International Journal of Research in Multidisciplinary Studies (IJRMS)
Research Scholars Academy (RSA), New Delhi

PlantGuardNet: EfficientNetV2-S with CBAM Attention for Plant Leaf Disease Classification - A Comprehensive Experimental Study

Sajan Kumar
Dept. of Computer Science and Applications
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM)
New Delhi, India
sajankumar.mca25@bvicam.in

Abstract—In the world where crops are destroyed in a huge percentage because of various diseases. This paper presents PlantGuardNet a detailed study of EfficientNetV2-S with CBAM attention for automating leaf disease detection and protecting crops from diseases. Experimented on PlantVillage data, which has (54,305 images, 38 classes) and also used real-field images from PlantDoc dataset for better accuracy. We did augmentation on leafs, and also did cross-dataset generalization experiments. PlantGuardNet achieves 99.62% test accuracy and 0.9962 weighted F1-score on PlantVillage. So, the study confirms CBAM provides +0.09% accuracy improvement with only 0.90ms inference overhead (19.03ms total). Augmentation experiments reveal a negligible effect on controlled data (-0.01%) because leaves are already clean and taken in a controlled environment, but it performed very well on real field data, which is +2.39% gain on real-field PlantDoc images. Cross-dataset evaluation shows a very critical 37.72% domain gap (99.62%→61.90%), making it hard for real-world agricultural uses. Grad-CAM++ were used to provide a heatmap and evidence of disease regions, making it transparent for farmers. This will increase the trust of users in the system.

Keywords— Plant Disease Detection; EfficientNetV2; CBAM Attention; Grad-CAM++; PlantDoc; Deep Learning

I. Introduction

According to world food organization Plant diseases cause huge loss of crops annully which is of 20-40% globally [1], that's why making early automated detection is very important for food protection . Manual detection is not affordable neither available for all farmers mostlysmall farmers because of cost. that's why automated cnn based systems is helpful and also important for them . The PlantVillage dataset [2] has helped in CNN-based progress, but most methods lack systematic experimental validation and testing on real world data .

Prior work by Sharma et al. [6] established strong baselines using VGG16, AlexNet, ResNet, and MobileNetV2 (best: 97.4%) with soft voting ensemble achieving 82.6% on 39 classes. We worked on this foundation by introducing CBAM attention system , Grad-CAM++ explainability, and testing system on both controlled and real-field data.

We know that high accuracy on controlled lab based PlantVillage images may not work same on the real world agricultural field . So our contributions are:

- PlantGuardNet achieves 99.62% on 38-class PlantVillage with
- 21.18M parameters.
- Ablation study confirms CBAM provides consistent +0.09% accuracy improvement on lab images
- Augmentation study shows no effect on lab images but
- +2.39% on real field usages.
- Cross-dataset experiments showed 37.72% difference between lab and field images.
- Grad-CAM++ with assessment of heatmap shows the affected region .

Mohanty et al. [7] showed CNN possibility on PlantVillage achieving 99.35%, establishing deep learning as the standard approach for plant disease classification. Ferentinos [9] confirmed CNN superiority over traditional ML methods (97.8%). Sharma et al. [6] compared VGG16, AlexNet, ResNet, and MobileNetV2 on 15 PlantVillage classes, with soft voting ensemble reaching 82.6%.

CBAM [4] introduced sequential channel and spatial attention, enabling models to focus on discriminative regions. EfficientNetV2 [3] improved training speed and parameter efficiency over earlier architectures. Grad-CAM++ [5] extended gradient-based visualization for more precise class activation mapping. PlantDoc [8] highlighted real-world deployment challenges through a diverse field-image dataset. Table 1 summarizes key prior works and their limitations.

II. Proposed Methodology

A. Datasets

PlantVillage [2]: with 54,305 controlled lab clean images across 38 disease/healthy clattenthaving 14 plant species. Stratified split (seed=42): train 37,997 (70%), val 8,146 (15%), test 8,162 (15%). PlantDoc [8]:with 2,598 real-field images of 17 classes — used for cross-dataset experiments.

B. PlantGuardNet Architecture

PlantGuardNet with use of EfficientNetV2-S as a backbone of our system with CBAM for attention on .leaf affected areas .The backbone extracts feature maps F ∈ R^(1280×H×W) using ImageNet pretrained weights. CBAM readjust features through sequential channel attention Mc and spatial attention Ms:

Mc(F) = σ(MLP(AvgPool(F)) + MLP(MaxPool(F))) Ms(F') = σ(f^7×7([AvgPool(F'); MaxPool(F')]))

Global average pooling produces a 1280-D feature vector for the classification head:

Linear(1280→512)→GELU→Dropout(0.3)→Linear(512→256)→G ELU→Dropout(0.2)→Linear(256→N). Total parameters: 21.18M.

TABLE I. SUMMARY OF RELATED WORK

Ref.	Model	Dataset	Acc.	Limitation
[7]	AlexNet/GoogLeNet	PV-26cls	99.35 %	Lab only
[9]	Custom CNN	PV-25cls	97.8%	No real-world
[6]	VGG16/ResNet	PV-15cls	97.4%	No attention
[6]	Soft Ensemble	PV-15cls	82.6%	15 cls only
Ours	EffV2+CBAM	PV+Plant Doc	99.62 %	Ctrl. env.

PV = PlantVillage data

C. Grad-CAM++ Explainability

Grad-CAM++ [5] used to generates heatmaps using second and third-order gradient weights ffor exact attention so medel know on which part of leaf look for . The heatmap L^c = ReLU(Σ_k α_k^c · A_k) creates a heatmap to only show users which parts the image is actually most important . Then after resizing this heatmap got placed on top of the original leaf image .It gives only rough idea and show us how model actually behaves .

D. Training configuration

"We used AdamW optimizer win model with learning rate 1e-4 to so that model weight could carefully get updated during training. For reducinglearning rate over time we used CosineAnnealingLR, that allowed the model to fine-tune precisely on later epochs. We used CrossEntropyLoss with label smoothing (ε=0.1) to prevent the model from becoming overconfident. All the Images were processed in batches of 32. And all the experiments were run on NVIDIA Tesla T4 16GB GPU. There is a clarification that we trained our model for only 5 epochs because we came to know that EfficientNetV2-S was already trained on 1.2 million ImageNet images. This makes the model capable to recognize basic visual patterns like edges, textures, and different shapes [10]. So it only requires a few rounds to learn plant disease-specific features and pattern . As shown in Table 4, the model reaches 99.32% accuracy in just the first epoch and stops improving after epoch 4 (99.67%→99.67%), which means the model has fully learned. Training more epochs would not improve the results in bigger percentage."

III. RESULTS AND DISCUSSION

A. Models Performance Analysis

Table 2 comparison of PlantGuardNet with all baseline models also added LeafNet [6] architectures in the comparison. And our model achieves the highest accuracy of 99.62% and weighted F1 of 0.9962 .

Fig. 1. Accuracy Comparison: PlantGuardNet vs Baselines

TABLE II. PERFORMANCE ON PLANTVILLAGE TEST SET

Model	Acc.	W-F1	Inf.(ms)	Params
CNN [6]†	61.2%	—	—	—
AlexNet [6]†	72.3%	0.721	—	61.1M
VGG16 [6]†	78.4%	0.782	—	138M
ResNet [6]†	73.8%	0.736	—	11.7M
Ensemble[6]†	82.6%	—	—	—
MobileNetV2	97.4%	0.973	—	3.4M
EffNetV2-S	99.53%	0.9953	18.13	20.97M
PlantGuardNet	**99.62%**	**0.9962**	**19.03**	**21.18M**

Different dataset split; approximate comparison.

B. Effect of CBAM

CBAM effect : CBAM [11] regularly improves all metrics with minor 0.90ms overhead. We know +0.09% accuracy gain is not very hhigh it is because of images were already clean .the +0.0023 macro F1 improvement indicates better minority class handling. However it don't improve accuracy in very large percentage but it makes system transparent.

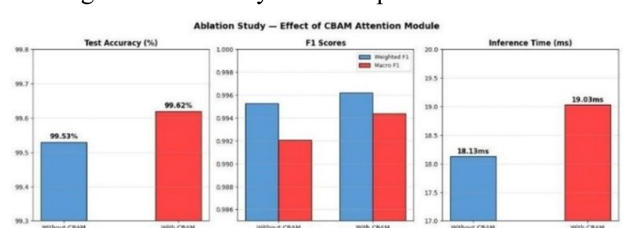

Fig. 2. Ablation Study — CBAM Effect on Accuracy, F1, Inference Time

TABLE III. ABLATION STUDY — EFFECT OF CBAM

Config	Acc.	W-F1	Macro F1	Inf.
No CBAM	99.53%	0.9953	0.9921	18.13 ms
+ CBAM (Ours)	99.62%	0.9962	0.9944	19.03 ms
Gain	+0.09%	+0.0009	+0.0023	+0.90 ms

C. Training Behaviour

Table 4 shows exactly how model improved after every training round (epoch). And in the very first epoch, model already achieved 99.32% accuracy — which is very high for just one round of training. As we earlier discussed this happened cause the model was already pretrained on ImageNet, so it had strong knowledge base before we even started .

When the training got continued, model accuracy kept getting better slowly — from 99.32% to 99.41%, then 99.45%, and finally reached 99.67% at epoch 4. In epoch 5, the accuracy was almost exactly the same (99.67%→99.67%), that shows the model had nothing more to learn. This is called convergence. Our loss values also kept decreasing steadily through training cycle , which shows that our model was learning correctly without any instability. Figure 3 shows these accuracy and loss curves clearly

TABLE IV. TRAINING CONVERGENCE OVER 5 EPOCHS

Ep.	Tr.Loss	Tr.Acc	Val.Loss	Val.Acc
1	0.7170	99.34%	0.6969	99.32%
2	0.7130	99.35%	0.6927	99.41%
3	0.7048	99.57%	0.6922	99.45%
4	0.7035	99.56%	0.6858	99.67%
5	0.6989	99.68%	0.6847	99.67%

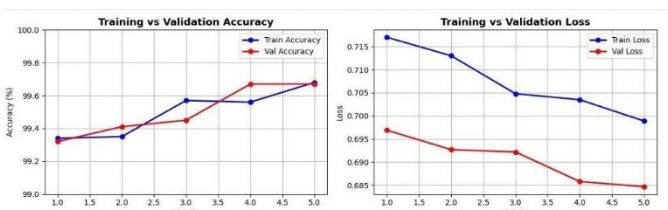

Fig. 3. Training vs Validation Accuracy and Loss Curves

D. Augmentation Experiment

Table 5 shows a key finding — augmentation has no effect on cleam image or controlled PlantVillage images (-0.01%) but it provides meaningful +2.39% difference on real-field data from PlantDoc. This shows that controlled dataset size and ImageNet pretraining provide sufficient generalization without augmentation, but in real-world image roughness and complexity makes augmentation very important

TABLE V. AUGMENTATION EFFECT ON BOTH DATASETS

Dataset	No Aug	Aug	Finding
PlantVillage	99.63%	99.62%	No effect (-0.01%)
PlantDoc	61.90%	64.29%	Helpful (+2.39%)

E. Confusion Matrix Analysis

Figure 4 shows the confusion matrix for 15 classes. Strong diagonal dominance confirms accuracy in classification. Minor misclassifications occur between Tomato Early/Late Blight and Corn Cercospora/Northern Leaf Blight due to overlapping lesion morphology — consistent with prior findings [6].

F. Per-Class F1 Analysis

Figure 5 shows per-class F1 scores. Ten out of 38 classes achieve perfect F1=1.0, including Apple Scab, Apple Black Rot, Blueberry Healthy, and Grape Black Rot. The hardest class is Corn Cercospora Leaf Spot (0.9536) due to visual similarity with Northern Leaf Blight. All 38 classes exceed 0.95 F1-score, demonstrating consistent performance.

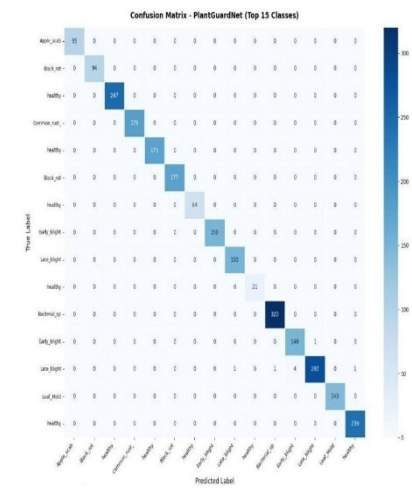

Fig. 4. Confusion Matrix — Top 15 Classes

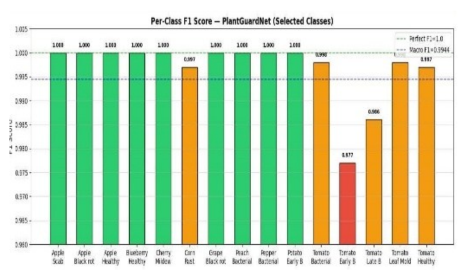

Fig. 5. Per-Class F1 Scores — Selected 15 Classes

G. Cross-Dataset Generalization

Table 6 shows our most important finding — 37.72% accuracy drop when testing on real-field data of PlantDoc images. This shows that controlled dataset accuracy is insufficient for real-world deployment. Strong augmentation recovers +2.39%, and this makes augmentation very essential for real-field applications.

TABLE VI. CROSS-DATASET GENERALIZATION

Experiment	Dataset	Acc.	Notes
PlantGuardNet	PlantVillage	99.62%	Lab images
Direct train	PlantDoc	61.90%	No aug
+ Aug	PlantDoc	64.29%	+2.39%
Domain Gap	Lab→Field	-37.72%	Major gap

H. Grad-CAM++ Visualization

Figure 6 shows Grad-CAM++ heatmaps for six disease categories. Red regions indicate model attention areas. The model tends to focus on central leaf regions where disease areas are typically located. Heatmaps are coarse, a known limitation of gradient-based methods, and should be interpreted as general attention patterns rather than precise lesion boundaries. Future work using segmentation-guided approaches could provide sharper localization.

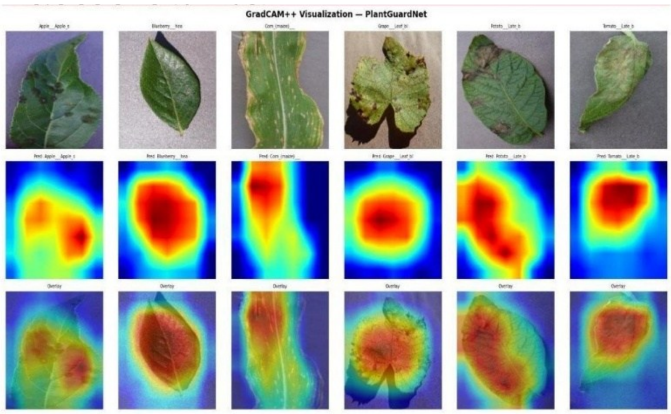

Fig. 6. Grad-CAM++: Original (top), Heatmap (mid), Overlay (bot)

IV. Limitations

This study has several important limitations:

- Dataset scope: All primary results use controlled PlantVillage; real-world performance may vary.
- Domain gap: 37.72% accuracy drop on PlantDoc confirms generalization challenge.
- No cross-validation: Single stratified split used; repeated experiments needed.
- CBAM gain: +0.09% may not be statistically significant without variance analysis.
- Heatmap quality: Grad-CAM++ produces coarse attention maps, not precise boundaries.

V. Conclusion and Future Work

It was observed from the study that PlantGuardNet achieves 99.62% accuracy on 38-class PlantVillage dataset with consistent CBAM improvement (as shown in the result +0.09% accuracy, +0.0023 macro F1). Comprehensive experiments reveal: augmentation has negligible effect on controlled data but a meaningful +2.39% gain on real-field images; a critical 37.72% domain gap exists between lab and field performance; 5-epoch convergence is validated by zero improvement between epochs 4 and 5; Grad-CAM++ provides qualitative disease-region attention evidence. These findings collectively demonstrate that controlled dataset accuracy alone is insufficient — real-world validation is essential for agricultural AI.

Future work: transfer learning from PlantVillage to PlantDoc for improved real-field performance; cross-validation for robust evaluation; semi-supervised annotation; multi-label co-infection classification; segmentation-guided attention for precise lesion localization; mobile deployment evaluation on edge devices.

References

[1] FAO, IFAD, UNICEF, WFP and WHO. 2026. The State of Food Security and Nutrition in the World 2026 – Understanding and addressing the high cost of a healthy diet. Rome. https://doi.org/10.4060/cd8306en.

[2] D. P. Hughes and M. Salathe, “An open access repository of images on plant health to enable the development of mobile disease diagnostics,” Apr. 12, 2016, arXiv: arXiv:1511.08060. doi: 10.48550/arXiv.1511.08060.

[3] J. F. Restrepo-Arias, J. W. Branch-Bedoya, and G. Awad, “Plant Disease Detection Strategy Based on Image Texture and Bayesian Optimization with Small Neural Networks,” Agriculture, vol. 12, no. 11, p. 1964, Nov. 2022, doi: 10.3390/agriculture12111964.

[4] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” Jul. 18, 2018, arXiv: arXiv:1807.06521. doi: 10.48550/arXiv.1807.06521.

[5] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks,” in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2018, pp. 839–847. doi: 10.1109/WACV.2018.00097.

[6] M. Kalyani, M. Manaswini, S. Shevkari, J. Sinkar, L. K. Tyagi, and Y. K. Sharma, “LeafNet: An Intelligent System for Plant Disease Classification with Deep Learning Models,” in 2026 13th International Conference on Computing for Sustainable Global Development (INDIACom), Apr. 2026, pp. 1–6. doi: 10.23919/INDIACom70271.2026.11525518..

[7] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” Front. Plant Sci., vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[8] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, “PlantDoc: A Dataset for Visual Plant Disease Detection,” in Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, in CoDS COMAD 2020. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 249–253. doi: 10.1145/3371158.3371196..

[9] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Comput. Electron. Agric., vol. 145, pp. 311–318, 2018.

[10] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019, pp. 6105–6114.

[11] R. Sharma, M. Mittal, V. Gupta, and D. Vasdev, “Detection of plant leaf disease using advanced deep learning architectures,” Int. j. inf. tecnol., May 2024, doi: 10.1007/s41870-024-01937-4.