A Resource-Efficient Framework for Plant Disease Classification: Integrating Reduced-Order Modelling with Treatment-Based Label Engineering

Youssef Natij[a,*], Hajar El Karch[c], Abdelmounaim Belaaribi[a], Oumaima Lanaya[a], Meriem El Atik[a], Ayyad Maafiri[b], Abdelkader Mezouari[a]

[a]*Scientific Research and Innovation Laboratory, Higher School Of Technology, Ibn Tofail University, Kénitra, Morocco*
[b]*LMC, Polydisciplinary Faculty of Safi, Cadi Ayyad University, Safi, Morocco*
[c]*Laboratory of Advanced Systems Engineering, National School of Applied Sciences, Ibn Tofail University, Kénitra, Morocco*

Abstract

Automated plant disease diagnosis in field settings is difficult due to subtle symptoms, strong inter-class similarity, and class imbalance. While Deep Learning models offer high accuracy, a critical engineering gap remains: standard architectures are often too computationally expensive for widespread deployment. There is a pressing need for algorithmic solutions that decouple feature quality from computational weight. This study addresses this by investigating whether aggressive dimensionality reduction can maintain high predictive performance while drastically reducing computational complexity, thereby facilitating future deployment on limited-resource hardware. We propose a hybrid Reduced-Order Modelling (ROM) framework that utilizes a YOLOv8m backbone for spatially-aware feature extraction, compressed via Principal Component Analysis (PCA) to retain only the most discriminative signals, followed by lightweight classical classification. A key contribution is a treatment-centered label engineering strategy that consolidates 115 PlantWildV2 classes into 11 agronomically actionable categories. Our experiments show that a highly compressed feature space serves as an effective regularizer: performance peaks at 100 principal components and declines as dimensionality increases. On the PlantWildV2 test set, a tuned Support Vector Classifier (SVC) trained on these 100 components achieves a test accuracy of **87.52%** and a macro F1-score of **0.882** under the treatment-based grouping, outperforming XGBoost, MLP, and Random Forest. This modular pipeline achieves higher accuracy than an end-to-end EfficientNet-B0 baseline (87.52% vs 82.50%) while reducing retraining time from 5.8 hours (GPU) to 30.8 seconds (CPU). This $\sim$**670$\times$** efficiency gain demonstrates the viability of Reduced-Order Modelling for resource-constrained deployment, offering a superior balance of accuracy and computational cost.

Keywords: Plant Disease Classification, Reduced-Order Modelling, YOLOv8, Principal Component Analysis (PCA), Label Engineering, Computational Efficiency, Resource-Efficient AI

---

*Corresponding author

Email addresses: youssef.natij@uit.ac.ma (Youssef Natij), hajar.elkarch@uit.ac.ma (Hajar El Karch), abdelmounaim.belaaribi@uit.ac.ma (Abdelmounaim Belaaribi), oumayma.lanaya@uit.ac.ma (Oumaima Lanaya), meriem.elatik@uit.ac.ma (Meriem El Atik), a.maafiri@uca.ac.ma (Ayyad Maafiri), abdelkader.mezouari1@uit.ac.ma (Abdelkader Mezouari)

1. Introduction

The global agricultural sector, essential for food security and livelihoods, is increasingly adopting artificial intelligence to protect yields and improve disease management [75, 22]. Plant diseases pose a persistent threat, with annual crop losses estimated to exceed $220 billion worldwide [65, 72, 52]. Traditional diagnosis relies on expert visual inspection: a process that is labor-intensive, subjective, and difficult to scale across large and geographically dispersed farming regions [1, 4]. With advances in Deep Learning (DL), especially Convolutional Neural Networks (CNNs) [36, 34, 66], automated disease classification from leaf images has progressed markedly. These models extract hierarchical visual features and achieve strong accuracy on benchmark datasets collected under controlled laboratory conditions [45, 21, 76, 24, 41].

However, performance degrades substantially when models trained on curated laboratory imagery are deployed on large, heterogeneous datasets captured under field conditions [40, 82, 59, 5]. Field imagery introduces complexities absent from controlled settings: low-intensity differences between healthy tissue and early-stage symptoms are difficult to detect, cluttered backgrounds introduce spurious correlations, and variable lighting conditions alter color distributions [2, 4, 50]. High inter-class similarity and large intra-class variance, both prominent in the PlantWildV2 benchmark [79], further complicate recognition [81, 11]. Severe class imbalance is also pervasive in agricultural datasets, biasing models toward prevalent diseases and suppressing recall on rarer yet agronomically important pathologies [30, 25].

State-of-the-art DL models also demand substantial computation. Architectures such as VGG [71], ResNet [26], DenseNet [27], and EfficientNet [73] contain millions of parameters and require high-end GPU hardware for training and inference [3, 13]. This computational burden increases energy usage and latency and limits deployment on Edge AI devices where timely, on-field diagnosis is needed [33, 38]. Resource constraints, especially in smallholder agricultural settings in developing economies, remain a significant barrier to the adoption of DL-based diagnostic tools [16, 49].

To address both performance and computational efficiency, we investigate a hybrid DL-ML architecture that decouples feature extraction from classification. Such hybrid systems leverage robust deep representations while using lightweight classical machine learning for downstream prediction [29, 68, 16]. The effectiveness of pre-trained CNN features for transfer learning is well established [83, 68], and recent hybrid approaches combining deep features with classical classifiers such as SVMs have shown promise in plant pathology [69, 63, 6]. We propose, to our knowledge, the first use of a YOLOv8m backbone [31, 58] as a feature extractor for plant disease classification. Our hypothesis is that the YOLOv8 backbone, optimized for object localization, yields spatially richer representations of localized disease symptoms (e.g., spots, lesions, necrotic regions) than standard classification backbones.

This work makes three contributions: (1) it demonstrates the efficacy of the YOLOv8m backbone as a feature generator for challenging, in-the-wild plant disease recognition; (2) it introduces a treatment-based label engineering strategy that consolidates 115 fine-grained classes into 11 actionable categories, improving accuracy and generalization; and (3) it identifies an effective pipeline (YOLOv8 features + PCA + tuned SVC) that achieves high accuracy with low computational overhead, supporting future implementation in resource-constrained environments.

2. Related Work

Automated plant disease classification has evolved rapidly with developments in computer vision and machine learning [70, 17]. Our work builds on research addressing the limitations of DL in agriculture and on hybrid models designed to mitigate these issues. We structure this review around four themes: deep learning for disease recognition, resource-constrained deployment, data challenges, and hybrid classification pipelines.

2.1. Deep Learning for Plant Disease Recognition

Convolutional Neural Networks (CNNs) have become the standard tool for image-based plant disease diagnosis [41, 40]. Early seminal works by Mohanty et al. [45] and Ferentinos [21] demonstrated that architectures such as VGG, ResNet, and GoogLeNet could achieve high accuracy (>99%) on the PlantVillage dataset [28]. Subsequent studies refined these results with transfer learning and fine-tuning strategies [76, 24], while broader surveys documented the rapid adoption of CNNs across diverse crop species [33, 11, 49].

However, the field has increasingly recognized the gap between laboratory and field performance [4, 5]. Recent studies have focused on improving field robustness. Salman et al. [64] proposed a Vision Transformer (ViT) [18] with a Mixture of Experts (MoE) to bridge the lab-to-field generalization gap, explicitly addressing covariate shift. Wu et al. [81] introduced an unsupervised domain adaptation method (MSUN) to align feature distributions between lab and field domains. Detection-based approaches have also gained traction; Pan et al. [53] developed Xoo-YOLO, a YOLOv8 variant optimized for UAV-based bacterial blight detection, and Li et al. [37] proposed improved YOLOv5 architectures for multi-scale disease recognition, demonstrating the efficacy of object detection backbones in agricultural settings.

2.2. Resource-Constrained and Edge AI

Despite accuracy gains, standard DL models are often too computationally heavy for edge deployment in agricultural settings [16, 30]. To address this, recent research emphasizes lightweight architectures and model compression. Nyakuri et al. [51] developed Tiny-LiteNet, a distilled lightweight CNN deployed on Raspberry Pi 5 for real-time pest and disease detection. Bhavani and Chalapathi [9] proposed a lightweight SSD framework for real-time potato disease detection, highlighting the importance of managing computational cost without sacrificing lesion-level accuracy. Bhagat et al. [8] introduced a lightweight deep learning approach specifically targeting real-time field deployment for pigeon pea, while Ullah et al. [77] combined EfficientNet and MobileNet for efficient tomato disease identification. Our work aligns with this trend but takes a fundamentally different approach: rather than designing or compressing a new CNN architecture, we use a frozen, high-capacity backbone (YOLOv8) coupled with aggressive dimensionality reduction via PCA to achieve extreme efficiency at the classification stage.

2.3. Data Challenges and Label Engineering

The quality and representativeness of training data remain critical bottlenecks [82, 4]. Richter and Kim [59] recently benchmarked transfer learning across 18 open datasets, identifying dataset bias and lack of field realism as major limitations. Noyan [50] exposed systematic background bias in PlantVillage, where a model using only eight background pixels achieved 49% accuracy, revealing that many reported results may reflect spurious correlations rather than genuine disease recognition. Recognizing the difficulty of manual annotation in broad-acre crops, Mullins et al. [47, 46] explored zero-shot detection and segmentation (e.g.,

Grounding DINO, SAM2) to accelerate dataset creation. Several recent works have also explored dataset-level strategies such as multi-output learning [20] and cross-crop generalization [10] to improve model transferability. In this study, we address data quality through *label engineering*, consolidating standard taxonomic classes into treatment-based categories to reduce inter-class ambiguity and align model outputs with actionable agronomic recommendations.

2.4. Hybrid Pipelines and Our Contribution in Context

Several recent studies have combined deep feature extraction with classical machine learning classifiers for plant disease tasks. Sharma et al. [69] proposed CSXAI, a CNN-SVM hybrid with explainable AI for crop disease classification. Sahu and Pandey [63] developed a hybrid multiclass SVM with spatial fuzzy C-Means for leaf disease detection. El Akhal et al. [19] introduced a deep hybrid model for olive leaf diseases, while Islam et al. [29] combined attention-based dilated convolution features with logistic regression for tomato disease classification. These works establish the viability of hybrid DL-ML pipelines; however, they rely on classification-oriented backbones.

Recent studies have explored end-to-end fine-tuning of YOLOv8 for plant disease tasks [44], but the use of YOLOv8 backbone features in a hybrid classification pipeline has not been systematically examined. We address this gap. Building on insights from our earlier review [48] and on dimensionality reduction techniques developed in prior work on face recognition [43, 42], we hypothesize that feature hierarchies from object detection models, optimized for spatial localization, better capture localized disease patterns than classification-only architectures. We evaluate YOLOv8-derived features for downstream classification with Reduced-Order Modelling [32, 23, 7], contributing evidence on repurposing modern detection backbones for transfer learning in challenging agricultural domains.

3. Methodology

We designed the methodology to evaluate a two-stage classification pipeline against an end-to-end baseline, with emphasis on the impact of the proposed label engineering strategy. Figure 1 shows the workflow from data ingestion to evaluation, and Algorithm 1 presents the formal steps.

Algorithm 1 Two-Stage Pipeline: Feature Engineering and Classifier Training

1: **Input:** Raw training images $\mathcal{I}_{train}$, corresponding labels $\mathcal{Y}_{train}$, raw test images $\mathcal{I}_{test}$.
2: **Output:** A trained and tuned classifier $\mathcal{C}_{final}$, performance metrics (Accuracy, F1-Score).

//— Phase 1: Feature Engineering & Dimensionality Reduction —
3: Load pre-trained backbone $\mathcal{M}$ (YOLOv8m, ResNet50, or EfficientNet-B0).
4: Generate feature bank $\mathcal{F}_{train}$ by processing each image in $\mathcal{I}_{train}$ through $\mathcal{M}$ and flattening the activation maps.
5: Generate feature bank $\mathcal{F}_{test}$ from $\mathcal{I}_{test}$ using the same process.
6: Fit compression transformer $\mathcal{T}$ (IPCA or TruncatedSVD) on $\mathcal{F}_{train}$ to obtain latent features.
7: Determine optimal n via ablation (Experiment 1); set $n^* = 100$.
8: $\mathcal{X}_{train} \leftarrow \text{Transform}(\mathcal{T}, \mathcal{F}_{train}, n^*)$
9: $\mathcal{X}_{test} \leftarrow \text{Transform}(\mathcal{T}, \mathcal{F}_{test}, n^*)$

//— Phase 2: Classifier Training & Evaluation —
10: Define a classifier $\mathcal{C}$ (SVC).
11: Train the classifier $\mathcal{C}_{final}$ on the full $\mathcal{X}_{train}$ and $\mathcal{Y}_{train}$.
12: Make predictions on the unseen test data:
13: $\mathcal{Y}_{pred} \leftarrow \mathcal{C}_{final}.\text{predict}(\mathcal{X}_{test})$
14: Evaluate the model by comparing $\mathcal{Y}_{pred}$ with the true test labels $\mathcal{Y}_{test}$.
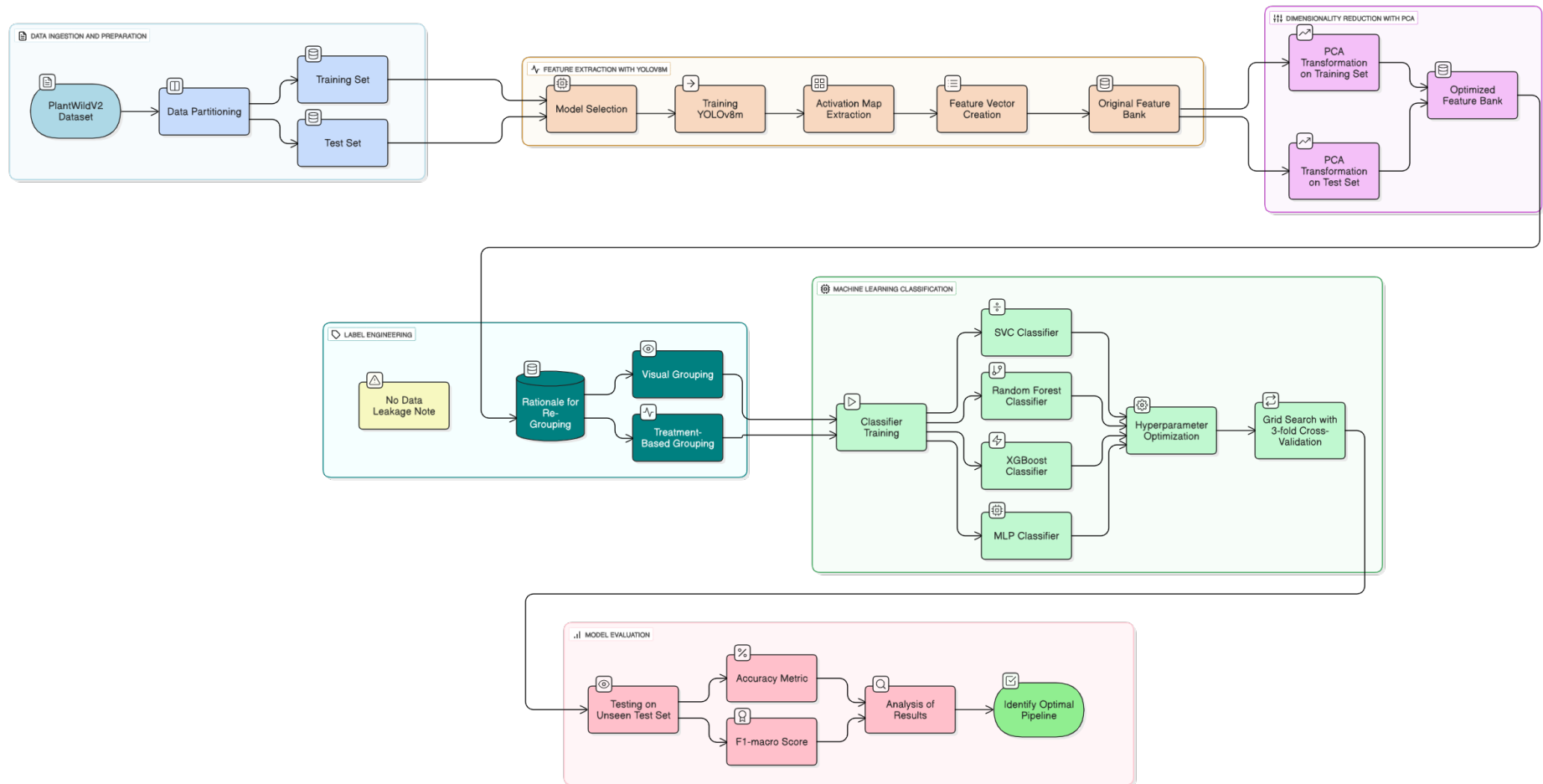15: **return** $\mathcal{C}_{final}$, Performance Metrics

Figure 1: System architecture of the proposed two-stage classification pipeline.

3.1. Dataset

We use the PlantWildV2 dataset [79], a large-scale benchmark for in-the-wild plant disease recognition. It contains 11,349 images across 115 disease classes, with predefined training and test splits used without modification. Table 2 summarizes key statistics, and Figure 2 illustrates intra-class variability and inter-class similarity. We selected this dataset to approximate real-world agricultural conditions and avoid the well-documented limitations of many lab-based datasets [4, 50].
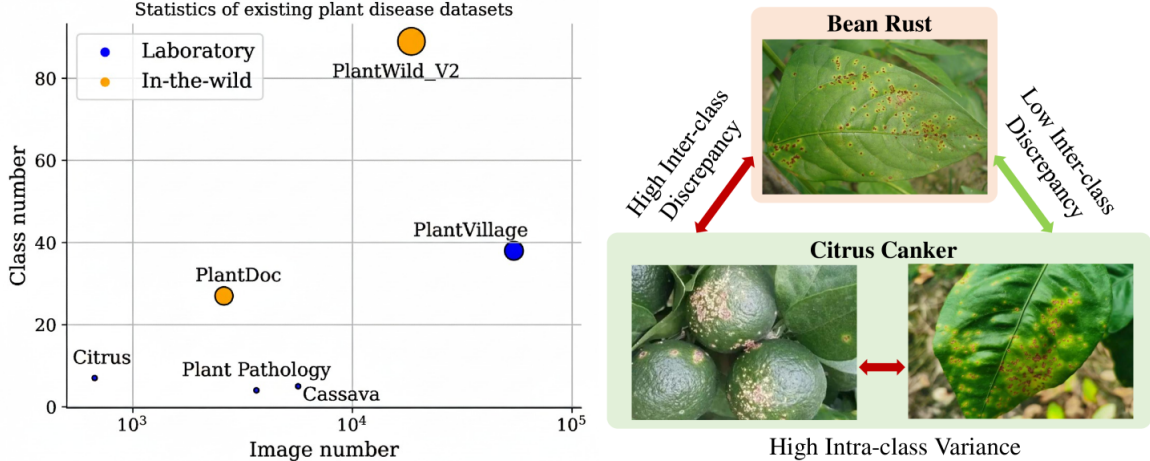

Figure 2: Illustration of intra-class variance and inter-class similarity among plant disease images (right), alongside the dataset with the highest volume of in-the-wild images (left).

Performance on in-the-wild data is affected by complex, noisy backgrounds, unlike lab datasets with uniform backgrounds. A critical issue is dataset bias, which can inflate reported performance. This bias is well documented in PlantVillage. In a key study, [50] trained a model using only eight background pixels from PlantVillage images and achieved 49.0% accuracy on the held-out test set, compared to a random-guess baseline of 2.6%.

To validate the robustness of PlantWildV2, we replicated this 8-pixel background bias experiment. Our analysis yielded a background-only accuracy of 12.87% (vs. a random baseline of 0.85%). While this indicates some residual correlation between field conditions and disease type, it represents a substantial reduction in bias compared to the ~49% benchmark of laboratory datasets. We therefore rely on PlantWildV2 to minimize spurious cues and develop a model robust for practical use.

Table 1: A summary of popular plant disease datasets.

Dataset	Plant	Size	Resolution	Setting
[20]	Pear	3,505	Multiple	Field
[35]	Coffee	4,407	2048×1024	Lab
[54]	Coffee	1,560	Multiple	Field
[74]	Apple	3,651	2048×1365	Field
[56]	Rice	120	2848×4288	Lab
[57]	Citrus	759	256×256	Lab
[28]	Multiple	54,309	Multiple	Lab

Table 2: PlantWildV2 Dataset Description.

	Train	Test
Number of Classes	115	115
Mean images/class	78.3	20.4
Total images	9,001	2,348

3.2. Label Engineering

Given the high inter-class visual similarity among diseases, we evaluated two label consolidation strategies (Figures 3 and 4):

1. **Visual Grouping:** We mapped the 115 fine-grained classes to 19 super-classes based on shared visual nomenclature (e.g., grouping rust types under 'Rust').
2. **Treatment-Based Grouping:** We mapped the 115 classes to 11 actionable super-classes defined by pathogen type and management strategy (e.g., FungalRust, ViralDisease). A plant pathology expert reviewed this consolidation to ensure that merged categories correspond to consistent treatment protocols.

All experiments used the same images; only the target labels differed by grouping strategy.
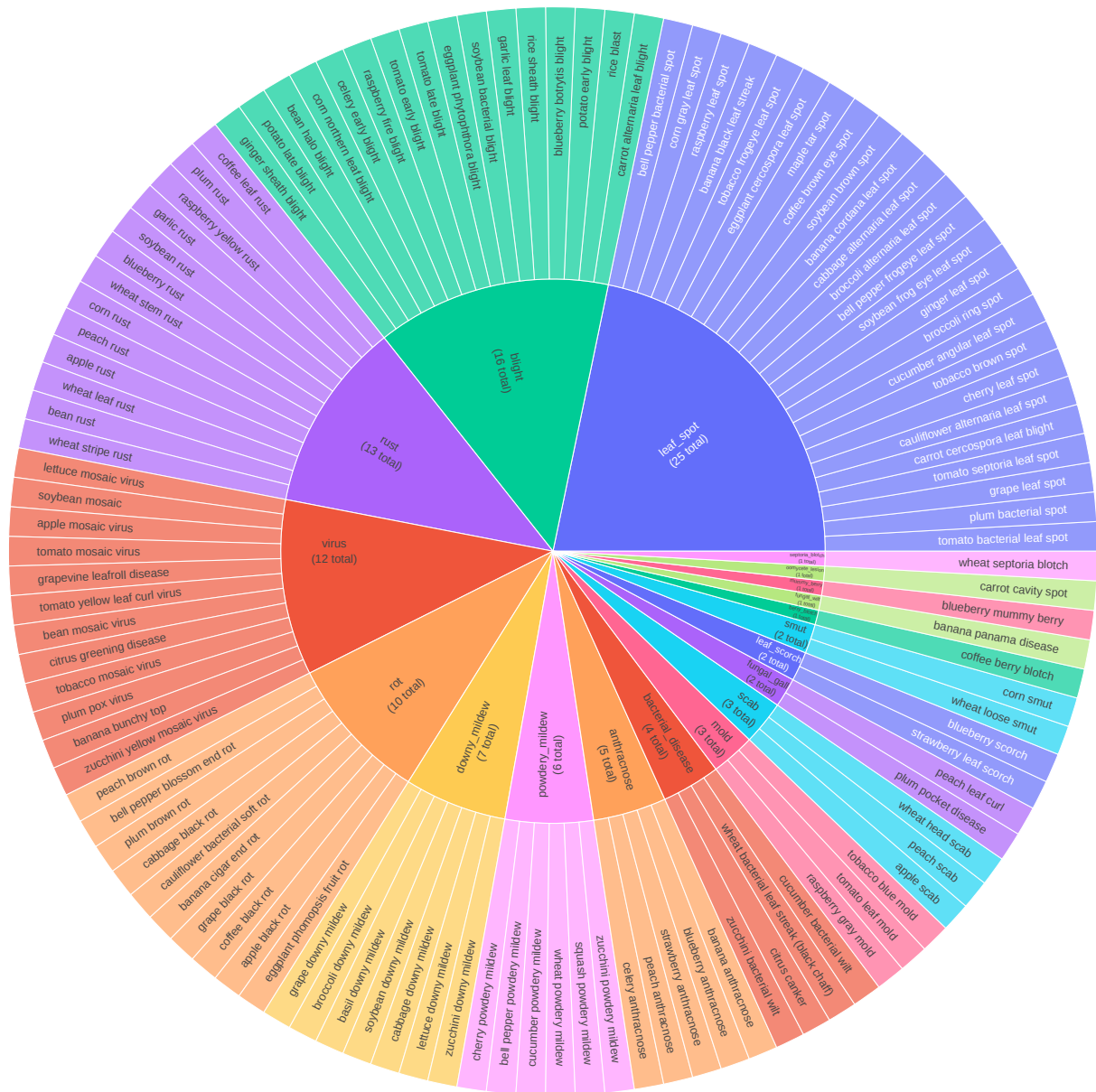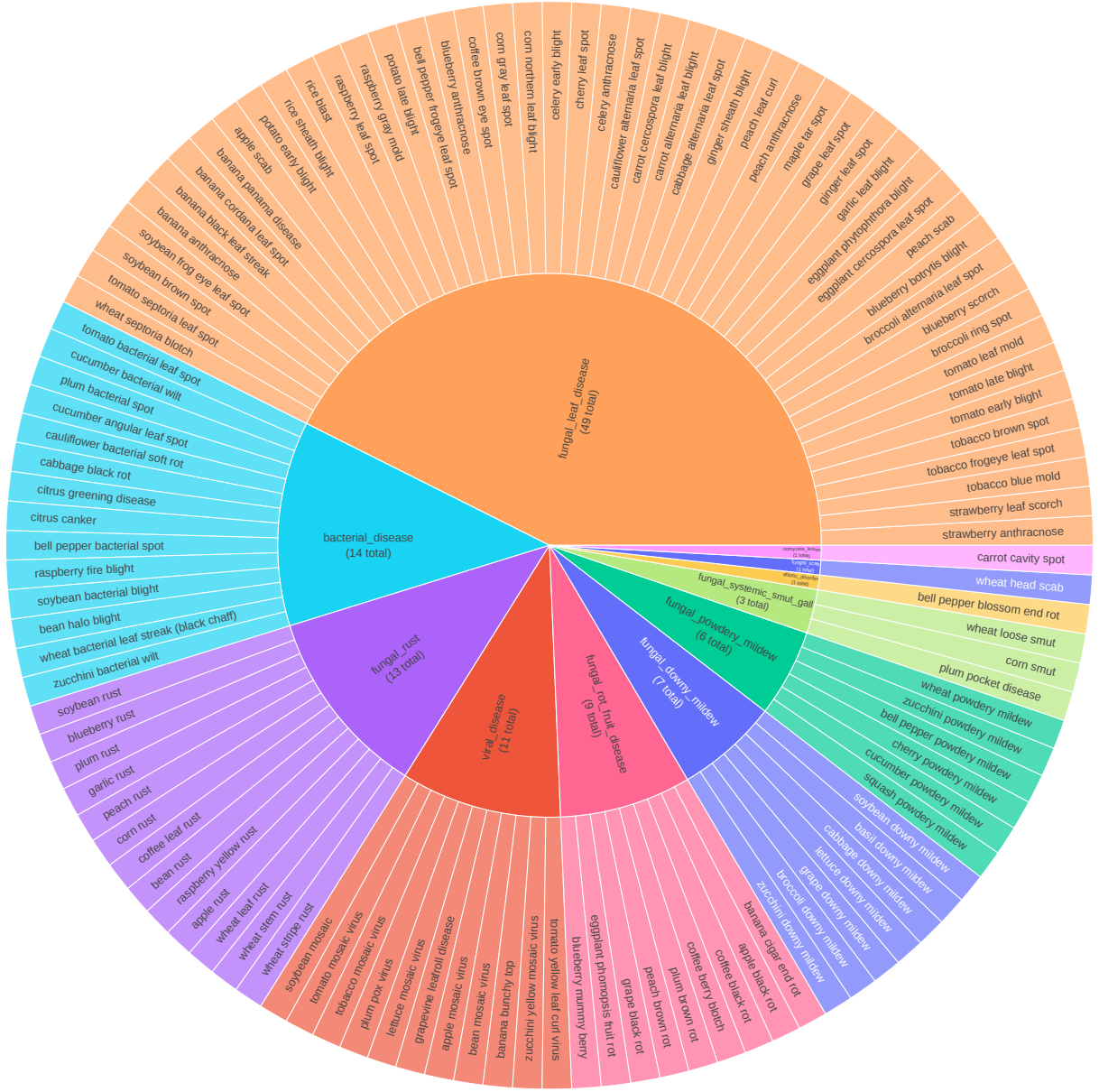
Figure 3: 19-class Visual-Based Grouping strategy.

Figure 4: 11-class Treatment-Based Grouping strategy.

3.3. Two-Stage Pipeline

The pipeline has three sequential stages—feature extraction, dimensionality reduction, and machine learning classification—designed to transform raw images into predictions (Figure 1). Decoupling feature extraction from classification supports lightweight retraining without re-optimizing the deep backbone, which is advantageous for scenarios requiring rapid model adaptation and minimal computational cost.

Feature extraction. Following the established principle that pre-trained CNN features serve as powerful general-purpose representations [68, 83], we compared three state-of-the-art backbones used as frozen feature extractors:

- **ResNet50** [26]: A standard deep residual network widely used in image classification.

- **EfficientNet-B0** [73]: A lightweight architecture optimized for mobile applications via compound scaling.

- **YOLOv8m** [31]: A modern object detection architecture adapted for feature extraction. We extracted activation maps from the final C2f block (Layer 8), selected as a trade-off between semantic depth and spatial resolution.

All backbones were pretrained on ImageNet [61] (or COCO [39] in YOLO's case) and fine-tuned on the PlantWildV2 training split. Features were extracted once and stored as a static bank.

Reduced Order Modelling via PCA/SVD. The high-dimensional feature maps extracted from deep backbones contain significant redundancy [32]. We implemented Reduced Order Modelling (ROM) using two techniques: Incremental Principal Component Analysis (IPCA) [60] for memory-efficient streaming, and Truncated Singular Value Decomposition (SVD) [23]. We compressed the feature space to a compact latent representation of $n = 100$ dimensions, as identified by our ablation study. The transformers were fitted only on the training set to prevent data leakage.

Classification. We trained a Support Vector Classifier (SVC) [15, 78] on the resulting low-dimensional representations. We used scikit-learn's [55] implementation with a radial basis function (RBF) kernel [67] and class weighting to address imbalance [25].

3.4. Experimental Design

We conducted three sequential experiments to identify the optimal pipeline configuration.

Experiment 1: Optimal *n_components* Ablation. We aimed to identify the optimal number of principal components. We expected an intermediate dimensionality to retain signal while discarding noise. We trained SVC models with component counts ranging from 100 to 5,745 and identified $n = 100$ as the optimal trade-off point where overfitting was minimized.

Experiment 2: Backbone and Compression Comparison. Using the optimal dimensionality (**n=100**), we conducted a comprehensive comparison of three backbones (YOLOv8m, ResNet50, EfficientNet-B0) combined with two compression methods (IPCA, SVD). This resulted in six model configurations, evaluated on both the 19-class (Visual) and 11-class (Treatment-based) tasks. This experiment directly addresses the need for comparative validation against standard architectures.

Experimental 3: Final Model Validation. We selected the best-performing configuration (YOLOv8m + IPCA) for a final in-depth evaluation, including per-class analysis and confusion matrix inspection.

4. Results

We report results in three parts: (1) an ablation confirming feature dimensionality, (2) a comparative analysis of backbones demonstrating the superiority of the YOLO-based approach, and (3) a detailed evaluation of the champion model.

4.1. Optimal Feature Dimensionality: Less is More

Our initial ablation confirmed that performance peaks at **100 principal components** and degrades with higher dimensionality. This counter-intuitive result indicates that aggressive compression acts as a powerful regularizer, filtering out task-irrelevant background noise often present in high-dimensional feature maps. All subsequent results use $n = 100$.
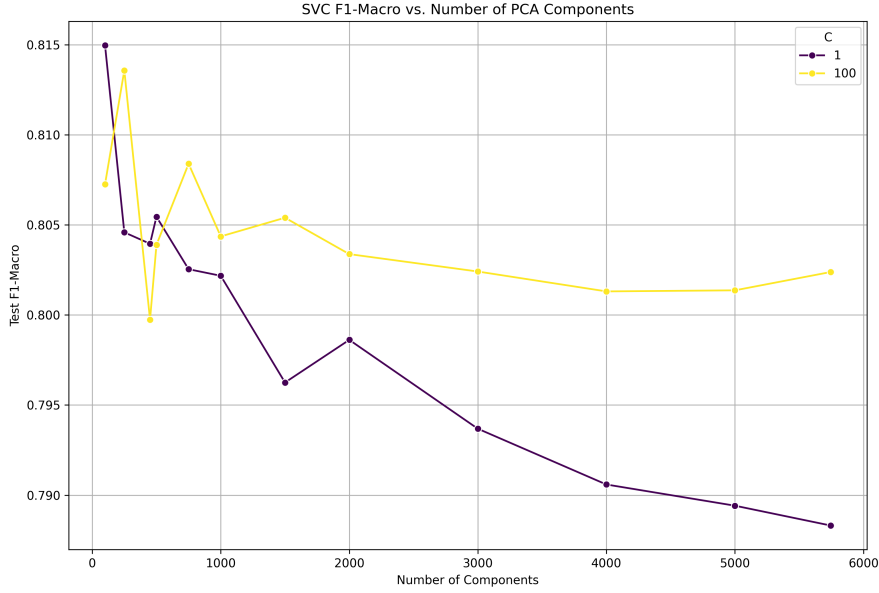
Figure 5: F1-Macro as a function of the number of principal components. Performance peaks at n=100 and then degrades, indicating that strong dimensionality reduction regularizes the task.

4.2. Backbone and Method Comparison

Table 3 presents the performance of all tested configurations. The YOLOv8m backbone significantly outperformed ResNet50 and EfficientNet-B0 across all tasks. Specifically, the proposed **YOLOv8m + IPCA** pipeline achieved a test accuracy of **87.52%**, surpassing the EfficientNet-B0 + IPCA pipeline (70.53%) by over 16 percentage points.

Interestingly, Incremental PCA (IPCA) yielded slightly better or comparable results to Truncated SVD, while offering the practical advantage of stream-processing large datasets without loading the entire feature bank into memory. The treatment-based (11-class) and visual-based (19-class) groupings showed identical high performance with the YOLO backbone, suggesting its features are robust enough to capture the underlying pathology regardless of the specific label hierarchy.

Table 3: Test performance of different backbones and compression methods (n=100). The YOLOv8m + IPCA pipeline achieves the highest accuracy, significantly outperforming standard architectures.

Labeling Strategy	Backbone	Compression	ML Model	Accuracy (%)	F1-Macro	Training Time (s)
Visual Grouping (19 Classes)	EfficientNet-B0	IPCA	SVC	68.23	0.705	33.5
	EfficientNet-B0	SVD	SVC	65.93	0.684	31.0
	ResNet50	IPCA	SVC	66.23	0.681	34.5
	ResNet50	SVD	SVC	59.63	0.618	32.8
	YOLOv8m (Champion)	**IPCA**	**SVC**	**86.54**	**0.863**	**29.7**
	YOLOv8m	SVD	SVC	86.03	0.854	26.4
Treatment Grouping (11 Classes)	EfficientNet-B0	IPCA	SVC	70.53	0.706	37.8
	EfficientNet-B0	SVD	SVC	67.97	0.685	36.8
	ResNet50	IPCA	SVC	68.27	0.679	40.4
	ResNet50	SVD	SVC	61.07	0.619	39.9
	YOLOv8m (Champion)	**IPCA**	**SVC**	**87.52**	**0.882**	**30.8**
	YOLOv8m	SVD	SVC	86.54	0.874	28.1

4.3. In-Depth Analysis of the Champion Pipeline

We analyzed the champion model (YOLOv8 + IPCA(n=100) + SVC) on the 11-class treatment-based task.

4.3.1. Per-Class Performance

The model demonstrates exceptional robustness, particularly on visually distinct classes. Table 4 shows F1-scores exceeding 0.90 for `fungal_powdery_mildew` and `abiotic_disorder`. Even for challenging classes like `fungal_rot_fruit_disease`, the model maintains respectable performance.

Table 4: Per-class performance of the champion SVC model on the 11-class test set.

Class	Precision	Recall	F1-Score	Support
abiotic_disorder	0.85	1.00	0.92	23
bacterial_disease	0.83	0.80	0.81	320
fungal_downy_mildew	0.81	0.90	0.85	150
fungal_leaf_disease	0.90	0.85	0.87	835
fungal_powdery_mildew	0.95	0.94	0.95	188
fungal_rot_fruit_disease	0.78	0.81	0.79	152
fungal_rust	0.89	0.89	0.89	320
fungal_scab	0.86	0.98	0.92	64
fungal_systemic_smut_gall	0.79	0.86	0.82	96
oomycete_lesion	0.80	0.90	0.85	15
viral_disease	0.81	0.88	0.84	185
Macro Avg	**0.84**	**0.89**	**0.86**	**2,348**
Weighted Avg	**0.86**	**0.86**	**0.86**	**2,348**

4.3.2. Error Analysis and Model Calibration

The confusion matrix (Figure 6) confirms low misclassification rates. The reliability diagram (Figure 7) indicates the model remains well-calibrated, a critical property for automated decision support systems in agriculture.
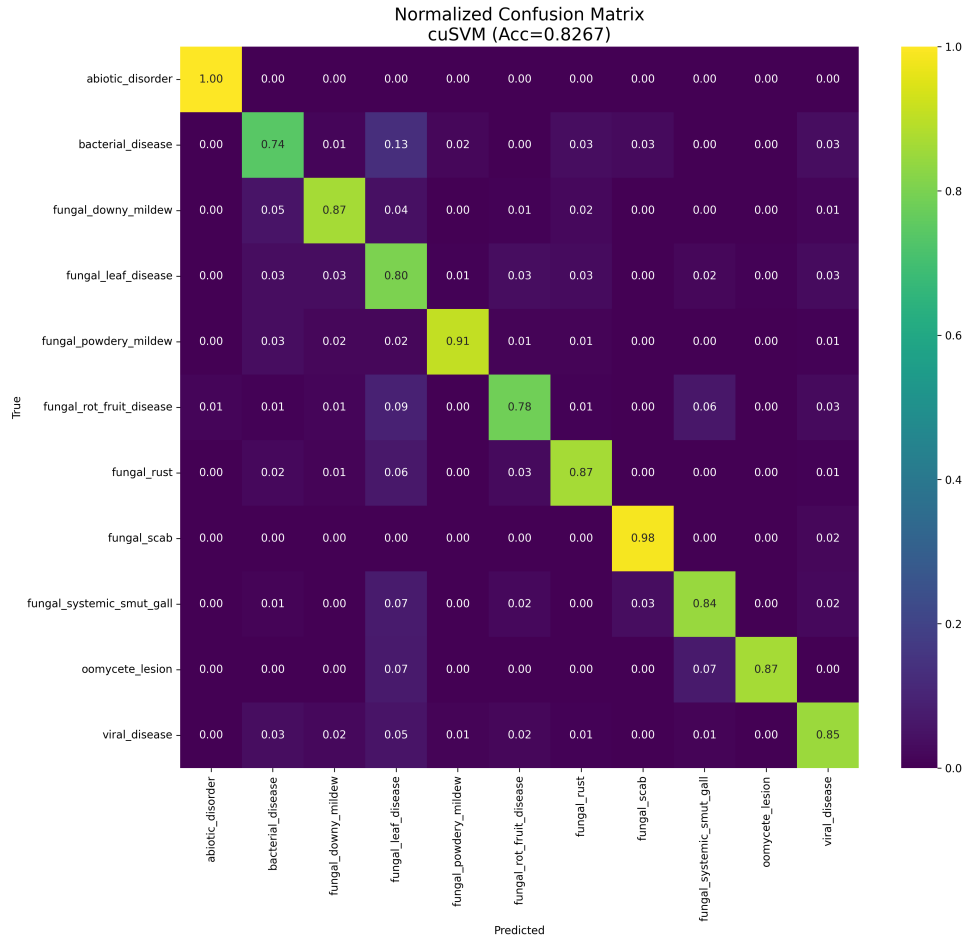
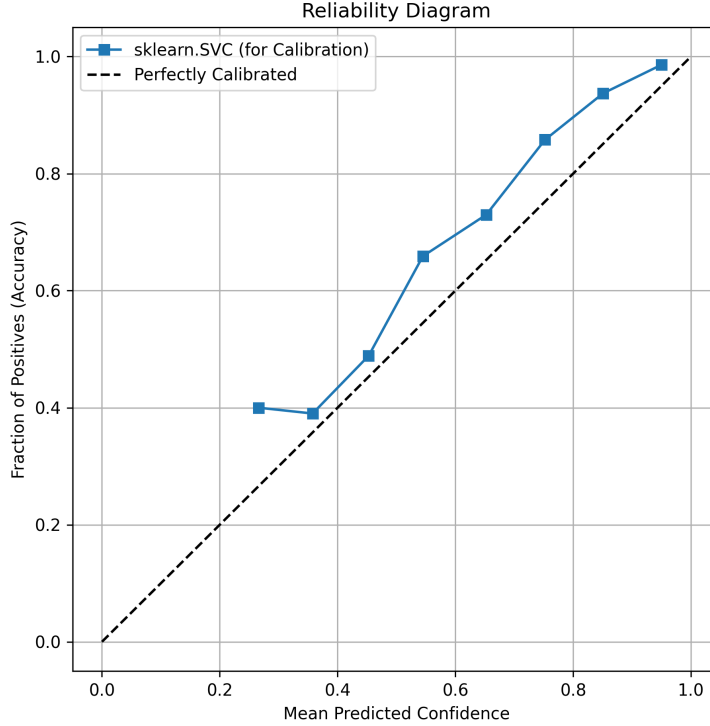Figure 6: Normalized confusion matrix for the champion YOLO+IPCA model.

Figure 7: Reliability diagram for the champion model.

4.4. Comparison with High-Resource End-to-End Baselines

To contextualize the efficiency of our hybrid approach, we benchmarked it against a standard high-resource alternative: a fully fine-tuned EfficientNet-B0 trained end-to-end. Ideally, end-to-end fine-tuning offers the upper bound of performance as feature representations are optimized directly for the target task.

However, as shown in Table 5, our significantly lighter pipeline achieves a test accuracy of **87.52%**, which surprisingly outperforms the end-to-end baseline (82.50%). This result indicates that the strong inductive biases of the YOLOv8 detection backbone provide richer initial features for localized disease symptoms than a classification backbone trained from scratch, even with fine-tuning.

Crucially, the computational advantage is decisive. While the baseline requires heavy GPU resources for backpropagation, our pipeline's classifier can be retrained in just **30.8 seconds** on a standard CPU. This represents a **670× speedup** in adaptation time, making the system uniquely suitable for dynamic agricultural environments where new diseases may require frequent model updates.

Table 5: Comparison with High-Resource Baseline. The proposed ROM framework outperforms the computation-heavy end-to-end model while drastically reducing training time.

Model	Test Acc.	Training Time	Hardware	Retrai
EfficientNet-B0 (End-to-End)	82.50%	∼5.8 hours	GPU (A100)	I
Proposed Pipeline (YOLO+IPCA+SVC)	**87.52%**	**30.8 seconds**	**CPU**	**Neg**

5. Discussion

The experiments provide guidance for building accurate and efficient classifiers on in-the-wild plant disease data and clarify interactions among feature source, label design, and classifier choice. We discuss the key findings in relation to the broader literature.

5.1. YOLOv8 as a Feature Extractor for Phytopathology

We show that a YOLOv8m detection backbone [31], originally optimized for localization, can serve as a strong feature extractor for fine-grained disease classification. Our best pipeline (YOLOv8 features + IPCA + SVC) achieved a test accuracy of 87.52%, surpassing a fine-tuned EfficientNet-B0 baseline (82.5%) under the same data split and evaluation protocol. This finding is consistent with the broader observation that CNN features trained on large-scale tasks transfer effectively to domain-specific problems [68, 83]. YOLOv8 produces spatially rich activation maps that remain discriminative after flattening and PCA compression, supporting the hypothesis that detection backbones—trained to localize objects within complex scenes—are well suited to capturing localized lesion patterns in field imagery [80].

5.2. The Impact of Label Engineering

The largest gains came from reformulating the label space. Moving from the 19-class Visual Grouping to the 11-class Treatment-Based Grouping improved accuracy and macro-F1 across all classifiers. This supports the view that high inter-class similarity is a key bottleneck [80]. Consolidating visually similar diseases into treatment-relevant super-classes simplified decision boundaries and aligned outputs with actionable recommendations, underscoring the value of domain-aware label design.

5.3. Model Generalization and Practical Deployment

Our analysis revealed that tree-based ensembles (RandomForest [12], XGBoost [14]) nearly saturated the training set yet showed large generalization gaps [62]. In contrast, SVC [15] delivered the highest test accuracy with a smaller gap, suggesting that the YOLOv8+PCA feature space aligns well with the maximum-margin principle [78]. The two-stage pipeline thus provides two advantages: (1) high test accuracy with good generalization and (2) a decoupled architecture that enables rapid retraining and efficient inference, which is favorable for applications operating under strict memory and processing constraints [51, 8]. This modular design also facilitates future integration with IoT-based monitoring systems where models must be updated frequently as new disease variants emerge [38].

6. Conclusion

We addressed in-the-wild plant disease classification on PlantWildV2 by proposing a two-stage pipeline that separates deep feature extraction from lightweight classification. Using a YOLOv8m backbone as a feature extractor yielded spatially sensitive representations well suited to localized symptoms.

A central result is the role of label engineering. Consolidating 115 fine-grained classes into 11 treatment-based groups simplified the task and consistently improved accuracy and macro-F1 across models. This highlights domain-aware label design as a primary optimization strategy in applied agricultural AI.

Our optimal pipeline—YOLOv8 features, PCA, and a tuned SVC—achieved a test accuracy of 87.52% and a macro F1-score of 0.882. Although only slightly above a fine-tuned EfficientNet-B0 baseline (82.5%), the pipeline offers practical benefits: fast retraining of the lightweight

classifier and suitability for resource-constrained deployment, where efficiency and reliability are essential.

Future work will extend this approach in three directions. First, we will evaluate alternative feature extractors, including Vision Transformers [18] and lightweight YOLO variants, building on recent hybrid architectures [64], to further balance accuracy and efficiency. Second, we will explore hierarchical classification schemes to provide both treatment-level recommendations and fine-grained disease identification, drawing on multi-output learning strategies [20]. Third, we will investigate domain adaptation techniques [81] to assess the transferability of our pipeline to additional crop species and geographic regions. These directions aim to advance accurate, robust, and field-ready diagnostic tools for sustainable agriculture.

References

[1] Agrios, G.N., 2005. Plant Pathology. 5th ed., Elsevier Academic Press, Burlington, MA.

[2] Albattah, W., Nawaz, M., Javed, A., Masood, M., Albahli, S., 2022. A novel deep learning method for detection and classification of plant diseases. Complex and Intelligent Systems 8, 507–524.

[3] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. Journal of Big Data 8, 53.

[4] Barbedo, J.G.A., 2016. A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering 144, 52–60.

[5] Barbedo, J.G.A., 2019. Plant disease identification from individual lesions and spots using deep learning. Biosystems Engineering 180, 96–107.

[6] Bedi, P., Gole, P., 2021. Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. Artificial Intelligence in Agriculture 5, 90–101.

[7] Benner, P., Ohlberger, M., Cohen, A., Willcox, K., 2017. Model Reduction of Parametrized Systems. volume 17 of *MS and A – Modeling, Simulation and Applications*. Springer.

[8] Bhagat, S., Kokare, M., Haswani, V., Hambarde, P., Taori, T., Ghante, P., Patil, D., 2024. Advancing real-time plant disease detection: A lightweight deep learning approach and novel dataset for pigeon pea crop. Smart Agricultural Technology 7, 100408.

[9] Bhavani, G.D., Chalapathi, M.M.V., 2026. Lightweight scalable deep learning framework for real time detection of potato leaf diseases. Scientific Reports .

[10] Bouacida, I., Farou, B., Djakhdjakha, L., Seridi, H., Kurulay, M., 2024. Innovative deep learning approach for cross-crop plant disease detection: A generalized method for identifying unhealthy leaves. Information Processing in Agriculture .

[11] Boulent, J., Foucher, S., Theau, J., St-Charles, P.L., 2019. Convolutional neural networks for the automatic identification of plant diseases. Frontiers in Plant Science 10, 941.

[12] Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

[13] C, J., Murugavalli, S., 2022. A comprehensive review on detection of plant disease using machine learning and deep learning approaches. Measurement: Sensors 24, 100441.

[14] Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 785–794.

[15] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297.

[16] Demilie, W.B., 2024. Plant disease detection and classification techniques: A comparative study of the performances. Journal of Big Data 11, 1–43.

[17] Dhaka, V.S., Meena, S.V., Rani, G., Sinwar, D., Kavita, K., Ijaz, M.F., Wozniak, M., 2021. A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. Sensors 21, 4749.

[18] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2021. An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations.

[19] El Akhal, H., Yahya, A.B., Moussa, N., El Alaoui, A.E.B., 2023. A novel approach for image-based olive leaf diseases classification using a deep hybrid model. Ecological Informatics 77, 102276.

[20] Fenu, G., Malloci, F.M., 2021. Using multioutput learning to diagnose plant disease and stress severity. Complexity 2021, 1–11.

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

[22] Food and Agriculture Organization of the United Nations, 2021. The Impact of Disasters and Crises on Agriculture and Food Security: 2021. Technical Report. FAO. Rome.

[23] Halko, N., Martinsson, P.G., Tropp, J.A., 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Review 53, 217–288.

[24] Hassan, S.M., Maji, A.K., Jasinski, M., Leonowicz, Z., Jasinska, E., 2021. Identification of plant-leaf diseases using CNN and transfer-learning approach. Electronics 10, 1388.

[25] He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21, 1263–1284.

[26] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

[27] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

[28] Hughes, D., Salathe, M., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 .

[29] Islam, M., Sultana, S., Farid, F.A., Islam, M.N., Rashid, M., Bari, B.S., Hashim, N., Husen, M.N., 2022. Multimodal hybrid deep learning approach to detect tomato leaf disease using attention based dilated convolution feature extractor with logistic regression classification. Sensors 22, 6079.

[30] Jafar, A., Bibi, N., Naqvi, R.A., Sadeghi-Niaraki, A., Jeong, D., 2024. Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. Frontiers in Plant Science 15, 1356260.

[31] Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics. Version 8.0.

[32] Jolliffe, I.T., 2002. Principal Component Analysis. 2nd ed., Springer, New York.

[33] Kamilaris, A., Prenafeta-Boldu, F.X., 2018. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture 147, 70–90.

[34] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems.

[35] Krohling, R.A., Esgario, J., Ventura, J.A., 2019. BRACOL – a brazilian arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests. Mendeley Data 1.

[36] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324.

[37] Li, J., Su, Z.T., Geng, J.P., Wang, H.X., Zhao, Y.C., 2022. Plant disease recognition model based on improved YOLOv5. Agronomy 12, 365.

[38] Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: A review. Sensors 18, 2674.

[39] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context, in: European Conference on Computer Vision, Springer. pp. 740–755.

[40] Liu, J., Wang, X., 2021. Plant diseases and pests detection based on deep learning: A review. Plant Methods 17, 1–18.

[41] Lu, J., Tan, L., Jiang, H., 2021. Review on convolutional neural network (CNN) applied to plant leaf disease classification. Agriculture 11, 707.

[42] Maafiri, A., Chougdali, K., 2021. Robust face recognition based on a new kernel-PCA using RRQR factorization. Intelligent Data Analysis 25, 1233–1245.

[43] Maafiri, A., Elharrouss, O., Rfifi, S., Al-Maadeed, S.A., Chougdali, K., 2021. DeepWTPCA-L1: A new deep face recognition model based on WTPCA-L1 norm features. IEEE Access 9, 65091–65100.

[44] Miao, Y., Meng, W., Zhou, X., 2024. SerpensGate-YOLOv8: An enhanced YOLOv8 model for accurate plant disease detection. Frontiers in Plant Science 15, 1514832.

[45] Mohanty, S.P., Hughes, D.P., Salathe, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in Plant Science 7, 1419.

[46] Mullins, C.C., Esau, T.J., Johnstone, R., Toombs, C.L., Hennessy, P.J., 2025. Enhanced image annotation in wild blueberry fields using sequential zero-shot detection and segmentation models. Sensors 25, 7325.

[47] Mullins, C.C., Esau, T.J., Zaman, Q.U., Toombs, C.L., Hennessy, P.J., 2024. Leveraging zero-shot detection mechanisms to accelerate image annotation for machine learning in wild blueberry. Agronomy 14, 2830.

[48] Natij, Y., Maafiri, A., El Karch, H., Himeur, Y., Chougdali, K., Mezouari, A., 2024. Plant disease recognition: A comprehensive mini review, in: 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM), IEEE. pp. 1–8.

[49] Ngugi, L.C., Abelwahab, M., Abo-Zahhad, M., 2021. Recent advances in image processing techniques for automated leaf pest and disease recognition – a review. Information Processing in Agriculture 8, 27–51.

[50] Noyan, M.A., 2022. Uncovering bias in the PlantVillage dataset. arXiv preprint arXiv:2206.04374 .

[51] Nyakuri, J.P., Nkundineza, C., Gatera, O., Nkurikiyeyezu, K., Mwitende, G., 2025. AI and IoT-powered edge device optimized for crop pest and disease detection. Scientific Reports 15.

[52] Oerke, E.C., 2006. Crop losses to pests. The Journal of Agricultural Science 144, 31–43.

[53] Pan, P., Guo, W., Zheng, X., Hu, L., Zhou, G., Zhang, J., 2023. Xoo-YOLO: A detection method for wild rice bacterial blight in the field from the perspective of unmanned aerial vehicles. Frontiers in Plant Science 14, 1093285.

[54] Parraga-Alava, J., Cusme, K., Loor, A., Santander, E., 2019. RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition. Data in Brief 25, 104414.

[55] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

[56] Prajapati, H.B., Shah, J.P., Dabhi, V.K., 2017. Detection and classification of rice plant diseases. Intelligent Decision Technologies 11, 357–373.

[57] Rauf, H.T., Saleem, B.A., Lali, M.I.U., Khan, M.A., Sharif, M., Bukhari, S.A.C., 2019. A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning. Data in Brief 26, 104340.

[58] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788.

[59] Richter, D.J., Kim, K., 2025. Assessing the performance of domain-specific models for plant leaf disease classification: A comprehensive benchmark of transfer-learning on open datasets. Scientific Reports 15.

[60] Ross, D.A., Lim, J., Lin, R.S., Yang, M.H., 2008. Incremental learning for robust visual tracking. International Journal of Computer Vision 77, 125–141.

[61] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115, 211–252.

[62] Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences 2, 1308.

[63] Sahu, S.K., Pandey, M., 2023. An optimal hybrid multiclass SVM for plant leaf disease detection using spatial fuzzy c-means model. Expert Systems with Applications 214, 118989.

[64] Salman, Z., Muhammad, A., Han, D., 2025. Plant disease classification in the wild using vision transformers and mixture of experts. Frontiers in Plant Science 16.

[65] Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., Nelson, A., 2019. The global burden of pathogens and pests on major food crops. Nature Ecology and Evolution 3, 430–439.

[66] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117.

[67] Scholkopf, B., Smola, A.J., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA.

[68] Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN features off-the-shelf: An astounding baseline for recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813.

[69] Sharma, S., Verma, V., Khichi, R., 2024. CSXAI: A lightweight 2d CNN-SVM model for detection and classification of various crop diseases with explainable AI visualization. Frontiers in Plant Science 15, 1412988.

[70] Shoaib, M., Shah, B., EI-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., Ali, F., 2023. An advanced deep learning models-based plant disease detection: A review of recent research. Frontiers in Plant Science 14, 1158933.

[71] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

[72] Strange, R.N., Scott, P.R., 2005. Plant disease: A threat to global food security. Annual Review of Phytopathology 43, 83–116.

[73] Tan, M., Le, Q.V., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: Proceedings of the 36th International Conference on Machine Learning, pp. 6105–6114.

[74] Thapa, R., Zhang, K., Snavely, N., Belongie, S., Khan, A., 2020. The plant pathology challenge 2020 data set to classify foliar disease of apples. Applications in Plant Sciences 8, e11390.

[75] Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences 108, 20260–20264.

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

[77] Ullah, Z., Alsubaie, N., Jamjoom, M., Alajmani, S.H., Saleem, F., 2023. EffiMob-Net: A deep learning-based hybrid model for detection and identification of tomato diseases using leaf images. Agriculture 13, 737.

[78] Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

[79] Wei, T., Chen, Z., Huang, Z., Yu, X., 2024. Benchmarking in-the-wild multimodal plant disease recognition and a versatile baseline, in: ACM International Conference on Multimedia.

[80] Wei, X.S., Song, Y.Z., Mac Aodha, O., Wu, J., Peng, Y., Tang, J., Yang, J., Belongie, S., 2021. Fine-grained image analysis with deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 8927–8948.

[81] Wu, X., Fan, X., Luo, P., Choudhury, S.D., Tjahjadi, T., Hu, C., 2023. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. Plant Phenomics 5, 0038.

[82] Xu, M., Kim, H., Yang, J., Fuentes, A., Meng, Y., Yoon, S., Kim, T., Park, D.S., 2023. Embracing limited and imperfect training datasets: Opportunities and challenges in plant disease recognition using deep learning. Frontiers in Plant Science 14, 1225409.

[83] Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Advances in Neural Information Processing Systems 27.