Al-Nahrain Journal for Engineering Sciences

Al-Nahrain J. Eng. Sci. 29(1) 103-110, 2026
Original Article

AI-driven crop disease detection with efficient NetB3 hybrids for sustainable agriculture

Rituraj Jain[1], Kuldeep Tapodhan [2], Shubham Bhalala [3], Yash Jotangiya [4] and Amar Davda[5].

1,2,3,4,5Department of Information Technology, Marwadi University, Rajkot, Gujarat, India

These authors contributed equally to this work

DOI: http://doi.org/10.29194/NJES.29010103

Received: September 11, 2025 **Revised:** December 11, 2025 **Accepted:** December 19, 2026 **Published:** March 20, 2026

Abstract

In precision agriculture, crop disease detection can be a highly valuable undertaking in which scalable and correct solutions may save considerable amounts of money and loss of yield. This paper introduces a comparative analysis of state-of-the-art deep learning models with special attention to EfficientNetB3 hybrids, which are trained on a balanced subsample of the PlantVillage dataset with 33 classes based on nine crops. To overcome the shortcomings of the previous studies, which used unbalanced sample, a leakage-free balancing approach was used, resulting in 13,200 training and 3,300 validation samples. Custom head transfer learning was used where it was tested using two strategies; FreezeUnfreeze fine-tuning, and Singlephase training. MobileNetV2, InceptionV3, DenseNet121, GhostNet, in addition to other baseline CNNs, were compared to baseline Convolutional Neural Networks (CNNs). The findings indicate that EfficientNetB3 hybrids are superior with an accuracy of ≥99.5% and 99.9% Area Under the Curve (AUC) and specificity than the previous CNN-based systems. The paper logically defines a performance ladder between model options and real-life deployment demands, such as lightweight mobile applications to precision agriculture systems, and points out future trends in the field-based validation.

Keywords: *Convolutional Neural Networks, Data Augmentation, Precision Agriculture, Sustainability, Transfer Learning*

Corresponding author: Provide the corresponding author information and publisher here. E-mail address: jainrituraj@yahoo.com

1. Introduction

Crop diseases have continued to pose challenges to the agricultural sector since crop diseases drastically lower the world yields and endanger food security. Precision agriculture requires early and reliable detection of these diseases to facilitate timely intervention of the diseases and reduce economic losses as well as environmental hazards. Deep learning (DL) has become an innovative solution to plant disease classification with the advent of artificial intelligence (AI) and offers automated, precise, and scalable solutions [1].

The usefulness of convolutional neural networks (CNNs) including ResNet, DenseNet, and Visual Geometry Group (VGG) has been proven by earlier studies, with reported accuracy values of 94% and higher on benchmark datasets, like PlantVillage [2–10]. Nevertheless, such studies are methodologically limited. Most of them depended on raw and unbalanced data sets with dominant disease classes biasing assessment. Performance reporting was also commonly confined to accuracy to the exclusion of holistic measures like specificity and Area Under the Curve (AUC), which are important in an agricultural setting where false positives may result in interventions that are unwarranted. Additionally, although lightweight CNNs and mobile-based solutions have been investigated, there is little systematic benchmark on different architectures utilizing the same training and evaluation settings.

To overcome such gaps, this paper proposes a balanced dataset approach to PlantVillage, in which 33 classes of disease and healthy for nine crops are standardized to have equal and balanced evaluation. A comparative benchmarking of the four baseline CNNs (MobileNetV2, InceptionV3, DenseNet121, GhostNet) and the higher-end EfficientNetB3 hybrids trained with transfer learning and regularization methods performed in the proposed work. FreezeUnfreeze fine-tuning and Single-phase training are two sets of training that are systematically evaluated to observe the stability and adaptability of performance.

The main contributions of this work are four-fold: (i) a balanced dataset pipeline to have a robust evaluation, (ii) a full benchmarking between lightweight, dense, and hybrid architectures, (iii) state-of-the-art performance with EfficientNetB3 hybrids, and (iv) the correspondence of findings with real-world deployment scenarios. The study will take a step further in the classification of crop diseases over the previous CNN-based methods and provide practical insights for real-world precision agriculture applications.

Al-Nahrain J. Eng. Sci. is a Scopus indexed **open access Journal** with **ISSN 2521-9154** and **eISSN 2521-9162**

2. Literature Review

2.1 Overview of AI and Deep Learning in Crop Disease Detection

DL and AI have also played a significant role in advancing crop disease detection to a significant level, offering solutions that are not only scientifically valid but also applicable in practice to farmers. The CNNs most popular model to classify leaf diseases. Indicatively, the researchers in [2] tested ten pre-trained CNNs with transfer learning on the PlantVillage dataset, obtaining 97.38% accuracy, but the DenseNet201 did the best. The strength of ResNet was also verified by [3], which reached over 95 % on a 12,500-image dataset, and [4], which achieved 98.6 % accuracy with DenseNet121. Authors of [5] used segmentation together with CNNs, which enhanced classification to over 98 % of ten disease classes.

The usefulness of AI has also been applied in predictive as well as in diagnosis. Authors of [11] created an AI-based prognostication system, which demonstrated over 85 % accuracy in predicting crop diseases, enabling preventive agriculture. Authors of [7] scaled up to a 15,000-image dataset, achieving 97.35% accuracy using ResNet34. New architectures have also been proposed, including DCNN-19, which was introduced by [6] and reached 97.2% accuracy on a Tamil Nadu data set, and [12] verified the usefulness of ResNet and VGG on mixed datasets.

The recent trend in deployment-oriented studies is to emphasize on accessibility. Authors of [13] and [10] created mobile-friendly CNN applications, with 94% accuracy, and [9] went further to real time, mobile deployment, with >95% accuracy on community datasets. The broader role of ML/DL in precision agriculture was highlighted by complementary studies by [14], whereas in controlled settings VGG-16 was validated and showed over 96% accuracy by [8].

2.2 Traditional ML and Image Processing Approaches

Prior to the emergence of DL, machine learning methods were predominantly involved in plant disease detection in classical approach. Common algorithms used included Support Vector Machines (SVMs), Random Forests (RF), and k-Nearest Neighbors (kNN) on images of leaves following the stage of preprocessing, including denoising, segmentation and feature extraction. Authors of [15] used SVMs that are optimized with Particle Swarm Optimization (PSO) to find the best feature selection, compared to [16] who underlined the use of feature engineering to extract discriminative features. Though they worked well with smaller data sets, these methods had problems with generalization with larger, more complex and imbalanced data sets, limiting their applicability to precision agriculture.

2.3 Deep Learning with CNN Architectures

Deep learning has transformed the process of detecting plant diseases because it does not require the use of handcrafted features. The CNNs rapidly gained popularity as the classification systems, being able to automatically extract features and perform well. For instance, the AgriScan system proved to be cross-platform in the diagnosis of leaf disease through CNNs [17]. An accuracy of 98.6% was reported using DenseNet121 [4] whereas custom architectures including the Deep Convolutional Neural Network-19 (DCNN-19) [6], and further studies using transfer learning [12] further illustrated the flexibility and adaptability of CNN models. Although successful, the majority of CNN-based studies use unbalanced datasets like raw PlantVillage that favors results to dominant classes but underrepresents minority diseases. Also, the assessment criteria frequently used relied on accuracy and not on such measures as sensitivity, specificity, and AUC, which are important to sound agricultural decision-making.

2.4 Advanced Hybrid Models and Vision Transformers

The application of hybrid architecture and advanced imaging technology is now also being explored by researchers to evade CNN limitations. As reported by [18], the fusion of CNNs and Vision Transformers (ViTs) significantly enhanced robust, especially fine-grained, disease characteristics. This found its solidification in [19] who fused multispectral imaging with DL, and demonstrated that ViTs can be deployed to bridge CNNs on complex data. One such effective method, hyperspectral imaging discussed by [20], was a nondestructive tool that detected early disease early on by recording spectral variations triggered by pathogens. Crowd-sourced platforms are also considered amongst the innovative strategies and may be used to integrate expert knowledge or farmer knowledge with the ML model to improve the classification of diseases in local contexts [21]. Although these solutions are promising, they typically have challenges to scalability, standardization and computational costs.

2.5 EfficientNet and Emerging Trends

EfficientNet is a novel generation of CNN design by multiply scaling network depth, width and resolution. Authors of [22] demonstrated that EfficientNetB3 with multimodal data and advanced augmentation was superior to the traditional CNNs in terms of detection accuracy. Overcoming imbalance in the dataset, [23] used Generative Adversarial Networks (GAN)-augmentation and neural style transfer to produce synthetic images, which enhanced the generalization. Generative GAN-augmentation is the method for generating synthetic training images by the GANs to increase the diversity of datasets, and to boost model robustness especially when data imbalance exists. In hybrid CNNGAN pipelines, [24] weighted their loss functions to focus on the minority classes. Authors of [25] proved that further adaptability to imbalanced datasets was promoted by progressive learning in EfficientNetV2.

These works underscore EfficientNet's promise and the importance of augmentation strategies. However, evaluations remain limited by imbalanced datasets, lack of systematic benchmarking under consistent protocols, and a narrow reliance on accuracy, leaving generalizability questions unresolved. Table 1 summarizes key studies on plant and leaf disease detection, comparing methods, datasets, and reported accuracies.

While prior works demonstrate the effectiveness of CNN-based approaches for plant disease classification, several methodological and practical gaps remain unaddressed. Table 2 summarizes these research gaps and outlines how the present study addresses them through balanced dataset construction, systematic benchmarking, hybrid EfficientNetB3 customization, and robust evaluation protocols.

Table 1. Comparative summary of recent deep learning and machine learning approaches for plant and leaf disease detection, highlighting methods, datasets, findings, and reported accuracies.

Ref.	*Methods Used*	*Findings*	*Dataset*	*Accuracy / Results*
[2]	10 pre-trained CNNs, transfer learning, fine-tuning	DenseNet201 best, 97.38% accuracy	PlantVillage + augmentation	Best accuracy with DenseNet201
[3]	CNN + 12,500 images	ResNet >95% accuracy	12,500 images (38 classes)	~95% classification accuracy
[4]	DenseNet121 + preprocessing	98.6% accuracy	PlantVillage	High reliability
[5]	Segmentation + CNN	98.08% accuracy	PlantVillage	98%+ for 10 classes
[6]	Custom CNN (19 layers)	97.2% accuracy	Tamil Nadu dataset	Strong performance
[7]	ResNet34 + regional dataset	97.35% accuracy	15,000 images	ResNet outperformed
[8]	VGG-16 + classifier	>96% accuracy	15,915 samples	Accurate test results
[9]	Mobile-optimized CNN	Real-time detection	Public dataset	>95% accuracy mobile
[10]	CNN backend + UI	>94% accuracy, low-latency	4,000 images	Accurate & responsive
[11]	AI, image processing, web app	Early detection & prevention, 85%+	PlantVillage	Accurate crop prediction
[12]	Transfer learning (ResNet, VGG)	Effective across crops	Mixed images	Promising results
[13]	DCNN + mobile app	Functional app validated	PlantVillage	High test scores
[14]	Ensemble models, comparative study	Survey, broad applicability	Mixed datasets	High ML/DL impact

Table 2. Identified research gaps in crop disease classification and corresponding contributions of the present study.

Identified Research Gap	*How This Study Addresses It*	*Proposed Solution / Contribution*
Reliance on imbalanced datasets leading to biased evaluation	Curated balanced dataset (500 images/class across 33 classes)	Leakage-free balancing and augmentation pipeline for fairer training and validation
Most studies only report accuracy, neglecting holistic metrics	Evaluated with Accuracy, Precision, Recall, Specificity, F1, AUC (macro/micro)	Multi-metric evaluation ensures robust and reproducible results
Limited benchmarking, often only a few CNNs tested (e.g., ResNet, DenseNet)	Systematic benchmarking of five state-of-the-art architectures, including lightweight, dense, and hybrid designs	Comparative framework identifies trade-offs between performance and efficiency
Lack of focus on advanced hybrid architectures beyond plain CNNs	Introduced EfficientNetB3 hybrids with custom heads, tested with Freeze–Unfreeze and Single-phase training	Demonstrated superior classification (≥99.5%) with strong generalization
Inconsistent augmentation practices; sometimes biologically implausible	Designed biologically plausible augmentation pipeline (rotations, shifts, flips, brightness)	Improved generalization without distorting leaf morphology
Limited discussion on deployment contexts (farm vs. research vs. mobile)	Established hierarchy of suitability (MobileNetV2 = lightweight, DenseNet121 = research, EfficientNetB3 = high-stakes precision agriculture)	Aligns models with practical agricultural deployment scenarios

3. Methodology

This work takes the comparative experimental approach to conduct a systematic study on the state-of-the-art deep learning architectures in the classification of crop diseases. The methodology focuses on the reproducibility and fairness of the results through the standardization of the data preparation pipeline, training policies, and evaluation processes in all the models. In comparison between CNNs and hybrid EfficientNet-based models, the study will determine model architectures and training strategies that offer the best predictive-computational efficiency trade-offs. Figure 1 summarizes the general methodological approach to be used in this research, including the collection of data, the construction of a balanced dataset, preprocessing, model selection, the adaptation of transfer learning, training strategies, and evaluation to the comparative analysis.

3.1 Dataset and Data Preprocessing

In this study, the dataset used consisted of a multi-crop collection of plant-leaf images with 33 categories, both healthy and diseased leaves of different species [26]. PlantVillage dataset is a publicly accessible set of 54,305 images of healthy and diseased plant leaves that were taken in controlled conditions. It initially includes 14 types of crop species such as apple, blueberry, cherry, grape, orange, peach, pepper, potato, raspberry, soybean, squash, strawberry, and tomato. In the process of dataset inspection, it was noted that five crops, blueberry, orange, raspberry, soybean and squash had only one class (healthy or diseased), which was not representative enough to classify them robustly. These crops were left out of the study to achieve a methodological consistency and a significant comparative assessment across various disease groups. The study thus narrowed down to a select group of nine crops that contained a total of 33 classes with both healthy and diseased ones (Table 3). In order to deal further with the issue of class imbalance in the remaining crops, a balanced dataset construction approach was utilized. All classes were standardized to contain 500 images (400 to train and 100 to validate), which resulted in an ultimate dataset of 13,200 training samples and 3,300 validation samples. This equalized dataset construction directly responds to the shortcomings of previous studies (Table 2) who utilized unbalanced raw PlantVillage data, a situation that provides a fair and impartial assessment.

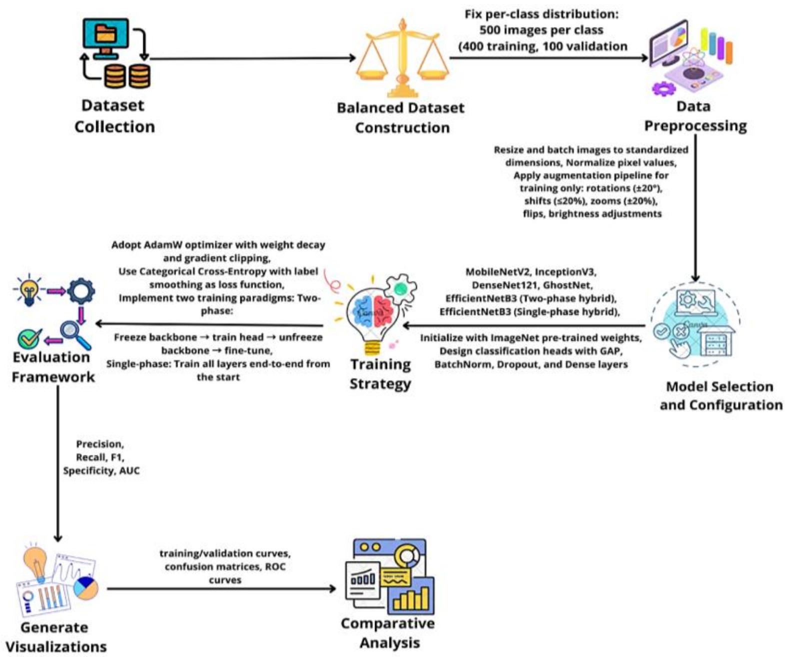


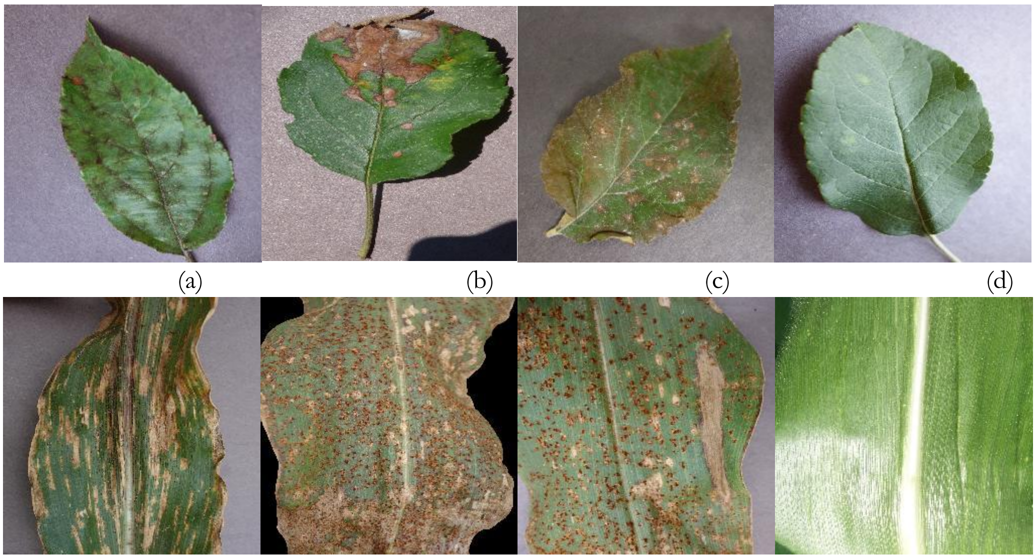
Figure (1): Methodological workflow for comparative analysis of deep learning models in crop disease classification.

Table 3. Distribution of the PlantVillage dataset used in this study, covering nine selected crops and their associated 33 classes of diseased and healthy leaves.

Crops	*Leaf Disease Category*
Apple Leaf DataSet	Apple_Scab Black_rot Cedar_apple_rust Healthy
Cherry Leaf Data Set	(including_sour)_powdery_mildew (including_sour)_healthy
Corn(maize) Leaf Data Set	Cercospora_leaf_spotGray_leaf_spot Common_rust Northern_Leaf_Blight Healthy
Grape Leaf Data Set	Black_rot Esca_(Black Measles) Leaf_blight_(Isariopsis_Leaf_Spot) Healthy
Peach Leaf Data Set	Bacterial_spot Healthy
Pepper Leaf Data Set	bell_bacterial_spot bell_healthy
Potato Leaf Data Set	Early_blight Late_blight Healthy
Strawberry Leaf Data Set	Leaf_Scorch Healthy
Tomato Leaf Data Set	Bacterial_Spot Early_blight Late_blight Leaf_Mold Septoria_leaf_spot Spider_mites Two-spotted_spider_mite Target_Spot Tomato_Yellow_Leaf_Curl_Virus Tomato_Mosaic_Virus Healthy

The balancing process was carried out in a rigorous and leakage-free manner. For validation, if ≥100 images were available, a random subset was selected; otherwise, the shortfall was filled using non-duplicated samples from the training pool. For training, if ≥400 images remained, a random subset was retained; when fewer were available, on-the-fly data augmentation was applied to achieve the target count. For classes containing a lower number of samples synthetic diversity was created using controlled on-the-fly augmentations of random rotations (+/-20 degrees), width and height shifts (less than equal to 20 percent), zooming (+/-20 percent), horizontal flip, and brightness change. These transformations ensured that the biological realism of the leaf texts would be preserved, while overcoming the shortage of data present in the minority classes. This strategy made sure that all classes had equal contribution to the training process which helped in preventing model bias and increased generalization to field-like variability. Importantly, augmentation was excluded from the validation set to preserve unbiased evaluation.

Representative examples of healthy and diseased leaves from the PlantVillage dataset are presented in Figure 2. These samples highlight the visual diversity and challenges in classification, as disease symptoms often manifest in subtle patterns such as discoloration, lesion boundaries, and texture variations.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure (2): Representative sample images from the dataset Apple leaf and Maize (Corn) leaf categories: (a) Apple Scab, (b) Apple Black Rot, (c) Apple Cedar Rust, (d) Apple Healthy (f) Common Rust, (g) Northern Leaf Blight, (h) Corn Healthy.

The augmentation pipeline was designed strategically to optimize generalization of models and ensure the biological plausibility of disease patterns. Changes were random rotations (up to 20 degrees), width and height contraction and expansion (up to 20%), zoom, horizontal flips and brightness. They were applied to run-time in a stochastic manner in that they change between epochs without the necessity to store augmented images. Such a methodology did not introduce biologically highly implausible orientations of the leaves or distortions of disease symptoms.

This balanced mode of dataset construction and preprocessing allowed equal representation of classes, prevented training-validation leakage, and a realistic, augmentation-free validation distribution, which led to robust and reproducible model testing.

3.2 Transfer Learning Approach and Customization

One of the fundamental concepts employed in this study was transfer learning, which capitalized on the capability of CNN backbones that train on ImageNet and utilize their feature extraction capacities. MobileNetV2, InceptionV3, DenseNet121, GhostNet and EfficientNetB3 already possessed the hierarchical features of visual representations, trained with large-sized natural images, and thus were easily adapted to the domain of identifying plant diseases. In order to adapt these pre-trained backbones to the 33-class crop disease classification problem, the original classification layers were deleted and a new Custom Head was created to suit this particular task. The Custom Head tended to follow a stratified order which included:

- Global Average Pooling (GAP) to compress spatial feature maps into a fixed-size vector.
- Batch Normalization to stabilize and accelerate training.
- Dropout layers as regularization to reduce overfitting.
- Dense (Fully Connected) layers with Rectified Linear Unit (ReLU) activations and L2 regularization to learn higher-level feature interactions.
- A final Dense Softmax layer with 33 units, providing class probability distributions.

The last custom head architecture of EfficientNetB3 Hybrid model included a GlobalAveragePooling2D layer and a BatchNormalization + Dropout layer (Dropout rate = 0.5), and a dense layer of 512 nodes (activation = ReLU, kernel regularizer = L2 1e-4) with a BatchNormalization + Dropout layer (Dropout rate = 0.6) and a 33-unit softmax output layer. Training has been done using the AdamW (weight decay = 1e-4 / 1e-5, global clip norm = 1.0) optimizer with batch size = 32, image resolution = 300x300 pixels, Categorical Cross-Entropy loss with label smoothing = 0.05. The learning rate was 2e-4 for the 10 epochs warm up (frozen backbone) and 3e-5 for the 40-epoch fine tuning (unfrozen backbone). These values were set up through iterative validation experiments starting from baseline set ups for EfficientNetB3 transfer learning models and fine-tuning them for stability and generalization. The baseline (network) included a 512-unit dense layer (activation = ReLU, L2 = 1e-4) and dropout layers (0.5 and 0.6) and batch normalization, and a 33-unit softmax output. The ReLU activation was kept in order to ensure good gradient propagation and faster convergence across all the training phases.

Two strategies were taken during training. During the warm-up stage, Custom Head was trained but the backbone layers were frozen so that the classifier can adjust to crop-specific features. At the fine-tuning step, the backbone layers were thawed and minimized with reduced learning rate in order to optimize feature extraction. In the single-phase model, all layers including backbone were initially trainable, with a tradeoff between faster adaptation and increased computational requirements.

A diverse suite of models was implemented to capture architectural variability in depth, connectivity, and efficiency. All models were initialized with ImageNet pre-trained weights. Their architectural details and training configurations are summarized in Table 4.

3.3 Training Strategy

The training was conducted on a standardized regime so that comparability across models could be made and that there would be the ability to make architectural changes depending on augmentation and freezing strategies. The AdamW optimizer with weight decay and gradient clipping were all used to optimize all models. The reason for choosing the AdamW optimizer is because of its better weight decay regularization to avoid overfitting and stabilize the convergence as compared to the traditional Adam optimizer. It had been adopted with adaptive learning rate schedule and gradient clipping to control the learning rate during transfer learning and fine-tuning phases and ensure that the optimization is smoother from both baseline CNNs and EfficientNet-B3 hybrids. Categorical Cross-Entropy loss function, which uses label smoothing was introduced to reduce model overconfidence to enhance overall generalization. The stability of the training was also aided with ModelCheckpoint (to maintain the most successful weights), ReduceLROnPlateau (to stochastically reduce the learning rate), and EarlyStopping (to stop training once the validation improvement levels off).

There were two training paradigms that were used. The backbone layers in the two-phase strategy were frozen initially to enable the Custom Head to fit and the backbone was fine-tuned progressively with a learning rate reduced. Conversely, the single-phase approach enabled all layers to be initially trainable and facilitated by the dynamic learning rate scheduling. Collectively, these paradigms have given a fair assessment of stability and adaptability in transfer learning.

To be transparent and reproducible, the particular training settings used per model, as well as the use of augmentations, label smoothing, dropout rates, L2 regularization, and callback settings, are all summarized in Table 5.

3.4 Evaluation Framework

All the models were evaluated on the validation dataset using a demanding evaluation protocol after training. Detailed classification reports were produced at the class level, giving Precision, Recall, and F1-scores per of the 33 categories, and macro- and weighted-averaged. Confusion matrices were built to have a visual representation of how the predicted and actual classes correlated with each other, and systematic misclassifications were identified. Also, per-class statistics, such as Accuracy, Sensitivity, Specificity, F-score, and AUC (via one-vs-rest), were calculated to have a full-scale performance profile. At the aggregate level, the overall indicators of Accuracy, Macro-averaged Precision, Sensitivity, Specificity, F1-score and Macro/Micro AUCs were provided. Training/validation curves, confusion matrices, and Receiver Operating Characteristic (ROC) curves were the visualization results that gave both quantitative and qualitative insights into performance. Unlike previous studies which only gave accuracy, this work uses an extensive range of measurements (Precision, Recall, Specificity, F1-score and AUC) to represent a comprehensive review (Table 2).

Table4. Summary of model architectures and training configurations for 33-class crop disease classification

Model	*Key Features*	*Total Parameters*	*Trainable Parameters*
MobileNetV2 Baseline	Inverted residuals, depthwise separable convolutions, Global Average Pooling (GAP), Dropout, Dense Softmax	~3.5 M	~150k–200k
InceptionV3 Baseline	Inception modules, GAP, Dropout, Dense Softmax	~23.8 M	~150k–200k
DenseNet121 Baseline	Dense blocks, transition layers, GAP, Dropout, Dense Softmax	~8 M	~150k–200k
GhostNet Baseline	Ghost modules for efficient feature generation, GAP, Dropout, Dense Softmax	~5 M	~150k–200k
EfficientNetB3 Hybrid (Two-phase)	Global Average Pooling, BatchNorm, Dropout, Dense (L2), Dense Softmax	~11.6 M	~808k
EfficientNetB3 Hybrid (Single-phase)	All layers trainable; GAP, BatchNorm, Dropout, Dense (L2), Dense Softmax	~11.6 M	~11.5 M

Table 5. Training configurations of baseline and hybrid models for 33-class crop disease classification

Model	*Label Smoothing*	*Dropout Rates (Custom Head)*	*L2 Regularization (Custom Head Dense)*	*ReduceLROnPlateau (monitor, factor, patience, min_lr)*	*EarlyStopping (monitor, patience, restore_best_weights)*
MobileNetV2 Baseline	0.1	0.5	1e-4	val_loss, 0.5, 5, 1e-7	val_loss, 10, True
InceptionV3 Baseline	0.1	0.5	1e-4	val_loss, 0.5, 5, 1e-7	val_loss, 10, True
DenseNet121 Baseline	0.1	0.5	1e-4	val_loss, 0.5, 5, 1e-7	val_loss, 10, True
GhostNet Baseline	0.1	0.5	1e-4	val_loss, 0.5, 5, 1e-7	val_loss, 10, True
EfficientNetB3 Hybrid + Custom Head (Two-phase)	0.05	0.5, 0.6	1e-4	val_loss, 0.5, 5 (warm), 7 (ft), 1e-7 (warm), 1e-8 (ft)	val_loss, 10 (warm), 15 (ft), True
EfficientNetB3 Hybrid + Custom Head (Single-phase)	0.05	0.5, 0.6	1e-4	val_loss, 0.5, 6, 1e-7	val_loss, 12, True

3.5 Comparative Analysis

Finally, a performance table prepared on comparative basis, which summarized the overall assessment metrics of all the models. Such combination allowed to directly compare lightweight CNNs to deeper and hybrid ones with significant insights on predictive capability, generalization behavior, and trade-offs of its computations. The findings do not only identify the most feasible approaches that will be involved in categorizing crop diseases but also the consequences of the augmentation strategies, transfer learning solutions and training paradigms in the real-life agricultural applications.

4. Results

4.1 Training Dynamics

Figure 3 to 8 demonstrate training and validation curves that highlight the optimization trend of the tested architectures. InceptionV3 demonstrated a fairly low convergence rate, stabilizing at less than 90 percent accuracy in validation, meaning that inception modules did a low job in capturing the fine details of the crop disease. GhostNet showed highly unstable learning behavior with swinging training and validation curves, which reflected its sensitivity to parameter initialization and reduced richness of its features due to lightweight ghost modules.

Nevertheless, MobileNetV2 and DenseNet121 demonstrated a certain trend of optimization because the validation accuracy was above 95% after 15-20 epochs. The high degree of connectivity of the DenseNet121, in particular, led to smoother loss curves and improved reliability of the convergence.

The EfficientNetB3 hybrid variants demonstrated the most consistent and the most optimistic training behavior. The FreezeUnfreeze approach as well as the Singlephase approach remained more aligned between the training and validation curves with validation accuracy reaching 100%. This implies extensive generalization and minimal overfitting, which can be attributed to compound scaling, balanced regularization, and very structured custom heads. Interestingly, single-phase version exhibited a little earlier convergence and a little better final accuracy revealing the advantage of full end-to-end fine-tuning when computer resources are at hand.

In all the models, label smoothing, dropout, and balanced augmentation were the factors that contributed to minimal overfitting. This can be most visualized with EfficientNetB3-trained models in that small-train-validation gaps ensure the absence of the learning dynamics and excellent generalization of models to all 33 disease categories.

4.2 Classification Performance

Confusion matrices (Figures 9 and 10) were created based on the two highest-performing architectures, EfficientNetB3 Hybrid (Freeze-Unfreeze) and EfficientNetB3 Hybrid (Single-phase). Both the models exhibited better classification among the 33 crop disease classes. Misclassifications were uncommon and mostly happened between morphologically similar classes of diseases, like Potato Late Blight and Tomato Late Blight which are identical in visual expression in lesion formation.

EfficientNetB3 (single-phase) had slightly higher consistency in per-class accuracy than its freezeunfreeze counterpart, indicating that end-to-end fine-tuning can improve the ability of the model to learn fine-grained variations. However, the two versions had extremely high class-specific recall, precision, and F1-scores, which highlights their strength in real-life applications

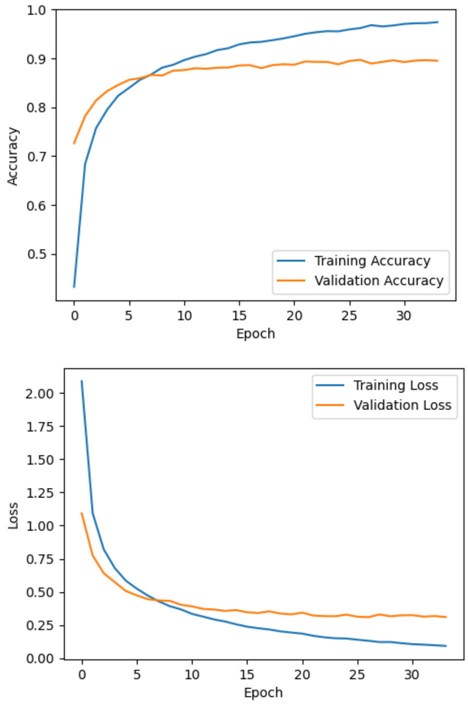


Figure 3. Training And Validation Curves of Accuracy and Loss for InceptionV3 Model

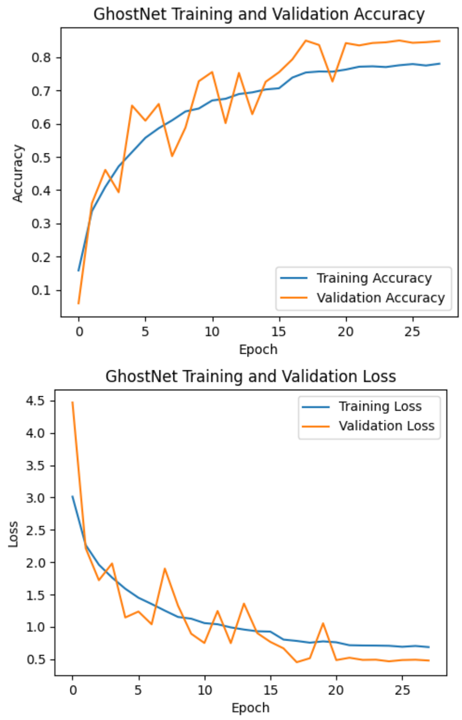


Figure 4. Training And Validation Curves of Accuracy and Loss for GhostNet Model

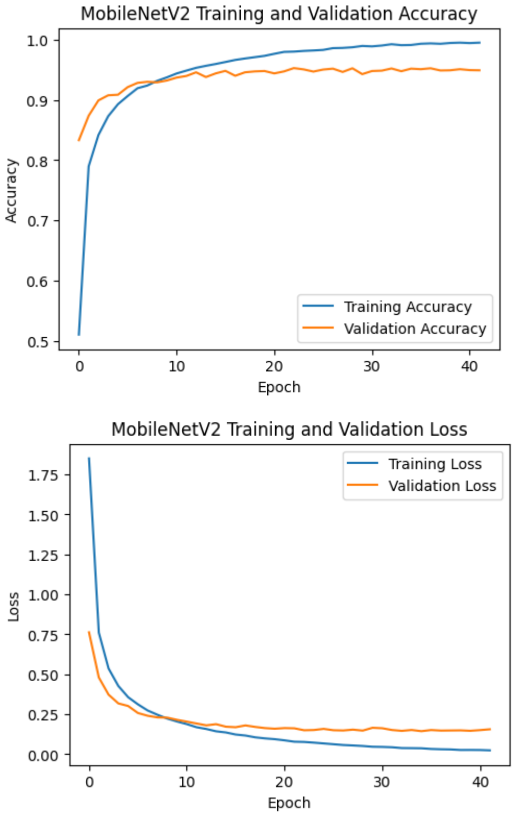


Figure 5. Training And Validation Curves of Accuracy and Loss for MobileNetV2 Model

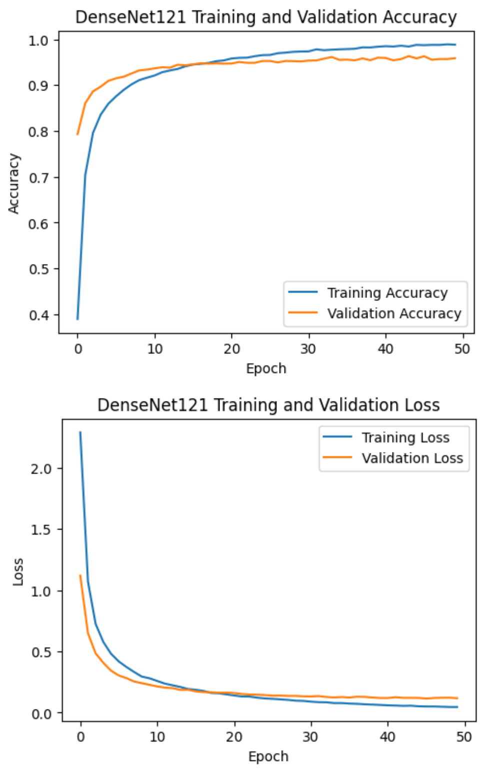


Figure 6. Training And Validation Curves of Accuracy and Loss for DenseNet121 Model

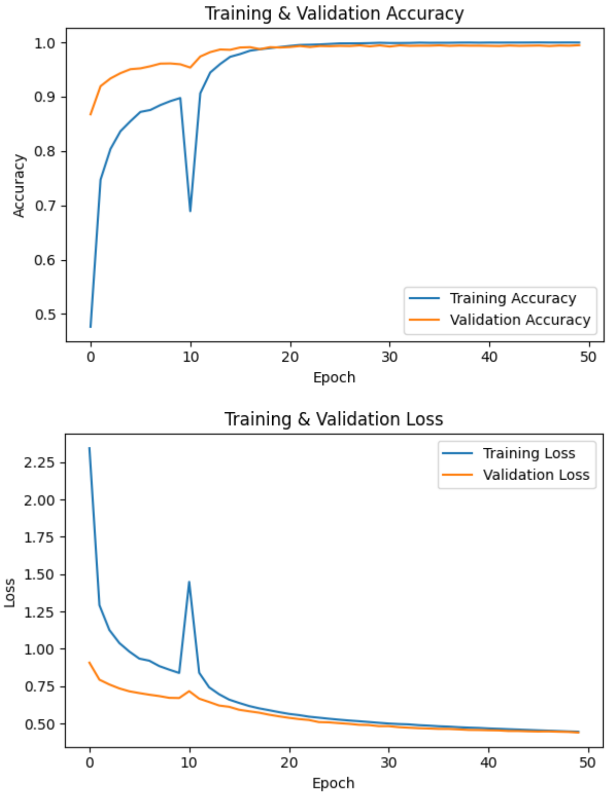


Figure 7. Training And Validation Curves of Accuracy and Loss for EfficientNetB3 Hybrid-FreezUnfreez Model

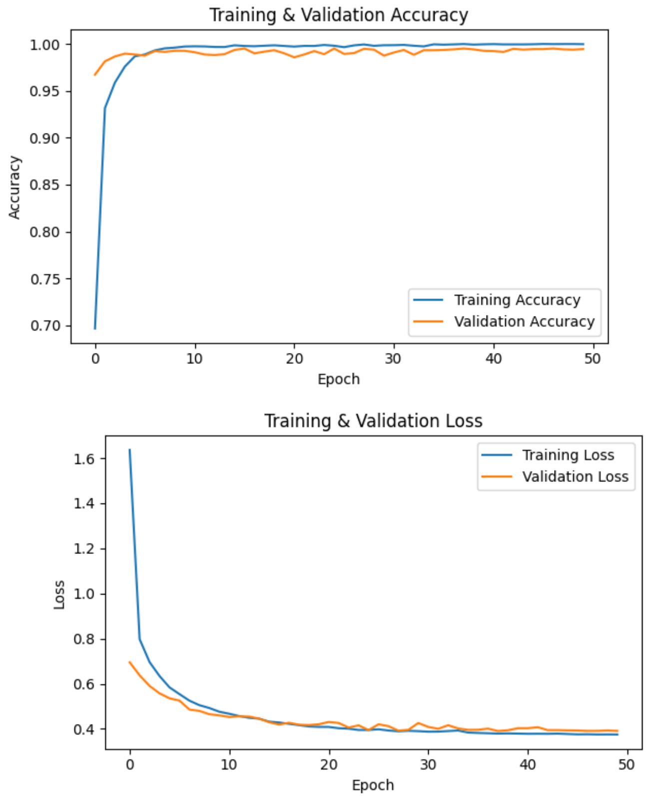


Figure 8. Training And Validation Curves of Accuracy and Loss for EfficientNetB3 Hybrid- single-phase training Model

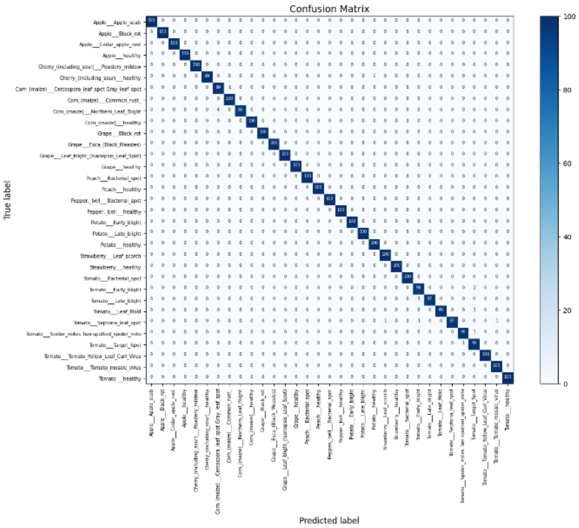


Figure 9. Confusion Metrix for EfficientNetB3 Hybrid- FreezUnfreez Model

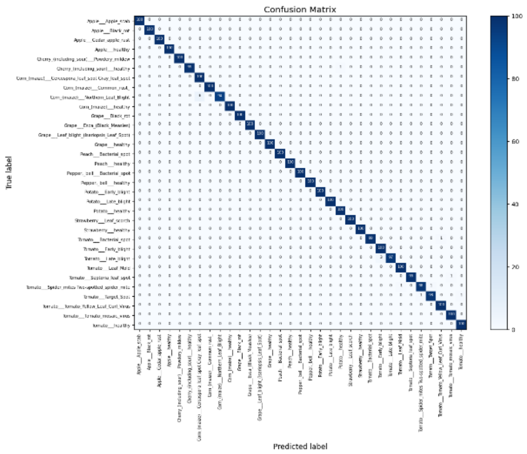


Figure 10. Confusion Metrix for EfficientNetB3 Hybrid- single-phase training Model

4.3 Comparative Analysis of Models

Table 6 summarizes the comparative performance of all of the evaluated models. The highest accuracy was obtained with DenseNet121 in the baseline CNNs (96.3), with a similar improvement in accuracy in terms of precision (96.39%), sensitivity (96.3%), and F1-score (96.29%). Its high connectivity appears to facilitate the feature propagation and reuse, which can enable more accurate classification compared to other baselines. MobileNetV2 was also noted to be performing well with an accuracy of 95.2% and this leaves no doubt that it is an effective and light weight model which can be applied in resource limited agricultural pursuits.

Comparatively, InceptionV3 was also poorly performing in comparison to other CNNs with the accuracy of 89.7%. The findings suggest that its architecture may not be most appropriate to structural differences in patterns of crop diseases, but its inception modules can represent multi-scale features. GhostNet was computationally efficient with a middle-level 91.7% accuracy. Despite the low parameter count, according to the performance gap, a trade-off is exhibited between efficiency and predictive strength, in particular, in fine-grained disease recognition.

The EfficientNetB3 hybrid models showed a great improvement over baselines. FreezeUnfreeze was the most accurate with 99.48% and Single-phase a little higher with 99.52%. The macro and micro-AUC in both models were very high (>=99.96%), which means that they possessed high discriminating power across all the 33 categories of crop diseases. Their ability to minimize false positives is also supported by a value of 99.98% specificity, a very important aspect of practical-based agricultural diagnostics where a false interpretation may lead to costly intervention.

On the whole, the findings indicate a strong performance hierarchy. Classical CNNs, especially DenseNet121, still have a good level of reliability, good accuracy, and generalization ability.

Nevertheless, the EfficientNetB3 hybrids exhibit the market-leading efficiency, as they are more effective in classification with high precision, recall, and specificity. The lack of a significant difference in the performance between freezeunfreeze and single-phase training strategies is indicative of the natural robustness of EfficientNetB3, but that in a situation with adequate computational resources, single-phase training can yield slight improvements.

These results make EfficientNetB3 hybrids the most appropriate models to be used in practice in the field of precision agriculture, and MobileNetV2 and DenseNet121 are promising when balancing their accuracy and efficiency.

Table 6. Comparative performance of baseline CNNs and EfficientNetB3 hybrids on the 33-class crop disease dataset

Models/ Metrics	*Overall Accuracy (%)*	*Precision Macro (%)*	*Sensitivity Macro (%)*	*Specificity Macro (%)*	*F1 Macro (%)*	*AUC Macro (%)*	*AUC Micro (%)*
MobileNetV2	95.21	95.29	95.21	99.85	95.23	99.93	99.96
InceptionV3	89.7	89.86	89.7	99.68	89.69	99.75	99.83
DenseNet121	96.3	96.39	96.3	99.88	96.29	99.95	99.97
GhostNet	91.7	92.33	91.7	99.74	91.58	99.89	99.89
EfficientNetB3 Hybrid-FreezUnfreez	99.48	99.5	99.48	99.98	99.49	99.98	99.98
EfficientNetB3 Hybrid- single-phase training	99.52	99.53	99.52	99.98	99.51	99.96	99.96

4.4 Comparison with Prior Literature

Table 7 will give the results of our performance in comparison with the recent literature on crop disease classification to put the proposed system in perspective. Previous CNN-based experiments have indicated up to 98.6% accuracy on the PlantVillage and related datasets, commonly on unbalanced data. Our hybrids, EfficientNetB3, in contrast, when trained on a balanced 33-class dataset performed with a minimum accuracy of 99.5% and a higher sensitivity, specificity and AUC, a new standard in the area of crop disease detection.

5. Discussion

5.1 Performance of Baseline CNN Models

DenseNet121 was the best of the baseline architectures, with an accuracy of 96.3%, then MobileNetV2 and GhostNet (95.2 and 91.7%) and InceptionV3 (89.7%). The dense connectivity of denseNet121 enables reuse of features and better gradient flow, enabling it to better acquire complex patterns related to disease. This is in line with previous studies in which DenseNet variants have demonstrated better performance than other CNNs on PlantVillage [2,4], but our balanced dataset produced more valid classification results, relative to earlier experiments, which used imbalanced conditions. Although marginally less precise, MobileNetV2 is appealing because of its compact architecture, which is quite suitable in mobile and embedded agricultural devices, which is emphasized in previous works with a focus on mobile devices [9,13]. Despite its efficiency, GhostNet exhibited lower robustness in fine-grained disease categories, probably because its ghost modules only have a restricted representative capacity. InceptionV3 yielded the worst performance implying that although inception modules are efficient capturing multi-scale features, they are less skilled in capturing the localized and subtle lesion patterns of crop leaf diseases.

The comparative performance among baseline CNNs highlights that architectural depth and connectivity play a decisive role in feature learning. For example, DenseNet121's dense block connectivity enables more effective gradient flow and feature reuse, resulting in superior performance over shallower models like MobileNetV2. In contrast, InceptionV3's modular structure captures multi-scale features but struggles with fine-grained texture variations common in crop diseases. These observations confirm that model architecture design directly governs robustness, convergence behavior, and class-wise discrimination accuracy.

5.2 Superiority of EfficientNetB3 Hybrids

Both the EfficientNetB3 hybrid models were significantly better than all baselines, with the FreezeUnfreeze model having an accuracy of 99.48% and the Single-phase training marginally higher at 99.52%. These scores outperform previous scholarship, in which simultaneous precisions usually lie between 94% and 98.6% among CNN-based techniques [2–10]. The superior performance of EfficientNetB3 hybrids can be attributed to three interrelated factors:

1. Compound Scaling – balancing network depth, width, and resolution to capture both global structures and fine-grained disease features.
2. Transfer Learning – leveraging ImageNet pre-trained weights for strong low- and mid-level feature representations, refined through task-specific fine-tuning.
3. Regularization – dropout, batch normalization, label smoothing, L2 regularization, and biologically plausible augmentations ensured stable training and prevented overfitting.

The above strengths allowed EfficientNetB3 to differentiate visually similar diseases (e.g., Tomato Late Blight vs. Potato Late Blight), an area where baseline CNNs often failed. This makes EfficientNetB3 a new architecture in agricultural disease classification.

Table 7. Comparative performance of the proposed system against recent literature on crop disease classification

Study / Model	*Methods / Models*	*Reported Accuracy (%)*	*Other Metrics*
[2]	10 pre-trained CNNs (DenseNet201 best)	97.38	–
[3]	ResNet	>95	–
[4]	DenseNet121	98.6	–
[5]	Segmentation + CNN	98.08	–
[7]	ResNet34	97.35	–
[8]	VGG-16	>96	–
[9]	Mobile-optimized CNN	>95	–
[10]	CNN	>94	Functional test scores
Proposed System	(EfficientNetB3 Hybrid – Single Phase)	99.52	Precision: 99.53, Sensitivity: 99.52, Specificity: 99.98, F1: 99.51, AUC (macro/micro): 99.96
Proposed System	(EfficientNetB3 Hybrid – Freeze–Unfreeze)	99.48	Precision: 99.50, Sensitivity: 99.48, Specificity: 99.98, F1: 99.49, AUC (macro/micro): 99.98

5.3 Impact of Training Strategies

Training paradigms have been compared and it was revealed that the Freeze-Unfreeze and Single-phase strategies proved to be highly effective and only the minimal differences in accuracy (99.48% and 99.52% respectively). FreezeUnfreeze strategy is more effective and consistent since the initial step is to train the custom head and polish the backbone slowly. By contrast, the Single-phase method permits full end-adaptation in the first place, at a minor cost of increased computational needs. This implies that EfficientNetB3 hybrids are stable to training regimes and the choice of strategy can be informed more by computational resources than performance trade-offs.

5.4 Class-wise Performance and Specificity

EfficientNetB3 versions in confusion matrices exhibited minimum misclassification with almost similar visual appearance diseases. Both models achieved a macro and micro-AUC of more than 99.9% and this denotes good discriminating capability in the 33 categories. Their capacity to minimize false positives that is especially critical in agriculture because unnecessary treatments make it expensive and risky are underscored by the specificity of 99.98. Notably, the balanced dataset design and augmentation plan ensured the similarity in the performance of the majority and minority classes, which formed a significant constraint of the previous research that was predisposed to the dominant groups.

5.5 Practical Implications

The findings establish a hierarchy of model suitability for real-world deployment:

- MobileNetV2: Ideal for low-resource or mobile applications where efficiency and portability are prioritized over maximum accuracy.
- DenseNet121: A strong option for controlled research environments, providing balanced accuracy and interpretability.

• EfficientNetB3 hybrids: The most robust solution for precision agriculture, delivering near-perfect classification accuracy and reliability across diverse crop disease categories.

This comparative perspective demonstrates that integrating transfer learning with compound-scaled architectures not only advances the state-of-the-art but also provides practical pathways for deployment in farming contexts, from edge devices to high-performance research systems.

5.6 Future Directions

Although the current research shows excellent performance on a balanced PlantVillage dataset, future research needs to be expanded to field-acquired images, that show natural lighting, occlusion, and background variations. These extensions are essential to fill the difference between the controlled datasets and the real-life farming conditions. Also, need to should explore the use of lighter versions like EfficientNetLite to find a middle ground between computational efficiency and predictive performance that is consistent with previous mobile-based designs [9,10]. Such actions would allow scalable and accessible AI-based solutions to farmers, particularly in agricular settings with resource limitations.

6. Conclusion

This study demonstrates that balancing datasets, customizing transfer learning architectures, and employing rigorous evaluation protocols significantly improve the robustness of crop disease classification models. By curating a leakage-free balanced dataset from PlantVillage and benchmarking five state-of-the-art architectures, this work provides a fair comparative framework that overcomes biases prevalent in prior research.

Among baseline CNNs, DenseNet121 emerged as the strongest, while MobileNetV2 offered efficiency for resource-limited settings. However, the EfficientNetB3 hybrids consistently outperformed all models, achieving ≥99.5% accuracy, macro/micro-AUC above 99.9%, and 99.98% specificity. These results surpass the performance of previously reported CNN-based methods, underscoring the value of compound scaling, biologically plausible augmentation, and robust transfer learning customization. Importantly, the study establishes a hierarchy of model suitability, MobileNetV2 for lightweight deployment, DenseNet121 for controlled research, and EfficientNetB3 hybrids for high-stakes precision agriculture.

The work highlights that accurate and fair disease detection is achievable when dataset imbalances are addressed and advanced architectures are employed under consistent protocols.

7. References:

[1] A. Bilal, J. A. Khan, A. Alzahrani, K. Almohammadi, M. Alamri, and X. Liu, "Fuzzy deep learning architecture for cucumber plant disease detection and classification," J. Big Data, vol. 12, no. 1, p. 117, May 2025.
https://doi.org/10.1186/s40537-025-01156-z

[2] D. Sutaji and H. Rosyid, "Convolutional Neural Network (CNN) Models for Crop Diseases Classification," Kinetik: Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control, Jun. 2022.
https://doi.org/10.22219/kinetik.v7i2.1443

[3] Y. Bhattania, P. Singhal, and T. Agarwal, "Plant Leaf Disease Detection Using Deep Learning," Int. J. Res. Appl. Sci. Eng. Technol., vol. 10, no. 5, pp. 2518-1523, May 2022.
https://doi.org/10.22214/ijraset.2022.42892

[4] S. Y. Goy, Y. F. Chong, T. K. K. Teoh, C. C. Lim, and V. Vijean, "Recognition of plant diseases by leaf image classification using deep learning approach," in AIP Conf. Proc., 2023, p. 020001.
https://doi.org/10.1063/5.0112725

[5] R. M. J. Al-Akkam and M. S. M. Altaei, "Plants Leaf Diseases Detection Using Deep Learning," Iraqi J. Sci., pp. 801-816, Feb. 2022.
https://doi.org/10.24996/ijs.2022.63.2.34

[6] M. Nagaraju and P. Chawla, "Plant Disease Classification using DCNN-19 Convolutional Neural Networks," in Proc. 2021 9th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Future Directions) (ICRITO), IEEE, Sep. 2021, pp. 1-6.
https://doi.org/10.1109/ICRITO51393.2021.9596200

[7] Md. M. Rana, T. A. Tithy, N. R. Mamun, and H. K. Sharker, "Plant Leaf Diseases Identification in Deep Learning," Comput. Sci. Eng.: Int. J., vol. 12, no. 5, pp. 1-13, Oct. 2022.
https://doi.org/10.5121/cseij.2022.12501

[8] A. A. Alatawi, S. M. Alomani, N. I. Alhawiti, and M. Ayaz, "Plant Disease Detection using AI based VGG-16 Model," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 4, 2022.
https://doi.org/10.14569/IJACSA.2022.0130484

[9] A. M. Ahmed, T. Qiu, F. Xia, B. Jedari, and S. Abolfazli, "Event-Based Mobile Social Networks: Services, Technologies, and Applications," IEEE Access, vol. 2, pp. 500-513, 2014.
https://doi.org/10.1109/ACCESS.2014.2319823

[10] R. B. S., T. A. Shriram, J. S. Raju, M. Hari, B. Santhi, and G. R. Brindha, "Farmer-Friendly Mobile Application for Automated Leaf Disease Detection of Real-Time Augmented Data Set using Convolution Neural Networks," J. Comput. Sci., vol. 16, no. 2, pp. 158-166, Feb. 2020.
https://doi.org/10.3844/jcssp.2020.158.166

[11] B. S. Eleena, M. Mangipudi, and K. Apoorva, "Study on the Prognostication of Crop Diseases using Artificial Intelligence," Asian J. Res. Comput. Sci., pp. 1-11, May 2022.
https://doi.org/10.9734/ajrcos/2022/v13i430318

[12] T. Gupta, Titunath, and V. Jain, "Plant Disease Detection using Deep Learning," in Proc. 2023 Int. Conf. Sustain. Comput. Smart Syst. (ICSCSS), IEEE, Jun. 2023, pp. 202-206.
https://doi.org/10.1109/ICSCSS57650.2023.10169268

[13] J. Jyotsna, P. Ramteke, and P. Baxla, "Plant Disease Prediction Using Deep Learning," Int. J. Comput. Electron. Asp. Eng., vol. 3, no. 2, Aug. 2022.
https://doi.org/10.26706/ijceae.3.2.arset1002

[14] R. Kumar, N. Shukla, and Princee, "Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning," in Proc. 2022 Int. Mobile Embedded Technol. Conf. (MECON), IEEE, Mar. 2022, pp. 168-172.
https://doi.org/10.1109/MECON53876.2022.9752173

[15] S. Alzoubi, M. Jawarneh, Q. Bsoul, I. Keshta, M. Soni, and M. A. Khan, "An advanced approach for fig leaf disease detection and classification: Leveraging image processing and enhanced support

vector machine methodology," Open Life Sci., vol. 18, no. 1, Nov. 2023.
https://doi.org/10.1515/biol-2022-0764

[16] S. M. Javidan, A. Banakar, K. Rahnama, K. A. Vakilian, and Y. Ampatzidis, "Feature engineering to identify plant diseases using image processing and artificial intelligence: A comprehensive review," Smart Agric. Technol., vol. 8, p. 100480, Aug. 2024.
https://doi.org/10.1016/j.atech.2024.100480

[17] T. A. Seyam and A. Pathak, "AgriScan: Next.js powered cross-platform solution for automated plant disease diagnosis and crop health management," J. Electr. Syst. Inf. Technol., vol. 11, no. 1, p. 45, Oct. 2024.
https://doi.org/10.1186/s43067-024-00169-7

[18] M. De Silva and D. Brown, "Multispectral Plant Disease Detection with Vision Transformer-Convolutional Neural Network Hybrid Approaches," Sensors, vol. 23, no. 20, p. 8531, Oct. 2023.
https://doi.org/10.3390/s23208531

[19] A. Upadhyay et al., "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," Artif. Intell. Rev., vol. 58, no. 3, p. 92, Jan. 2025.
https://doi.org/10.1007/s10462-024-11100-x

[20] L. Wan, H. Li, C. Li, A. Wang, Y. Yang, and P. Wang, "Hyperspectral Sensing of Plant Diseases: Principle and Methods," Agronomy, vol. 12, no. 6, p. 1451, Jun. 2022.
https://doi.org/10.3390/agronomy12061451

[21] W. Haider, A.-U. Rehman, N. M. Durrani, and S. U. Rehman, "A Generic Approach for Wheat Disease Classification and Verification Using Expert Opinion for Knowledge-Based Decisions," IEEE Access, vol. 9, pp. 31104-31129, 2021.
https://doi.org/10.1109/ACCESS.2021.3058582

[22] D. Senanu Ametefe et al., "Enhancing leaf disease detection accuracy through synergistic integration of deep transfer learning and multimodal techniques," Inf. Process. Agric., Sep. 2024.
https://doi.org/10.1016/j.inpa.2024.09.006

[23] O. Khare, S. Mane, H. Kulkarni, and N. Barve, "LeafNST: an improved data augmentation method for classification of plant disease using object-based neural style transfer," Discover Artif. Intell., vol. 4, no. 1, p. 50, Jul. 2024.
https://doi.org/10.1007/s44163-024-00150-3

[24] P. Prashant, S. Sharma, J. V. N. Ramesh, P. K. Pareek, P. K. Shukla, and S. V. Pandit, "Enhancing Tomato Leaf Disease Detection through Generative Adversarial Networks and Genetic Algorithm based Convolutional Neural Network," Fusion: Pract. Appl., vol. 16, no. 2, pp. 147-177, 2024.
https://doi.org/10.54216/FPA.160210

[25] R. S. Sandhya Devi, V. R. Vijay Kumar, and P. Sivakumar, "EfficientNetV2 Model for Plant Disease Classification and Pest Recognition," Comput. Syst. Sci. Eng., vol. 45, no. 2, pp. 2249-2263, 2023.
https://doi.org/10.32604/csse.2023.032231

[26] 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.
https://doi.org/10.3389/fpls.2016.01419