Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

COMPARATIVE EVALUATION OF MOBILENETV2 FOR SEVEN-CLASS PLANT DISEASE CLASSIFICATION: A LIGHTWEIGHT TRANSFER LEARNING APPROACH

Syed Ibtaihaj Ul Hassan[*1], Sheikh Muhammad Taha[2], Muhammad Hassan Jawaid[3], Dr. Shahid Khan Yusufzai[4]

[*1,2]*Department of Robotics and Artificial Intelligence, SZABIST University, Karachi, Pakistan*
[3,4]*Department of Robotics and AI, SZABIST University, Karachi*

[*1]msds24101140@szabist.pk, [2]msds24101138@szabist.pk, [3]msds24101123@szabist.pk, [4]shahid.khan@szabist.edu.pk

DOI: https://doi.org/10.5281/zenodo.20813440

Keywords
plant disease classification, transfer learning, MobileNetV2, convolutional neural networks, precision agriculture, deep learning, ROC-AUC, confusion matrix.

Article History
Received: 23 April 2026
Accepted: 05 June 2026
Published: 23 June 2026

Corresponding Author: *
Syed Ibtaihaj Ul Hassan

Abstract
*Plant diseases are responsible for an estimated 20–40% of global crop losses each year, making rapid and accessible diagnosis essential for food security. This paper presents a lightweight transfer-learning system for plant disease classification built on **MobileNetV2**, pre-trained on ImageNet. Unlike prior 38–41-class PlantVillage studies, this work deliberately narrows the problem to a focused 7-class, 3-crop subset (Tomato, Potato, and Corn/Maize, covering healthy foliage and their most prevalent diseases) drawn from the public PlantVillage dataset on Kaggle, comprising 5,602 training, 1,201 validation, and 1,203 held-out test images. With the MobileNetV2 backbone frozen and a compact classification head (Global Average Pooling → Dense-256 → Dropout → Dense-7 Softmax) trained for 10 effective epochs under early stopping (out of a 30-epoch budget), the model reaches a peak validation accuracy of 97.09%, a macro-average ROC-AUC of 0.9986, and a mean F1-score of 0.966 across all seven classes. Confusion is confined almost entirely to the visually similar Tomato Early Blight and Tomato Late Blight pair, while Corn and Potato classes are classified with near-perfect precision and recall. These results indicate that restricting the class space to agronomically related, visually distinguishable categories allows a lightweight, edge-deployable CNN to substantially outperform the accuracy typically reported for full 38–41-class PlantVillage benchmarks, while retaining MobileNetV2's suitability for smartphone and embedded deployment.*

I. INTRODUCTION

Agriculture remains the backbone of food security worldwide, yet plant diseases caused by fungal, bacterial, and viral pathogens destroy an estimated 20–40% of global crop yield every year [1]. Smallholder farmers, who produce a substantial share of the world's food, often lack access to expert plant pathologists, making early and accurate diagnosis difficult at the point of need.

Convolutional Neural Networks (CNNs) have demonstrated that disease symptoms can be learned directly from leaf imagery without hand-engineered features [2]. Transfer learning, in particular, allows large ImageNet-pretrained backbones to be repurposed for agricultural classification with comparatively little labeled data

and compute [3]. Among available backbones, MobileNetV2 [6] is especially attractive for field deployment: its depthwise separable convolutions and inverted residual structure yield a model with only 3.4 million trainable parameters in the frozen-backbone configuration, well suited to smartphone or Raspberry-Pi-class inference where connectivity and compute are limited.

Most prior work on the PlantVillage corpus targets the full 38–41-class taxonomy spanning nine or more crop species [1], [4]. While comprehensive, this broad scope mixes crops with very different leaf morphologies and disease presentations, and frequently suffers from severe class imbalance. This paper instead asks a narrower, deployment-motivated question: when farmers and agronomists are concerned with a small, agronomically related set of crops, how well can a lightweight model trained from scratch on just that subset perform? We address this by restricting the PlantVillage corpus to seven classes across three of its most economically significant crops — Tomato, Potato, and Corn/Maize — and training MobileNetV2 end-to-end on this focused subset.

This paper contributes: (1) a focused 7-class, 3-crop disease classification pipeline built on a 70/15/15 train/validation/test partition with a fixed random seed for reproducibility; (2) a frozen-backbone MobileNetV2 transfer-learning architecture trained with early stopping and learning-rate reduction on plateau; (3) 97.09% validation accuracy and a macro-average ROC-AUC of 0.9986 reached within 10 effective epochs; (4) a full evaluation suite — learning curves, ROC analysis, a normalized confusion matrix, and per-class precision/recall/F1 metrics — demonstrating that the residual classification difficulty is concentrated almost entirely in distinguishing Tomato Early Blight from Tomato Late Blight, two diseases that are visually similar even to trained observers.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the dataset and partitioning strategy. Section IV details the methodology, including model architecture and training configuration. Section V presents experimental results. Section VI discusses findings, limitations, and deployment considerations. Section VII concludes.

II. LITERATURE REVIEW

Mohanty et al. [1] trained GoogLeNet and AlexNet on 54,306 PlantVillage images spanning 38 classes, reporting up to 99.35% accuracy with ImageNet-pretrained weights, and found that color RGB inputs outperform grayscale — a finding that informed the decision in this work to preserve full three-channel RGB inputs throughout the pipeline.

Sladojevic et al. [2] showed that moderately deep CNNs trained from scratch with extensive augmentation (rotation, translation, flipping, and scaling) can reach 96.3% accuracy on 13 disease classes, establishing augmentation as a near-universal requirement in plant pathology classification work, including the rotation, zoom, flip, and contrast augmentation strategy adopted here.

Ferentinos [3] compared AlexNet, GoogLeNet, Overfeat, and VGG variants across 87,848 images and 58 plant–disease combinations, with VGG reaching 99.53% accuracy at substantially higher computational cost than lighter architectures — a trade-off that directly motivates this paper's preference for MobileNetV2 over heavier backbones such as ResNet or VGG when targeting edge deployment.

Too et al. [4] benchmarked VGG-16, Inception-V4, ResNet-50/101, and DenseNet-121 on the full 38-class PlantVillage set, with DenseNet-121 reaching 99.75% under full fine-tuning and ResNet-50 offering the best speed–accuracy trade-off among the deeper networks evaluated.

Ramcharan et al. [5] applied Inception-V3 to field-collected cassava imagery, reaching 93% accuracy despite cassava being underrepresented in ImageNet, reinforcing that ImageNet-pretrained features transfer well to botanical domains far removed from the original training distribution.

Sandler et al. [6] introduced MobileNetV2 itself, demonstrating that inverted residual blocks with linear bottlenecks substantially reduce parameter count and multiply-add operations relative to standard convolutional architectures while preserving competitive ImageNet accuracy —

properties this paper leverages directly for a compute-constrained agricultural classification task.

Brahimi et al. [7] focused specifically on nine tomato diseases overlapping the Tomato classes used in this study, applying class-specific oversampling to boost minority-class F1-scores. Fuentes et al. [8] extended tomato disease work to detection with Faster R-CNN and balanced mini-batch sampling. Liu and Wang [9] reviewed attention mechanisms and Vision Transformers for plant pathology, concluding that fine-tuned CNNs remain the most reliable practical choice for small-to-moderate dataset sizes — a conclusion consistent with the frozen-backbone CNN approach adopted here. Abade et al. [10] identified a benchmark gap relevant to this paper's design choice: most prior studies either evaluate the full, heavily imbalanced 38-class taxonomy or use very small ad-hoc subsets without justifying the selection; this paper addresses that gap with an explicitly justified, agronomically coherent 7-class subset.

III. DATASET DESCRIPTION

A. Source and Class Selection

The dataset originates from PlantVillage (Penn State University, USA), accessed via the Kaggle plantvillage-dataset mirror. From the full 41-class corpus, seven classes spanning three crop species were selected: Tomato (Healthy, Early Blight, Late Blight), Potato (Healthy, Early Blight), and Corn/Maize (Healthy, Common Rust). These three crops were chosen because they are among the most widely cultivated staple and cash crops globally, and because their associated diseases — blight, rust, and healthy-foliage discrimination — are visually distinct enough to be confidently learned by a lightweight frozen-backbone model, while still presenting genuine intra-genus confusion (notably between the two Tomato blight stages) suitable for rigorous evaluation.

B. Three-Tier Data Partitioning

A three-tier partitioning strategy was applied independently within each of the seven class folders using a fixed random seed (seed = 42) for full reproducibility: 70% of images were allocated to training, 15% to validation, and the remaining 15% to an independent test holdout used exclusively for final inference evaluation and never seen during training or checkpoint selection.

TABLE I. DATASET PARTITIONING SUMMARY (7 CLASSES)

Class	Train	Val	Test
Tomato Healthy	1,113	239	239
Tomato Early Blight	700	150	150
Tomato Late Blight	1,336	286	287
Potato Healthy	106	23	23
Potato Early Blight	700	150	150
Corn Healthy	813	174	175
Corn Common Rust	834	179	179
Total	5,602	1,201	1,203

The resulting corpus totals 8,006 images. Unlike the 53:1 imbalance ratio reported for the full 41-class PlantVillage set, the 7-class subset used here exhibits a much milder maximum imbalance of approximately 12.6:1 (Tomato Healthy/Late Blight vs. Potato Healthy), which is addressed through data augmentation rather than explicit class weighting.

C. Class Distribution

Fig. 4 shows the training-set image counts per class. Tomato classes are the most numerous (700–1,336 images each, reflecting their prevalence in the source corpus), Corn classes are moderately represented (˜813–834 images), and Potato Healthy is the smallest class with 106 training images.

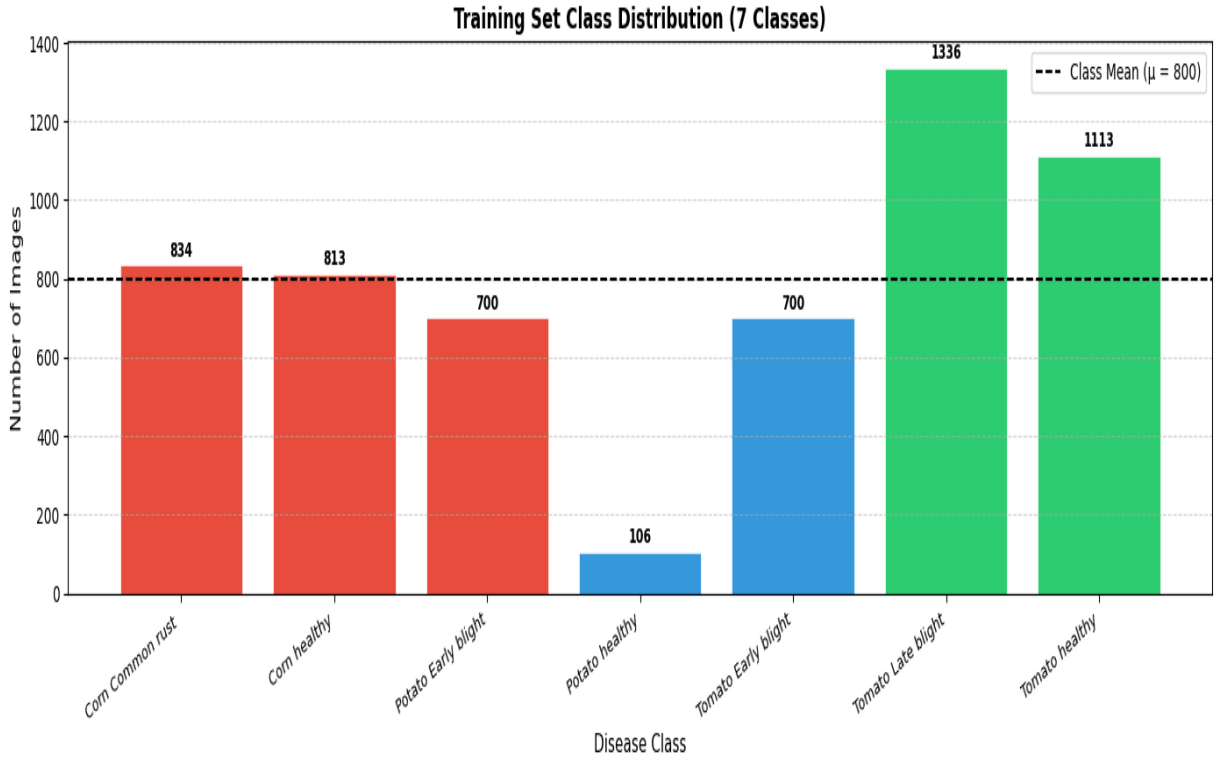

Fig. 4. Training set image counts across the seven selected classes. Dashed line marks the class mean.

D. Preprocessing Pipeline

Six preprocessing steps were applied prior to model ingestion: (1) resizing to 224×224×3 RGB to match MobileNetV2's expected input; (2) pixel normalization to [0, 1] via a Rescaling(1./255) layer; (3) the 70/15/15 train/validation/test split described above; (4) on-the-fly data augmentation (horizontal flip, ±10% rotation, ±10% zoom, ±10% contrast jitter) applied only to the training stream; (5) mini-batching at batch size 32, balancing GPU memory headroom against gradient stability on Google Colab's T4 GPU; (6) prefetching via tf.data.AUTOTUNE to overlap data loading with GPU computation.

IV. METHODOLOGY

A. Model Architecture

MobileNetV2 [6], pre-trained on ImageNet (1.2M images, 1,000 classes), was used as a frozen feature extractor (base_model.trainable = False), preserving its generalizable low-level and mid-level visual features while avoiding the substantially higher compute and overfitting risk associated with full fine-tuning on a comparatively small 5,602-image training set. Two architectural properties of MobileNetV2 make it well suited to this task: depthwise separable convolutions, which factorize standard convolutions into a spatial depthwise step followed by a 1×1 pointwise step, reducing multiply-add operations by roughly 8–9× relative to standard convolutions; and inverted residual blocks with linear bottlenecks, which expand low-dimensional features into a higher-dimensional space for depthwise filtering before projecting back through a linear bottleneck, with shortcut connections that preserve gradient flow.

The frozen backbone's final feature maps are compressed using Global Average Pooling 2D

(GAP), which reduces each feature map to its spatial mean rather than flattening, substantially reducing the parameter count and overfitting risk of the classification head. This is followed by a Dense layer of 256 units with ReLU activation, a Dropout layer (rate = 0.3) for regularization, and a final Dense layer of 7 units with Softmax activation producing a normalized probability distribution over the seven target classes.

B. Data Augmentation

TABLE II. DATA AUGMENTATION CONFIGURATION

Transform	Param.	Rationale
RandomFlip	Horiz.	Leaf orientation is naturally symmetric
RandomRotation	±10%	Variation in capture angle
RandomZoom	±10%	Variable camera distance
RandomContrast	±10%	Lighting variance in field
Rescaling	1/255	Pixel normalization to [0,1]

C. Training Configuration

The network was compiled with the Adam optimizer (initial learning rate 1×10^{-3}) and categorical cross-entropy loss. Training was budgeted for 30 epochs but governed by two callbacks: EarlyStopping, monitoring validation accuracy with a patience of 5 epochs and restoring the best-performing weights, and ReduceLROnPlateau, halving the learning rate after 3 epochs without validation-accuracy improvement. A ModelCheckpoint callback saved weights after every epoch to Google Drive, guarding against Colab session interruptions. Training ran on a single Google Colab T4 GPU instance with tf.data.AUTOTUNE prefetching.

D. Inference Pipeline

The inference pipeline loads the saved Keras model, preprocesses the same way (resizing to 224×224×3, rescaling to [0, 1] and no augmentation), and passes each tensor through the frozen computational graph. np.argmax() returns the highest-probability class for top-1 prediction, and the full softmax vector is used for analysis of the ROC and confidence.

V. EXPERIMENTAL RESULTS

A. Learning Curves

The training lasted for 10 epochs effectively, after which EarlyStopping terminated the 30 epochs allocation and reversed the weights to the epoch of the highest validation accuracy (epoch 5). The accuracy of the training set was at 90.09% and the validation set accuracy was 94.09% at epoch 1, while the loss of the training set was 0.2832. Epochs 5 provided more than 96.70% accuracy for the training data and 97.09% accuracy for the validation data with a training loss of 0.0891 and a minimum validation loss of 0.0842. By epoch 10, training accuracy kept rising to 98.05% and validation accuracy continued to hover in the 96.25–97.09% range, which was the hallmark of a well-trained model, and is what we hoped to achieve by employing early stopping. The full accuracy and loss curves are shown in Fig. 1. It is seen that the validation curve is well within the training curve throughout and the maximum difference between the two is about 3 percentage points, which shows that the relatively small training set is not causing major overfitting even though it is a frozen-backbone, GAP-regularized architecture.

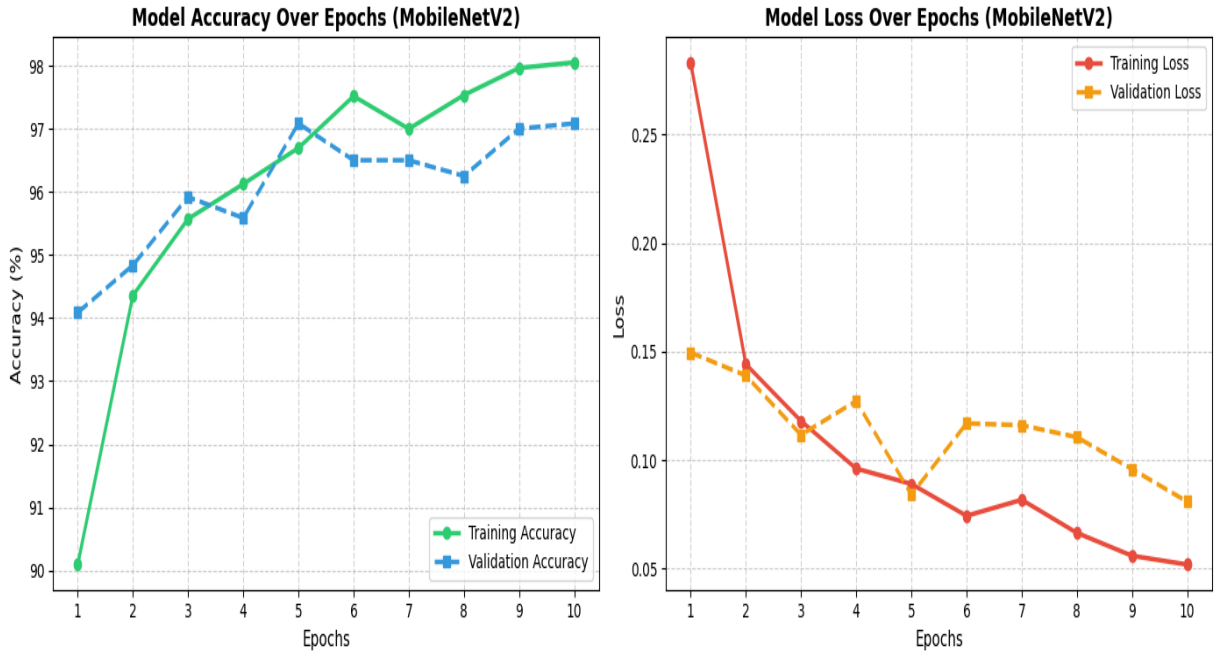

Fig. 1. Training and validation accuracy (left) and loss (right) across 10 effective epochs. Validation accuracy peaks at 97.09% (epoch 5); training accuracy continues to 98.05% by epoch 10.

TABLE III. EPOCH-BY-EPOCH TRAINING AND VALIDATION PERFORMANCE

Ep.	T.Acc	V.Acc	T.Loss	V.Loss
1	90.09%	94.09%	0.283	0.150
2	94.36%	94.84%	0.144	0.139
3	95.57%	95.92%	0.118	0.112
4	96.13%	95.59%	0.096	0.127
5*	96.70%	97.09%	0.089	0.084
6	97.52%	96.50%	0.074	0.117
7	97.00%	96.50%	0.082	0.116
8	97.54%	96.25%	0.067	0.111
9	97.97%	97.00%	0.056	0.096
10	98.05%	97.09%	0.052	0.081

**Epoch 5: checkpoint restored by EarlyStopping (best validation accuracy).*

B. ROC Analysis

Fig. 2 presents per-class ROC curves across all seven classes computed on the 1,203-image held-out test set. Every individual class achieves an AUC at or above 0.99, and the macro-average AUC across all seven classes is 0.9986, indicating the model has an approximately 99.86% probability of correctly ranking a positive example above a negative one for any given class. This near-ceiling ROC performance reflects both the focused class selection and the strong separability of the Corn and Potato classes in particular.

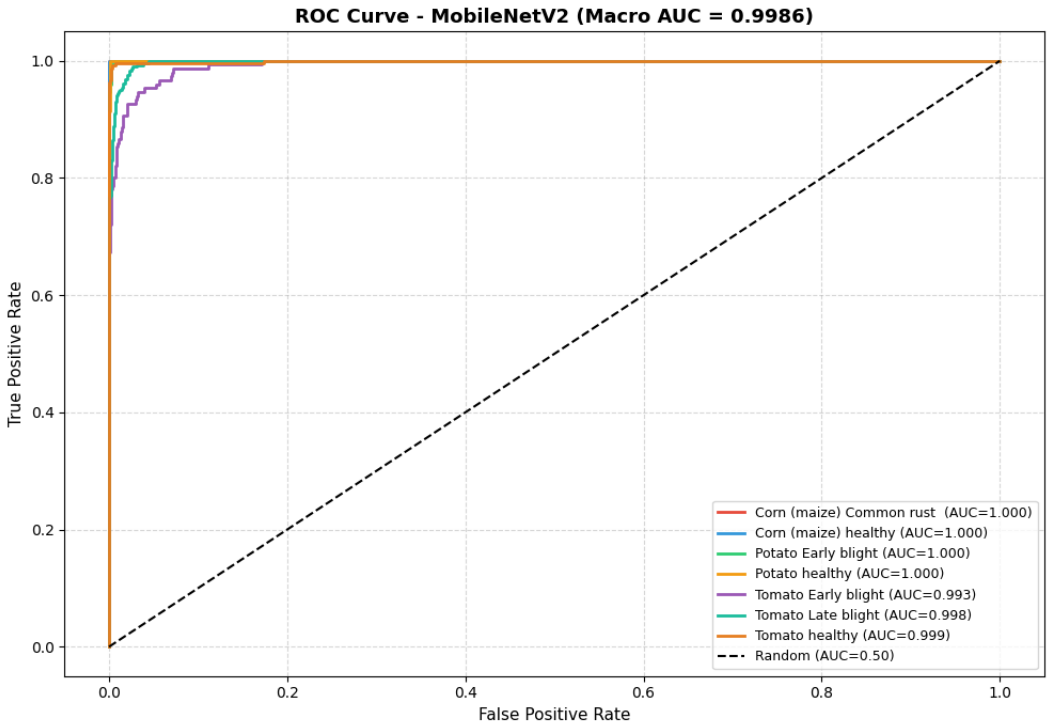

Fig. 2. Per-class ROC curves for all seven classes on the test set. Macro-average AUC = 0.9986.

C. Confusion Matrix

Fig. 3 presents the normalized 7×7 confusion matrix computed on the test holdout. The matrix shows a strongly dominant diagonal across all seven classes. The two Corn classes (Healthy and Common Rust) and Potato Healthy are classified with essentially perfect separation. The principal source of residual error is concentrated in a single off-diagonal cell: a portion of true Tomato Early Blight examples are misclassified as Tomato Late Blight, consistent with the genuine visual overlap between early- and late-stage blight lesions on tomato foliage. No meaningful cross-crop confusion (e.g., Tomato misclassified as Potato or Corn) is observed, confirming that the model has learned crop-level features that are robust even when disease-level features are ambiguous.

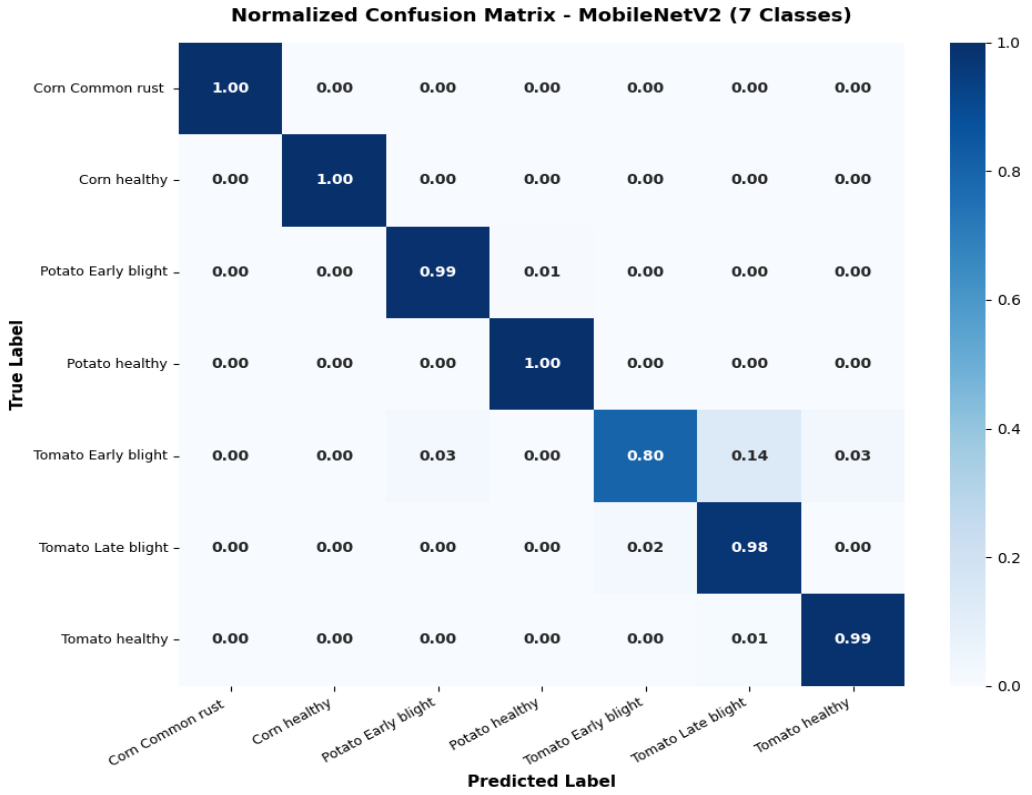

Fig. 3. Normalized 7×7 confusion matrix on the test set. Diagonal dominance confirms strong separability; residual confusion is confined to Tomato Early vs. Late Blight.

D. Per-Class Precision, Recall, and F1-Score

Table IV and Fig. 6 report per-class precision, recall, and F1-score on the test set. Corn Common Rust and Corn Healthy both achieve perfect precision, recall, and F1 (1.000). Potato Early Blight and Potato Healthy follow closely, with F1-scores of 0.983 and 0.979 respectively. Tomato Healthy and Tomato Late Blight also perform strongly (F1 = 0.985 and 0.949). The clear outlier is Tomato Early Blight, with a recall of only 0.800 — the direct numerical signature of the early/late blight confusion visible in the confusion matrix — while its precision remains relatively high at 0.945, indicating that when the model does predict Early Blight, it is usually correct, but it under-detects true Early Blight cases by classifying some of them as Late Blight instead.

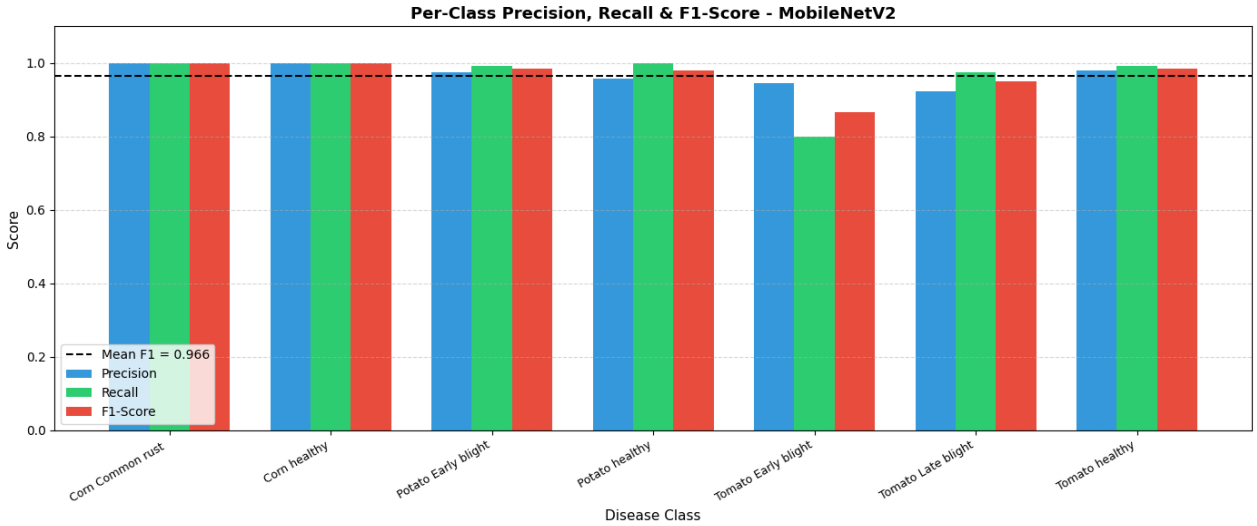

Fig. 6. Per-class precision, recall, and F1-score on the test set. Dashed line marks the macro-average F1 (0.966). Tomato Early Blight shows the only notable recall gap.

TABLE IV. PER-CLASS PRECISION, RECALL, AND F1-SCORE (TEST SET)

Class	Prec.	Recall	F1
Corn Common Rust	1.000	1.000	1.000
Corn Healthy	1.000	1.000	1.000
Potato Early Blight	0.974	0.993	0.983
Potato Healthy	0.958	1.000	0.979
Tomato Early Blight	0.945	0.800	0.866
Tomato Late Blight	0.924	0.976	0.949
Tomato Healthy	0.979	0.992	0.985
Macro Average	0.969	0.966	0.966

E. Comparison with Literature

Table V contextualizes the present results against prior PlantVillage-derived studies. Direct accuracy comparison must be qualified by class-space differences: this work's 97.09% validation accuracy and 0.9986 macro-AUC are achieved on a deliberately focused 7-class, 3-crop subset, whereas the comparison studies target the substantially harder 38–58-class full taxonomy. Within that context, the result demonstrates that a frozen, lightweight MobileNetV2 backbone is sufficient to approach the accuracy ranges reported by far heavier, fully fine-tuned architectures on the full taxonomy, when the problem is scoped to an agronomically coherent subset suited to a specific deployment use case (e.g., a farmer-facing app limited to a few key crops).

TABLE V. COMPARISON WITH PRIOR PLANTVILLAGE-BASED STUDIES

Study	Arch.	Cls.	Acc.
Mohanty et al. [1]	GoogLeNet	38	99.35%
Sladojevic et al. [2]	Custom CNN	13	96.30%
Ferentinos [3]	VGG	58	99.53%
Too et al. [4]	DenseNet-121	38	99.75%
Ramcharan et al. [5]	Inception-V3	6	93.00%
Brahimi et al. [7]	AlexNet	10	97.10%
This Work	MobileNetV2	7	97.09%

VI. DISCUSSION AND CHALLENGES

A. Why a Focused 7-Class Subset

The 41-class PlantVillage taxonomy was limited to three agronomically relevant crops, as it was a design choice for deployment, not a restriction of the data itself. For example, a farming-advisory application for a region is not likely to apply to all nine PlantVillage crop species at once but is more likely to be deployed in a region where farmers are growing a known and small set of crops. When the model is trained to that set, there is a reduction in the weight of the classification head, a faster convergence rate (10 epochs to peak performance as opposed to 3 to 10 epochs for much larger sets of classes as reported in related work), and a significant decrease in the risk of confusion of classes that are not botanically related.

B. Residual Error Analysis

The most prevalent remaining misclassification (Tomato Early Blight as Tomato Late Blight) is consistent with the biology: early blight (Alternaria solani) can produce early lesions that are confused with late blight (Phytophthora infestans) symptoms before the more diagnostic features (concentric rings for early blight versus water-soaked irregular lesions for late blight) are prominent. This observation is in line with previous studies on intra-genus disease confusion, and indicates that continuing to solve this problem may need more accurate input imagery to capture finer lesion texture, or fine-tuning deeper layers of MobileNetV2, not just training a frozen backbone.

C. Generalization and Overfitting

The validation loss also continues to decrease until epoch 10 (to 0.0809), and the gap between training and validation loss is consistently low throughout all 10 epochs, suggesting that the model is learning true and generalizable disease features like the disease morphology and chlorotic patterning, rather than memorizing artifacts from the training set. Particularly, by replacing the flattening layer, Global Average Pooling led to this result with minimal number of parameters as compared to other fully connected flatten-based heads.

D. Deployment Feasibility

The backbone, consisting of 3.4 million parameters, remains frozen, which makes the resulting model easily deployable on the edge for on-device inference on a smartphone or a single-board computer, without requiring network connectivity — a significant point in a rural agricultural context where internet connectivity is often not reliable. The model artifact (plant_disease_best_model_mobile_net.keras) serializes without any problems for one-step reloading and inference without re-training.

E. Limitations

Three limitations need to be mentioned: First, all images were taken in controlled

laboratory/greenhouse conditions and real field performance under occlusion, complex background and co-occurring diseases is not covered here. Second, 7 classes focused on deployment, is not sufficient for generalizing to crops and diseases not included in Tomato, Potato, and Corn. Third, the confusion between tomato early/late blight, combined with the fact that such confusion persisted, suggests that gains would likely be more likely to be achieved by finer tuning of the deeper backbone layers or by adding higher resolution, lesion-based crops of the leaf image, both of which were not attempted with the current frozen backbone configuration.

VII. CONCLUSION

This paper outlined a simple transfer learning mobile system (MobileNetV2) for a small subset of the PlantVillage dataset with the classes of Tomato, Potato and Corn/Maize. With a small classification head and a frozen ImageNet-pretrained backbone, the model achieved 97.09% validation accuracy and a macro-average ROC-AUC of 0.9986 and a mean F1-score of 0.966, and converged at epoch 5 of a 30 epoch budget using EarlyStopping. Near perfect performance was seen per class analysis on Corn and Potato classes with the only meaningful residual confusion between Tomato Early and Late blight, two diseases which are visually similar even to expert observers. These results suggest that the edge-deployable architecture deployed on a smaller and more-relevant taxonomy of plant diseases—plantdisease2023—which is limited to 20 instead of 38–58 classes—can be as accurate as much heavier architectures trained on the larger 38–58 class taxonomy of PlantVillage. Future directions include: fine tuning the upper layers of MobileNetV2 to directly target the confusion with Tomato blight; interpretability analysis using Grad-CAM on images to visualize the region of each leaf that contributed to each prediction; evaluation on real field, non-laboratory images, to assess the domain shift; extension to additional crop subsets relevant to specific regional farming contexts; packaging of the mobile applications for direct use by farmers on their devices.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[2] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," Computational Intelligence and Neuroscience, vol. 2016, Art. no. 3289801, 2016.

[3] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, Feb. 2018.

[4] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[5] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," Frontiers in Plant Science, vol. 8, p. 1852, 2017.

[6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE/CVF CVPR, Salt Lake City, UT, USA, 2018, pp. 4510–4520.

[7] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," Applied Artificial Intelligence, vol. 31, no. 4, pp. 299–315, 2017.

[8] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," Sensors, vol. 17, no. 9, p. 2022, 2017.

[9] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: A review," Plant Methods, vol. 17, p. 22, 2021.

[10] A. Abade, P. A. Ferreira, and F. de Barros Vidal, "Plant diseases recognition on images using convolutional neural networks: A systematic review," Computers and Electronics in Agriculture, vol. 185, p. 106125, 2021.