Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

Plant Disease Detection from Leaf Images using CNN Models

Sathya S[1] , Samiksha Mukund[2] ,Bhargavi Jagadesh[3]
Student, Department of BCA, St Francis College, Bengaluru[1]
Student, Department of BCA, St Francis College, Bengaluru[2]
Assistant Professor, Department of BCA, St Francis College, Bengaluru[3]

Abstract - Early and effective detection of plant diseases is vital for crop handling and food security. While deep learning models perform well on controlled datasets like PlantVillage, they often underperform on real-world leaf images due to variations in background, lighting, orientation, colour and image quality. This study thoroughly analyzed how well convolutional neural networks (CNNs) classify plant diseases across these environments by comparing a baseline CNN, fine-tuned ResNet18, fine-tuned MobileNetV2 and a three-model heterogeneous ensemble of these networks using probability averaging (soft voting). Models were trained on PlantVillage with data augmentation and tested on the PlantVillage test split and on an independent set of 236 PlantDoc field images spanning 27 classes matched to the PlantVillage labels. Classifiers were measured using accuracy, precision, recall, macro F1-score and confusion matrices, as well as model size and inference time. On PlantVillage, the ensemble achieved 95.41% accuracy. On PlantDoc, however, performance dropped sharply: MobileNetV2 performed best with 24.58% accuracy, compared with 12.29% for the ensemble. These results show that ensembling did not improve generalisation to field images and reveal a substantial domain gap between controlled and real-world conditions. Because the external test set is small and limited to classes shared between the two datasets, the findings should be interpreted cautiously, and no claim of farm-deployment readiness is made. The work provides insight into the accuracy vs cost trade-off of plant disease detection models and the importance of training with field-condition data.

Keywords: Plant disease classification, Domain shift, Field images, Ensemble learning, Convolutional Neural Networks (CNN), PlantDoc.

1. INTRODUCTION

Plant diseases can reduce crop health and quality. Visual diagnosis by experts is useful but may be time-consuming and difficult to scale. Computer vision and deep learning offer a path to classify plant images automatically, potentially supporting earlier screening and crop handling [2, 7, 14]. A key limitation is that a model trained using clean, controlled photographs may learn cues that do not transfer to farm images. Field photographs can differ in background, lighting, image quality, leaf orientation, occlusion, and symptom appearance. Thus, a high score based on a controlled dataset the model independently does not establish reliable field performance [2, 17].

This project compares three individual architectures—basic CNN, ResNet18, and MobileNetV2—with a three-model soft-voting ensemble. The central comparison is whether combining model predictions improves classification on field-style images, not merely whether an ensemble raises accuracy on the controlled test set. Prior research reports very high scores on controlled plant-image benchmarks, while

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

several studies and reviews identify domain shift between laboratory and field conditions as a continuing problem. Mohanty et al. pointed out high PlantVillage performance but a substantial decline on images from other sources [2]. Tian et al. pointed out lower accuracy as evaluation moved from laboratory images toward field images [17]. These findings motivate evaluating on data collected under conditions different from training.

Ensemble methods combine model outputs from multiple classifiers and may improve performance when component errors differ. Prior plant-disease ensembles have combined pretrained CNN architectures, but PlantVillage-only evaluation does not establish that ensembling improves external field performance [12, 13]. The present comparison uses PlantDoc as an external field-image test set and includes MobileNetV2 because its lightweight design is relevant to resource-constrained deployment [3, 6]. The project's contribution is its direct comparison of individual models and their ensemble across a controlled test set and a field-image set. The experiment is small and exploratory.

2. RESEARCH GAP

Deep learning models achieve high accuracy on controlled plant-image benchmarks such as PlantVillage [2, 14], yet studies that evaluate on images from other sources report a marked drop in performance as conditions move from the laboratory toward the field [2, 17]. Ensemble methods are often proposed to improve robustness, but the plant-disease ensembles reviewed here were evaluated on PlantVillage only [12], so it remains unclear whether combining models improves performance on external field images. In addition, few comparisons place a basic CNN, a deeper residual network (ResNet18) and a lightweight architecture (MobileNetV2) alongside their soft-voting ensemble on both a controlled test set and an independent field-image set. This study resolves that gap by evaluating these three models and their ensemble on the PlantVillage test split and on PlantDoc field images [3], and by measuring how much performance falls between the two settings.

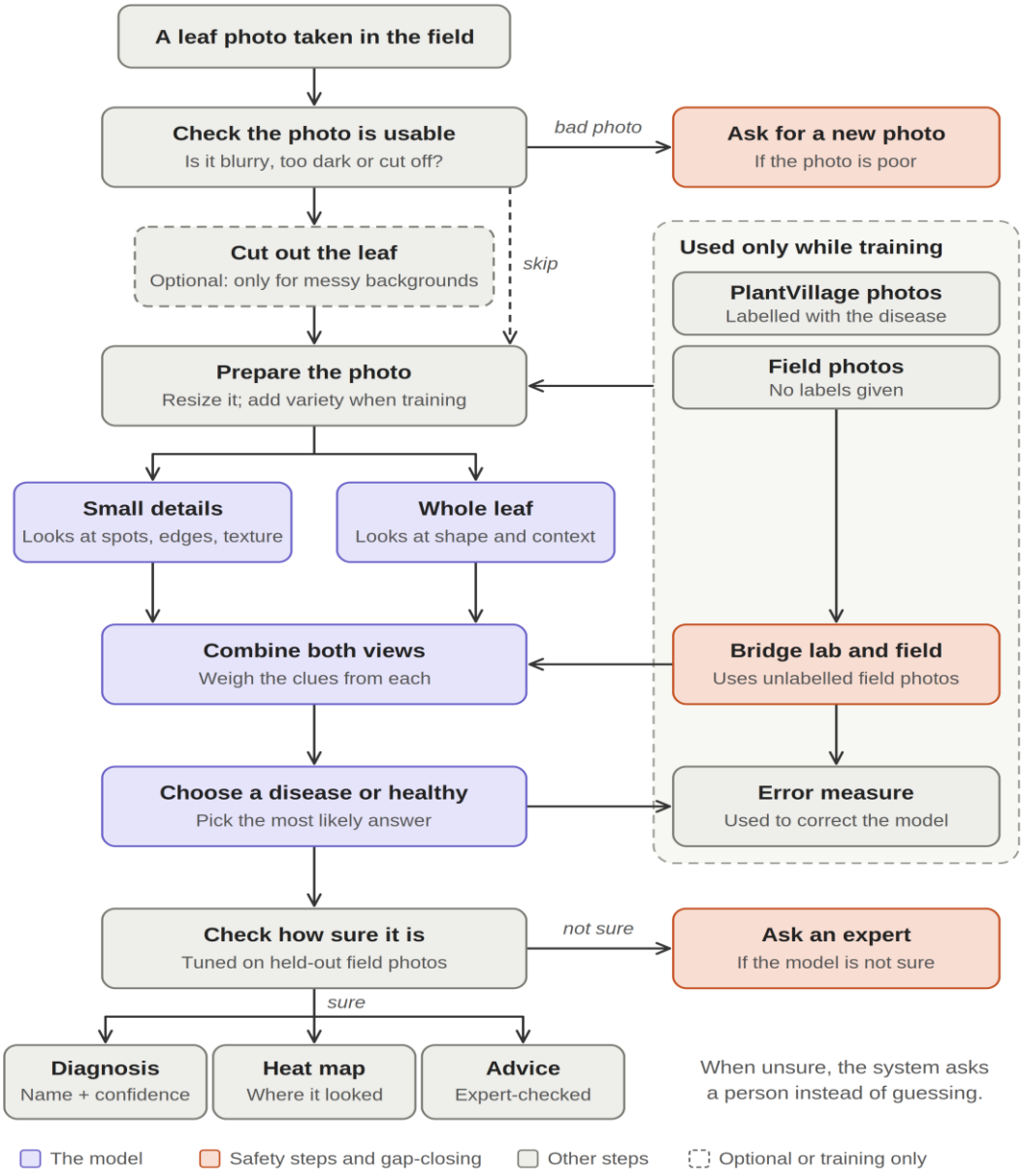
2.1 Related Work

Study	Approach	Relevance
Mohanty et al. (2016) [2]	CNN classification on PlantVillage; external images also considered	Shows that strong controlled-set results may not transfer to images from other sources.
Singh et al. (2020) [3]	Introduced the PlantDoc field-image dataset	Provides an external, less-controlled image source for evaluation.
Too et al. (2019) [14]	Compared fine-tuned CNN architectures on PlantVillage	Demonstrates high performance of transfer learning on a controlled benchmark.
Tian et al. (2024) [17]	Comparison of performance in laboratory, mixed and field conditions	Reports lower performance with more field-like images.
Talaat et al. (2026) [12]	Confidence-weighted ensemble of pretrained models	Relevant to ensembles, evaluation on PlantVillage only.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

3. PROPOSED WORKFLOW



4. AIM AND OBJECTIVES

Aim

To compare individual deep-learning models and a three-model ensemble for plant-leaf disease classification, with specific attention to transfer from PlantVillage images to PlantDoc field images.

Objectives

1. Train and evaluate basic CNN, ResNet18, and MobileNetV2 on the selected PlantVillage classes.
2. The soft-voting ensemble was built by averaging the class-probability outputs from the three individual models.
3. Compare all four approaches based on PlantVillage test data and the PlantDoc external image set using

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

accuracy, macro precision, macro recall, macro F1-score, and confusion matrices.
4. Quantify the change in performance between controlled and field-image evaluation and identify likely failure patterns.
5. Report the limits of the dataset and determine whether results justify further field data collection or domain adaptation.

5. METHODOLOGY

5.1 Study design and datasets

This is a comparative image classification experiment. PlantVillage was used for model design and controlled testing obtained from Kaggle. 54,284 PlantVillage images across 38 classes, with 8,176 images in the test set. PlantDoc is obtained from Github, used as an external field-image evaluation source that has 236 images across 27 classes that is matched to PlantVillage labels [1–3].

5.2 Image preprocessing

Images are reshaped to 224 × 224 pixels and converted to RGB. Preprocessing used the exact same normalization values were used for each model's pretrained weights. Augmentation techniques like rotation, flipping, zoom, and brightness variation were applied only to training images and not to validation or test images [20].

5.3 Models

- Basic CNN: three convolution/pooling blocks followed by fully connected classification layers; trained with Adam and categorical cross-entropy for up to 10 epochs.
- ResNet18: A residual CNN trained and evaluated separately. ResNet18 was trained using the selected pretrained weights, optimizer configuration, learning rate, batch size, and stopping criteria.
- MobileNetV2: MobileNetV2 was evaluated using the selected pretraining and fine-tuning configuration.
- Three-model ensemble: soft voting averages the three models' probability vectors and selects the class showing the highest average probability.

5.4 Evaluation

Model performance was reviewed by accuracy, precision (macro average), recall (macro average), F1-score (macro average) and confusion matrices. Include per-class support (number of test examples). Macro metrics treat all classes equally. These are important when the frequencies of classes are different. Define each metric using the standard equations and specify the averaging convention. Evaluation was based on accuracy, precision, recall, and macro F1-score. Report the train/validation/test split and random seed for a fair evaluation. Keep near-duplicate images, images from the same source, and augmented versions of an image in a single partition. The PlantDoc results are external evaluation results, not another random split of PlantVillage.

6. RESULTS

The values below summarize the experimental results from the conducted evaluations.

6.1 PlantVillage test data

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

Model	Accuracy %	Macro precision %	Macro recall %	Macro F1 %
Basic CNN	90.73	89.09	87.03	87.69
ResNet18	83.62	85.96	82.85	81.91
MobileNetV2	90.36	90.07	87.59	88.18
Three-model ensemble	95.41	95.06	94.44	94.58

Based on the PlantVillage test data, the ensemble has the highest pointed out accuracy (95.41%) among the four approaches. This finding applies to the controlled test set and does not establish superiority on field images.

6.2 Plant Doc field-image test

Model	Accuracy %	Macro precision %	Macro recall %	Macro F1 %
Basic CNN	7.20	8.24	5.30	4.93
ResNet18	5.08	5.80	4.37	3.68
MobileNetV2	24.58	23.62	19.86	18.60
Three-model ensemble	12.29	11.37	9.83	8.69

On PlantDoc, MobileNetV2 has the highest reported accuracy (24.58%), while the ensemble reaches 12.29%. The large performance decrease relative to PlantVillage is the central result. The report should not describe the ensemble as the best overall model; it is best on the controlled test set, while MobileNetV2 is best among these models based on the PlantDoc subset.

7. DISCUSSION

The results shows that the ensemble combines useful predictions on PlantVillage but does not transfer well to the reported PlantDoc sample. MobileNetV2 performs best among the evaluated models on PlantDoc, although its 24.58% accuracy remains low for practical diagnosis. This indicates that the present experiment demonstrates a domain-generalization problem rather than a farm-ready classifier.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

The PlantDoc sample has only 236 images across 27 classes in the draft. With so few examples, class-wise scores can be unstable, and one or two predictions may substantially change a class metric. The very low results may reflect a true domain shift, a label-mapping/output-index error, class mismatch, or some combination. The ensemble's underperformance on PlantDoc could occur if component models make correlated errors or if their confidence scores are poorly calibrated on the external domain. Further analysis using per-class confusion matrices, top-k predictions, probability calibration, and reviews of misclassified images could provide additional insight. The findings indicate that model combination independently can't necessarily improve robustness to domain shift.

8. STUDY LIMITATIONS

• The PlantDoc subset is small (236 images) which limits precision and generalizability.
• The 38 PlantVillage outputs to 27 PlantDoc matched classes require a clearly defined label-mapping rule with class index verification.
• The draft does not report train/validation split proportions, random seed, pretrained weights, batch size, learning rate or training settings for each model.
• Preprocessing and augmentation to look for duplicate or near-duplicate images, and to prevent test information leakage. In the PlantDoc table, each value should be clearly labeled (macro precision/recall/F1 especially) and the confusion matrix matches the accuracy.

9. CONCLUSION

We compare a simple CNN, ResNet18, MobileNetV2, and a soft-voting ensemble for classification of plant-leaf diseases in this work. The ensemble achieves the highest reported PlantVillage accuracy (95.41%) and MobileNetV2 achieves the highest reported PlantDoc accuracy (24.58%). The performance on PlantDoc is much lower than PlantVillage and the field-image result is not improved by the ensemble. Thus, the results obtained on controlled datasets do not always indicate similar performance under field conditions. Since the PlantDoc subset is small and class mapping needs to be explicitly verified, these results should be considered preliminary. Deployment claims need more images from the field, a crosswalk with verified labels, reproducible experimental settings, and independent field testing.

10. FUTURE WORK

1. Expand the field dataset across more images, sites, lighting conditions, and disease stages; obtain expert labels where possible.

2. Validate class crosswalk and output mapping prior to re-running PlantDoc inference.
3. Fine-tune using a training-only subset of target-domain images, then evaluate on a separate held- out field set.
4. Compare simple augmentation, class-balanced training, calibration, and domain-adaptation approaches against the current baselines.

5. Add leaf detection/segmentation only if error analysis shows background clutter is a major source of failure.
6. Build a simple web/mobile demonstration only after the model's limits and confidence behavior are clearly communicated.

11. REFERENCES

[1] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," arXiv:1511.08060, 2015.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

[2] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, article 1419, 2016. doi:10.3389/fpls.2016.01419.
[3] D. Singh et al., "PlantDoc: A dataset for visual plant disease detection," Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, 2020, pp. 249–253. doi:10.1145/3371158.3371196.
[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.
[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of IEEE CVPR, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.
[6] M. Sandler et al., "MobileNetV2: Inverted residuals and linear bottlenecks," Proceedings of IEEE/CVF CVPR, 2018, pp. 4510–4520. doi:10.1109/CVPR.2018.00474.
[7] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018. doi:10.1016/j.compag.2018.01.009.
[8] A. S. Abade, P. A. Ferreira, and F. de B. Vidal, "Plant diseases recognition on images using convolutional neural networks: A systematic review," Computers and Electronics in Agriculture, vol. 185, article 106125, 2021.
[9] P. S. Thakur et al., "Trends in vision-based machine learning techniques for plant disease identification: A systematic review," Expert Systems with Applications, vol. 208, article 118117, 2022.
[10] I. Pacal et al., "A systematic review of deep learning techniques for plant diseases," Artificial Intelligence Review, vol. 57, article 304, 2024.
[11] R. George et al., "Past, present and future of deep plant leaf disease recognition: A survey," Computers and Electronics in Agriculture, vol. 234, article 110128, 2025.
[12] F. M. Talaat, M. Tawfik, and W. M. Shaban, "GreenAid: A confidence-weighted ensemble deep learning system for real-time plant disease detection and management," Scientific Reports, vol. 16, article 19193, 2026. doi:10.1038/s41598-026-57979-0.
[13] A. H. Ali et al., "An ensemble of deep learning architectures for accurate plant disease classification," Ecological Informatics, vol. 81, article 102618, 2024. doi:10.1016/j.ecoinf.2024.102618.
[14] E. C. Too et al., "A comparative study of fine-tuning deep learning models for plant disease identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019. doi:10.1016/j.compag.2018.03.032.
[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations, 2015.
[16] I. Khandelwal and S. Raman, "Analysis of transfer and residual learning for detecting plant diseases using images of leaves," 2018. doi:10.1007/978-981-13-1135-2_23.
[17] Q. Tian et al., "Enhancing practicality of deep learning for crop disease identification under field conditions: Insights from model evaluation and crop-specific approaches," Pest Management Science, vol. 80, no. 11, pp. 5864–5875, 2024. doi:10.1002/ps.8317.
[18] T. G. Dietterich, "Ensemble methods in machine learning," Multiple Classifier Systems, LNCS 1857, 2000, pp. 1–15. doi:10.1007/3-540-45014-9_1.
[19] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009. doi:10.1016/j.ipm.2009.03.002.
[20] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, article 60, 2019. doi:10.1186/s40537-019-0197-0.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)