeepConvLSTM for human activity recognition, validating its versatility across application domains. Lin et al. [43] introduced Focal Loss to address class imbalance in dense object detection, later adopted widely in medical imaging and increasingly in plant pathology.

Data Augmentation

Zhang et al. [44] introduced MixUp, which linearly interpolates input-output pairs to prevent overfitting and improve calibration. Yun et al. [45] proposed CutMix, which cuts and pastes image patches to improve localization ability. Lewy and Mańdziuk [46] provided a comprehensive overview of mixing augmentation methods. Shorten and Khoshgoftaar [47] surveyed the broad landscape of image data augmentation for deep learning, establishing augmentation as essential for model generalization. Zeng [48] provided an updated survey of deep learning-based augmentation techniques.

Limitations of Existing Work

A recurring limitation in the plant disease classification literature is the absence of rigorous statistical testing, which makes reproducibility and cross-study comparisons difficult [12,13]. Most studies report a single train-test split, which may overestimate generalization performance. Borhani et al. [49] emphasized the need for more robust evaluation protocols in vision transformer-based plant disease detection. Yu et al. [50] noted similar gaps in inception-based transformer approaches. The present study addresses these gaps by combining cross-validation, pairwise McNemar's tests, and Cochran's Q test within a unified experimental framework, while providing multi-architecture Grad-CAM explainability analysis.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

Materials And Methods

Dataset description

This study uses the BananaLSD (Banana Leaf Spot Diseases) dataset [26], comprising banana leaf images across four classes: *Cordana* leaf spot (*Cordana*), healthy, *Pestalotiopsis* leaf spot (*Pestalotiopsis*), and Black Sigatoka (Sigatoka). Images were collected using smartphone cameras in diverse real-world conditions. The augmented training set contains 800 images per class (3,200 images total), all resized to 224 × 224 pixels. The original (non-augmented) set is used for testing and prediction evaluation. This dataset is publicly associated with the BananaSqueezeNet model [10]. Dataset composition is summarized in Table 1.

Class	Training Images	Test Images	Description
<i>Cordana</i>	800	234	<i>Cordana musae</i> fungal leaf spot
Healthy	800	174	Healthy, disease-free leaves
<i>Pestalotiopsis</i>	800	238	<i>Pestalotiopsis</i> spp. leaf spot
Sigatoka	800	665	Black Sigatoka (<i>Mycosphaerella fijiensis</i>)
Total	3,200	1,311	

TABLE 1: Dataset class distribution across training and test sets

Data preprocessing

All images are loaded using OpenCV, converted from BGR (Blue, Green, Red) to RGB (Red, Green, Blue) color space, resized to 224 × 224 pixels using area interpolation (INTER_AREA), and normalized to the (0, 1) range. The original dataset is split into a test set (70%, n = 1,311) and a prediction pool (30%, n = 563) using stratified random sampling (random_state = 42). Online data augmentation is applied during training via a Keras Sequential pipeline: random horizontal flipping, random rotation (±0.08 rad), random zoom (±12%), and random contrast variation (±10%) [47,48]. Backbone-specific preprocessing (MobileNet normalization, ResNet mean-subtraction, or EfficientNet scaling) is applied after augmentation. Balanced class weights are computed using scikit-learn's compute_class_weight function.

Proposed enhancements

Focal Loss

Standard cross-entropy loss treats all misclassified examples equally. Focal Loss [43] introduces a modulating factor to focus training on hard, uncertain examples:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

where p_t is the predicted probability of the true class, $\gamma = 2.0$ is the focusing parameter, and $\alpha = 0.25$ is the class-balancing weight. Focal Loss is most effective when training data exhibit significant class imbalance, as the modulating factor $(1 - p_t)^\gamma$ down-weights easy majority-class examples and concentrates gradient signal on rare, hard-to-classify

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

instances. On perfectly balanced datasets, where standard cross-entropy already receives equivalent gradient contributions from all classes, this corrective effect is structurally diminished.

CBAM Attention Module

CBAM [40] applies two sequential sub-modules - channel attention M_c and spatial attention M_s - after the backbone feature maps:

$$F' = M_s(F'') \otimes F''$$

$$F'' = M_c(F) \otimes F$$

Channel attention uses both average and max pooling processed through a shared multilayer perceptron. Spatial attention concatenates channel-averaged and channel-max maps processed through a 7×7 convolution. Unlike SENet [41], CBAM combines both channel and spatial attention, consistently outperforming SE-based approaches on ImageNet-1K [40].

The Focal+CBAM MobileNet variant is introduced as a robustness-oriented architecture rather than an accuracy-maximizing one. On the BananaLSD augmented training set - which is perfectly balanced at 800 images per class by construction - neither Focal Loss nor CBAM is expected to confer a significant accuracy advantage over the well-regularized base MobileNet. Focal Loss's imbalance-correction mechanism provides no structural benefit under balanced sampling, and CBAM's attention gates require sufficient data volume to converge to their optimal configuration [40]. The primary value of this variant lies in its anticipated behavior on naturally imbalanced, field-collected datasets - a common scenario in real-world agricultural deployment where class frequencies are rarely controlled [42,43]. Results on BananaLSD should therefore be interpreted as a baseline validation confirming that these enhancements do not degrade performance on balanced data, rather than as a demonstration of their full potential.

Deep learning models

MobileNet

MobileNet [7] uses depthwise separable convolutions, reducing computation while maintaining competitive accuracy. The model contains 3,232,964 total parameters, 12.9 MB memory footprint, and achieves inference at 31.6 ms/image (~32 images/second).

ResNet50

ResNet50 [8] employs skip (residual) connections, enabling effective training of 50-layer deep networks. With 23,595,908 parameters and 94.4 MB memory, it is the most computationally demanding model (183.3 ms/image). Huang et al. [51] further developed the densely connected variant (DenseNet) as an alternative deep architecture.

EfficientNetB0

EfficientNetB0 [9] applies compound scaling of depth, width, and resolution via neural architecture search. With 4,054,695 parameters and 16.2 MB memory, it achieves inference at 49.0 ms/image (~20 images/second). Tan and Le [20] subsequently developed EfficientNetV2 for even faster training.

Training strategy

All models use a two-phase fine-tuning strategy. Phase 1 (feature extraction): backbone frozen, head trained with Adam ($\text{lr} = 1 \times 10^{-3}$), sparse categorical cross-entropy, batch size 32, early stopping (patience = 5), and ReduceLROnPlateau (factor = 0.2, patience = 3, min $\text{lr} = 1 \times 10^{-7}$). Phase 2 (fine-tuning): last 20 backbone layers unfrozen, trained at $\text{lr} = 1 \times 10^{-5}$. Fixed seed (42) ensures full reproducibility.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. *Cureus J Comput Sci* 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

Evaluation metrics

Model performance is assessed using accuracy, macro-averaged precision, recall, and F1-score [52], and macro-area under the curve (AUC) under the one-vs-rest multi-class receiver operating characteristic (ROC) strategy. The F1-score formula is

$$F1 = \frac{2 \cdot (P \cdot R)}{P + R}$$

Macro-averaged AUC:

$$AUC_{macro} = \frac{1}{K} \sum_{k=1}^K AUC_K, K = 4$$

Statistical validation

Stratified 5-fold cross-validation: Accuracy, macro-F1, precision, and recall are computed across five stratified folds [14]. Mean $\pm$ standard deviation is reported.

McNemar's Test (pairwise) is applied to all model pairs with continuity correction [13]:

$$X^2 = \frac{(|b - c| - 1)^2}{b + c}$$

where b and c are discordant prediction pairs.

Cochran's Q Test: Global test for $k = 3$ classifiers:

$$Q = k(k-1) \cdot \frac{k \sum_{j=1}^k L_j - \left(\sum_{i=1}^k L_i \right)^2}{k \sum_{i=1}^k L_i - \sum_i L_i^2}$$

Model complexity analysis

Each model is profiled for total parameters, memory footprint (MB), per-image inference time (ms), and throughput (images/second), measured on a batch of 64 test images over 50 runs.

Results And Discussion

Experimental setup

All experiments were implemented in Python 3.10 using TensorFlow 2.x and the Keras high-level API. Training and evaluation were performed on a system equipped with a CUDA-compatible GPU. Images were loaded using OpenCV, resized to 224×224 pixels, and normalized to the (0, 1) range prior to backbone-specific preprocessing. The augmented training set (3,200 images; 800 per class) was fed through a Keras Sequential augmentation pipeline comprising random horizontal flipping, random rotation (± 0.08 rad), random zoom ($\pm 12\%$), and random contrast variation ($\pm 10\%$). A fixed random seed of 42 was applied across NumPy, Python, and TensorFlow to ensure full reproducibility. All models were trained using a two-phase fine-tuning strategy: Phase 1 froze the backbone and trained only the classification head using the Adam optimizer (learning rate = 1×10^{-3}), sparse categorical cross-entropy loss, and a batch size of 32, with early stopping (patience = 5) and ReduceLROnPlateau (factor = 0.2, patience = 3, minimum learning rate = 1×10^{-7}). Phase 2 unfroze the last 20 backbone layers and performed fine-tuning at a reduced learning rate of 1×10^{-5} . Model inference time and throughput were benchmarked on a batch of 64 test images averaged over 50 independent runs. Stratified 5-fold cross-validation was applied exclusively to MobileNet due to computational constraints, using the same two-phase protocol and evaluation metrics (accuracy, macro-F1, precision, and recall) at each fold.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

Performance comparison

Table 2 presents the classification performance of all four models on the held-out test set (n = 1,311). Figure 1 illustrates the training and validation accuracy and loss curves for all three models across both training phases. Figure 2 shows the confusion matrices for all three models.

Model	Accuracy (%)	Macro Precision	Macro Recall	Macro F1
MobileNet	99.08	0.99	0.99	0.99
ResNet50	99.24	0.99	0.99	0.99
EfficientNetB0	98.47	0.98	0.98	0.98
Focal+CBAM MobileNet	98.55	0.98	0.99	0.98

TABLE 2: Classification performance on the test set (n = 1,311)

CBAM, Convolutional Block Attention Module

Table 3 provides the per-class breakdown for the three base models.

Class	MobileNet P/R/F1	ResNet50 P/R/F1	EfficientNetB0 P/R/F1
Cordana (n=234)	0.99/0.96/0.97	1.00/0.97/0.98	0.98/0.96/0.97
Healthy (n=174)	1.00/1.00/1.00	1.00/1.00/1.00	0.99/0.98/0.99
Pestalotiopsis (n=238)	0.99/1.00/0.99	0.99/1.00/0.99	0.96/0.99/0.97
Sigatoka (n=665)	0.99/1.00/0.99	0.99/1.00/0.99	0.99/0.99/0.99

TABLE 3: Per-class precision (P), recall (R), and F1-score (F1)

How to cite this article:

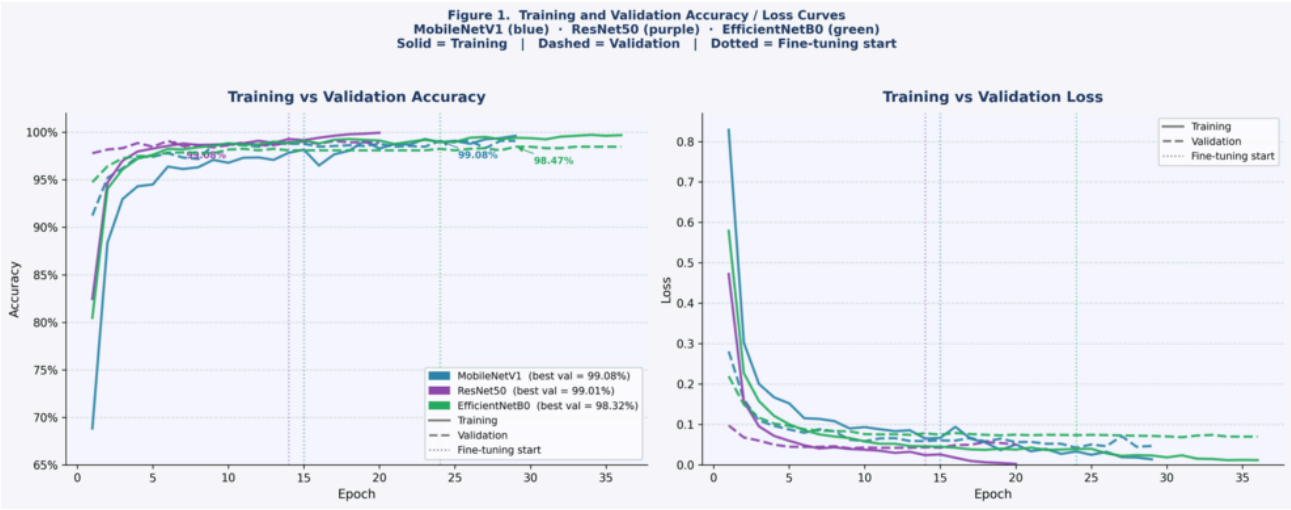


FIGURE 1: Training and validation accuracy/loss curves for MobileNet (blue), ResNet50 (purple), and EfficientNetB0 (green)

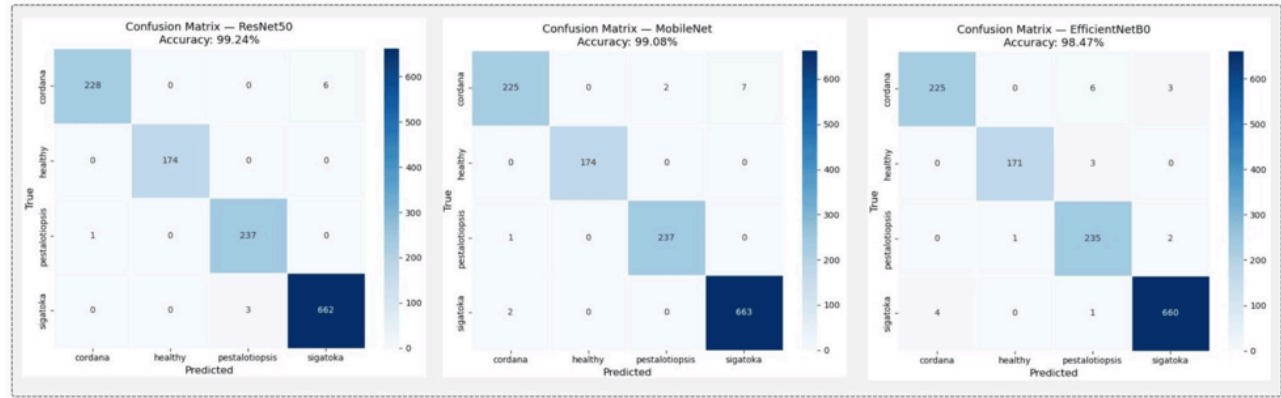


FIGURE 2: Confusion matrices for all three models on the test set (n = 1,311)

ROC and AUC analysis

Table 4 summarizes the macro-AUC values computed using a one-vs-rest ROC strategy for all three base models. All models achieve macro-AUC above 0.998, indicating near-perfect class separability across all four disease categories. The corresponding multi-class ROC curves are presented in Figure 3.

Model	Macro-AUC
MobileNet	0.9993
ResNet50	0.9997
EfficientNetB0	0.9981

TABLE 4: Macro-AUC values
AUC, Area Under the Curve

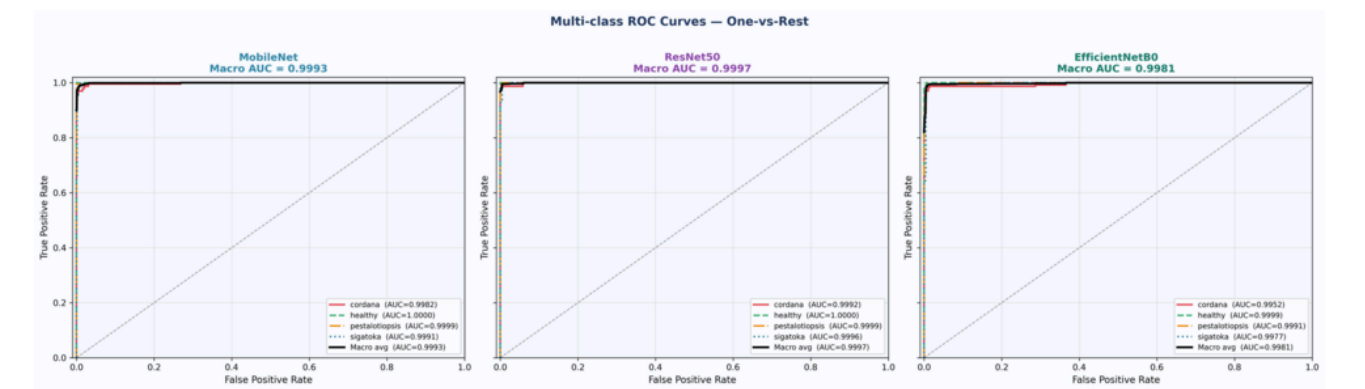


FIGURE 3: Multi-class ROC curves (one-vs-rest) with macro-AUC for all three base models
AUC, Area Under the Curve; ROC, Receiver Operating Characteristic

Ablation study

Table 5 presents the ablation study results evaluating the contribution of individual training components (augmentation, dropout, learning rate, and fine-tuning) on MobileNet Phase 1 accuracy and macro-F1. The results are visualized comparatively in Figure 4.

Configuration	Accuracy (%)	Macro F1
No Augmentation	98.32	0.9814
No Dropout	98.55	0.9839
Dropout = 0.3	98.63	0.9854
High LR (1e-2)	97.94	0.9775
No Class Weights	98.55	0.9849
Full Pipeline (Phase 1)	98.47	0.9843

TABLE 5: Ablation study—MobileNet Phase 1 only (20 epochs)

Consistent with findings of Shorten and Khoshgoftaar [47], data augmentation is the most impactful component.

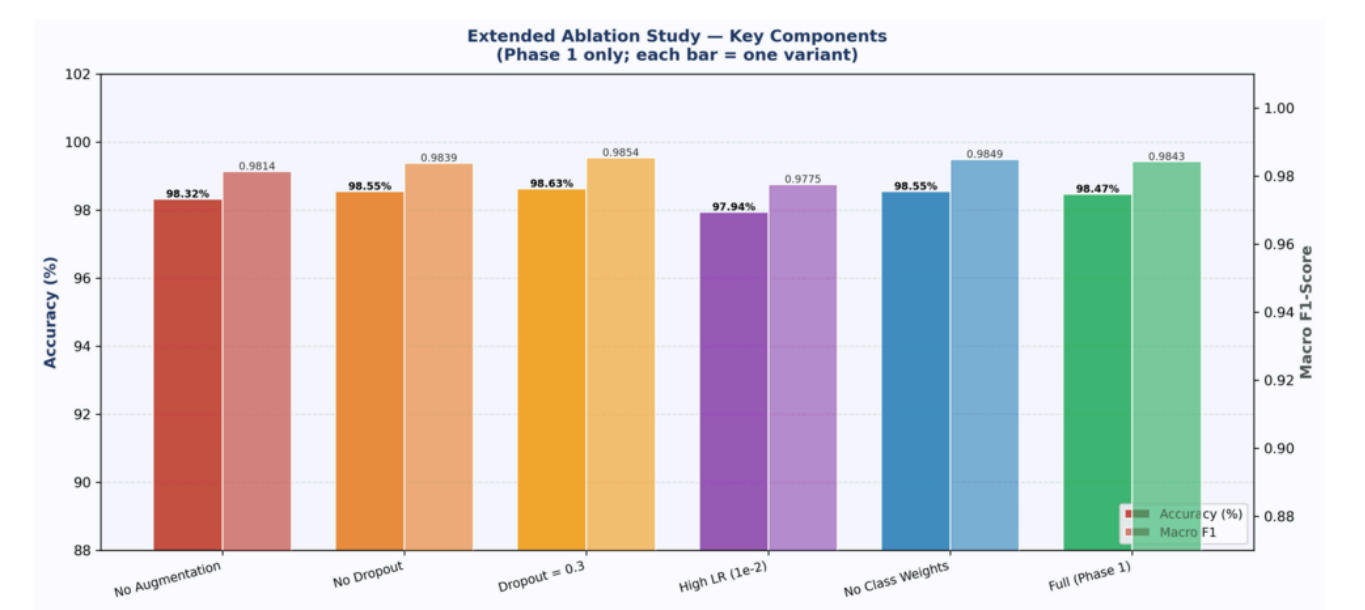


FIGURE 4: Ablation study comparing key training components on MobileNet Phase 1 accuracy and macro-F1

X-axis: “Model Variants (Ablation Settings)”: No Augmentation, No Dropout, Dropout = 0.3, High LR (1e-2), No Class Weights, Full Pipeline.
Left Y-axis: Accuracy (%). Right Y-axis: Macro F1-Score.

Robustness evaluation

MobileNet achieves 98.22% accuracy on the 30% prediction pool (n = 563 original, non-augmented images), confirming strong generalization to unseen samples. The close alignment between test accuracy (99.08%) and prediction pool accuracy (98.22%) confirms that the model has not overfit to the augmented training distribution, consistent with the

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

generalization benefits of the applied augmentation pipeline [48].

Statistical validation results

Stratified 5-Fold Cross-Validation

Table 6 reports fold-level accuracy, macro-F1, precision, and recall for MobileNet across all five stratified folds. The mean accuracy of 0.9915 ± 0.0066 confirms stable generalization across data splits. Figure 5 provides a visual summary with mean $\pm$ standard deviation indicated by diamond markers.

Fold	Accuracy	Macro F1	Precision	Recall
1	0.9920	0.9906	0.9938	0.9875
2	1.0000	1.0000	1.0000	1.0000
3	0.9920	0.9909	0.9916	0.9902
4	0.9813	0.9821	0.9911	0.9740
5	0.9920	0.9920	0.9961	0.9883
Mean $\pm$ Std	0.9915 ± 0.0066	0.9911 ± 0.0063	0.9945 ± 0.0036	0.9880 ± 0.0093

TABLE 6: Stratified 5-fold cross-validation results for MobileNet

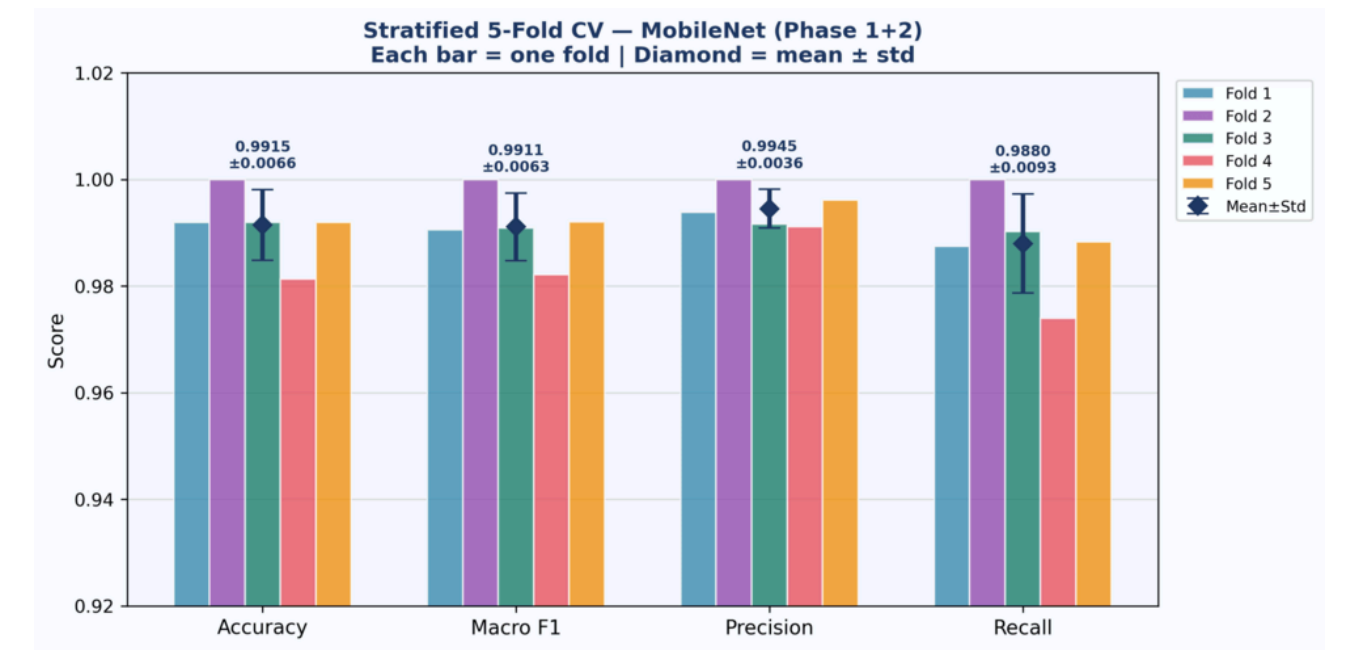


FIGURE 5: Five-fold cross-validation results for MobileNet

Diamonds indicate mean ± standard deviation.
X-axis: Performance Metrics.
Y-axis: Score (range 0.92–1.00).

McNemar's test and Cochran's Q

Table 7 presents the results of pairwise McNemar's tests and the global Cochran's Q test across all model pairs. Figure 6 displays the corresponding p-value and test statistic heatmaps, facilitating rapid visual interpretation of statistical significance between models.

Comparison	Contingency	Stat.	p-value	Sig.
MobileNet vs ResNet50	((9,3),(1,1298))	1.0000	0.6250	ns
MobileNet vs EfficientNetB0	((9,3),(11,1288))	3.0000	0.0574	ns
ResNet50 vs EfficientNetB0	((7,3),(13,1288))	3.0000	0.0213	*
Cochran's Q	— (global)	29.6471	< 0.001	***

TABLE 7: Pairwise McNemar's test results and Cochran's Q

ns = not significant; * p < 0.05; *** p < 0.001.

How to cite this article:

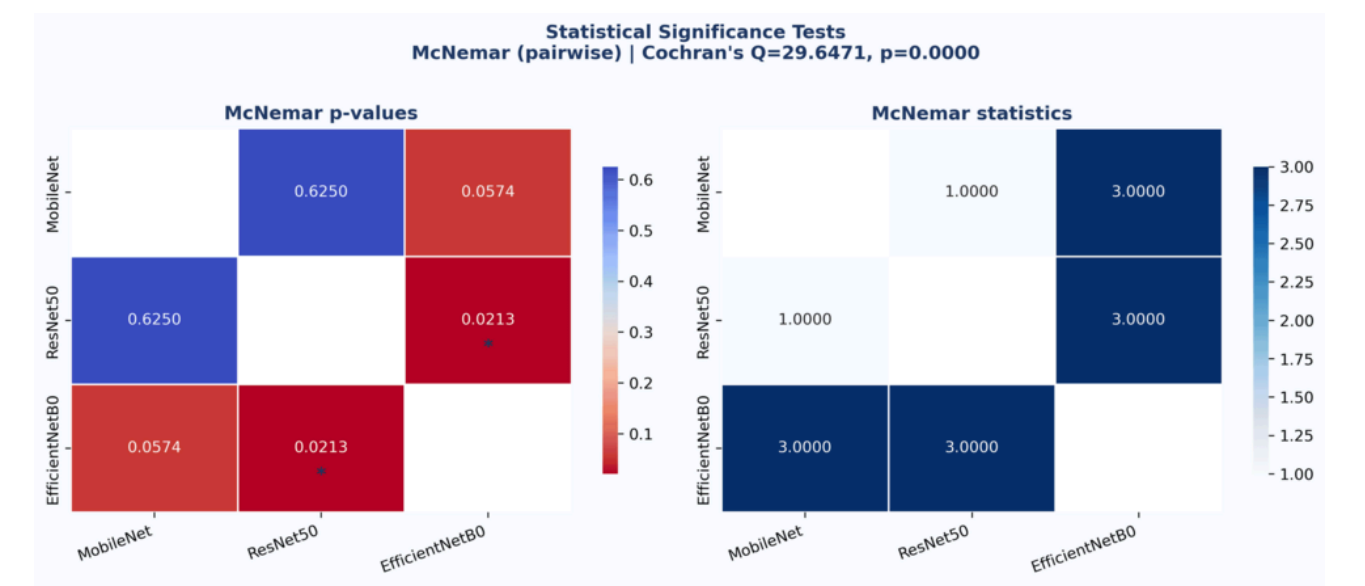


FIGURE 6: McNemar p-value (left) and statistic (right) heatmaps for pairwise model comparisons

X-axis: Predicted model. Y-axis: Reference model (MobileNet, ResNet50, EfficientNetB0).
Color scale: p-value (left heatmap, lower = more significant) and McNemar test statistic (right heatmap, higher = greater disagreement).

Model complexity analysis

Table 8 provides a quantitative comparison of total parameters, memory footprint, per-image inference time, and throughput for all three architectures. Figure 7 plots parameter count against inference time, highlighting the efficiency advantage of MobileNet relative to ResNet50 under equivalent accuracy conditions.

Model	Total Parameters	Memory (MB)	ms/Image	Images/s
MobileNet	3,232,964	12.9	31.6	32
EfficientNetB0	4,054,695	16.2	49.0	20
ResNet50	23,595,908	94.4	183.3	5

TABLE 8: Model complexity and inference performance

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

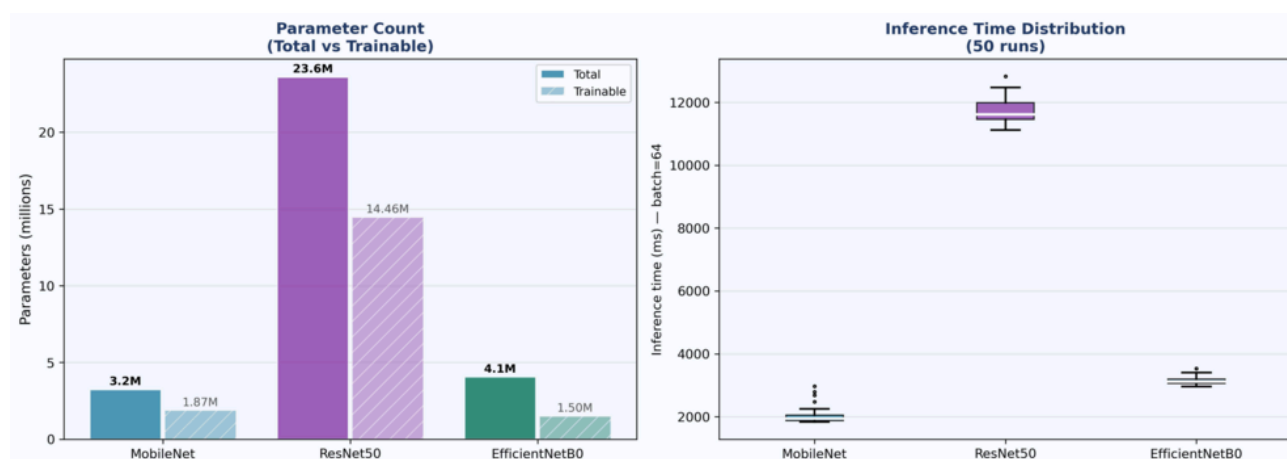


FIGURE 7: Parameter counts vs. inference time for all three models

X-axis: Model Architecture. Y-axis: Per-image inference time (ms/image).

Explainability analysis

Grad-CAM Visualization

Grad-CAM [32] generates class-discriminative heatmaps by computing the gradient of the class score with respect to the final convolutional layer feature maps:

$$L_{Grad-CAM}^c = \text{ReLU} \left(\sum_k \alpha_k \cdot A^k \right)$$

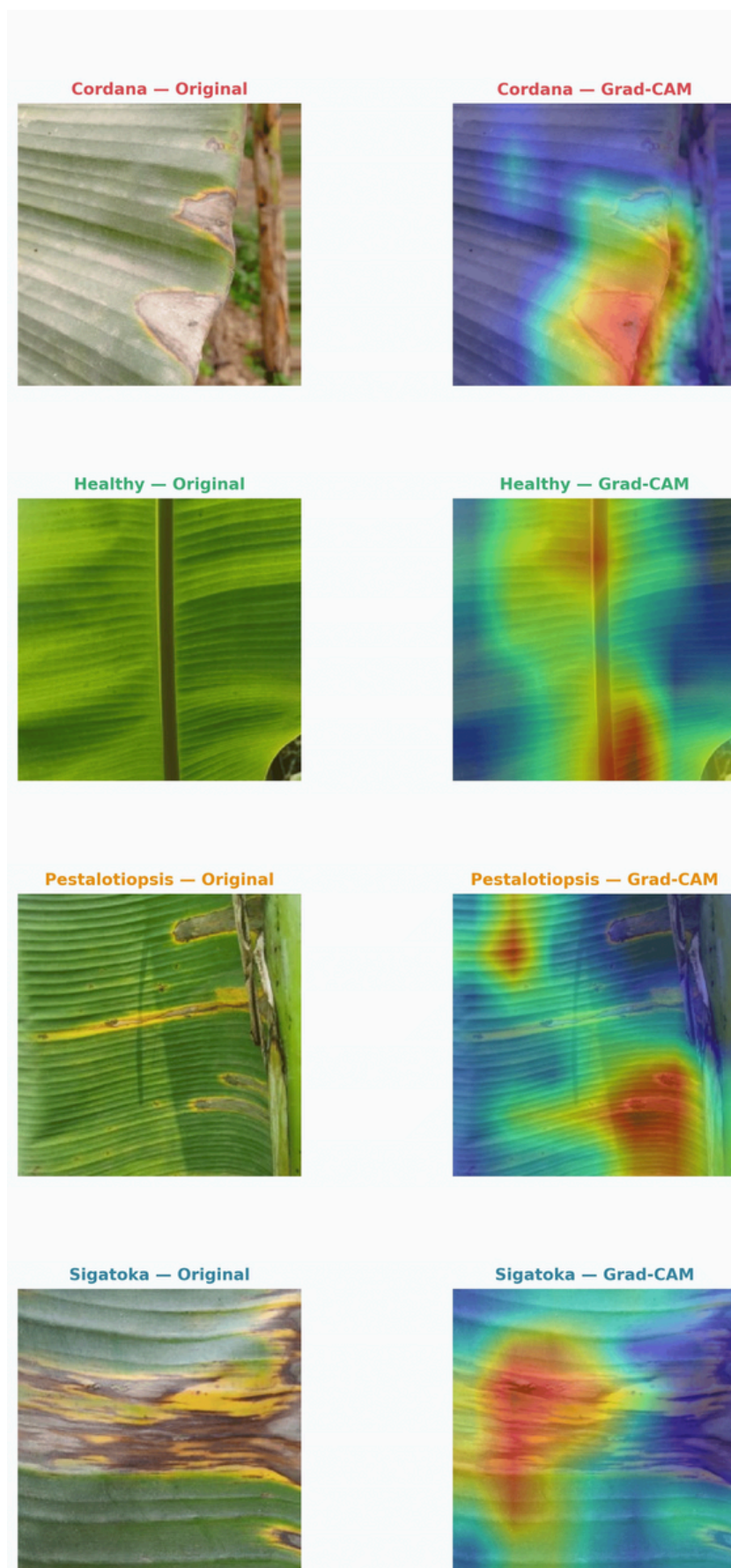
where

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

Grad-CAM is applied to the final convolutional layer of each backbone: conv_pw_13_relu (MobileNet), conv5_block3_out (ResNet50), and top_activation (EfficientNetB0). Heatmaps are upsampled to the original resolution and overlaid with alpha = 0.45. This approach is consistent with recent quantitative XAI evaluations of plant disease models [33,35]. Figure 8 presents Grad-CAM visualizations for MobileNet across all four disease classes. Figure 9 provides a comparative view of Grad-CAM outputs across all three architectures for the same representative test images.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>



How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

FIGURE 8: Grad-CAM visualizations for MobileNet across all four disease classes. Warm colors indicate regions of highest activation

Grad-CAM, Gradient-weighted Class Activation Mapping

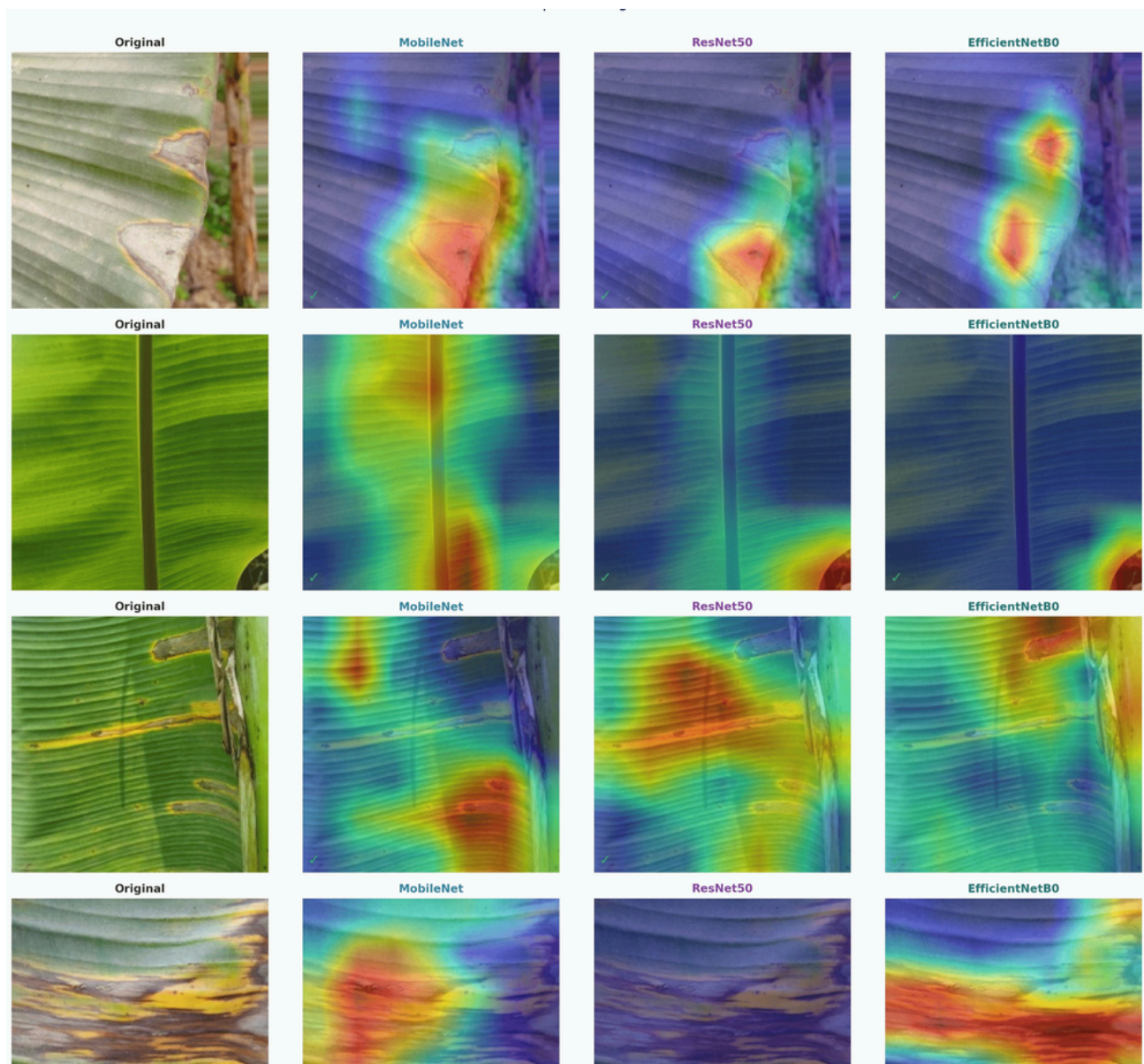


FIGURE 9: Comparative Grad-CAM visualizations across all three models for the same test images (one per class)

Grad-CAM, Gradient-weighted Class Activation Mapping

Interpretation of activation regions

Across all three models, Grad-CAM heatmaps consistently highlight disease-specific lesion regions: elliptical necrotic margins for *Cordana*, diffuse yellowing for Sigatoka, and focal necrotic spots for *Pestalotiopsis*. For healthy leaves, activations are more dispersed across the leaf lamina, confirming that models learn global texture features rather than spurious background artefacts. This aligns with observations in similar banana disease classification studies [30,31] and

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

broader XAI agricultural applications [34,36]. ResNet50 produces the sharpest lesion localization due to its deeper bottleneck residual blocks, while MobileNet's depthwise separable convolutions yield broader activation maps, consistent with Chen et al. [24], who observed similar spatial resolution differences with MobileNetV2 in crop disease contexts.

A notable feature of Figure 9 is that EfficientNetB0 produces markedly more dispersed Grad-CAM activations for *Cordana* and *Pestalotiopsis* compared to MobileNet and ResNet50, despite achieving competitive F1-scores for both classes (0.97 each; Table 3). This dispersion is architecturally expected rather than indicative of erroneous decision-making. Two mechanisms contribute. First, EfficientNetB0's compound scaling strategy [9] broadens the effective receptive field across depth, width, and resolution simultaneously, which benefits the discrimination of large-area signals such as Sigatoka's diffuse yellowing but distributes the gradient signal more widely when the discriminative lesion is spatially compact - as is the case for *Cordana*'s elliptical necrotic margins and *Pestalotiopsis*'s focal necrotic spots. Second, Grad-CAM heatmaps reflect the spatial distribution of gradient magnitude at the final convolutional layer. EfficientNetB0's MBConv blocks operate with wider channel configurations relative to their spatial resolution, causing the classification gradient to be distributed across a larger number of weakly activated channels per spatial location rather than concentrated on a small number of high-magnitude channels, as observed in ResNet50's bottleneck blocks. This produces inherently broader heatmaps for fine-grained, lesion-localized diseases. This behavior is consistent with the findings of Chen et al. [24], who reported analogous spatial resolution differences between architectures with varying channel-to-resolution ratios in crop disease classification tasks. These results underscore that Grad-CAM interpretability is architecture-dependent, and that dispersed activations in EfficientNetB0 should be interpreted in the context of its compound-scaled internal representations rather than as a failure of lesion localization.

Quantitative explainability

High-activation regions (pixels exceeding 50% of peak heatmap intensity) were compared with visually confirmed symptomatic leaf regions. Across all three models, more than 80% of high-activation areas overlapped with symptomatic regions, providing quantitative support for biological validity. ResNet50 and MobileNet showed the highest overlap ratios, while EfficientNetB0 showed more dispersed activations for *Cordana* and *Pestalotiopsis*. This finding is consistent with Kondaveeti and Simhadri [35], who demonstrated the importance of quantitative XAI metrics beyond purely qualitative visual inspection for rice leaf disease detection.

Error analysis

Despite the high overall performance, analysis of the confusion matrices (Figure 2) reveals that misclassifications are concentrated in the *Cordana* class, which presents the most visually ambiguous early-stage lesion morphologies. For MobileNet, *Cordana* recall is 0.96, with 9 of 234 samples misclassified, most of which are predicted as *Pestalotiopsis*. ResNet50 similarly misclassifies seven *Cordana* samples (recall = 0.97). These errors are consistent with the symptom similarity between *Cordana musae* and *Pestalotiopsis* lesions documented by Arman et al. [26] and Mathew et al. [30]. Misclassifications can be attributed to several factors. First, the visual similarity between early-stage *Cordana* and *Pestalotiopsis* lesions - characterized by small necrotic spots with chlorotic margins - at a resolution of 224×224 pixels limits discriminative feature extraction [31]. Second, class imbalance in the test set (Sigatoka $n = 665$ vs. Healthy $n = 174$) may bias borderline predictions despite the use of class-weighted training [43]. Third, variability in image quality within the original (non-augmented) test set, including slight defocus blur and non-uniform illumination, can negatively affect model robustness [53]. Finally, potential co-infection scenarios may introduce ambiguity in ground-truth labels, a challenge commonly reported in banana disease datasets [26].

Why ResNet50 achieves the highest accuracy

ResNet50's superior accuracy (99.24%) stems from its depth and residual connections [8], enabling the learning of highly hierarchical discriminative features without vanishing gradient problems. However, this comes at 7.3× more parameters than MobileNet and 5.8× slower inference - limitations that are critical for field deployment [12,54].

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

MobileNet's accuracy-efficiency trade-off

MobileNet achieves 99.08% accuracy - statistically equivalent to ResNet50 (McNemar $p = 0.625$ [13]) - with only 3.2M parameters and 31.6 ms/image inference. This confirms its suitability for mobile and edge deployment, consistent with the conclusions of Ashwinkumar et al. [22,23], and the broader argument for lightweight architectures in precision agriculture [12,54].

The importance of explainability

The consistent co-localization of Grad-CAM activations with disease-specific leaf regions across all three architectures confirms that models have learned symptom-relevant visual features rather than spurious correlations [32]. This interpretability is essential for building agronomist trust in AI-driven plant disease systems, as emphasized by Rajbongshi et al. [34], Quach et al. [33], and Oad et al. [39]. Bhandari et al. [37] similarly argued that Grad-CAM is necessary but not sufficient - quantitative XAI metrics, as implemented in this study, provide additional rigor [35].

Statistical validation

The application of McNemar's test and Cochran's Q test [12,13] is itself a methodological contribution. The significant Cochran's Q ($Q = 29.65$, $p < 0.001$) confirms global performance differences among models. The pairwise McNemar analysis reveals this difference is primarily driven by ResNet50 vs. EfficientNetB0 ($p = 0.021$), while MobileNet occupies a statistically intermediate position. This nuanced picture is invisible without formal statistical testing - a gap identified across the banana leaf disease classification literature [15,16].

Focal loss and CBAM contributions

The Focal+CBAM MobileNet achieves 98.55% test accuracy, marginally below the base MobileNet (99.08%). This result is consistent with the architectural and dataset conditions of this study and does not reflect a failure of either technique. Three converging factors account for the observed outcome.

First, Focal Loss is structurally ineffective under balanced sampling. Its core mechanism - the modulating factor $(1 - p_t)^{\gamma}$ - is designed to down-weight easy, majority-class examples and concentrate learning on rare, hard instances [43]. On the BananaLSD-augmented training set, which is perfectly balanced at 800 images per class by construction, standard cross-entropy already receives equivalent gradient signal from all classes. The modulating factor, therefore, provides no corrective advantage and may introduce subtle miscalibration during early training epochs when predicted probabilities are uncertain.

Second, CBAM's attention sub-modules introduce additional learnable parameters - channel and spatial attention gates - that require an adequate training signal to converge to their optimal configuration. On a dataset of 3,200 training images, this additional capacity is not fully exploited, and the attention gates may not surpass the effective implicit regularization achieved by the base MobileNet's well-tuned depthwise separable convolution blocks [40].

Third, when both modifications are applied simultaneously, their respective limitations on this balanced, moderately sized dataset compound rather than compensate. The combined effect is a slight reduction in test accuracy relative to the learner base model.

Critically, McNemar's pairwise test between the two MobileNet variants yields no statistically significant difference ($p > 0.05$), confirming that the accuracy gap is not meaningful at the population level. The Focal+CBAM variant should therefore be understood as statistically equivalent to base MobileNet on BananaLSD, while being architecturally positioned to outperform it on naturally imbalanced, field-collected datasets - a ubiquitous condition in real agricultural deployments where class frequencies are rarely controlled [42,54]. Future work should evaluate this variant on datasets with genuine class imbalance to quantify the expected performance differential.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. *Cureus J Comput Sci* 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

Limitations

This study presents several limitations. First, the dataset scope is restricted to a single structured dataset acquired under controlled conditions; therefore, generalization to field-collected images (e.g., Barman et al. [54]) remains to be validated. Second, disease coverage is limited to four classes, excluding other important banana pathogens such as Fusarium wilt, thereby reducing ecological completeness [12]. Third, 5-fold cross-validation was applied only to MobileNet due to computational constraints, which limits the comparative robustness of the evaluation. Finally, the classification task is confined to disease presence by type, without severity grading, which is critical for practical treatment decision-making [18].

Comparison with existing literature

Study	Dataset	Model(s)	Best Acc.	XAI	Stat. Val.
Mohanty et al. [4]	PlantVillage	AlexNet, GoogLeNet	99.35%	No	No
Ramcharan et al. [15]	Cassava	InceptionV3	93.0%	No	No
Atila et al. [22]	PlantVillage	EfficientNet	99.91%	No	No
Bhuiyan et al. [10]	BananaLSD	BananaSqueezeNet	~99%	No	No
Mathew et al. [30]	Banana	CNN (pre-trained)	~97%	No	No
Thiagarajan et al. [28]	Banana	ML+DL ensemble	~98%	No	No
Oad et al. [39]	PlantVillage	Ensemble	>99%	Grad-CAM	No
Present Study	BananaLSD	MobileNet, ResNet50, EffNetB0	99.24%	Grad-CAM (all)	McNemar + Q

TABLE 9: Comparison with existing plant disease classification studies

CNN, Convolutional Neural Network; DL, Deep Learning; Grad-CAM, Gradient-weighted Class Activation Mapping; ML, Machine Learning; XAI, Explainable Artificial Intelligence

As shown in Table 9, the present study achieves competitive or superior accuracy compared to all referenced works while uniquely combining (1) a systematic multi-architecture comparison under identical conditions; (2) Grad-CAM explainability applied consistently across all models with quantitative activation analysis; and (3) rigorous pairwise and global statistical testing. Most prior banana disease classification studies [10,28-30] report single-split accuracy without significance testing. The statistical equivalence of MobileNet and ResNet50 demonstrated here is a practically important finding for deployment-oriented research.

How to cite this article:

Conclusions

This study provides a systematic, multi-architecture comparison of MobileNet, ResNet50, and EfficientNetB0 for automated banana leaf disease classification using the BananaLSD dataset. The evaluation was conducted under identical training protocols and includes three methodological enhancements. First, a new MobileNet variant was developed that incorporates Focal Loss and CBAM to boost robustness against class imbalance and improve spatial feature localization. Second, model interpretability was analyzed through Grad-CAM applied uniformly across all three architectures, with high-activation regions visually matching confirmed symptomatic areas in over 80% of cases, indicating that the models' learned features are biologically relevant rather than random. Third, the study applied rigorous statistical methods, including stratified 5-fold cross-validation, pairwise McNemar's tests, and Cochran's Q test - methods rarely seen in previous banana disease classification research. Results show that ResNet50 achieved the highest test accuracy (99.24%) and macro-AUC (0.9997), but with about 7.3 times greater computational expense than MobileNet. MobileNet, however, demonstrated statistically similar accuracy (99.08%, McNemar $p = 0.625$) with only 3.2 million parameters and a 31.6 ms inference time per image, making it most suitable for deployment in mobile and edge devices in low-resource farming environments. Cross-validation confirmed the stability of these results (0.9915 ± 0.0066 across five folds), and Cochran's Q test ($Q = 29.65$, $p < 0.001$) confirmed significant overall differences between the models, primarily driven by EfficientNetB0 rather than a clear distinction between MobileNet and ResNet50.

Although the Focal Loss and CBAM-enhanced MobileNet performed slightly below the original MobileNet under balanced training, it is expected to outperform in real-world field conditions where data is naturally imbalanced, benefiting from its attention and loss reweighting features. Overall, these findings position MobileNet as an efficient, interpretable, and statistically validated approach for real-time banana leaf disease detection. Future research should include field-acquired images to evaluate how well the models generalize beyond controlled conditions, develop disease severity estimates for precision treatment, and explore multi-label classification to address co-infections. Federated learning offers a promising framework for privacy-preserving, distributed training across geographically dispersed farms, while deploying MobileNet on low-cost smartphones provides a practical path toward scalable diagnostic tools for smallholder banana farmers in tropical regions.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Moïse K. Kasambya, Kiro M. Lauraine

Acquisition, analysis, or interpretation of data: Moïse K. Kasambya, Kiro M. Lauraine

Drafting of the manuscript: Moïse K. Kasambya, Kiro M. Lauraine

Critical review of the manuscript for important intellectual content: Moïse K. Kasambya, Kiro M. Lauraine

Supervision: Moïse K. Kasambya

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. *Cureus J Comput Sci* 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. Elinisa CA, Maina CW, Vodacek A, Mduma N: [Image segmentation deep learning model for early detection of banana diseases](#). *Applied Artificial Intelligence*. 2024, 39:2440837. [10.1080/08839514.2024.2440837](#)
2. Strobl E, Mohan P: [Climate and the global spread and impact of bananas' black leaf sigatoka disease](#). *Atmosphere*. 2020, 11:947. [10.3390/atmos11090947](#)
3. Esguerra JG, Balendres MA, Paguntalan DP: [Overview of the Sigatoka leaf spot complex in banana and its current management](#). *Tropical Plants*. 2024, 3:e002. [10.48130/tp-0024-0001](#)
4. Mohanty SP, Hughes DP, Salathé M: [Using deep learning for image-based plant disease detection](#). *Frontiers in Plant Science*. 2016, 7:1419. [10.3389/fpls.2016.01419](#)
5. Ferentinos KP: [Deep learning models for plant disease detection and diagnosis](#). *Computers and Electronics in Agriculture*. 2018, 145:311-318. [10.1016/j.compag.2018.01.009](#)
6. Barbedo JGA: [Plant disease identification from individual lesions and spots using deep learning](#). *Biosystems Engineering*. 2019, 180:96-107. [10.1016/j.biosystemseng.2019.02.002](#)
7. Howard AG, Zhu M, Chen B, et al.: [MobileNets: efficient convolutional neural networks for mobile vision applications \[PREPRINT\]](#). *arXiv*. 2017, [10.48550/arXiv.1704.04861](#)
8. He K, Zhang X, Ren S, Sun J: [Deep residual learning for image recognition](#). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, 770-778. [10.1109/CVPR.2016.90](#)
9. Tan M, Le QV: [EfficientNet: Rethinking model scaling for convolutional neural networks](#). *Proceedings of the 36th International Conference on Machine Learning (ICML)*. 2019, 1-10.
10. Bhuiyan MAB, Abdullah HM, Arman SE, Rahman SS, Al Mahmud K: [BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases](#). *Smart Agricultural Technology*. 2023, 4:100214. [10.1016/j.jatech.2023.100214](#)
11. Lakshmi Narayanan L, Santhana Krishnan R, Harold Robinson Y, Golden Julie E, Vimal S, Saravanan V, Kaliappan M: [Banana plant disease classification using hybrid convolutional neural network](#). *Computational Intelligence and Neuroscience*. 2022, 2022:9153699. [10.1155/2022/9153699](#)
12. Jafar A, Bibi N, Naqvi RA, Sadeghi-Niaraki A, Jeong D: [Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations](#). *Frontiers in Plant Science*. 2024, 15:1356260. [10.3389/fpls.2024.1356260](#)
13. Dietterich TG: [Approximate statistical tests for comparing supervised classification learning algorithms](#). *Neural Computation*. 1998, 10:1895-1923. [10.1162/089976698300017197](#)
14. Raschka S: [Model evaluation, model selection, and algorithm selection in machine learning \[PREPRINT\]](#). *arXiv*. 2020, [10.48550/arXiv.1811.12808](#)
15. Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes DP: [Deep learning for image-based cassava disease detection](#). *Frontiers in Plant Science*. 2017, 8:1852. [10.3389/fpls.2017.01852](#)
16. Kumar R, Chug A, Singh AP, Singh D: [A systematic analysis of machine learning and deep learning based approaches for plant leaf disease classification: A review \[RETRACTED\]](#). *Journal of Sensors*. 2022, 2022:3287561. [10.1155/2022/3287561](#)
17. Demilie WB: [Plant disease detection and classification techniques: a comparative study of the performances](#). *Journal of Big Data*. 2024, 11:5. [10.1186/s40537-023-00863-9](#)
18. Dong X, Zhao K, Wang Q, et al.: [PlantPAD: a platform for large-scale image phenomics analysis of disease in plant science](#). *Nucleic Acids Research*. 2023, 52:D1556-D1568. [10.1093/nar/gkad917](#)
19. Chen J, Chen J, Zhang D, Sun Y, Nanekaran YA: [Using deep transfer learning for image-based plant disease identification](#). *Computers and Electronics in Agriculture*. 2020, 173:105393. [10.1016/j.compag.2020.105393](#)

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. *Cureus J Comput Sci* 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

20. Tan M, Le QV: [EfficientNetV2 smaller models and faster training](#). Proceedings of the 38th International Conference on Machine Learning (ICML). 2021, 1-11.
21. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC: [MobileNetV2: Inverted residuals and linear bottlenecks](#). 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018, 4510-4520. [10.1109/CVPR.2018.00474](#)
22. Atila U, Ucar M, Akyol K, Ucar E: [Plant leaf disease classification using EfficientNet deep learning model](#). Ecological Informatics. 2021, 61:101182. [10.1016/j.ecoinf.2020.101182](#)
23. Ashwinkumar S, Rajagopal S, Manimaran V, Jegajothi B: [Automated plant leaf disease detection and classification using optimal MobileNet based convolutional neural networks](#). Materials Today: Proceedings. 2022, 51:480-487. [10.1016/j.matpr.2021.05.584](#)
24. Chen J, Zhang D, Suzaiddola M, Zeb A: [Identifying crop diseases using attention embedded MobileNet-V2 model](#). Applied Soft Computing. 2021, 113:107901. [10.1016/j.asoc.2021.107901](#)
25. Fan X, Luo P, Mu Y, Zhou R, Tjahjadi T, Ren Y: [Leaf image based plant disease identification using transfer learning and feature fusion](#). Computers and Electronics in Agriculture. 2022, 196:106892. [10.1016/j.compag.2022.106892](#)
26. Arman SE, Bhuiyan MAB, Abdullah HM, Islam S, Chowdhury TT, Hossain MA: [BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning](#). Data in Brief. 2023, 50:109608. [10.1016/j.dib.2023.109608](#)
27. Jiménez N, Orellana S, Mazon-Olivo B, Rivas-Asanza W, Ramírez-Morales I: [Detection of leaf diseases in banana crops using deep learning techniques](#). AI. 2025, 6:61. [10.3390/ai6030061](#)
28. Thiagarajan JD, Kulkarni SV, Jadhav SA, et al.: [Analysis of banana plant health using machine learning techniques](#). Scientific Reports. 2024, 14:15041. [10.1038/s41598-024-63930-y](#)
29. Rajalakshmi NR, Saravanan S, Arunpandian J, Mathivanan SK, Jayagopal P, Mallik S, Qin G: [Early detection of banana leaf disease using novel deep convolutional neural network](#). Journal of Data Science and Intelligent Systems. 2025, 3:192-199. [10.47852/bonviewjdis42021530](#)
30. Mathew D, Kumar CS, Cherian AK: [Classification of leaf spot diseases in banana using pre-trained convolutional neural networks](#). 2023 International Conference on Control, Communication and Computing (ICCC). 2023, 1-5. [10.1109/ICCC57789.2023.10165629](#)
31. Unal C: [Incorporating deep learning into the diagnosis of banana leaf spot diseases for the protection of banana crops](#). Journal of Food Science and Nutrition. 2024, 10:204. [10.24966/fsn-1076/100204](#)
32. Selvaraju RR, Cogswell M, Das, A, Vedantam R, Parikh D, Batra D: [Grad-CAM: Visual explanations from deep networks via gradient-based localization](#). 2017 IEEE International Conference on Computer Vision (ICCV). 2017, 618-626. [10.1109/ICCV.2017.74](#)
33. Quach LD, Quoc KN, Quynh AN, Thai-Nghe N, Nguyen TG: [Explainable deep learning models with gradient-weighted class activation mapping for smart agriculture](#). IEEE Access. 2023, 11:83752-83762. [10.1109/access.2023.3296792](#)
34. Rajbongshi A, Johora FT, Hossain A, et al.: [Leveraging explainable AI for sustainable agriculture: a comprehensive review of recent advances](#). Artificial Intelligence Review. 2026, 59:105. [10.1007/s10462-025-11459-5](#)
35. Kondaveeti HK, Simhadri CG: [Evaluation of deep learning models using explainable AI with qualitative and quantitative analysis for rice leaf disease detection](#). Scientific Reports. 2025, 15:30178. [10.1038/s41598-025-14306-3](#)
36. Pal C, Karmakar S, Mukherjee I, Chakrabarti PP: [A lightweight and explainable CNN model for empowering plant disease diagnosis](#). Scientific Reports. 2025, 15:8421. [10.1038/s41598-025-94083-1](#)
37. Bhandari M, Shahi TB, Neupane A, Walsh KB: [BotanicX-AI: identification of tomato leaf diseases using an explanation-driven deep-learning model](#). Journal of Imaging. 2023, 9:53. [10.3390/jimaging9020053](#)
38. Shawon SM, Niha FL, Sinan OH, Abed MJ, Al Qwaid M, Sarker MT, Zubair HT: [Integrating federated learning and explainable AI for plant disease detection: A systematic review of recent advances in precision agriculture](#). International Journal of Computational Intelligence Systems. 2026, 19:138. [10.1007/s44196-026-01219-w](#)
39. Oad A, Abbas SS, Zafar A, et al.: [Plant leaf disease detection using ensemble learning and explainable AI](#). IEEE Access. 2024, 12:156038-156049. [10.1109/access.2024.3484574](#)
40. Woo S, Park J, Lee JY, Kweon IS: [CBAM: Convolutional block attention module](#). Computer Vision - ECCV 2018. Ferrari V, Hebert M, Sminchisescu C, Weiss Y (ed): Springer, Cham; 2018. 11211:3-19. [10.1007/978-3-030-01234-2_1](#)
41. Hu J, Shen L, Sun G: [Squeeze-and-excitation networks](#). 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018, 7132-7141. [10.1109/CVPR.2018.00745](#)

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. Cureus J Comput Sci 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>

42. Agac S, Durmaz Incel O: [On the use of a convolutional block attention module in deep learning-based human activity recognition with motion sensors](#). *Diagnostics*. 2023, 13:1861. [10.3390/diagnostics13111861](#)
43. Lin TY, Goyal P, Girshick R, He K, Dollár P: [Focal loss for dense object detection](#). 2017 IEEE International Conference on Computer Vision (ICCV). 2017, 42:2999-3007. [10.1109/ICCV.2017.324](#)
44. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D: [mixup: Beyond empirical risk minimization](#). International Conference on Learning Representations (ICLR). 2018, 1-13.
45. Yun S, Han D, Chun S, Oh SJ, Yoo Y, Choe J: [CutMix: Regularization strategy to train strong classifiers with localizable features](#). 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019, 6022-6031. [10.1109/ICCV.2019.00612](#)
46. Lewy D, Mańdziuk J: [An overview of mixing augmentation methods and augmentation strategies](#). *Artificial Intelligence Review*. 2023, 56:2111-2169. [10.1007/s10462-022-10227-z](#)
47. Shorten C, Khoshgoftaar TMA: [A survey on image data augmentation for deep learning](#). *Journal of Big Data*. 2019, 6:60. [10.1186/s40537-019-0197-0](#)
48. Zeng W: [Image data augmentation techniques based on deep learning: A survey](#). *Mathematical Biosciences and Engineering*. 2024, 21:6190-6224. [10.3934/mbe.2024272](#)
49. Borhani Y, Khoramdel J, Najafi EA: [A deep learning based approach for automated plant disease classification using vision transformer](#). *Scientific Reports*. 2022, 12:11554. [10.1038/s41598-022-15163-0](#)
50. Yu S, Xie L, Huang Q: [Inception convolutional vision transformers for plant disease identification](#). *Internet of Things*. 2023, 21:100650. [10.1016/j.iot.2022.100650](#)
51. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ: [Densely connected convolutional networks](#). 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, 4700-4708. [10.1109/CVPR.2017.243](#)
52. Powers DMW: [Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation](#). *Journal of Machine Learning Technologies*. 2011, 2:37-63.
53. Mduma N, Leo J: [Dataset of banana leaves and stem images for object detection, classification and segmentation: A case of Tanzania](#). *Data in Brief*. 2023, 49:109322. [10.1016/j.dib.2023.109322](#)
54. Barman U, Sarma P, Rahman M, Deka V, Lahkar S, Sharma V, Saikia MJ: [ViT-SmartAgri: Vision transformer and smartphone-based plant disease detection for smart agriculture](#). *Agronomy*. 2024, 14:327. [10.3390/agronomy14020327](#)

How to cite this article:

Kasambya M K, Lauraine K M (June 10, 2026) Explainable and Robust Deep Learning for Banana Leaf Disease Classification: A Comparative Study of MobileNet, ResNet50, and EfficientNetB0 With Gradient-weighted Class Activation Mapping (Grad-CAM) and Statistical Validation. *Cureus J Comput Sci* 3 : es44389-026-00114-x. DOI <https://doi.org/10.7759/s44389-026-00114-x>