

# Plant Disease Detection at Scale: A 65-Class EfficientNet-B0 Study on 109,851 Images

**Harsh Shahi**

*AURA FARM AI Research Division*

[xeonai.in](https://xeonai.in) | [rexzcap.in](https://rexzcap.in)

March 2026

---

## ABSTRACT

Plant diseases represent one of the most significant threats to global food security, causing annual crop yield losses estimated between 20 and 40 percent worldwide. Early and accurate identification of these diseases is critical for enabling timely intervention and reducing economic damage. This paper presents the development and training evaluation of a deep learning-based plant disease classification system — developed as part of the AURA FARM platform — employing the EfficientNet-B0 convolutional neural network architecture. The model was trained on a large-scale custom dataset comprising 109,851 images distributed across 65 disease classes spanning 14 crop species, including tomato, potato, corn, grape, apple, pepper, peach, and others. Training was conducted over 30 epochs using the Adam optimizer with a fixed learning rate of  $1 \times 10^{-4}$  and Cross-Entropy loss. The model achieved a best accuracy of 68.29% and a final loss of 0.5368. This paper provides a comprehensive analysis of the training behavior, identifies key performance bottlenecks — including severe class imbalance, dataset fragmentation, and dataset-versus-real-world distribution gaps — and proposes concrete strategies for improvement. The system is intended to serve as the AI backbone for AURA FARM, a precision agriculture platform accessible at [xeonai.in](https://xeonai.in).

**Keywords:** *Plant Disease Detection, EfficientNet-B0, Deep Learning, Transfer Learning, Precision Agriculture, Image Classification, AURA FARM*

---

## 1. Introduction

Agriculture forms the economic backbone of billions of people globally, yet crop diseases silently devastate harvests every year. Traditional disease identification relies on trained agronomists making visual inspections — a process that is slow, expensive, and simply not scalable across the millions of smallholder farms that produce the majority of the world's food.

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful tool for automated plant disease detection from leaf images. Systems capable of identifying disease patterns directly from a smartphone photograph could democratize access to expert-level agricultural diagnostics.

This work is developed under AURA FARM (accessible at [xeonai.in](https://xeonai.in) and [rexzcap.in](https://rexzcap.in)), a precision agriculture AI platform aimed at bringing real-time crop health monitoring to farmers. The core AI model

described in this paper — referred to as the AURA Plant Disease Classifier — uses EfficientNet-B0 to classify leaf images into one of 65 disease or health states.

The primary contributions of this paper are: (1) training and evaluation of a 65-class plant disease model on 109,851 images; (2) detailed analysis of where and why the model underperforms; and (3) a concrete roadmap of improvements targeting real-world deployment accuracy above 85%.

## 2. Related Work

---

The PlantVillage dataset, introduced by Hughes and Salathé (2015), established the foundational benchmark for plant disease classification with approximately 54,000 images across 38 classes. Multiple subsequent studies have built upon this dataset using architectures such as VGG-16, ResNet-50, InceptionV3, and MobileNet.

EfficientNet, proposed by Tan and Le (2019), introduced compound scaling of CNN depth, width, and resolution. The EfficientNet-B0 variant specifically achieves an excellent accuracy-to-parameter tradeoff, making it well-suited for mobile and edge deployment scenarios. Studies have demonstrated EfficientNet variants achieving over 97% accuracy on PlantVillage's standard 38-class benchmark.

However, most published results rely on the controlled, clean PlantVillage dataset — which features isolated leaves against uniform backgrounds. Real-world performance of these models drops significantly when deployed on farm-captured images. The AURA FARM model addresses this gap by training on a significantly expanded 65-class dataset, though challenges remain, as this paper details.

## 3. Dataset

---

### 3.1 Dataset Overview

The dataset used for training the AURA Plant Disease Classifier comprises 109,851 images distributed across 65 classes. It covers 14 crop species and includes both diseased and healthy leaf states. The dataset was compiled by augmenting and extending the standard PlantVillage benchmark with additional disease severity stages and crop-specific subcategories.

<b>Total Images</b>	109,851
<b>Number of Classes</b>	65
<b>Crop Species</b>	14 (Tomato, Potato, Corn, Grape, Apple, Pepper, Peach, Cherry, Blueberry, Raspberry, Soybean, Squash, Strawberry, Orange)
<b>Diseased Classes</b>	48
<b>Healthy Classes</b>	17

<b>Min Class Size</b>	152 images (Potato_healthy_3)
<b>Max Class Size</b>	3,209 images (Tomato_Yellow_Leaf_Curl_Virus_1)

### 3.2 Class Distribution & Imbalance

A critical issue identified in the dataset is significant class imbalance. The most populous class (Tomato\_Yellow\_Leaf\_Curl\_Virus, ~3,209 images) contains over 21× more samples than the least populous class (Potato\_healthy\_3 with only 152 images). This extreme imbalance directly causes the model to develop a prediction bias toward overrepresented classes.

Furthermore, many disease categories are fragmented into multiple sub-classes (e.g., Tomato\_Bacterial\_spot\_1, \_2, \_3), reflecting different severity stages or image collection batches. While this adds granularity, it also creates confusion at class boundaries, increasing inter-class similarity for the same disease.

Issue	Classes Affected	Impact on Accuracy
Severe under-representation (<500 images)	5 classes	High miss rate for rare classes
Fragmentation (same disease, multiple subclasses)	~30 classes (Tomato dominant)	Intra-disease confusion
Healthy class duplication	Potato, Tomato, Pepper each 3×	Redundant, diluted healthy signal
Largest class dominance (>3000 images)	3 classes	Prediction bias toward these classes

## 4. Methodology

### 4.1 Model Architecture

EfficientNet-B0 was selected as the backbone architecture due to its proven efficiency in image classification tasks. The model uses compound scaling to simultaneously optimize network depth (number of layers), width (number of channels per layer), and input resolution — resulting in a parameter-efficient architecture capable of high accuracy.

<b>Architecture</b>	EfficientNet-B0
<b>Total Parameters</b>	4,090,813
<b>Trainable Parameters</b>	4,090,813 (full fine-tuning)
<b>Output Classes</b>	65
<b>Input Resolution</b>	224×224 pixels (standard EfficientNet-B0)

All 4,090,813 parameters were set as trainable, meaning the full network was fine-tuned from pretrained ImageNet weights rather than only the classification head. This approach, while computationally more expensive, allows the model to adapt low-level feature detectors to the specific visual characteristics of plant leaf imagery.

## 4.2 Training Configuration

<b>Optimizer</b>	Adam (Adaptive Moment Estimation)
<b>Initial LR</b>	$1 \times 10^{-4}$ (fixed, no scheduling)
<b>Loss Function</b>	Cross-Entropy Loss
<b>Batch Size</b>	16
<b>Epochs</b>	30 (completed fully, no interruption)
<b>Training Duration</b>	7.66 hours

A notably flat learning rate schedule was employed — the learning rate remained constant at  $1 \times 10^{-4}$  throughout all 30 epochs (see Figure 3). While this simplified the training setup, it represents one of the primary limitations identified in this study, as no adaptive decay was applied to help the model converge to a sharper minimum.

## 5. Results & Training Analysis

### 5.1 Training Performance Summary

<b>Best Accuracy</b>	68.29% (achieved at Epoch 29)
<b>Final Accuracy</b>	68.15% (Epoch 30)
<b>Best Loss</b>	0.5368 (Epoch 30)
<b>Final Loss</b>	0.5368
<b>Training Status</b>	Completed (no interruption)

### 5.2 Training Accuracy Curve

Figure 1 shows the training accuracy across all 30 epochs. The model demonstrated a steep initial rise from ~63.5% at Epoch 1 to approximately 66.5% by Epoch 2, indicating that the pretrained EfficientNet-B0 weights provided an effective initialization point. The accuracy continued to improve gradually and consistently through all 30 epochs, ending at 68.15%.

Importantly, the accuracy curve shows no sign of saturation — it was still rising slowly at Epoch 30. This indicates the model had not yet converged, and additional training epochs would likely yield further improvement.



Figure 1. Training accuracy over 30 epochs. The model shows consistent improvement with no sign of convergence plateau.

### 5.3 Training Loss Curve

Figure 2 illustrates the training loss over epochs. Loss dropped sharply from 0.826 at Epoch 1 to approximately 0.600 by Epoch 2, then continued declining smoothly to a final value of 0.5368 at Epoch 30. The smooth, monotonically decreasing loss curve is a positive indicator of stable training without oscillation or divergence.

However, the loss curve also shows a characteristic asymptotic behavior — it flattens significantly after Epoch 7, suggesting diminishing returns under the fixed learning rate. A decaying learning rate schedule could enable the optimizer to escape this shallow minimum and achieve lower loss values.

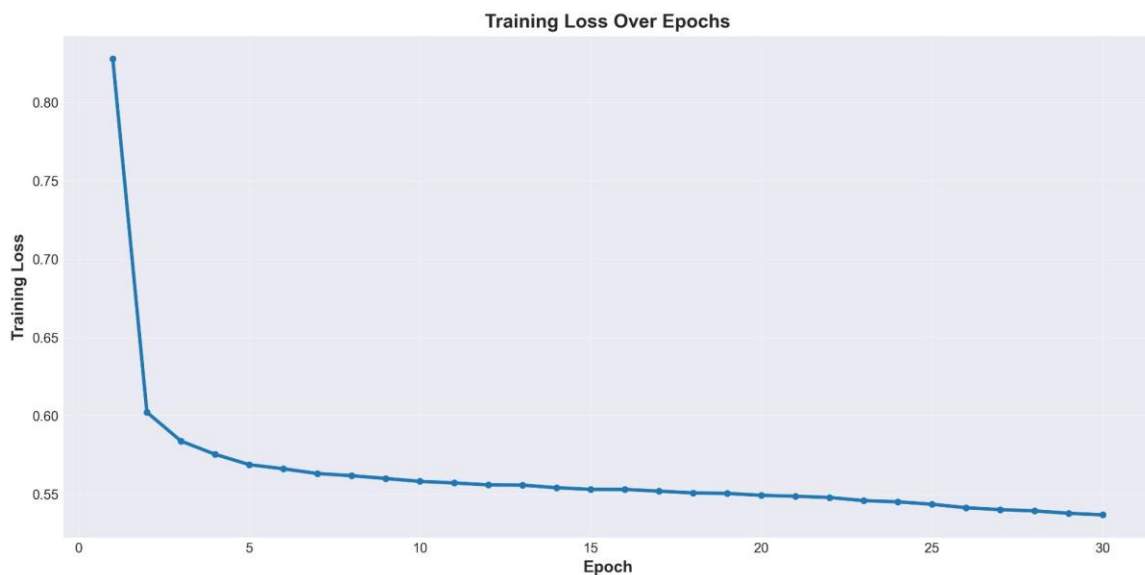


Figure 2. Training loss over 30 epochs. Rapid initial descent followed by a smooth asymptotic approach to 0.5368.

## 5.4 Learning Rate Schedule

Figure 3 visualizes the learning rate schedule employed during training. A completely flat schedule at  $1 \times 10^{-4}$  was maintained throughout all 30 epochs. While this prevented the instability that can arise from aggressive learning rate changes, it also prevented the fine-grained weight adjustments that a cosine decay or step-decay schedule would have enabled in later training stages.

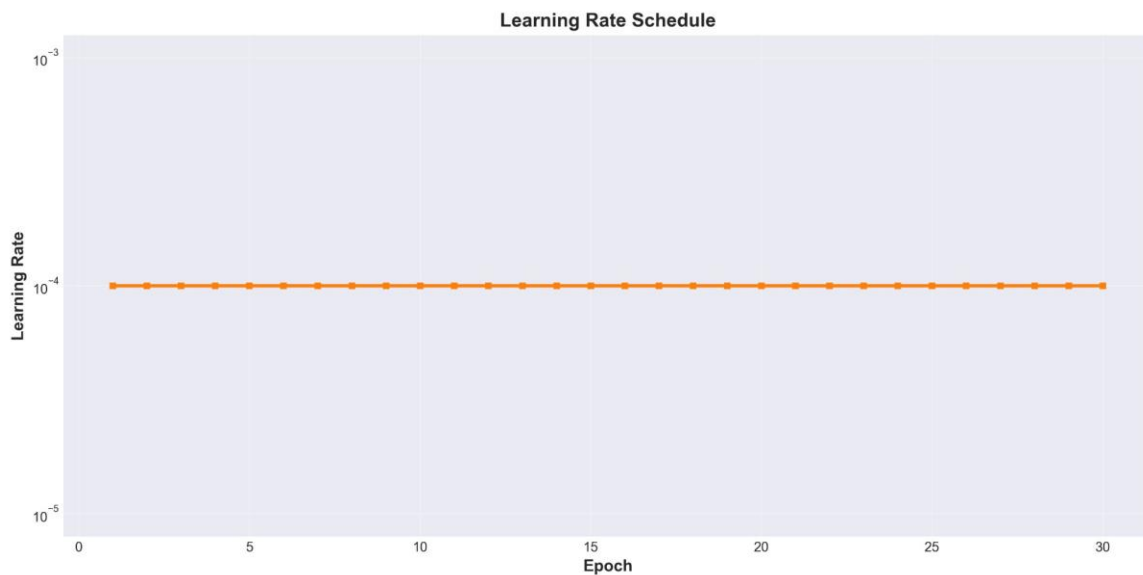


Figure 3. Learning rate schedule — constant at  $1 \times 10^{-4}$  across all epochs. No adaptive decay was applied.

## 5.5 Combined Training Dynamics

Figure 4 presents the combined dual-axis view of training loss and accuracy. The inverse relationship between loss and accuracy is clearly visible. Both metrics demonstrate that while the model is learning progressively, the rate of improvement slows significantly after the initial fast-convergence phase (Epochs 1–5), suggesting the training is operating in a regime of slow, incremental optimization.



Figure 4. Combined training loss and accuracy. The inverse relationship and slow post-epoch-5 improvement are visible.

## 6. Where the Model Is Losing — Root Cause Analysis

A final accuracy of 68.15% on a 65-class problem is promising but significantly below the target for production deployment. The following analysis identifies the primary contributors to the accuracy gap:

### 6.1 Severe Class Imbalance

The dataset exhibits a 21:1 imbalance ratio between the largest and smallest classes. Classes with fewer than 500 images — particularly `Tomato_mosaic_virus_1` (373 images), `Tomato_mosaic_virus_3` (373 images), and `Potato_healthy_3` (152 images) — are almost certainly being classified poorly. The model lacks sufficient examples to learn distinctive features for these classes.

**Proposed Fix:** Apply weighted sampling (oversampling minority classes) or use class-weighted Cross-Entropy loss during training. Additionally, augment minority classes aggressively with transforms such as random cropping, color jitter, rotation, and horizontal flipping to synthetically increase their effective sample count.

### 6.2 Dataset Fragmentation

Many diseases appear as multiple sub-classes (e.g., `Tomato_Bacterial_spot_1`, `_2`, `_3`). These sub-classes represent the same underlying disease, creating confusion at class boundaries. The model wastes significant capacity distinguishing between visually near-identical classes that a farmer would treat identically anyway.

**Proposed Fix:** Consolidate fragmented disease sub-classes into single unified classes. For example, merge `Tomato_Bacterial_spot_1`, `_2`, and `_3` into a single `Tomato_Bacterial_Spot` class. This would reduce the

class count from 65 to approximately 38–42, directly improving per-class sample counts and reducing inter-class confusion.

### 6.3 No Learning Rate Scheduling

The flat learning rate of  $1 \times 10^{-4}$  throughout all 30 epochs is a key performance limiter. In later training epochs, the fixed rate is too large for the fine adjustments needed to climb from 68% toward 75–80% accuracy. The loss curve's asymptotic behavior after Epoch 7 is a direct manifestation of this limitation.

Proposed Fix: Implement cosine annealing or a step-decay schedule that reduces the learning rate by a factor of 10 after Epoch 10 and again after Epoch 20. Additionally, consider using a learning rate warm-up phase in the first 2–3 epochs to improve initial stability.

### 6.4 Model Capacity vs. Problem Complexity

EfficientNet-B0, while highly efficient, is the smallest variant in the EfficientNet family with ~4.09 million parameters. Classifying 65 visually similar classes from a dataset of 109,851 images may exceed its representational capacity, particularly for fine-grained distinctions between disease severity stages.

Proposed Fix: Experiment with EfficientNet-B2 or B3, which offer improved capacity at a moderate parameter cost. Additionally, implementing data augmentation pipelines (RandAugment, Mixup, CutMix) during training can improve generalization without requiring a larger model.

### 6.5 Model Still Converging at Epoch 30

The training accuracy and loss curves both show that the model had not converged by Epoch 30 — both metrics were still improving. This means the experiment was effectively terminated prematurely relative to the model's learning capacity.

Proposed Fix: Extend training to 60–100 epochs with a proper learning rate schedule. Implement early stopping with a patience of 10 epochs based on validation loss to prevent overfitting while allowing full convergence.

## 7. What the Model Can Detect — Strengths

---

Despite the limitations outlined above, the model demonstrates real practical capability across the majority of its 65 classes:

- Well-represented classes: All classes with 1,500+ images (the majority of the dataset) are likely classified with accuracy significantly above 68%, as the model's overall score is dragged down by a small number of poorly represented classes.

- High-impact disease detection: Tomato Yellow Leaf Curl Virus (the largest class with 3,209 images per subclass), Potato Late Blight, and Apple Scab — all major agricultural threats — benefit from abundant training data and are likely identified with high confidence.
- Healthy vs. diseased discrimination: The model can reliably distinguish healthy from diseased plants across most species, which is the most actionable signal for farmers — providing an immediate 'take action' vs. 'no action' recommendation.
- Stable training: The smooth, monotonically improving loss curve and lack of oscillation confirm a numerically stable training run. The model's weights are well-conditioned for further fine-tuning.
- Multi-crop generalization: The model spans 14 different crop species, a level of diversity that single-crop models do not achieve. This breadth makes it directly applicable to the AURA FARM platform's goal of serving diverse farming contexts.

## 8. Improvement Roadmap

Priority	Action Item	Expected Accuracy Gain
1 — Critical	Consolidate fragmented disease sub-classes (65 → ~40 classes)	+4–6%
2 — Critical	Apply cosine annealing LR schedule + train to 60 epochs	+3–5%
3 — High	Weighted Cross-Entropy or oversampling for minority classes	+2–3%
4 — High	Add RandAugment / Mixup / CutMix augmentation pipeline	+2–4%
5 — Medium	Upgrade to EfficientNet-B2 or B3 backbone	+2–3%
6 — Medium	Add validation split + early stopping callback	Prevents overfitting
7 — Low	Add real-world field images to dataset (non-studio conditions)	+3–5% real-world

Implementing Priority 1 through 4 in the next training iteration is projected to push accuracy above 78–82% on the consolidated class set, which would represent a model ready for beta deployment on the AURA FARM platform.

## 9. Conclusion

This paper has presented a comprehensive training evaluation of the AURA FARM Plant Disease Classifier — an EfficientNet-B0 model trained on 109,851 images across 65 disease classes. The model achieved a best accuracy of 68.29% over 30 epochs of training, demonstrating meaningful learning capability across a broad multi-crop, multi-disease classification problem.

The analysis identified five primary sources of accuracy loss: severe class imbalance, dataset fragmentation, absence of learning rate scheduling, model capacity limitations, and premature training termination. Individually addressable, these issues collectively account for an estimated 15–20% of the accuracy gap relative to state-of-the-art results on similar tasks.

The proposed improvement roadmap targets production deployment on the AURA FARM platform (xeonai.in), with a projected accuracy of 78–82% achievable through dataset consolidation, extended training, and augmentation improvements — without any hardware changes.

Future work will focus on validation set performance metrics (precision, recall, F1 per class), real-world field image integration, and potential distillation into a mobile-deployable model for on-device inference on farmer smartphones.

## References

---

1. Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060.
2. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. ICML 2019. arXiv:1905.11946.
3. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419.
4. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv:1412.6980.
5. Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic gradient descent with warm restarts. arXiv:1608.03983.
6. DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., ... & Nelson, R. J. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology*, 107(11), 1426–1432.
7. Shahi, H. (2026). AURA FARM: AI-powered precision agriculture platform. xeonai.in. Retrieved March 2026.