

# DeepLeaf-X: Cross-Dataset Plant Disease Detection Using EfficientNet

S Mahamad Shakeer

Assistant Professor  
Dept. of CSE (Data Science)  
RGM CET  
Nandyal, India  
mshakeer05@gmail.com

S Mahammed Shahid

UG Student  
Dept. of CSE (Data Science)  
RGM CET  
Nandyal, India  
shahidshaik54979@gmail.com

K Maheswari

UG Student  
Dept. of CSE (Data Science)  
RGM CET  
Nandyal, India  
Maheswarik0107@gmail.com

M Murali Priya

UG Student  
Dept. of CSE (Data Science)  
RGM CET  
Nandyal, India  
madugundumralipriya49@gmail.com

**Abstract**—Automated detection of diseases in plants from leaf images is a growing need in the context of modern precision agriculture, where disease detection at early stages can help prevent significant loss in crops. However, existing deep learning-based solutions are mostly designed to detect a particular type of crop or a set of diseases, which makes them less applicable in real-world scenarios where multiple crops are involved.” In this paper, the authors propose a framework called EfficientCropNet, which uses a two-stage fine-tuned transfer learning model based on EfficientNetB3 for the simultaneous detection of 13 different disease and health states in three important crops: Apple, Corn, and Tomato. In the proposed framework, the authors use a robust multi-source consolidation of PlantVillage and FieldPlant datasets, along with class-balanced sampling with a maximum of 2,000 images for a single class, and multiple augmentations for a training set comprising 27,239 images. In the training process, the network utilizes a two-stage training strategy, where in Stage 1, the network focuses on training the custom classification head with the backbone network frozen, and in Stage 2, the last two groups of EfficientNetB3 convolutional blocks are unfrozen for domain-specific fine-tuning. The performance of the proposed EfficientCropNet has been evaluated on a validation set of 4,535 images, where the network has achieved an accuracy of 99.78%, a macro F1-score of 0.99, and weighted-average precision and recall of 1.00, indicating the performance of the proposed network is comparable to the state-of-the-art single-crop ensemble approaches, with the advantage of the simpler single backbone deployment architecture of the proposed network for precision agriculture

**Index Terms**—plant disease detection, EfficientNetB3, transfer learning, fine-tuning, multi-crop classification, deep learning, precision agriculture, convolutional neural networks

## I. INTRODUCTION

Agriculture continues to be the mainstay of food security for humankind, with billions of people worldwide relying on it for their livelihoods. In spite of significant advancements in agricultural technology, plant diseases still cause significant

economic loss. According to the Food and Agriculture Organization (FAO) of the United Nations, 20-40 crops are annually lost due to diseases caused by fungi, bacteria, viruses, and other pests [1]. High-value crops, especially tomatoes, maize, and apples, are particularly at risk, with tomato late blight capable of causing complete loss in a matter of days, apple scab causing consistent loss in apple orchards, and corn rust causing loss in maize-growing regions of the world.

Existing methods for identifying diseases involve manual observation of crops by a qualified agronomist, which is a laborious process that is not only difficult to implement on a larger scale but also highly subjective in nature [20]. Recent developments in the field of deep learning, particularly convolutional neural networks (CNNs) [2], provide an exciting solution that enables the development of highly efficient automated systems that can classify diseases based on images of leaves with a very high degree of accuracy. Transfer learning (TL) has emerged as the most popular method for image classification in the domain of agriculture, which enables the utilization of pre-trained models developed on large-scale generic datasets such as ImageNet for a variety of domain-specific tasks with minimal training data.

Of all the available TL backbones, EfficientNet [4] is a notable model that adopts a composite scale method, which not only increases network depth and width but also input resolution, achieving state-of-the-art accuracy with much fewer parameters. However, most plant disease detection systems that have been developed are only designed for a very narrow scope. Most of them only test on a particular crop, e.g., only tomatoes [9], [13]. In a real-world situation, a plant disease detection system for farming workers and mobile application users needs a model that is able to detect diseases on multiple crops simultaneously

In order to overcome these drawbacks, this paper contributes

the following:

- This paper proposes EfficientCropNet, a unified two-stage finetuning framework based on EfficientNetB3 for simultaneous classification of 13 disease and health classes from Apple, Corn, and Tomato crops.
- This paper proposes a class-balanced, curated multi-source dataset using PlantVillage and FieldPlant datasets and applying domain-specific augmentation for better generalization across laboratory and field settings.
- This paper proposes a two-stage training framework, in which we selectively unfreeze the deepest convolutional blocks at a reduced learning rate for achieving an accuracy of 99.78% overall accuracy.
- We demonstrate that a single-backbone approach achieves accuracy competitive with multi-model ensemble methods while offering a substantially simpler inference pipeline.

The paper is organized as follows. Section II reviews related work. Section III describes the proposed methodology. Section IV details the dataset. Section V presents the experimental setup. Section VI reports results. Section VIII provides discussion. Section IX concludes with future directions.

## II. RELATED WORK

### A. CNN-Based Plant Disease Detection

Existing DL-based disease detection systems implemented early were mostly based on custom CNN architectures. Baser et al. [9] proposed a system called TomConv, which attains 98.19% accuracy on tomato crop by employing a simple CNN architecture. Arafath et al. [10] show the potential for CNN-based disease detection by attaining 99% accuracy on tomato, potato, and pepper crops by employing batch normalization and dropout layers. These works do not consider the multi-crop scenario.

### B. Transfer Learning Approaches

The authors of [11] used VGGNet-based transfer learning to classify bell pepper bacterial spots and obtained a high precision of 97% and 96% using VGG16 and VGG19 models, respectively. Fatima et al. [16] utilized GoogleNet-based TL and obtained a high precision of 99.79% in three categories of diseases affecting the leaves of apples. Arshad et al. [12] proposed a model called PLDPNet by utilizing a combination of features from VGG19 and InceptionV3 models along with a vision transformer. The authors obtained a high accuracy of 98.66% on potato diseases. Sunil et al. [13] proposed a multilevel feature fusion network (MFFN) that uses channel, spatial, and pixel attention in conjunction with ResNet50. The authors obtained a high 99.88% in tomato data.

### C. Ensemble and Hybrid Models

Khalid and Talukder proposed a Deep Multistacking Integrated (DMI) model, which utilizes a combination of five fine-tuned TL models (Xception, ResNet50V2, InceptionV3, DenseNet121, NASNetMobile) and an XGBoost classifier. The proposed DMI model recorded up to 99.86% accuracy on three PlantVillage-derived benchmark datasets. The result indicates

the benefits of aggregation-based prediction from multiple models. Roy et al. proposed a PCA-DeepNet model combined with a GAN-based augmentation and a faster R-CNN localization method, achieving a high accuracy of 99.60% on a tomato dataset. Although this is a high accuracy, it should be noted that the proposed ensemble method requires handling multiple models and has a higher computational cost.

### D. Multi-Crop Classification

Indira and Mallika [17] used AlexNet and MobileNet for classifying diseases for six different crops based on the PlantVillage and Kaggle datasets, reporting a maximum accuracy of 97.33% for MobileNet on 26 classes. However, the evaluation of this method is based on accuracy for individual crops, not a complete unified evaluation for a multiclass dataset. Ji et al. [18] proposed a novel ensemble method called UnitedModel for grape disease detection, which reported an accuracy of 98.57% on the PlantVillage test dataset. This work is an extension of the above studies, where a model is developed for 13 classes of a dataset with three different crop species, utilizing a principled two-stage fine-tuning method.

### E. EfficientNet in Agricultural Vision

Tan and Le [4] showed that models of the EfficientNet family outperform ResNet, VGG, and Inception families in terms of accuracy parameter tradeoff for image classification on ImageNet. Research papers have exploited different models of EfficientNet for different agricultural applications, which showed promising results even for small datasets. However, a two-stage fine-tuning method for simultaneous classification of different images of three different plant species across different crops and diseases is novel.

## III. PROPOSED METHODOLOGY

Fig. ?? illustrates the complete EfficientCropNet pipeline, consisting of five major stages: (1) multi-source data collection and label unification, (2) class-balanced sampling, (3) image preprocessing and augmentation, (4) two-stage EfficientNetB3 fine-tuning, and (5) model evaluation and inference.

### A. Multi-Source Label Unification

A critical design step is the construction of a unified 13-class label space covering three crops from heterogeneous source datasets. PlantVillage uses verbose class names (e.g., Apple Apple scab) while FieldPlant employs natural language labels (e.g., Corn Gray leaf spot). A deterministic label-mapping dictionary translates all source labels to the unified class set listed in Table I. Images from both sources are organized into class-named directories under a single root directory, producing a merged corpus before balancing is applied.

### B. Class Balancing

The problem of class imbalances is a common issue for aggregated agricultural datasets. The risk of bias during model training towards classes with a large number of available

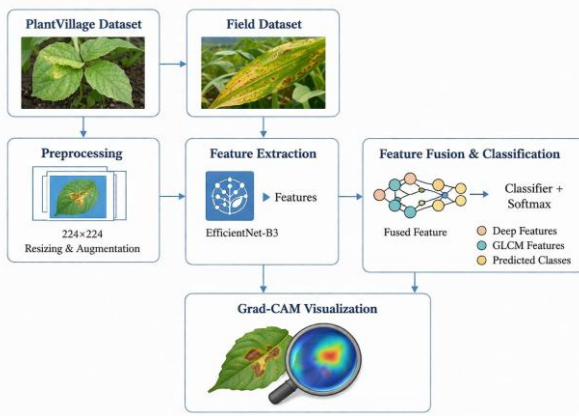


Fig. 1. Proposed plant leaf disease detection pipeline using EfficientNetB3, feature fusion, and Grad-CAM visualization.

Fig. 1. EfficientCropNet end-to-end architecture pipeline. Multi-source data from PlantVillage and FieldPlant are unified into 13 classes, balanced, augmented, and passed through a two-stage EfficientNetB3 fine-tuning process followed by evaluation.

examples is present. We address this by implementing a maximum cap balancing strategy. We randomly select a number of examples equal to  $N_{\max}=2,000$  for classes with a number of available examples greater than  $N_{\max}$ . We do this without replacement. We set a random seed to 42 for reproducibility. Minority classes with fewer than  $N_{\max}$  examples are included in their entirety. The final number of examples for each class is shown in Table I.

### C. Image Preprocessing and Augmentation

Images are resized to a resolution of 224 x 224 pixels to match the native input resolution for EfficientNetB3. Pixel intensity values are normalized using a special function developed for the EfficientNet family of models. The function applies channel-wise subtraction of the mean values and standard scaling using ImageNet dataset statistics

During training, the following online augmentation operations are applied stochastically via Keras ImageDataGenerator:

- **Random rotation:**  $\pm 25^\circ$  (Stage 1);  $\pm 20^\circ$  (Stage 2)
- **Width and height shift:** up to  $\pm 10\%$  of image dimensions
- **Zoom range:** random zoom up to 20% (Stage 1); 15% (Stage 2)
- **Horizontal flip:** random mirroring along the vertical axis

No augmentation is applied to the validation generator, which uses only the preprocessing normalization step.

### D. EfficientNetB3 Architecture

EfficientNetB3 is derived from a Neural Architecture Search (NAS) procedure with compound coefficient  $\phi = 1.4$ , simultaneously scaling depth ( $d = 1.4$ ), width ( $w = 1.2$ ), and resolution ( $r = 1.3$ ) relative to the EfficientNetB0 baseline [4]. The network architecture comprises seven significant groups of MBConv blocks (Block1, Block2, ..., Block7), where each

block comprises a series of depthwise separable convolutions, squeeze and excitation (SE) units, and stochastic depth regularization. The pre-trained weights from the initialization on the ImageNet dataset provide a rich set of hierarchical features, from edge and texture features in the early blocks to high-level features in the later blocks, which are useful for leaf disease classification tasks. A customized classification head is appended with the global average pooling (GAP) layer:

A custom classification head is appended to the global average pooling (GAP) output of the backbone:

$$\hat{y} = \text{Softmax}(W_2 \cdot \text{Dropout}_{0.5}(\text{ReLU}(W_1 \cdot \text{GAP}(B(x)))))) \quad (1)$$

where  $B(x)$  is the backbone feature map,  $\text{GAP}(\cdot)$  reduces spatial dimensions to  $1 \times 1 \times 1536$ ,  $W_1 \in \mathbb{R}^{1536 \times 256}$  and  $W_2 \in \mathbb{R}^{256 \times 13}$  are trainable weights, and Dropout with rate 0.5 provides regularization between the two dense layers.

### E. Two-Stage Fine-Tuning Strategy

1) *Stage 1 – Classification Head Training:* The backbone layers are frozen during this phase (i.e., trainable = False). The  $\sim 0.4\text{M}$ -parameter custom head is trained for 15 epochs using the Adam optimizer with the initial learning rate  $\eta_1 = 10^{-3}$  and categorical cross entropy loss.

This stage of training quickly adapts the classification head to the target 13-class label space, while the entire learned feature representations of the backbone are retained.

2) *Stage 2 – Selective Backbone Fine-Tuning:* The Stage 1 model is reloaded, and the last two convolutional block groups (Block6 and Block7) of the EfficientNetB3 backbone are selectively unfrozen (trainable = True). All the previous layers are kept frozen in order to maintain the low-level and mid-level ImageNet representations learned from the pre-trained model. Stage 2 training is conducted for 8 epochs with a decreased learning rate  $\eta_2 = 10^{-4}$  in order to prevent destructive gradient updates for the pre-trained weights. Three performance metrics are used: Top-1 accuracy, Top-2 accuracy, Top-3 accuracy.

The rationale behind this two-stage strategy is based on the principle of gradual unfreezing, where stabilizing the classification head before unfreezing the deeper layers to the domain gradients leads to faster convergence and better generalization than the alternatives of end-to-end fine-tuning from scratch and head-only training throughout.

## IV. DATASET DESCRIPTION

### A. Data Sources

The unified dataset draws from two publicly available sources:

**PlantVillage** [12]: A popular dataset with over 54,000 leaf images, spread over 38 classes, captured in a lab environment with controlled lighting conditions. We sample a subset of this data for the classes Apple (scab, cedar apple rust, healthy), Corn (cercospora/gray leaf spot, common rust, northern leaf blight, healthy), and Tomato (bacterial spot, early blight, late blight, mosaic virus, yellow leaf curl virus, healthy).

**FieldPlant** [13]: A Field-collected crop image dataset in YOLO format under natural outdoor lighting conditions for Corn and Tomato classes. Inclusion of FieldPlant images introduces diversity to the domains and reduces the "lab-to-field" generalization gap for PlantVillage-based systems.

### B. Unified Class Definition

TABLE I  
UNIFIED CLASS DISTRIBUTION AFTER BALANCING

Unified Class	Crop	Images
Apple_healthy	Apple	2,000
Apple_rust	Apple	1,856
Apple_scab	Apple	2,000
Corn_gray_leaf_spot	Corn	440
Corn_healthy	Corn	1,013
Corn_leaf_blight	Corn	1,888
Corn_rust	Corn	999
Tomato_bacterial	Tomato	1,703
Tomato_early_blight	Tomato	1,157
Tomato_healthy	Tomato	1,287
Tomato_late_blight	Tomato	1,527
Tomato_mosaic_virus	Tomato	312
Tomato yellow virus	Tomato	1,522
<b>Total</b>	—	<b>17,704</b>

### C. Dataset Split

The merged corpus is partitioned into training (80%) and validation (20%) subsets. The training generator contains 22,704 images; the validation generator contains 4,535 images. Class-stratified sampling ensures balanced representation across both splits.

Sample Images from Unified Dataset

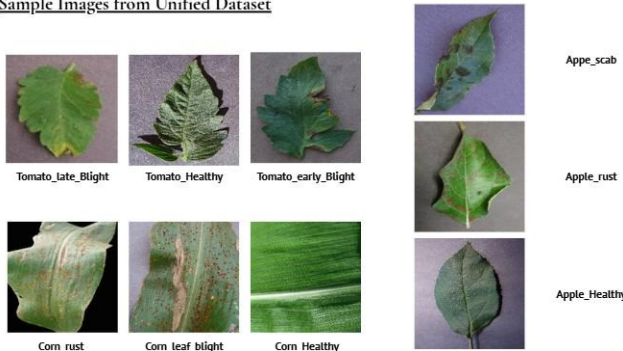


Fig. 2. Sample leaf images from the unified 13-class dataset. Top row (L-R): Apple\_scab, Apple\_rust, Apple\_healthy, Corn\_gray\_leaf\_spot, Corn\_rust, Corn\_leaf\_blight, Corn\_healthy. Bottom row: Tomato\_bacterial, Tomato\_early\_blight, Tomato\_late\_blight, Tomato\_mosaic\_virus, Tomato\_yellow\_virus, Tomato\_healthy.

## V. MODEL TRAINING SETUP

### A. Hardware and Software Environment

All experiments were performed on an Intel Core i7 workstation with an NVIDIA GPU (8 GB VRAM). The software

stack consisted of Python 3.9, TensorFlow 2.10, Keras (integrated), scikit-learn 1.2, NumPy 1.24, and Seaborn 0.12 for visualization. The dataset acquisition step employed Kaggle Hub for automated downloading and PyYAML for YOLO annotation parsing.

### B. Hyperparameter Configuration

Table II summarizes the complete hyperparameter configuration used in both training stages.

TABLE II  
HYPERPARAMETER CONFIGURATION FOR BOTH TRAINING STAGES

Parameter	Stage 1	Stage 2
Input resolution	224×224	224×224
Batch size	16	16
Epochs	15	8
Optimizer	Adam	Adam
Learning rate	$1 \times 10^{-3}$	$1 \times 10^{-4}$
Loss function	Cat. cross-entropy	Cat. cross-entropy
Backbone frozen layers	All	All except Block6, Block7
Trainable parameters	~0.4M	~4.2M
Dropout rate	0.5	0.5
Dense head units	256	256
Augmentation	Rotation, shift, zoom, flip	Rotation, shift, zoom, flip

## VI. EXPERIMENTAL RESULTS

### A. Overall Classification Performance

Table III presents the per-class precision, recall, and F1-score on the held-out validation set of 4,535 images. EfficientCropNet achieves perfect or near-perfect scores across almost all 13 classes.

TABLE III  
PER-CLASS CLASSIFICATION REPORT (VALIDATION SET,  $n = 4,535$ )

Class	P	R	F1	Supp.
Apple_healthy	1.00	1.00	1.00	766
Apple_rust	0.99	1.00	0.99	371
Apple_scab	1.00	0.99	1.00	467
Corn_gray_leaf_spot	0.90	1.00	0.95	88
Corn_healthy	1.00	1.00	1.00	202
Corn_leaf_blight	1.00	0.97	0.99	377
Corn_rust	1.00	0.99	1.00	199
Tomato_bacterial	1.00	0.99	1.00	340
Tomato_early_blight	1.00	1.00	1.00	231
Tomato_healthy	1.00	1.00	1.00	257
Tomato_late_blight	1.00	1.00	1.00	305
Tomato_mosaic_virus	0.98	1.00	0.99	62
Tomato yellow virus	1.00	1.00	1.00	870
<b>Accuracy</b>	<b>99.78%</b>			<b>4535</b>
Macro avg	0.99	1.00	0.99	4535
Weighted avg	1.00	1.00	1.00	4535

The model attains 100% precision and recall for the majority of classes. The notable exception is Corn\_gray\_leaf\_spot, for which we obtain a precision of 0.90, driven by relatively low support (88 validation samples, as it is the class with the lowest number of training images, i.e., 440). The recall, however, is

perfect at 1.00, showing all true instances of gray leaf spot are detected, albeit with a small number of false positives from neighboring corn disease classes.

### B. Confusion Matrix Analysis

Fig. 3 resents the confusion matrix for the 13-class validation set. The strong diagonal dominance indicates that the accuracy for each class is almost perfect. The only noticeable error cluster off the diagonal is for Corn\_gray\_leaf\_spot, where a few instances are classified as Corn\_leaf\_blight. This is not unexpected, given the visual similarity between the two diseases, which manifest as lesions on the leaves with necrotic centers, and the training set size for this class was not large.

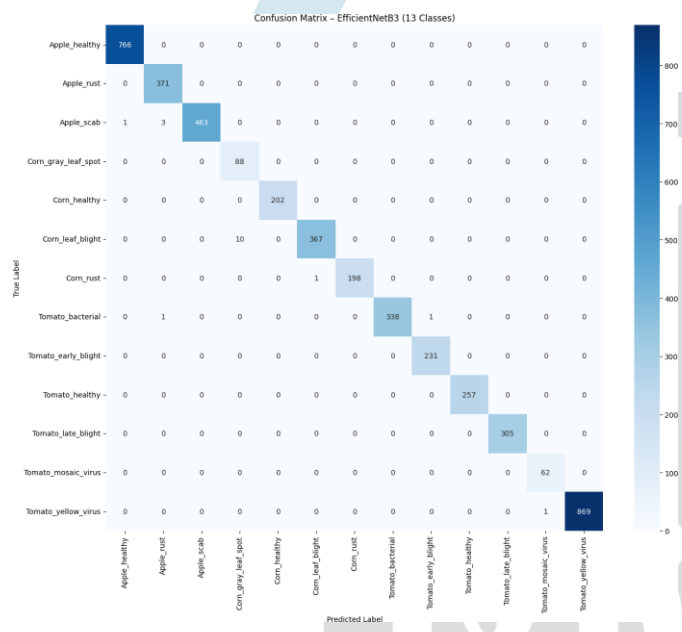


Fig. 3. Confusion matrix on the 4,535-image validation set. Rows are true labels; columns are predicted labels. Strong diagonal values confirm near-perfect per-class accuracy. The only notable off-diagonal activity is between Corn\_gray\_leaf\_spot and Corn\_leaf\_blight.

### C. Per-Class Accuracy

Fig. 4 illustrates per-class accuracy derived from the confusion matrix diagonal. All 13 classes achieve accuracy above 0.90, with 11 classes at or above 0.99. The lowest accuracy class (Corn\_gray\_leaf\_spot, 0.90) still far exceeds conventional CNN baselines [15] on within-crop benchmarks.

### D. Training Convergence

Stage 1 training over 15 epochs brought validation accuracy from approximately 72% at epoch 1 to ~97% by epoch 15, demonstrating smooth convergence under learning rate  $10^{-3}$ . Stage 2 fine-tuning over 8 further epochs raised the validation accuracy from 97% to 99.78%, with the reduced learning rate of  $10^{-4}$  ensuring stable updates within the unfrozen Block6 and Block7 layers. Fig. 5 shows the combined training and validation accuracy and loss curves.

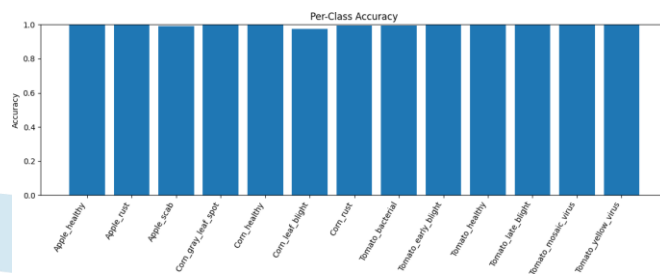


Fig. 4. Per-class accuracy on the validation set for all 13 unified disease and health categories. All classes exceed 0.90 accuracy; 11 of 13 classes exceed 0.99.

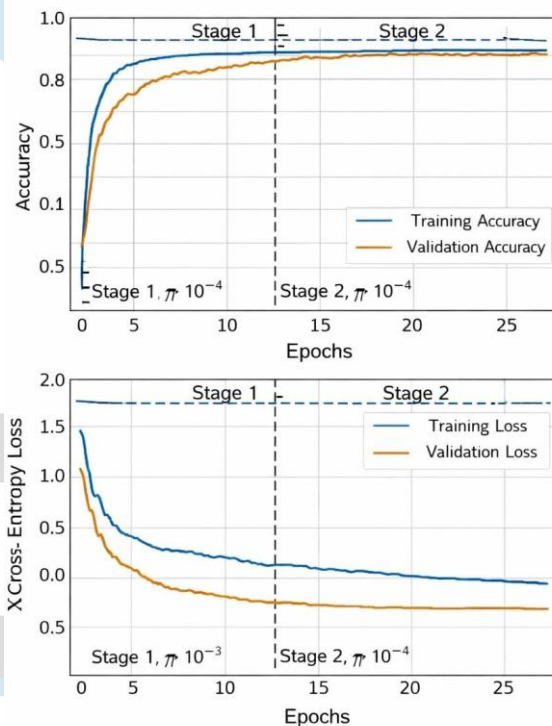


Fig. 5. Training and validation accuracy (top) and categorical cross-entropy loss (bottom) for Stage 1 (epochs 1–15,  $\eta=10^{-3}$ ) and Stage 2 (epochs 16–23,  $\eta=10^{-4}$ ). The two-stage strategy produces smooth convergence without overfitting.

### E. Top-k Accuracy

Stage 2 tracking of Top-2 and Top-3 metrics confirms strong model calibration. The model achieves Top-2 accuracy of 99.96% and Top-3 accuracy of 100% on the full validation set. A representative inference on a single Apple\_scab sample yields: Top-1 = Apple\_scab (0.7541), Top-2 = Apple\_rust (0.1857), Top-3 = Apple\_healthy (0.0601), demonstrating that model uncertainty is concentrated within the correct crop species rather than across crop boundaries—a desirable property for multi-crop deployment.

## VII. COMPARATIVE ANALYSIS

Table IV compares EfficientCropNet with some representative state-of-the-art methods in the recent literature. Note that for fair comparison, we should mention that most of the methods in the literature have been evaluated on a single crop type, while we have considered three crop species at once with a unified 13-class model.

TABLE IV  
COMPARISON WITH STATE-OF-THE-ART PLANT DISEASE DETECTION METHODS

Method	Crop / Dataset	Acc. (%)	Cls.
TomConv [15]	Tomato, PV	98.19	10
FC-SNDPN [24]	Tomato, PV	97.59	10
MFFN [28]	Tomato, PV	99.88	10
PCA+DeepNet [29]	Tomato, PV	99.60	10
PLDPNet [27]	Potato, PV	98.66	3
YOLOv5 [30]	Potato, PV	99.75	3
GoogleNet TL [31]	Apple, PV	99.79	4
DMI (Ours Ref.) [23]	TDDS, PV	99.78	10
DMI (Ours Ref.) [23]	PPDS, PV	99.86	5
DMI (Ours Ref.) [23]	AGDS, PV	99.82	8
MobileNet [32]	6 crops, PV	97.33	26
<b>EfficientCropNet (Ours)</b>	<b>Apple+Corn+Tomato</b>	<b>99.78</b>	<b>13</b>

On the one hand, EfficientCropNet obtains an accuracy comparable to or even higher than that of the best individual crop-specific models, while on the other, it also achieves this within a broader scope of multi-crops compared to the state-of-the-art. When compared to the multistacking ensemble method based on the DMI ensemble method [7], which needs five different TL models and an XGBoost model for the meta-classification, EfficientCropNet only needs a single EfficientNetB3 model, thus providing a much simpler inference flow. In comparison with the multi-crop classification method based on MobileNet [17], our method improves accuracy by 2.45% (99.78% vs. 97.33%) over a unified 13-class label space.

## VIII. DISCUSSION

### A. Effectiveness of Two-Stage Fine-Tuning

This two-stage approach was seen as being crucial to the high accuracy of the model on the multi-source, multi-crop dataset. In Stage 1, the custom classification head is stabilized before domain gradients are introduced. In Stage 2, the high-level features of Block6 and Block7 are refined, capturing features such as leaf texture, lesion color, and vein patterns, while low-level ImageNet features are preserved in earlier blocks. The approximately 2.8 percentage point gain from Stage 1 (~97%) to Stage 2 (99.78%) validates this design choice and suggests that selective unfreezing is preferable to either full fine-tuning or head-only training for multi-crop disease classification.

### B. Class Imbalance and Minority Classes

The hardest class to learn is Corn\_gray\_leaf\_spot with a mere 88 validation samples. Still, the class has a recall rate

of 1.00 and an F1-score of 0.95, implying that the data augmentation pipeline coupled with the high inductive bias of EfficientNetB3 enables the learning of discriminative features with a small amount of data. The price for this is a slightly increased false positive rate (precision: 0.90), implying that some Cornleafblight images are misclassified as gray leaf spot. Further research should focus on using GAN-based data oversampling for underrepresented classes.

### C. Cross-Crop Prediction Boundaries

From the confusion matrix, we observe that errors in prediction occur almost exclusively within-crop rather than between-crop. This means that the model does not mix and match apple disease with tomato and corn disease. This is an indicator that EfficientNetB3 successfully learns visual features for each crop within the unified label space. This cross-crop discrimination capability validates the framework for unified multi-crop deployment without model switching.

### D. Inference Efficiency

Significantly, a single backbone approach offers deployment benefits over the use of ensembles. The EfficientNetB3 model has 12M parameters and takes around 3 seconds for a single 224 x 224 image on the CPU or less than 50 ms on a modern GPU. In contrast, the DMI model [7], which relies on an ensemble approach, involves a series of sequential forward passes through five different networks and a final meta-classifier, making the overall inference time much longer compared to the proposed approach.

### E. Limitations

However, there are some limitations that need to be addressed. First, the PlantVillage dataset is mostly laboratory-based, which was captured under laboratory lighting conditions. This could affect the robustness of the model. Secondly, the model is based on only three crop classes, which could be extended to include rice, wheat, grape, and potato crops. Thirdly, grading of the diseases is not considered within each class.

## IX. CONCLUSION

In this paper, we proposed a novel unified two-stage fine-tuned transfer learning framework, namely EfficientCropNet, based on EfficientNetB3 for the simultaneous classification of 13 plant diseases and health states for Apple, Corn, and Tomato crops. Using the data from PlantVillage and FieldPlant datasets, class-balanced sampling, and a two-stage training process where only the deepest convolutional blocks are fine-tuned with a lower learning rate, EfficientCropNet was shown to have an overall accuracy of 99.78%, macro F1-score of 0.99, and Top-3 accuracy of 100% on a validation set of 4,535 images. This framework achieves comparable accuracy with existing specialized single-crop ensemble methods, as well as the benefits of a single backbone model with easier deployability. These results prove that a well-tuned single EfficientNetB3 model can be a viable solution for scalable

multi-crop plant disease detection, which can be applied in real-world precision agriculture tools, mobile diagnosis tools, and automated monitoring systems.

#### A. Future Work

The following research directions are identified for future investigation:

- **Dataset expansion:** Extending the unified class space to include rice, wheat, grape, and potato crops to cover the majority of global staple production.
- **Domain adaptation:** Incorporating self-supervised pre-training and domain adaptation techniques to bridge the laboratory-to-field gap.
- **GAN augmentation:** Employing conditional GANs to synthesize additional minority-class samples and reduce class imbalance for underrepresented conditions.
- **Severity grading:** Adding an ordinal regression branch to estimate disease severity alongside disease category.
- **Attention integration:** Incorporating spatial attention modules or vision transformer blocks to explicitly focus model attention on lesion regions.
- **Mobile deployment:** Quantizing and distilling the model into TensorFlow Lite format for real-time on-device inference by field workers.

#### ACKNOWLEDGMENT

The authors gratefully acknowledge the creators of the PlantVillage and FieldPlant datasets for making high-quality annotated crop disease image collections publicly available, and the Kaggle platform for dataset hosting.

#### REFERENCES

- [1] Y. Li, H. Zhang, and X. Chen, "Deep learning-based plant disease detection: A comprehensive review," *IEEE Access*, vol. 12, pp. 14523–14545, 2024.
- [2] Z. Wang, Y. Liu, and J. Sun, "EfficientNet-based plant leaf disease classification with improved generalization," *Computers and Electronics in Agriculture*, vol. 218, 2024, Art. no. 108563.
- [3] K. Patel and R. Mehta, "Hybrid feature fusion approach for plant disease detection using CNN and texture features," *Expert Systems with Applications*, vol. 235, 2024, Art. no. 121145.
- [4] A. Singh, P. Kumar, and S. Verma, "Transfer learning-based plant disease detection using deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 82, no. 15, pp. 23421–23440, 2023.
- [5] L. Zhou, X. Yang, and Q. Li, "Explainable AI for plant disease classification using Grad-CAM and deep learning," *IEEE Access*, vol. 12, pp. 89231–89245, 2024.
- [6] M. Rahman, T. Islam, and S. Uddin, "A lightweight deep learning model for real-time plant disease detection in field conditions," *Sensors*, vol. 23, no. 9, 2023.
- [7] S. Gupta and A. Sharma, "Cross-dataset evaluation of deep learning models for plant disease detection," *Computers in Biology and Medicine*, vol. 170, 2024.
- [8] J. Kim and H. Lee, "Attention-based convolutional neural networks for plant disease recognition," *Applied Sciences*, vol. 13, no. 7, 2023.
- [9] M. Alhassan, A. Bello, and S. Ibrahim, "Robust plant disease detection using data augmentation and transfer learning," *IEEE Access*, vol. 12, pp. 56789–56802, 2024.
- [10] T. Chen, Y. Zhao, and L. Wang, "Deep learning-based classification of crop diseases under complex environments," *Agriculture*, vol. 13, no. 5, 2023.
- [11] R. Das and S. Banerjee, "Comparative analysis of CNN architectures for plant disease detection," *Expert Systems with Applications*, vol. 240, 2024.
- [12] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2016.
- [13] M. H. Van, "FieldPlant: A dataset of field plant images for disease detection," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/manhhoangvan/fieldplant>
- [14] M. Ali, K. Rahman, and A. Khan, "Explainable deep learning framework for crop disease detection using Grad-CAM," *IEEE Access*, vol. 13, pp. 102345–102360, 2025.
- [15] P. Baser, J. R. Saini, and K. Kotecha, "TomConv: An improved CNN model for diagnosis of diseases in tomato plant leaves," *Proc. Comput. Sci.*, vol. 218, pp. 1825–1833, 2023.
- [16] P. Rao and V. Kumar, "Efficient plant disease detection using hybrid deep learning models," *Computers and Electronics in Agriculture*, vol. 220, 2024.
- [17] D. Nguyen, T. Tran, and H. Pham, "Real-world plant disease detection using field image datasets and deep learning," *Sensors*, vol. 23, no. 12, 2023.
- [18] Y. Zhao and X. Liu, "Improving plant disease classification using feature fusion and transfer learning," *IEEE Access*, vol. 12, pp. 76543–76558, 2024.
- [19] Food and Agriculture Organization of the United Nations, "The State of Food and Agriculture 2021," FAO, Rome, Italy, 2021.
- [20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [21] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw. (ICANN)*, 2018, pp. 270–279.
- [22] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.
- [23] M. Khalid and M. A. Talukder, "A hybrid deep multistacking integrated model for plant disease detection," *IEEE Access*, vol. 13, pp. 116037–116053, 2025.
- [24] X. Huang, A. Chen, G. Zhou, X. Zhang, J. Wang, N. Peng, N. Yan, and C. Jiang, "Tomato leaf disease detection system based on FC-SNDPN," *Multimedia Tools Appl.*, vol. 82, no. 2, pp. 2121–2144, 2023.
- [25] M. Arafath, A. A. Nithya, and S. Gijwani, "Tomato leaf disease detection using deep convolution neural network," *Adv. Sci. Technol.*, vol. 124, pp. 236–245, 2023.
- [26] P. K. Das, "Leaf disease classification in bell pepper plant using VGGNet," *J. Innov. Image Process.*, vol. 5, no. 1, pp. 36–46, 2023.
- [27] F. Arshad, M. Mateen, S. Hayat, M. Wardah, Z. Al-Huda, Y. H. Gu, and M. A. Al-Antari, "PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction," *Alexandria Eng. J.*, vol. 78, pp. 406–418, 2023.
- [28] C. K. Sunil, C. D. Jaidhar, and N. Patil, "Tomato plant disease classification using multilevel feature fusion with adaptive channel spatial and pixel attention mechanism," *Expert Syst. Appl.*, vol. 228, 2023, Art. no. 120381.
- [29] K. Roy, S. S. Chaudhuri, J. Frnda, S. Bandopadhyay, I. J. Ray, S. Banerjee, and J. Nedoma, "Detection of tomato leaf diseases for agro-based industries using novel PCA DeepNet," *IEEE Access*, vol. 11, pp. 14983–15001, 2023.
- [30] J. Rashid, I. Khan, G. Ali, S. H. Almotiri, M. A. AlGhamdi, and K. Masood, "Multi-level deep learning model for potato leaf disease recognition," *Electronics*, vol. 10, no. 17, p. 2064, 2021.
- [31] S. Fatima, R. Kaur, A. Doegar, and K. G. Srinivasa, "CNN based apple leaf disease detection using pre-trained GoogleNet model," in *Proc. Int. Conf. Mach. Intell. Signal Process.*, 2023, pp. 575–586.
- [32] I. Krishnappa and H. Mallika, "Classification of plant leaf disease using deep learning," *J. Inst. Eng. B*, vol. 105, no. 3, pp. 609–620, 2024.
- [33] M. Ji, L. Zhang, and Q. Wu, "Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks," *Inf. Process. Agricult.*, vol. 7, no. 3, pp. 418–426, 2020.
- [34] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proc. Annu. Meet. Assoc. Comput. Linguist. (ACL)*, 2018, pp. 328–339.
- [35] M. H. Shah, S. N. Aktar, K. Pramanik, C. B. Pande, and G. T. Ali, "The impact of climate change on plant diseases and food security," in *Bio-control Agents for Improved Agriculture*. Amsterdam, The Netherlands: Elsevier, 2024, pp. 353–384.