Abstract

This study systematically evaluates the efficacy of 56 pretrained neural network architectures, used without fine-tuning, as feature extractors for plant disease anomaly detection across laboratory and field environments. We compare convolutional and transformer-based networks in conjunction with various dimensionality reduction techniques and anomaly detection algorithms to address the performance gap between controlled and real-world imaging conditions. Using apple leaf disease datasets (Plant Village and Plant Pathology) containing identical disease classes, we implement two complementary evaluation strategies: anomaly detection trained solely on healthy samples and clustering-based classification to distinguish between specific disease types. Results reveal a consistent 5-10% performance reduction when transitioning from laboratory to field images, highlighting the challenge of developing robust field-deployable systems. The lightweight ShuffleNet_v2_x1_0 architecture (2.3M parameters) outperformed substantially larger models like DINOv2 (300M) and ViT (86M) in field conditions, challenging the assumption that larger models necessarily yield better performance for specialized tasks. Among dimensionality reduction techniques, t-SNE consistently outperformed others, while Local Outlier Factor demonstrated the most stable anomaly detection performance across datasets. For clustering, density-based DBSCAN with ShuffleNet_v2_x1_0 achieved

superior performance on field images. These findings provide practical insights for developing computationally efficient plant disease detection systems for resource-constrained environments, demonstrating that anomaly detection approaches with off-the-shelf pretrained models offer viable alternatives to supervised classification, especially when comprehensive labeled datasets are impractical.

2.3.1 Introduction

Plant diseases pose a significant threat to agricultural productivity, food security, and economic stability worldwide, with estimated global crop losses exceeding 20-40% annually due to pathogens [19]. Early and accurate detection of plant diseases is crucial for implementing timely interventions, reducing pesticide use, and preventing disease spread across agricultural landscapes [15]. Traditional disease detection methods rely heavily on visual inspection by trained experts, which is time-consuming, labor-intensive, and subject to human error [3].

In recent years, advances in computer vision and machine learning have enabled automated approaches to plant disease detection, offering the potential for more scalable, consistent, and objective diagnostic capabilities [8, 16]. Deep learning approaches, particularly convolutional neural networks (CNNs) and vision transformers, have demonstrated remarkable success in classifying plant diseases from leaf images [22]. However, these supervised approaches require large amounts of labeled training data for each disease class, which is often impractical to obtain for the diverse range of plant pathogens and their varying manifestations [26].

Anomaly detection presents a promising alternative paradigm that requires training only on healthy samples, identifying diseased specimens as deviations from the normal state [18, 6]. This approach aligns well with agricultural monitoring scenarios where healthy plants constitute the majority class, and various diseases represent anoma-

lous conditions [12]. Additionally, anomaly detection frameworks can potentially identify novel or previously unseen disease manifestations that supervised classifiers would struggle to recognize [5].

A critical challenge in developing robust plant disease detection systems is the significant performance gap between controlled laboratory environments and real-world field conditions [25]. Laboratory-acquired images typically feature isolated leaves against uniform backgrounds with consistent lighting, while field-acquired images contain variable illumination, complex backgrounds, and diverse perspectives that can dramatically affect feature extraction and classification performance [3].

This study addresses these challenges by comprehensively evaluating the efficacy of various neural network architectures as feature extractors for anomaly detection across both laboratory and field-acquired apple leaf disease datasets. By systematically comparing convolutional and transformer-based networks in conjunction with different dimensionality reduction techniques and anomaly detection algorithms, we aim to identify robust methodologies that can translate from controlled environments to practical field applications.

Our work makes several key contributions: (1) a systematic evaluation of 56 neural network architectures as feature extractors for plant disease anomaly detection; (2) comparative analysis of performance across laboratory and field imaging conditions using parallel datasets with matching disease classes; (3) identification of lightweight models that achieve benchmark accuracy while minimizing computational requirements; and (4) practical insights into the

most effective combinations of feature extraction, dimensionality reduction, and anomaly detection approaches for agricultural disease monitoring applications.

2.3.2 Materials and Method

2.3.2.1 Dataset

This study utilized two complementary apple leaf disease datasets to evaluate the robustness of feature extraction methods across different imaging conditions. The datasets represent controlled laboratory conditions and natural field environments, respectively, while containing the same disease classes.

The Plant Village dataset [11] consists of laboratory-acquired images of individual plant leaves photographed against controlled backgrounds. For this study, the apple leaf subset was utilized, which includes segmented images of single leaves. These images were captured under consistent lighting conditions with uniform backgrounds, resulting in standardized image dimensions of 256×256 pixels.

The Plant Pathology dataset [24], collected as part of a Kaggle competition, contains field-acquired images of apple leaves in their natural environment. Unlike the controlled Plant Village images, these photographs exhibit varying lighting conditions, backgrounds, perspectives, and image dimensions. The images capture leaves still attached to the tree or branch, providing a more challenging and realistic scenario for disease detection.

To enable fair comparison between the datasets, several preprocessing steps were implemented. First, the datasets were balanced to contain identical disease classes (healthy, apple scab, and cedar apple rust). Second, the number of samples per class was standardized by removing excess observations from either dataset where necessary. Both datasets were then processed to ensure consistent sample counts while maintaining their inherent characteristics regarding acquisition conditions.

Table 1: Summary of the standardized apple leaf disease datasets

Dataset	Class	Samples	Image Size	Acquisition
Plant Village	Healthy	516	256×256	Laboratory
	Cedar apple rust	275		
	Apple scab	583		
Plant Pathology	Healthy	516	Variable	Field
	Cedar apple rust	275		
	Apple scab	583		

The dataset combination provides an opportunity to assess how feature extraction methods perform across different imaging conditions while maintaining consistent disease classes. The laboratory-acquired Plant Village images offer an idealized, controlled scenario, while the field-acquired Plant Pathology images present a more challenging real-world testing environment.

2.3.2.2 Tested Backbones

This study evaluates a comprehensive set of neural network architectures as feature extractors for anomaly detection. We imple-

mented both convolutional neural networks (CNNs) and transformer-based architectures pre-trained on ImageNet to extract meaningful representations from input images.

Table 2: Overview of neural network backbone architectures evaluated in this study

Backbone	Param (M)	Input Size	Backbone	Param (M)	Input Size
densenet121	8.0	224×224	regnet_y_8gf	39.4	224×224
densenet161	28.7	224×224	resnet101	44.5	224×224
densenet169	14.1	224×224	resnet152	60.2	224×224
densenet201	20.0	224×224	resnet18	11.7	224×224
dinov2_vitb14	86.0	224×224	resnet34	21.8	224×224
dinov2_vitl14	300.0	224×224	resnet50	25.6	224×224
dinov2_vits14	21.0	224×224	resnext101_32x8d	88.8	224×224
googlenet	13.0	224×224	resnext101_64x4d	83.5	224×224
inception_v3	27.2	299×299	resnext50_32x4d	25.0	224×224
mobilenet_v3_large	5.5	224×224	shufflenet_v2_x0_5	1.4	224×224
mobilenet_v3_small	2.5	224×224	shufflenet_v2_x1_0	2.3	224×224
regnet_x_16gf	54.3	224×224	shufflenet_v2_x1_5	3.5	224×224
regnet_x_1_6gf	9.2	224×224	shufflenet_v2_x2_0	7.4	224×224
regnet_x_32gf	107.8	224×224	swin_b	87.8	224×224
regnet_x_3_2gf	15.3	224×224	swin_s	49.6	224×224
regnet_x_400mf	5.5	224×224	swin_t	28.3	224×224
regnet_x_800mf	7.3	224×224	swin_v2_b	87.9	224×224
regnet_x_8gf	39.6	224×224	swin_v2_s	49.7	224×224
regnet_y_16gf	83.6	224×224	swin_v2_t	28.4	224×224
regnet_y_1_6gf	11.2	224×224	vgg11	132.9	224×224
regnet_y_32gf	145.0	224×224	vgg11_bn	132.9	224×224
regnet_y_3_2gf	19.4	224×224	vgg13	133.0	224×224
regnet_y_400mf	4.3	224×224	vgg13_bn	133.0	224×224
regnet_y_800mf	6.4	224×224	vgg16	138.4	224×224
vgg16_bn	138.4	224×224	vit_l_16	304.3	224×224
vgg19	143.7	224×224	vit_l_32	306.5	224×224
vgg19_bn	143.7	224×224	wide_resnet101_2	126.9	224×224
vit_b_16	86.6	224×224	wide_resnet50_2	68.9	224×224

Convolutional Neural Networks

We investigated several CNN architecture families:

- **ResNet family:** ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, which utilize residual connections to enable training of deeper networks [9]. Additionally, we included variants with wider channels (Wide ResNet50, Wide ResNet101) and

grouped convolutions (ResNeXt50, ResNeXt101).

- **VGG family:** VGG11, VGG13, VGG16, VGG19, and their batch-normalized counterparts, representing traditional deep CNN architectures with sequential convolutional layers [21].
- **DenseNet family:** DenseNet121, DenseNet161, DenseNet169, DenseNet201, featuring dense connectivity patterns that strengthen feature propagation [10].
- **Efficient architectures:** EfficientNet (B0-B7), EfficientNetV2 (S, M, L), MobileNetV2, MobileNetV3, which are optimized for computational efficiency while maintaining high accuracy [23].
- **Other CNN architectures:** GoogleNet, Inception-v3, RegNet, ShuffleNet, and SqueezeNet variations, each with unique architectural innovations designed to improve performance or efficiency.

Transformer-based Architectures

We also examined vision transformers that have demonstrated strong performance in recent years:

- **Vision Transformer (ViT):** ViT-B/16, ViT-B/32, ViT-L/16, ViT-L/32, ViT-H/14, which apply the transformer architecture directly to image patches [7].
- **Swin Transformer:** Swin-T, Swin-S, Swin-B and their V2 variants, which incorporate hierarchical feature maps and shifted windows for more efficient attention computation [14].

- **DINOv2:** DINOv2-ViT-S/14, DINOv2-ViT-B/14, DINOv2-ViT-L/14, which are self-supervised vision transformers trained using distillation with no labels [17].

Feature Extraction Methodology

For all architectures, we removed the classification heads and extracted features from the penultimate layer. For CNNs, this typically corresponds to the output after global average pooling, while for transformers, we used the [CLS] token representation. All models were pre-trained on ImageNet and used without fine-tuning to evaluate their transfer learning capabilities for anomaly detection.

For standard torchvision models, we utilized the official pre-trained weights [1]. For DINOv2 models, we loaded weights directly from the official Facebook Research repository [2]. Input images were processed using the standard preprocessing pipeline recommended for each model, including resizing, normalization, and in some cases, center cropping.

2.3.2.3 Evaluation Strategies

We implemented two complementary strategies to evaluate the efficacy of extracted features:

Anomaly Detection Approach

The extracted features were used as input to anomaly detection algorithms. For each dataset, these algorithms were trained using

only the healthy samples and evaluated on the diseased samples within the same dataset. This approach allowed us to assess how well the feature extractors could separate normal from anomalous samples across different imaging conditions.

Clustering-based Classification

We also evaluated whether the dimensionality-reduced features preserved sufficient class-discriminative information for conventional clustering algorithms to recover the original disease classes. This approach differs from anomaly detection by attempting to distinguish between specific disease types rather than just identifying abnormalities.

We tested multiple clustering algorithms:

- **K-Means:** A centroid-based algorithm that partitions the data into k clusters, with each observation belonging to the cluster with the nearest mean.
- **Hierarchical Clustering:** An agglomerative approach that builds nested clusters by merging or splitting them successively.
- **Gaussian Mixture Models:** A probabilistic model that assumes data points are generated from a mixture of several Gaussian distributions.
- **DBSCAN:** A density-based clustering algorithm that groups together points that are closely packed in feature space.

To evaluate clustering performance, we mapped each cluster to its most common ground truth label and calculated Cohen's Kappa coefficient to measure the agreement between clustering assignments and original disease classifications.

2.3.2.4 Dimensionality Reduction

To visualize the extracted features and assess their separability, we applied dimensionality reduction techniques. We selected t-SNE, UMAP, and PCA for this purpose:

- **t-SNE (t-distributed Stochastic Neighbor Embedding):** A non-linear dimensionality reduction technique that is particularly effective for visualizing high-dimensional data in lower dimensions (typically 2D or 3D). It focuses on preserving local structures and is widely used for visualizing clusters in feature spaces.
- **UMAP (Uniform Manifold Approximation and Projection):** A manifold learning technique that preserves both local and global structures in the data. UMAP is often faster than t-SNE and can produce more interpretable embeddings, making it suitable for visualizing complex datasets.
- **PCA (Principal Component Analysis):** A linear dimensionality reduction method that transforms the data into a new coordinate system, where the greatest variance lies on the first coordinates (principal components). PCA is computationally efficient and provides a global view of the data structure.

These techniques were applied to the extracted features from both datasets, allowing us to visualize the distribution of healthy and diseased samples in lower-dimensional spaces. The visualizations provided insights into the separability of different classes and the effectiveness of the feature extractors in capturing relevant information for anomaly detection.

2.3.2.5 Anomaly Detection Algorithms

We implemented a range of anomaly detection algorithms to evaluate the performance of the extracted features. The algorithms were selected based on their popularity and effectiveness in various domains, including:

Statistical Methods

- **IQR with Confidence Interval:** This approach combines robust statistics with probabilistic bounds. First, we use the interquartile range (IQR) of healthy samples to identify potential outliers. We then calculate a confidence interval (95%) around the mean of the remaining inliers. Any sample falling outside this interval is classified as anomalous.

Machine Learning Methods

- **Isolation Forest:** This algorithm [13] isolates observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the

selected feature. Anomalies require fewer partitions to be isolated, resulting in shorter average path lengths. We used a contamination parameter of 0.1 for all experiments.

- **One-Class SVM:** This method [20] learns a boundary around normal data points in feature space. Samples outside this boundary are classified as anomalies. We employed the RBF kernel with nu=0.1, training only on healthy samples.

- **Local Outlier Factor (LOF):** This density-based algorithm [4] compares the local density of a point with the densities of its neighbors. Points with substantially lower density than their neighbors are considered anomalies. We configured LOF in novelty mode with optimal neighborhood size.

- **Gaussian Mixture Model (GMM):** This probabilistic model assumes that normal data points are generated from a mixture of Gaussian distributions. We fit a single-component GMM to healthy samples and identified anomalies as points with low probability density, using the 1st percentile of healthy samples' scores as the threshold.

Experimental Setup

We implemented two parallel experimental workflows to evaluate the feature representations:

Anomaly Detection Workflow

For evaluating anomaly detection capabilities, we followed this procedure for each dataset independently:

1. Extract features from the dataset using the backbone networks.
2. Apply dimensionality reduction techniques (t-SNE, UMAP, or PCA) to the extracted features.
3. Train the anomaly detection algorithms using only the healthy samples from the dataset.
4. Evaluate performance on the diseased samples from the same dataset, where all non-healthy classes should be detected as anomalies.

We assessed each algorithm using standard binary classification metrics: accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

Clustering-based Classification Workflow

For evaluating the class-discriminative information in the feature space, we implemented:

1. Extract features from the dataset using the backbone networks.
2. Apply dimensionality reduction techniques to reduce feature dimensionality.
3. Apply various clustering algorithms (K-Means, Hierarchical Clustering, GMM, DBSCAN) to the reduced features.
4. Map each resulting cluster to the most common ground truth label among its members.

5. Calculate Cohen's Kappa coefficient between the cluster assignments and the original disease classifications to measure agreement beyond chance.

The clustering approach provides insights into whether the feature extractors capture sufficient information to distinguish between specific disease classes, rather than just separating normal from abnormal samples. It also serves as a more challenging evaluation scenario that mimics unsupervised disease classification.

By applying both methodologies to the controlled laboratory images (Plant Village) and the variable field images (Plant Pathology), we could comprehensively evaluate how different feature extractors perform across varying imaging conditions, which is crucial for practical disease detection applications in agriculture.

2.3.3 Results and Discussion

Our analysis of anomaly detection and clustering performance across different backbone architectures, dimensionality reduction techniques, and anomaly detection algorithms revealed several combination achiving the accuracy benchmark. Figure 1 and Figure 2 show the distribution of anomaly detection accuracy across all tested configurations respectively for the Plant Village and the Plant Pathology datasets. In the same way, Figure 3 and Figure 4 show the distribution of clustering accuracy across all tested configurations respectively for the Plant Village and the Plant Pathology datasets.

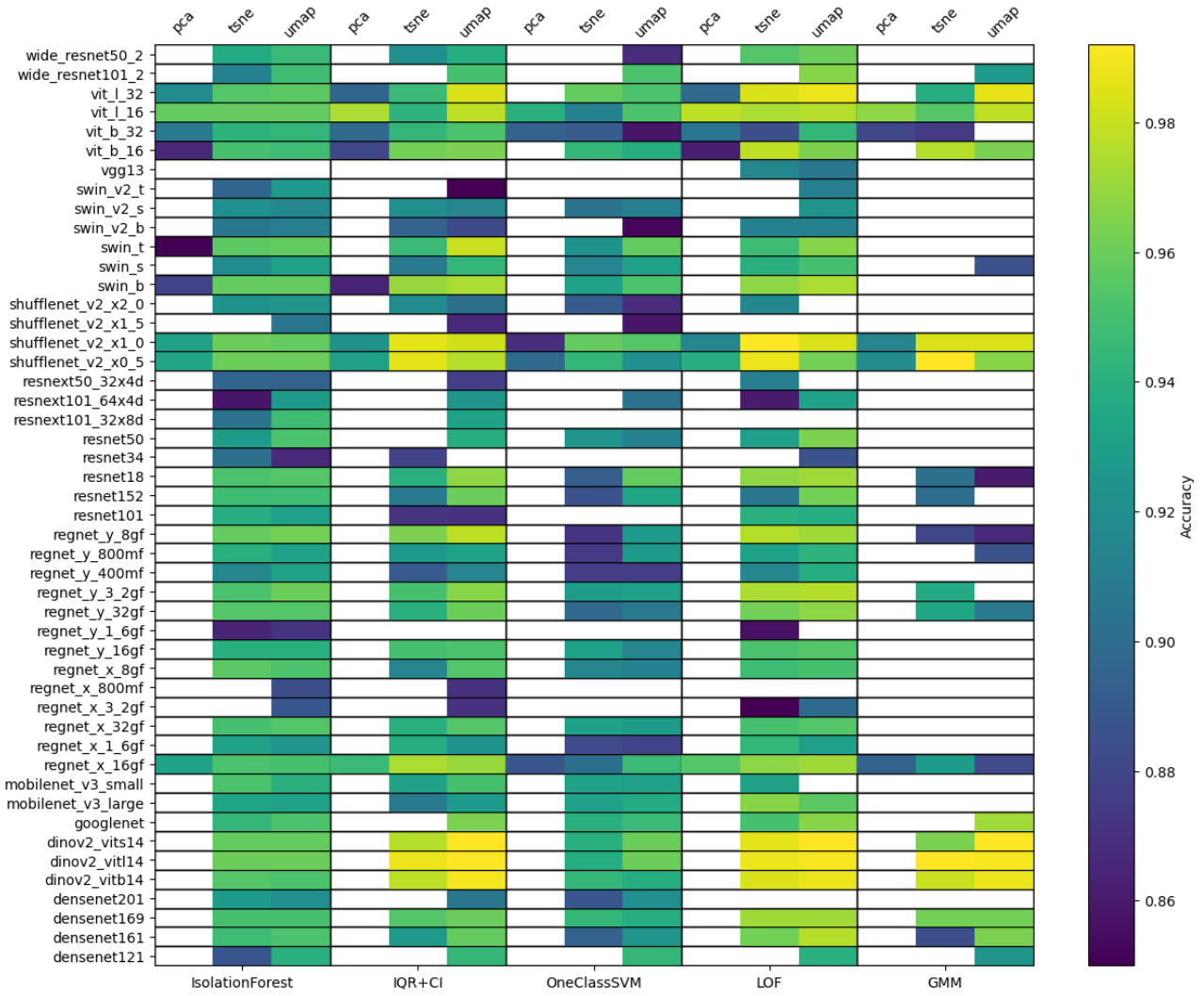

Figure 1: Anomaly detection performance across different backbone architectures and dimensionality reduction techniques on the Plant Village dataset. Backbones on y-axis, anomaly detetection algorithm on lower x-axis, dimensionality reduction method on top x-axis. The color indicates the accuracy for each backbone-detection-reduction combination.

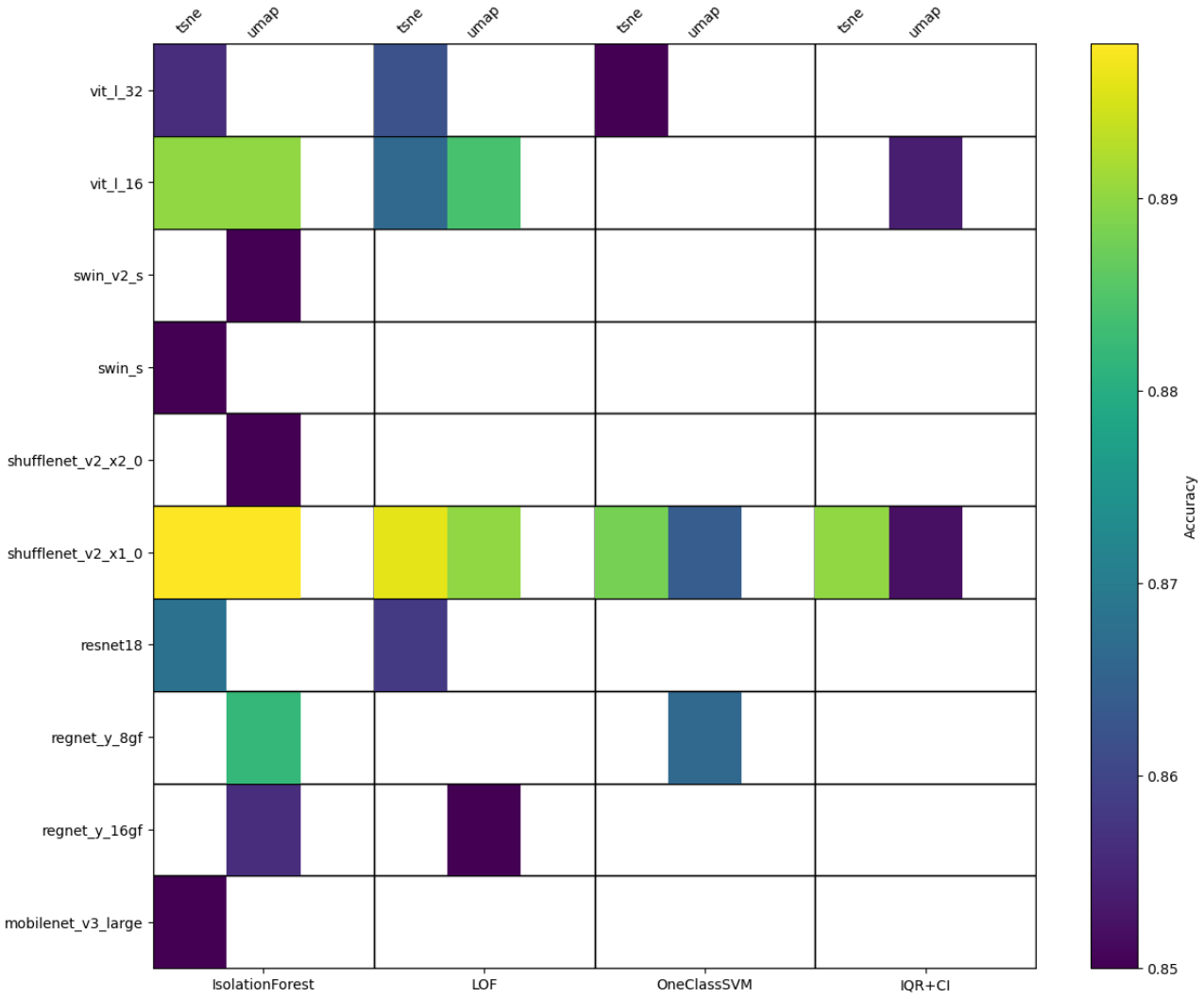

Figure 2: Anomaly detection performance across different backbone architectures and dimensionality reduction techniques on the Plant Pathology dataset. Backbones on y-axis, anomaly detetection algorithm on lower x-axis, dimensionality reduction method on top x-axis. The color indicates the accuracy for each backbone-detection-reduction combination.

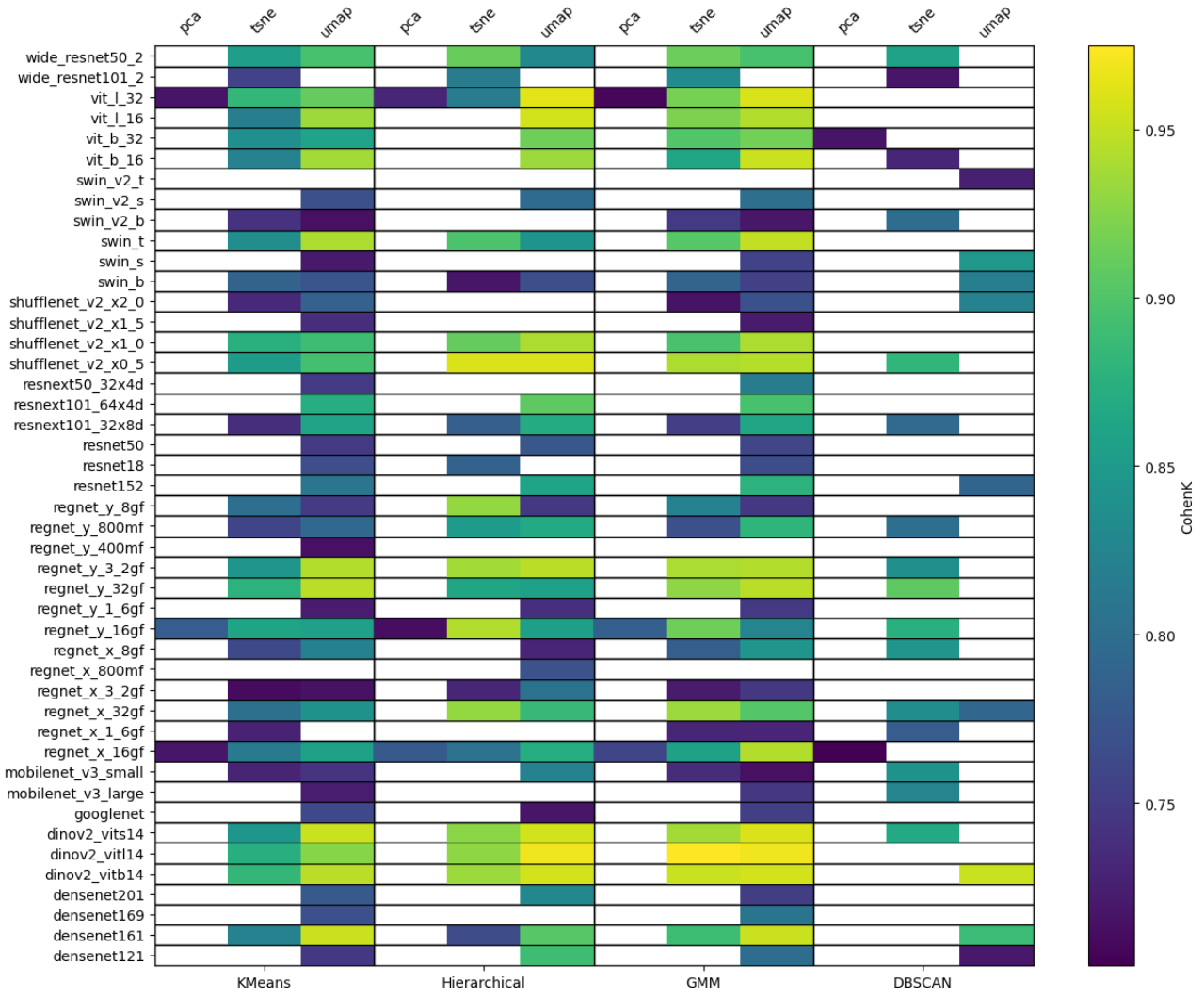

Figure 3: Clustering performance across different backbone architectures and dimensionality reduction techniques on the Plant Village dataset. Backbones on y-axis, clustering algorithm on lower x-axis, dimensionality reduction method on top x-axis. The color indicates the accuracy for each backbone-detection-reduction combination.

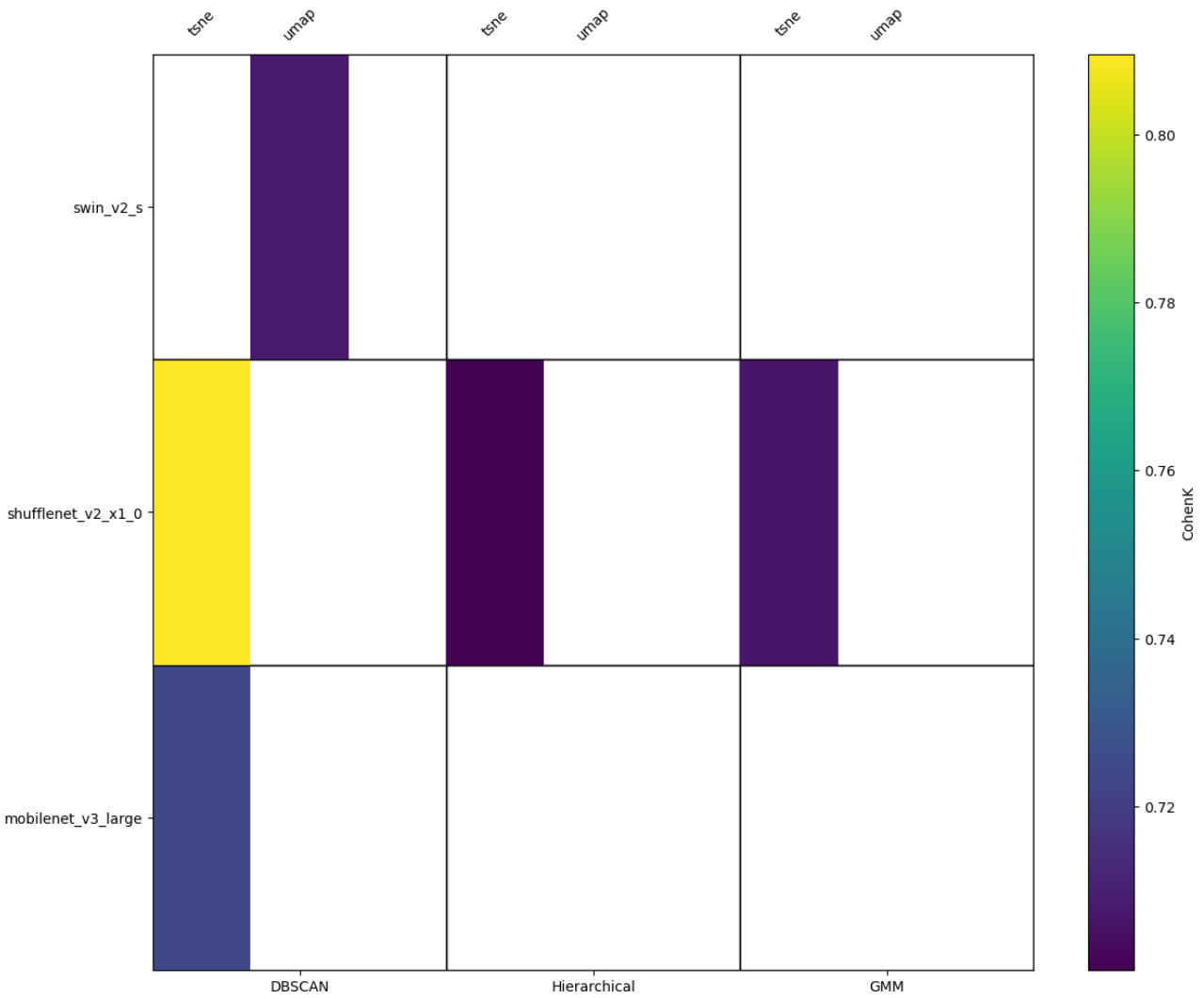

Figure 4: Clustering performance across different backbone architectures and dimensionality reduction techniques on the Plant Pathology dataset. Backbones on y-axis, clustering algorithm on lower x-axis, dimensionality reduction method on top x-axis. The color indicates the accuracy for each backbone-detection-reduction combination.

Dataset related performances

The analysis of performance metrics across datasets revealed sig-

nificant differences in model efficacy between laboratory-acquired and field-acquired images. The Plant Village dataset, consisting of controlled laboratory images with segmented leaves against uniform backgrounds, consistently enabled superior performance compared to the Plant Pathology dataset across all evaluation metrics.

For anomaly detection tasks, accuracy scores on the Plant Village dataset frequently exceeded 90%, as shown in Figure 1. In contrast, the same architectures applied to the Plant Pathology dataset did not achieved the benchmark or got a performance decrease of approximately 5-10%, as illustrated in Figure 2.

This performance gap was even more pronounced in clustering tasks, where Cohen's Kappa coefficients reached as high as 0.9 with 12 backbones on Plant Village data but peaked at 0.80 on Plant Pathology data and achived benchmark with only three backbones, as seen in Figures 3 and 4. The controlled environment and object-focused nature of the Plant Village dataset allowed models to concentrate on leaf characteristics directly relevant to disease detection without the confounding variables present in field conditions.

The Plant Pathology dataset's variable lighting, complex backgrounds, and inconsistent perspectives presented a substantially more challenging scenario for feature extraction focusing on diseseas sympthoms.

These findings highlight the significant challenge of transitioning plant disease detection systems from controlled laboratory environments to real-world field applications, where background elements and environmental variability introduce substantial complexity to the feature extraction and classification process.

Nethertheless, the benchmark was achieved also for the challenging field-acquired images, indicating that the selected architectures and methods are capable of generalizing to real-world conditions.

Effect of Feature Extraction Architecture

Among the tested architectures, the ShuffleNet_v2 feature extractor consistently demonstrated strong performance across both datasets, with the x1_0 size version achieving the benchmark for both anomaly and clusterization tasks. Other well performing architectures were DINOv2 and ViT on Plant Village dataset for both tasks, but they did not perform consistently on the Plant Pathology dataset. Remarkably ShuffleNet_v2_x1_0 is a lightweight architecture with only 2.3M parameters, in respect to the DINOv2 and ViT architectures which have 300M and 86M parameters respectively. This suggests that the selected feature extractors are capable of achieving high performance even with limited computational resources, making them suitable for deployment in resource-constrained environments.

Impact of Dimensionality Reduction

Different dimensionality reduction techniques showed varying effectiveness:

- **t-SNE** consistently yielded the highest performances on both tasks and datasets when in combination with ShuffleNet_v2_x1_0.
- **UMAP** performed competitively with t-SNE, in some cases surpassing it, but not the best option for all tasks and datasets.
- **PCA**, while computationally efficient, generally produced lower accuracy compared to t-SNE and UMAP, indicating that lin-

ear dimensionality reduction may not sufficiently preserve the complex structure necessary for plant disease detection.

The choice of dimensionality reduction technique significantly influenced the performance of both anomaly detection and clustering tasks. t-SNE and UMAP were particularly effective in preserving the local structure of the data, leading to better separability of classes in the reduced feature space.

Comparison of Anomaly Detection Algorithms

Among the tested anomaly detection algorithms:

- **Isolation Forest** exalled on Plant Pathology dataset while the performances with Plant Village were low in respect to the others methods.
- **One-Class SVM** did not exalled on any of the datasets.
- **LOF** showed the most stable performances.
- **GMM** demonstrated comparable performance with LOF on Plant Village, but it never reached the benchmark on Plant Pathology.
- **IQR with Confidence Interval**, while simpler than the machine learning approaches, still achieved respectable performance on both datasets, highlighting the effectiveness of statistical approaches for this task.

The Plant Pathology dataset generally yielded lower performance

compared to Plant Village across all algorithms, reflecting the greater difficulty of analyzing field-acquired images with variable conditions.

Clustering Performance

The clustering-based classification approach revealed complementary insights about the discriminative power of extracted features.

- **K-Means, Hierarchical, and GMM** Good results on Plant Village achieving benchmark with multiple backbones, but on Plant Pathology only K-Means and GMM reached the benchmark.
- **DBSCAN** achieved the benchmark on Plant Village with less backbones in respect the other methods, while on Plant Pathology achieved the best result with ShuffleNet_v2_x1_0.

The clustering-based classification approach revealed complementary insights about the discriminative power of extracted features, with significant differences in algorithm performance across datasets.

- **K-Means, Hierarchical, and GMM** achieved strong results on Plant Village, reaching the benchmark with multiple backbone architectures. However, on the more challenging Plant Pathology dataset, only K-Means and GMM reached the benchmark. This suggests that centroid and distribution-based approaches perform consistently when the number of clusters is explicitly defined to match the disease classes.
- **DBSCAN** exhibited a distinct behavior pattern, achieving the benchmark on Plant Village with fewer backbones compared

to other methods, while on Plant Pathology it achieved the best overall result specifically with ShuffleNet_v2_x1_0. This unique performance profile can be attributed to DBSCAN's density-based approach, which differs fundamentally from the other algorithms in several ways:

- Unlike parametric methods that assume specific cluster shapes, DBSCAN identifies arbitrarily shaped clusters based on density variations, potentially capturing the complex symptom patterns in field conditions more effectively.
- DBSCAN automatically estimates its critical epsilon parameter based on the nearest neighbor distances in the feature space, making it particularly responsive to the actual distribution characteristics rather than prior assumptions.
- Its built-in outlier detection capability, which labels points in low-density regions as noise, provides natural robustness against the variable imaging conditions present in the Plant Pathology dataset.
- The exceptional performance with ShuffleNet_v2_x1_0 suggests this lightweight architecture (2.3M parameters) produces feature distributions with clearer density gradients between disease classes, despite having significantly fewer parameters than transformer-based alternatives.

These findings indicate that while conventional clustering methods perform well in controlled environments, density-based approaches

may offer advantages for disease detection in variable field conditions, particularly when paired with efficient feature extractors that create well-separated density regions in the feature space.

2.3.4 Conclusions

This comprehensive evaluation of neural network architectures as feature extractors for plant disease anomaly detection has yielded several important findings with significant implications for agricultural monitoring applications.

Our first key finding revealed a consistent performance gap between laboratory and field-acquired images, with detection accuracy typically 5-10% lower on field images. This quantifies the substantial challenge of translating plant disease detection systems from controlled environments to practical field applications. Despite this gap, our study identified combinations of feature extractors and detection algorithms that achieved benchmark performance even in challenging field conditions, demonstrating that robust field-deployable systems are achievable.

The ShuffleNet_v2_x1_0 architecture emerged as the most consistently effective feature extractor across both datasets and evaluation methodologies. Remarkably, this lightweight network (2.3M parameters) outperformed substantially larger models like DINOv2 (300M parameters) and ViT (86M parameters) in field conditions. This finding challenges the common assumption that larger, more complex models necessarily yield better performance for specialized tasks.

Instead, it suggests that computational efficiency and targeted feature extraction may be more valuable than model capacity for plant disease detection, particularly in resource-constrained deployment scenarios.

Among dimensionality reduction techniques, t-SNE consistently yielded the highest performance across most configurations, with UMAP following closely. The substantially lower performance of PCA indicates that nonlinear dimensionality reduction techniques better preserve the complex feature relationships crucial for disease differentiation. This finding highlights the importance of maintaining local neighborhood structures in the reduced feature space for effective anomaly detection.

The comparison of anomaly detection algorithms revealed that LOF demonstrated the most stable performance across datasets, while Isolation Forest excelled specifically on field-acquired images. This suggests that different detection methodologies have complementary strengths depending on image acquisition conditions. For practical field applications, ensemble approaches combining multiple detection algorithms might prove beneficial.

For clustering-based classification, DBSCAN with ShuffleNet_v2_x1_0 achieved superior performance on field images compared to other combinations. This density-based approach appears particularly well-suited to handling the variable imaging conditions present in field settings, capturing the natural density variations between healthy and diseased samples in the feature space.

These findings have important practical implications for agricultural

disease monitoring systems. By selecting lightweight, efficient architectures like ShuffleNet_v2_x1_0, developers can create deployment-ready solutions for resource-constrained environments such as edge devices or mobile applications. The established benchmark performance on field-acquired images demonstrates that anomaly detection approaches are viable alternatives to supervised classification, especially in scenarios where obtaining comprehensive labeled datasets for every potential disease is impractical.

Future research directions should explore fine-tuning strategies specifically for agricultural domain adaptation, which may further close the performance gap between laboratory and field conditions. Additionally, investigating temporal anomaly detection for disease progression monitoring and extending the approach to multi-spectral or hyperspectral imagery could enhance detection capabilities, particularly for early-stage infections. Finally, developing integrated systems that combine anomaly detection with targeted classification for identified anomalies could create more comprehensive disease management solutions for practical agricultural applications.

In conclusion, this study establishes that computationally efficient feature extraction architectures, when combined with appropriate dimensionality reduction and anomaly detection algorithms, can effectively identify plant diseases across varying imaging conditions. These findings provide a foundation for developing practical, field-deployable systems for early disease detection that can contribute to sustainable agricultural practices and improved food security.

Bibliography

[1] Models and pre-trained weights — Torchvision main documentation.

[2] facebookresearch/dinov2, March 2025. original-date: 2023-03-29T16:00:37Z.

[3] Jayme G. A. Barbedo. Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, 172:84–91, August 2018.

[4] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104, May 2000.

[5] Samuele Bumbaca and Enrico Borgogno-Mondino. Supporting Screening of New Plant Protection Products through a Multispectral Photogrammetric Approach Integrated with AI. *Agronomy*, 14(2):306, February 2024. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[6] Raghavendra Chalapathy and Sanjay Chawla. Deep Learning for Anomaly Detection: A Survey, January 2019. arXiv:1901.03407 [cs].

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, June 2021. arXiv:2010.11929 [cs].

[8] Konstantinos P. Ferentinos. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145:311–318, February 2018.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015. arXiv:1512.03385 [cs].

[10] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. pages 4700–4708, 2017.

[11] David P. Hughes and Marcel Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics, April 2016. arXiv:1511.08060 [cs].

[12] Ryoya Katafuchi and Terumasa Tokunaga. Image-based Plant Disease Diagnosis with Unsupervised Anomaly Detection Based on Reconstructability of Colors, September 2021. arXiv:2011.14306 [cs].

[13] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, 2008.

[14] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, August 2021. arXiv:2103.14030 [cs].

[15] Federico Martinelli, Riccardo Scalenghe, Salvatore Davino, Stefano Panno, Giuseppe Scuderi, Paolo Ruisi, Paolo Villa, Daniela Stroppiana, Mirco Boschetti, Luiz R. Goulart, Cristina E. Davis, and Abhaya M. Dandekar. Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, 35(1):1–25, 2015. Publisher: Springer Verlag/EDP Sciences/INRA.

[16] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, 7, September 2016. Publisher: Frontiers.

[17] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning Robust Visual Features without Supervision, February 2024. arXiv:2304.07193 [cs].

[18] Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen,

Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. A Unifying Review of Deep and Shallow Anomaly Detection. *Proceedings of the IEEE*, 109(5):756–795, May 2021. arXiv:2009.11732 [cs].

[19] Serge Savary, Laetitia Willocquet, Sarah Jane Pethybridge, Paul Esker, Neil McRoberts, and Andy Nelson. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3(3):430–439, March 2019.

[20] Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 07 2001.

[21] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, April 2015. arXiv:1409.1556 [cs].

[22] Aadarsh Kumar Singh, Akhil Rao, Pratik Chattopadhyay, Rahul Maurya, and Lokesh Singh. Effective plant disease diagnosis using Vision Transformer trained with leafy-generative adversarial network-generated images. *Expert Systems with Applications*, 254:124387, November 2024.

[23] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, September 2020. arXiv:1905.11946 [cs].

[24] Ranjita Thapa, Noah Snavely, Serge Belongie, and Awais

Khan. The Plant Pathology 2020 challenge dataset to classify foliar disease of apples, April 2020. arXiv:2004.11958 [cs].

[25] Yosuke Toda and Fumio Okura. How Convolutional Neural Networks Diagnose Plant Disease. *Plant Phenomics (Washington, D.C.)*, 2019:9237136, 2019.

[26] Sasikala Vallabhajosyula, Venkatramaphanikumar Sistla, and Venkata Krishna Kishore Kolli. A novel hierarchical framework for plant leaf disease detection using residual vision transformer. *Heliyon*, 10(9):e29912, May 2024.