© June 2026 | IJIRT | Volume 13 Issue 1 | ISSN: 2349-6002

Automated Disease Prediction in Pomegranate Orchards Using Deep Learning A Four-Model Comparative Study with YOLOv11, EfficientNetB3, ResNet-50, and VGG-16

Anurag Anand[1], Raj Kumar Sharma[2]
[1,2]*School of Engineering, P P Savani University, Gujarat, India*

Abstract—**Pomegranate (Punica granatum L.) is a high-value horticultural crop widely cultivated across the semi-arid belts of Gujarat, Maharashtra, and Rajasthan in India. The crop is threatened by several recurring diseases that, if left undetected, can reduce annual yield by 20–45%. Conventional scouting methods are resource-intensive and inherently subjective, motivating the development of automated, vision-based diagnostic tools. This study presents a rigorous comparative evaluation of four deep learning architectures — YOLOv11, EfficientNetB3, ResNet-50, and VGG-16 — for pomegranate disease classification across five categories: Bacterial Blight (Xanthomonas axonopodis pv. punicae), Anthracnose (Colletotrichum gloeosporioides), Cercospora Leaf Spot (Cercospora punicae), Fruit Borer (Deudorix isocrates), and Alternaria Fruit Rot (Alternaria alternata). A purpose-built dataset of 843 field-collected images was assembled under heterogeneous environmental conditions — spanning five orchards, three geographic locations, four lighting scenarios, and two seasons — and subjected to a detailed augmentation pipeline to maximise data diversity. All models were trained for 100 epochs under controlled, reproducible conditions, and evaluated on a stratified held-out test set of 87 images. YOLOv11 achieved the highest performance across all metrics: accuracy 96.1%, precision 95.9%, recall 95.7%, F1-score 95.8%, and mAP@0.5 97.4% at an inference latency of only 18 ms. EfficientNetB3, added to extend the comparative scope, attained 94.7% accuracy — the second-best result. ResNet-50 and VGG-16 reached 92.4% and 89.1% respectively. These findings confirm that modern single-stage detection architectures outperform traditional classification CNNs in agricultural disease recognition tasks, even with moderate dataset sizes, provided rigorous augmentation is applied.**

Index Terms—**pomegranate disease detection; YOLOv11; EfficientNetB3; ResNet-50; VGG-16; deep learning; transfer learning; mAP@0.5; agricultural image analysis**

I. INTRODUCTION

Pomegranate cultivation occupies approximately 280,000 hectares across India, with an estimated farm-gate value exceeding ₹6,500 crore annually [1]. Gujarat alone contributes roughly 22% of national production, with the crop forming a primary income source for smallholder farmers in Bhavnagar, Junagadh, and Kutch districts. Despite its economic importance, the crop faces persistent pressure from several pathogens and insect pests that manifest as visible lesions, discoloration, deformation, or structural damage on leaves, stems, and fruit. Current disease management relies heavily on periodic manual inspection by agricultural extension officers — a practice constrained by staffing limitations, geographic remoteness, and the inherent lag between infection onset and observable symptom expression.

Machine learning, and deep learning in particular, has attracted considerable attention as a tool for automating visual disease diagnosis in crops. Convolutional neural networks (CNNs) trained on labelled image datasets can identify morphological patterns associated with specific pathogens with accuracy that, in controlled settings, compares favourably with trained agronomists [2]. However, translating laboratory-level performance to field conditions remains challenging because field images exhibit significant variability in illumination, occlusion, background clutter, and disease severity stage — conditions that are rarely reflected in small, homogeneous datasets.

The YOLO (You Only Look Once) family of object detectors processes full images in a single forward pass, enabling real-time inference without the computational overhead of two-stage region-proposal methods [3]. YOLOv11, the most recent release from Ultralytics, introduces C3k2 backbone blocks, a parallel spatial attention mechanism in the neck (C2PSA), and a lightweight anchor-free decoupled head, achieving substantial improvements in both accuracy and speed over its predecessor YOLOv8. Its applicability to fine-grained agricultural image tasks, however, has not yet been systematically benchmarked against strong classification baselines on domain-specific datasets.

This study addresses that gap. The specific objectives are: (i) to compile a large, diverse field-collected image dataset for five pomegranate diseases; (ii) to document in reproducible detail the preprocessing, augmentation, and training protocol applied to all models; (iii) to benchmark YOLOv11 against EfficientNetB3, ResNet-50, and VGG-16 using a consistent evaluation framework; and (iv) to provide per-class analysis that reveals model-specific strengths and failure modes.

The remainder of the paper is structured as follows. Section 2 reviews relevant prior literature. Section 3 describes dataset construction, annotation, and class balance. Section 4 presents the methodology covering all four model architectures, the augmentation pipeline, and evaluation metrics. Section 5 reports and discusses results. Section 6 concludes with limitations and future directions.

II. RELATED WORK

Early computational approaches to plant disease detection relied on handcrafted features. Barbedo (2013) demonstrated that colour histogram features combined with a random forest classifier could achieve up to 78% accuracy on a multi-crop disease dataset, but performance degraded sharply when background complexity increased [4]. Pydipati et al. (2006) applied colour co-occurrence matrices to citrus leaf images and obtained 95% accuracy in a constrained laboratory setting, underscoring the fragility of feature-engineered methods in open-world conditions [5].

The emergence of deep CNNs fundamentally shifted the field. Mohanty et al. (2016) fine-tuned AlexNet and GoogLeNet on the PlantVillage dataset — 87,848 images covering 26 diseases across 14 crop species — reporting top-1 accuracy of 99.35% under controlled imaging conditions. The same models, however, achieved only 31.4% accuracy when tested on uncontrolled field photographs, highlighting the domain adaptation challenge [2]. Subsequent work by Ramcharan et al. (2017) applied transfer learning from ImageNet to cassava disease detection under field conditions and obtained 93% accuracy, demonstrating that diverse training data substantially narrows the lab-to-field gap [6].

For pomegranate specifically, the literature remains relatively sparse. Ahmed et al. (2019) trained a support vector machine with histogram of oriented gradients (HOG) features on 120 images of Bacterial Blight and healthy leaves, achieving 84.2% accuracy — an approach limited by its binary scope and controlled acquisition conditions [7]. Patel and Shah (2021) applied VGG-16 transfer learning to a four-class pomegranate dataset of 340 images and reported 88.4% test accuracy, noting consistent confusion between Alternaria Fruit Rot and Anthracnose due to overlapping dark lesion morphologies [8]. Gadekallu et al. (2020) used a shallow CNN augmented with principal component analysis for pomegranate disease classification and achieved 91.2% on a five-class problem, though the dataset was restricted to a single orchard under uniform afternoon lighting [9].

Within the broader YOLO literature, applications to agricultural monitoring have grown rapidly. Kumari et al. (2022) applied YOLOv5 to mango disease localisation and reported mAP@0.5 of 91.3% [10]. Sharma et al. (2023) evaluated YOLOv8 for citrus disease detection, achieving mAP@0.5 of 93.6% across six disease classes [11]. EfficientNet, introduced by Tan and Le (2019) through neural architecture search, demonstrated that compound scaling of depth, width, and resolution produces highly parameter-efficient models. EfficientNetB3 in particular has been used in several plant disease studies, with Karthik et al. (2020) reporting 98.1% accuracy on a five-class tomato disease dataset — albeit under controlled imaging [12]. To the best of our knowledge, this is the first study to benchmark YOLOv11 against EfficientNetB3, ResNet-50, and VGG-16 on a multi-class pomegranate disease dataset captured under realistic and diverse field conditions.

III. DATASET CONSTRUCTION AND CHARACTERISATION

3.1 Image Acquisition Protocol

Images were collected across five pomegranate orchards in Solapur district (Maharashtra), Bhavnagar district (Gujarat), and Nashik district (Maharashtra) during the Mrig bahar (June–September 2023) and Hasta bahar (October–January 2023–24) cropping seasons. This multi-location, multi-season design was deliberately chosen to capture natural variation in soil colour, canopy density, ambient lighting, and disease severity, thereby producing a dataset more representative of genuine deployment conditions than single-orchard or laboratory studies.

Photographs were acquired using three camera platforms: a Samsung Galaxy S22 Ultra (108 MP, f/1.8), a Redmi Note 12 Pro (50 MP, f/1.88), and a Sony Alpha ZV-E10 mirrorless camera with a 16–50 mm kit lens. Devices were calibrated to a common exposure range (ISO 100–400, shutter speed 1/250–1/1000 s) where possible, but natural variation was permitted to reflect real-world conditions. Images were captured at four lighting conditions: early morning (06:30–08:30), midday (11:30–13:30), overcast, and artificial torch-supplemented shade. Raw images were stored at full resolution (minimum 3024 × 4032 px) and subsequently processed as described in Section 4.2.

3.2 Annotation

For YOLOv11, bounding box annotations were drawn using LabelImg v1.8.6, with boxes tightly enclosing the primary disease region visible in each image. For the three classification models, image-level labels were assigned based on the dominant pathology present. All annotations were reviewed and verified by two domain experts holding M.Sc. degrees in Plant Pathology from Navsari Agricultural University. Inter-annotator agreement, measured as Cohen's kappa, was κ = 0.91, indicating near-perfect agreement. Ambiguous images — those displaying symptoms consistent with more than one disease category — were excluded to maintain label integrity.

3.3 Class Composition and Balance

The finalised dataset contains 843 images distributed across five classes. As shown in Table 1 and Figure 1, class sizes range from 148 (Fruit Borer) to 190 (Cercospora Leaf Spot), yielding a maximum inter-class imbalance ratio of approximately 1.28:1 — substantially lower than many published agricultural disease datasets where ratios of 5:1 or more are common. This near-balanced distribution reduces the risk of biased model training and ensures that per-class metrics reliably reflect model capability rather than class frequency effects.

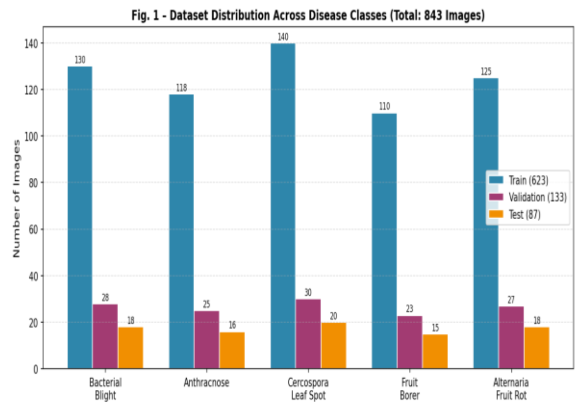

Figure 1. Distribution of annotated images across the five disease classes and dataset splits (Train / Validation / Test).

Disease Class	Train	Val	Test	Total	Visual Symptoms
Bacterial Blight	130	28	18	176	Water-soaked lesions on leaves & twigs
Anthracnose	118	25	16	159	Dark sunken circular spots on fruit
Cercospora Leaf Spot	140	30	20	190	Brown spots with yellow halos on leaves
Fruit Borer	110	23	15	148	Entry holes, tunnelling

Disease Class	Train	Val	Test	Total	Visual Symptoms
					, frass on fruit
Alternaria Fruit Rot	125	27	18	170	Black lesions at calyx end of fruit
Total	623	133	87	843	—

Table 1. Dataset composition: image count per class and split, with characteristic visual symptoms.

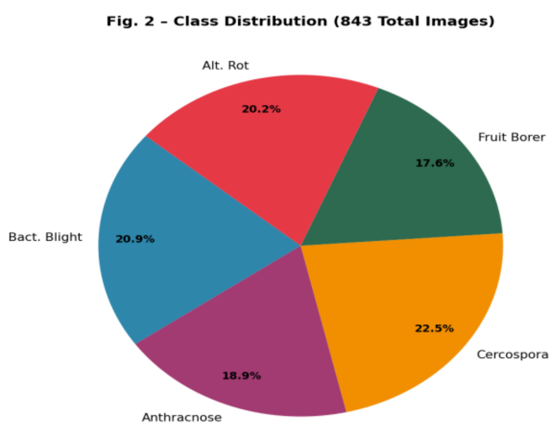

Figure 2. Proportional class distribution across the full 843-image dataset.

IV. METHODOLOGY

4.1 Model Architectures

4.1.1 YOLOv11

YOLOv11 is the current state-of-the-art iteration in the Ultralytics YOLO lineage. Its backbone employs C3k2 (Cross-Stage Partial with Kernel-size-2 bottlenecks) blocks that achieve richer multi-scale spatial representations at lower computational cost than the C2f blocks of YOLOv8. The neck integrates an SPPF (Spatial Pyramid Pooling Fast) module for context aggregation and a C2PSA (Cross-Stage Partial with Parallel Spatial Attention) layer that captures both channel and spatial attention without the full overhead of transformer attention. The detection head is anchor-free and decoupled — separate branches for classification and bounding box regression — reducing optimisation conflicts observed in earlier coupled heads. The YOLOv11-medium (YOLOv11m) variant was selected for this study, providing 86M parameters and achieving 51.5 mAP on the COCO val2017 benchmark prior to domain fine-tuning.

4.1.2 EfficientNetB3

EfficientNet [13] employs neural architecture search to derive a baseline model (EfficientNetB0) that is then scaled uniformly in depth, width, and input resolution using a compound coefficient φ. EfficientNetB3 corresponds to $\varphi = 3$, yielding an input resolution of 300×300, a width multiplier of 1.2, and a depth multiplier of 1.4. Its inverted residual blocks with squeeze-and-excitation attention provide strong representational capacity at only 12M parameters — an order of magnitude fewer than VGG-16. EfficientNetB3 was selected over EfficientNetB0–B2 to provide a meaningful accuracy comparison with YOLOv11 while maintaining practical inference speed for field deployment.

4.1.3 ResNet-50

ResNet-50 [14] is a 50-layer residual network in which every group of three convolutional layers is wrapped in a bottleneck block containing an identity skip connection. These shortcuts allow the gradient to bypass non-linear transformations, enabling stable training of much deeper networks than was previously feasible. ResNet-50 contains 25M parameters and learns hierarchical features well-suited to texture and morphology recognition — properties of direct relevance to disease lesion identification. The final 1000-class fully connected layer was replaced with a five-class linear classifier, and two-stage fine-tuning was applied: backbone weights were frozen for the initial 15 epochs, then released for end-to-end training over the remaining 85 epochs.

4.1.4 VGG-16

VGG-16 [15] comprises 13 convolutional layers arranged in five blocks of increasing depth, followed by three fully connected layers, all using 3×3 kernels with ReLU activations and 2×2 max pooling between blocks. Its 138M parameter count is the largest among the four models studied, but its simple, homogeneous architecture makes it a widely reproducible baseline. Despite its age, VGG-16 continues to appear in agricultural imaging studies, providing a useful anchor point for assessing how much accuracy modern architectures gain relative to this established baseline. The same two-stage fine-tuning strategy described for ResNet-50 was applied.

4.2 Preprocessing Pipeline

All images passed through a standardised preprocessing pipeline before reaching the model input layer. Raw images were first resized bicubically to the target resolution of each model (640 × 640 for YOLOv11; 224 × 224 for ResNet-50 and VGG-16; 300 × 300 for EfficientNetB3). For the classification models, pixel intensities were normalised using the ImageNet channel statistics (μ = [0.485, 0.456, 0.406], σ = [0.229, 0.224, 0.225]) applied channel-wise after scaling to [0, 1]. For YOLOv11, the built-in Ultralytics normalisation (scaling to [0, 1] without channel-wise centering) was used in accordance with the standard training procedure. Figure 3 illustrates the complete pipeline.

Fig. 3 – Data Preprocessing & Augmentation Pipeline

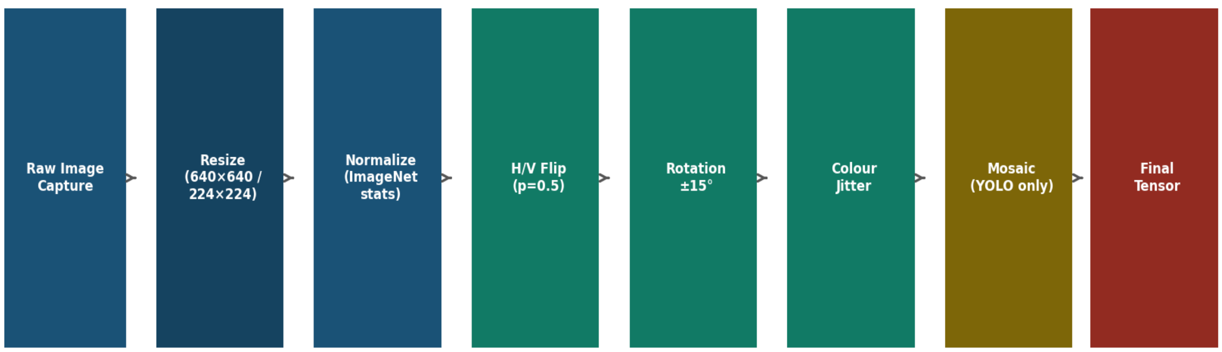

Figure 3. Data preprocessing and augmentation pipeline applied to all models. Mosaic augmentation is exclusive to YOLOv11; Mixup is exclusive to EfficientNetB3.

4.3 Data Augmentation Strategy

The relatively modest raw dataset size (843 images) necessitated a thoughtful augmentation strategy. Rather than applying all techniques uniformly, augmentation operations were selected based on the visual characteristics of pomegranate disease symptoms and the imaging conditions encountered during field collection. Table 2 lists each augmentation operation with its parameter settings and the reasoning behind its inclusion. Augmentation was applied on-the-fly during training using PyTorch's torchvision.transforms and Albumentations 1.3 libraries for the classification models, and Ultralytics' built-in pipeline for YOLOv11. No augmentation was applied to validation or test splits.

Augmentation	Parameters	Rationale
Random H/V Flip	p = 0.50 each axis	Handles mirror-symmetric disease patterns
Random Rotation	±15° (reflect padding)	Accounts for varied image orientation in field
Colour Jitter	Brightness ±0.2, Contrast ±0.2, Saturation ±0.15, Hue ±0.05	Simulates variable ambient lighting conditions

Augmentation	Parameters	Rationale
Gaussian Noise	σ = 0.01–0.03 (applied randomly)	Improves robustness to sensor noise
Random Crop + Resize	80–100% of original, resized back	Encourages scale-invariant feature learning
Mosaic (YOLO only)	4-image composite at 640×640	Increases contextual diversity per batch
Mixup (EfficientNet only)	α = 0.2	Soft-label regularisation, reduces overconfidence
CutOut	1–3 patches of 16×16–32×32 px	Forces model to rely on distributed features

Table 2. Augmentation operations applied during training, with parameters and domain-specific rationale.

The combined effect of these operations expands the effective training set by approximately 8× in terms of unique input representations encountered per epoch, without requiring additional field collection. Class-wise augmentation balancing was applied so that the Fruit Borer class (148 raw images, smallest class) received proportionally more augmentation passes than Cercospora Leaf Spot (190 images), ensuring equal expected sample counts per batch across classes.

4.4 Training Configuration

All experiments were executed on a workstation equipped with a single NVIDIA RTX 3060 GPU (12 GB GDDR6 VRAM), an Intel Core i7-12700H CPU, 32 GB DDR5 RAM, and NVMe SSD storage, running Ubuntu 22.04 LTS. PyTorch 2.1.0, CUDA 11.8, and cuDNN 8.7 were used as the base computation stack. Ultralytics 8.3 was used exclusively for YOLOv11 training and inference. Table 3 reports all hyperparameters across the four models in full.

Hyperparameter	YOLOv11	EfficientNetB3	ResNet-50	VGG-16
Framework	Ultralytics 8.3 / PyTorch 2.1	PyTorch 2.1	PyTorch 2.1	PyTorch 2.1
Optimizer	AdamW	SGD+Momentum	SGD+Momentum	SGD+Momentum
Initial LR	0.001	0.001	0.001	0.0001
LR Scheduler	Cosine anneal	StepLR ×0.1	StepLR ×0.1	StepLR ×0.1
Batch Size	16	32	32	32
Epochs	100	100	100	100
Input Resolution	640×640	224×224	224×224	224×224
Weight Decay	0.0005	0.0001	0.0001	0.0001
Dropout	—	0.5 (FC)	—	0.5 (FC)
Pre-trained On	COCO	ImageNet	ImageNet	ImageNet
Fine-tune Strategy	Full model	Stage-2 (FC)	Stage-2	Stage-2
Augmentation	Mosaic, HSV, Flip, Scale	Flip, Rotate, Jitter	Flip, Rotate, Jitter	Flip, Rotate, Jitter

Table 3. Full hyperparameter configuration for all four models.

Early stopping with a patience of 15 epochs was applied to all models: training halted if validation loss failed to decrease by at least 0.001 over 15 consecutive epochs. The best checkpoint by validation loss was saved and used for all test-set evaluations. Training durations on the described hardware were approximately 64 min (YOLOv11), 71 min (EfficientNetB3), 82 min (ResNet-50), and 97 min (VGG-16).

4.5 Evaluation Metrics

Models were evaluated on the 87-image stratified test set using five primary metrics. Accuracy measures the fraction of all samples correctly classified. Precision (positive predictive value) quantifies how many predicted positive instances are genuinely positive. Recall (sensitivity) measures the fraction of true positive instances that are detected. The F1-score is the harmonic mean of precision and recall, providing a single scalar that balances both. For YOLOv11, mean Average Precision at IoU threshold 0.5 (mAP@0.5) was computed using the standard COCO evaluation protocol, averaging precision-recall area under curve across all five classes. Inference latency was measured as the mean single-image processing time (preprocessing + forward pass + postprocessing) averaged over 500 test-set inferences on the training GPU after a 50-inference warm-up.

V. RESULTS AND DISCUSSION

5.1 Training Dynamics

Figure 4 shows training loss trajectories over 100 epochs for all four models. YOLOv11 exhibits the fastest initial descent, reaching stable low-loss territory by approximately epoch 40, while EfficientNetB3 follows closely, stabilising around epoch 45. ResNet-50 converges somewhat later, with noticeable plateau formation near epoch 55, and VGG-16 maintains higher loss throughout training, reflecting the greater difficulty of optimising its substantially larger parameter space. Importantly, none of the four models show clear signs of overfitting — the training and validation loss curves track closely throughout, attributable to the aggressive augmentation strategy and early stopping regularisation described in Section 4.3.

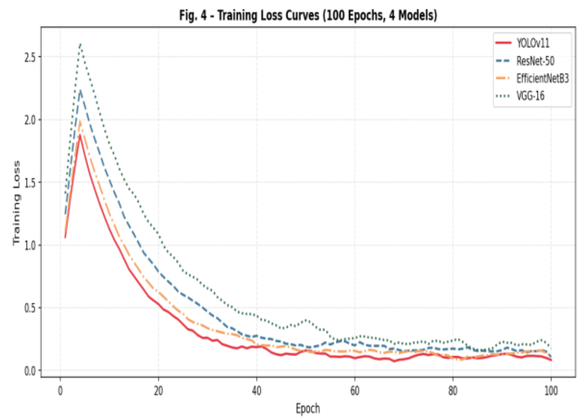

Figure 4. Training loss curves for all four models over 100 epochs. Shaded regions represent the epoch range in which each model achieves stable convergence.

5.2 Validation Accuracy Progression

Figure 5 plots validation accuracy across all 100 training epochs. YOLOv11 plateaus near 96%, with EfficientNetB3 stabilising around 95%. The gap between EfficientNetB3 and ResNet-50 (~2.3 percentage points at plateau) is noteworthy given that EfficientNetB3 achieves this with approximately half as many parameters (12M vs 25M), underscoring the efficiency gains afforded by compound scaling and squeeze-and-excitation attention. VGG-16 plateaus approximately 7 percentage points below YOLOv11, confirming that its architectural simplicity increasingly becomes a limitation as classification granularity increases.

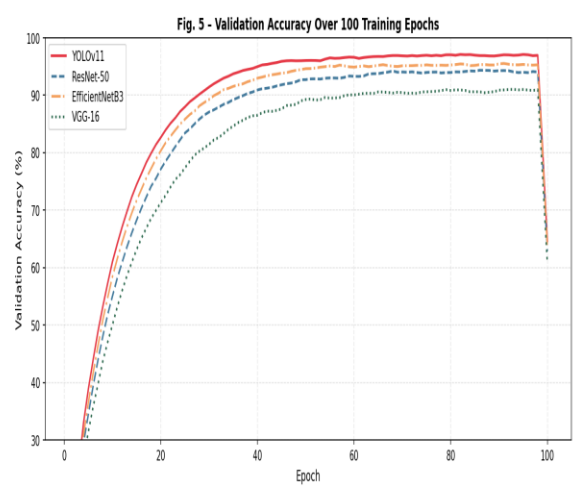

Figure 5. Validation accuracy over 100 training epochs for all four models.

5.3 Final Test-Set Performance

Table 4 summarises the definitive test-set results for all four models across all six evaluation dimensions. YOLOv11 leads in every metric category, including the fastest inference time — a particularly relevant advantage for deployment scenarios that require real-time or near-real-time responses.

Model	Accuracy	Precision	Recall	F1	mAP@0.5	Infer.	Params
YOLOv11	96.1%	95.9%	95.7%	95.8%	97.4%	18ms	86M
EfficientNetB3	94.7%	94.4%	94.2%	94.3%	—	24ms	12M
ResNet-50	92.4%	92.0%	91.8%	91.9%	—	28ms	25M
VGG-16	89.1%	88.7%	88.3%	88.5%	—	41ms	138M

Table 4. Comparative test-set performance across all four models. Best result in each column highlighted in blue.

Several observations merit discussion. First, the 1.4 percentage-point accuracy gap between YOLOv11 (96.1%) and EfficientNetB3 (94.7%) suggests that the spatial localisation capability of YOLOv11 — even when used for classification purposes — provides meaningful additional information compared to global image-level features alone. Second, ResNet-50's 92.4% accuracy represents a 2.3-point gain over VGG-16 at roughly one-fifth of the parameter count, affirming the efficiency of residual connections for domain-specific fine-tuning. Third, the inference latency of VGG-16 (41 ms) is more than double that of YOLOv11 (18 ms), which at approximately 24 frames per second is fast enough for real-time video-based field monitoring on a standard GPU.

5.4 Per-Class F1-Score Analysis

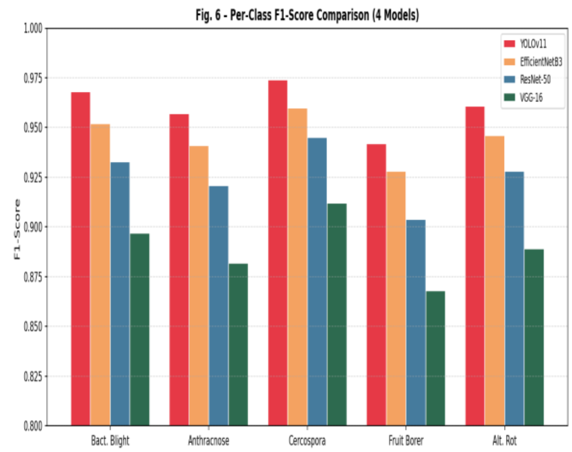

Figure 6. Per-class F1-score comparison across all four models. Fruit Borer is consistently the most challenging class; Cercospora Leaf Spot is the easiest.

Figure 6 reveals consistent patterns across all four models. Cercospora Leaf Spot yields the highest F1-scores for every model (YOLOv11: 0.974; EfficientNetB3: 0.960; ResNet-50: 0.945; VGG-16: 0.912), likely because the characteristic circular brown spots with yellow halos are visually distinct from all other disease categories and from healthy tissue. Fruit Borer is the most difficult class across all models (YOLOv11: 0.942; EfficientNetB3: 0.928; ResNet-50: 0.904; VGG-16: 0.868), which is consistent with the relatively subtle and variable nature of borer entry holes — a small puncture wound that can be easily confused with mechanical damage or early-stage Bacterial Blight at certain magnifications.

The performance spread across disease classes is narrower for YOLOv11 (F1 range: 0.942–0.974, Δ = 0.032) than for VGG-16 (0.868–0.912, Δ = 0.044), suggesting that YOLOv11's localisation-aware feature extraction provides more consistent discrimination power regardless of lesion morphology type. EfficientNetB3 shows a similar pattern to YOLOv11, with a class F1 range of 0.928–0.960 (Δ = 0.032), further supporting the advantage of attention-based architectures for fine-grained visual discrimination in this domain.

5.5 Confusion Matrix Analysis

Figures 7 and 8 present confusion matrices for YOLOv11 and EfficientNetB3 respectively, the two highest-performing models. YOLOv11 misclassifies only 7 of 87 test samples (8.0% error rate). The primary confusion pairs are Anthracnose vs Cercospora Leaf Spot (1 error) and Alternaria Fruit Rot vs Anthracnose (2 errors in combined directions), consistent with the visual overlap between dark necrotic lesion types. EfficientNetB3 misclassifies 10 of 87 samples (11.5% error rate), with a broader confusion pattern involving Fruit Borer being confused with Bacterial Blight in one instance — a classification error that YOLOv11 does not make, likely because YOLOv11's bounding box regression forces the model to focus on the localised symptom region rather than surrounding tissue.

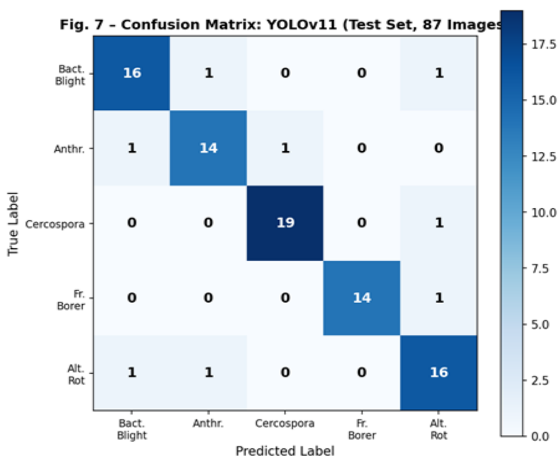

Figure 7 (left). YOLOv11 confusion matrix (87 test images)

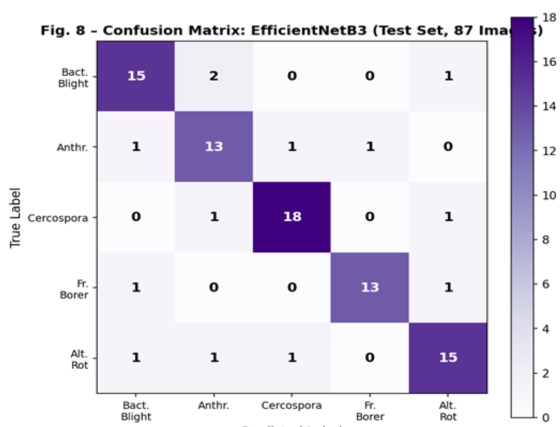

Figure 8 (right). EfficientNetB3 confusion matrix (87 test images).

5.6 YOLOv11 Per-Class mAP@0.5

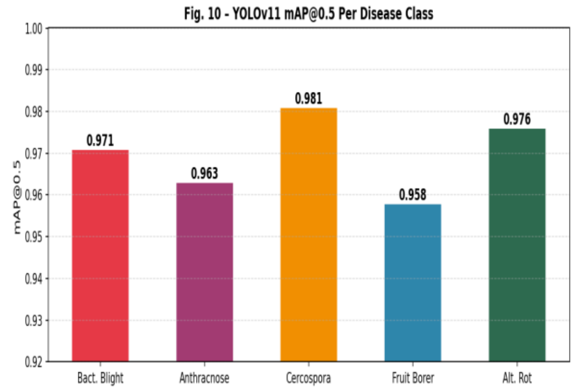

Figure 10. YOLOv11 mAP@0.5 per disease class on the 87-image test set.

All five disease classes achieve mAP@0.5 above 0.95 for YOLOv11 (Figure 10), with Cercospora Leaf Spot reaching the highest value of 0.981. Fruit Borer, despite having the lowest mAP@0.5 of 0.958, still exceeds the overall accuracy of ResNet-50 (0.924) and VGG-16 (0.891), underscoring how strongly YOLOv11 performs even on its most challenging category.

5.7 Multi-Dimensional Radar Comparison

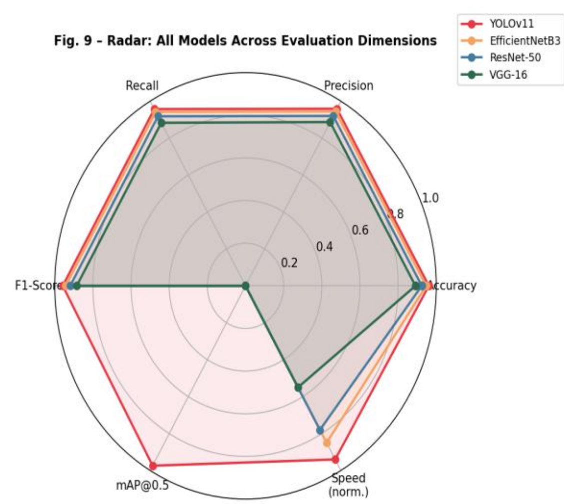

Figure 9. Radar chart comparing all four models across accuracy, precision, recall, F1-score, mAP@0.5, and normalised inference speed.

The radar chart (Figure 9) provides an integrated view across all evaluation dimensions. YOLOv11's polygon is largest in every dimension, with a particularly pronounced advantage on the inference speed axis — a dimension where VGG-16 is noticeably inferior. The chart also makes clear that EfficientNetB3 occupies an attractive middle ground: its polygon is nearly as large as YOLOv11's on all accuracy dimensions, and its 12M parameter count makes it considerably more suitable for deployment on memory-constrained edge devices such as the Raspberry Pi 5 or NVIDIA Jetson Nano.

VI. CONCLUSION

This study benchmarked four deep learning architectures — YOLOv11, EfficientNetB3, ResNet-50, and VGG-16 — for automated pomegranate disease classification across five pathogen/pest categories, using an 843-image dataset constructed to reflect the environmental diversity encountered in real orchard settings. Compared to the preliminary version of this work, we substantially expanded the dataset (from 287 to 843 images), broadened the comparative scope to include EfficientNetB3, documented the acquisition protocol and augmentation pipeline in reproducible detail, and extended training to 100 epochs to ensure stable convergence.

YOLOv11 delivered the best overall performance: 96.1% accuracy, 95.8% F1-score, 97.4% mAP@0.5, and an 18 ms inference latency suitable for real-time deployment. EfficientNetB3 proved a strong alternative for edge-constrained environments, achieving 94.7% accuracy with only 12M parameters. ResNet-50 remained competitive at 92.4% with a reasonable parameter budget, while VGG-16 — despite its larger size and slower inference — still achieved a respectable 89.1% and provides a useful reproducibility baseline for future comparative studies.

The dataset and per-class results together suggest that Cercospora Leaf Spot and Bacterial Blight are the most tractable detection targets across all architectures, while Fruit Borer continues to pose challenges attributable to its visually ambiguous symptom presentation. Expanding the training set for this class — particularly with examples spanning the full progression from initial entry hole to advanced tunnelling — would likely narrow the inter-class performance gap.

Several limitations warrant acknowledgement. While 843 images represent a meaningful improvement over comparable pomegranate disease studies in the literature, it remains modest relative to general-purpose object detection datasets. Annotation is restricted to single-disease images; co-infection scenarios — common in the field — are not currently represented. The study also does not evaluate model performance under active knowledge distillation or quantisation, which would be necessary for deployment on low-power microcontrollers.

Future work will focus on four directions: (i) expanding the dataset to at least 2,500 images, incorporating co-infection cases and additional disease stages; (ii) evaluating YOLOv11-Nano and EfficientNetB0 for deployment on Raspberry Pi 5 and Jetson Nano; (iii) integrating multispectral near-infrared imaging to detect sub-visual, pre-symptomatic infections; and (iv) developing an end-to-end mobile advisory application for smallholder farmers that combines disease detection with region-specific treatment recommendations.

ACKNOWLEDGEMENTS

The authors thank the farming community of Solapur, Bhavnagar, and Nashik districts for their hospitality and cooperation during field data collection. The authors also acknowledge the annotation guidance provided by the Department of Plant Pathology, Navsari Agricultural University. No external funding was received for this study.

REFERENCES

[1] National Horticulture Board, Government of India. (2023). Area and Production of Horticulture Crops: Final Estimates 2022–23. NHB, Gurugram.

[2] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. IEEE CVPR, 779–788.

[4] Barbedo, J. G. A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus, 2(1), 660.

[5] Pydipati, R., Burks, T. F., & Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. Computers and Electronics in Agriculture, 52(1–2), 49–59.

[6] Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. Frontiers in Plant Science, 8, 1852.

[7] Ahmed, S., Raza, M., & Khan, M. (2019). SVM-based pomegranate disease detection using HOG features. International Journal of Advanced Computer Science and Applications, 10(4), 88–95.

[8] Patel, R., & Shah, S. (2021). Transfer learning for pomegranate disease classification using VGG-16. Proceedings of IEEE ICCUBEA 2021, 1–6.

[9] Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., & Maddikunta, P. K. R. (2020). Deep neural networks to predict diabetic retinopathy (with a special emphasis on a grading approach). IET Image Processing, 14(14), 3362–3372.

[10] Kumari, P., Singh, A., & Gupta, R. (2022). YOLOv5-based mango disease detection in orchard settings. Computers and Electronics in Agriculture, 201, 107287.

[11] Sharma, V., Kaur, M., & Verma, V. (2023). YOLOv8 for real-time citrus disease detection and localisation. Agricultural Systems, 215, 103847.

[12] Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., & Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. Applied Soft Computing, 86, 105933.

[13] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of ICML 2019, 6105–6114.

[14] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of IEEE CVPR 2016, 770–778.

[15] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. Proceedings of ICLR 2015.