Executive Summary

The deployment of automated diagnostic systems in precision agriculture is contingent upon their reliability in open-world environments. While deep learning has achieved superhuman performance in closed-set classification of plant diseases, the operational reality of agricultural fields introduces a vast array of Out-of-Distribution (OOD) stimuli—ranging from novel pathogen strains and abiotic stressors to non-plant artifacts and sensor anomalies. The inability of standard classifiers to robustly reject these anomalies constitutes a critical safety vulnerability; a high-confidence misclassification of a novel, virulent blight as a benign nutrient deficiency can lead to catastrophic crop loss.

This report presents a comprehensive technical analysis of the state-of-the-art methodologies for OOD detection, specifically tailored to the architectural paradigm of Vision Transformers (ViTs) and the domain constraints of Fine-Grained Visual Categorization (FGVC). We critically examine the transition from classical Extreme Value Theory (EVT) approaches, such as OpenMax, to modern geometric methods operating in the latent spaces of Foundation Models like DINOv2. A central focus is placed on the interaction between Parameter-Efficient Fine-Tuning (PEFT), specifically Low-Rank Adaptation (LoRA), and uncertainty estimation. Our analysis synthesizes empirical evidence to demonstrate that unmerged LoRA embeddings, when analyzed via Mahalanobis Distance, offer a superior mechanism for detecting "near-OOD" samples—those semantically proximate anomalies that plague fine-grained disease classification. Furthermore, we explore the interpretability of these decisions through multi-scale attention mechanisms and Class Activation Mapping (CAM), proposing a unified framework that balances sensitivity, robustness, and explainability.

---

1. The Operational Imperative: Reliability in Fine-Grained Agricultural AI

1.1 The Fine-Grained Visual Categorization (FGVC) Challenge in Pathology

Plant disease diagnosis is a quintessential Fine-Grained Visual Categorization (FGVC) problem. Unlike generic object recognition (e.g., distinguishing a cat from a dog), FGVC requires discriminating between subclasses that share a high degree of structural similarity. In the context of phytopathology, a "healthy" tomato leaf, a leaf with "Early Blight" (*Alternaria solani*), and a leaf with "Septoria Leaf Spot" (*Septoria lycopersici*) share the same global geometry, color palette, and biological morphology. The discriminative features are often minute, localized textural anomalies—concentric rings in a necrotic spot versus water-soaked lesions.

This high intra-class variance (due to varying growth stages, lighting, and leaf angles) and low

inter-class variance creates a perilous landscape for OOD detection. A standard Convolutional Neural Network (CNN) or Vision Transformer (ViT) trained on a closed set of 10 diseases will essentially partition the high-dimensional feature space into 10 regions. When presented with an 11th, unknown disease (OOD), the model forces the input into one of the existing partitions. Because the global structure (a leaf) is In-Distribution (ID), the model often assigns high confidence to the prediction based on shared background features, a phenomenon known as the "overconfidence" problem.

1.2 The Shift from CNNs to Vision Transformers

Historically, CNNs dominated this field, leveraging their inductive bias for local texture to identify lesions. However, the field is rapidly shifting toward Vision Transformers (ViTs). ViTs, which process images as sequences of patch tokens, utilize self-attention mechanisms capable of modeling long-range dependencies. This is theoretically advantageous for plant disease detection, where the spatial distribution of lesions (e.g., scattered spots vs. marginal necrosis) is diagnostic.

However, ViTs introduce new complexities for OOD detection.

- **Feature Uniformity:** The Layer Normalization (LayerNorm) inherent in ViTs tends to produce feature representations that are more uniform in magnitude and distribution compared to the unnormalized activations of CNNs.
- **Shape Bias:** ViTs exhibit a stronger "shape bias" compared to the "texture bias" of CNNs. While this improves robustness to occlusion, it can make the model less sensitive to the textural anomalies that distinguish different pathologies, potentially clustering distinct diseases closer together in the latent space.
- **OOD Behavior:** Empirical studies suggest that standard post-hoc detection methods optimized for CNNs (like ODIN or React) often degrade in performance when applied directly to ViTs without modification.

1.3 The Scope of Analysis

This report dissects specific mechanisms to address these challenges:

1. **Geometric Methods:** The Mahalanobis Distance and its variants (RMD, Mahalanobis++), which leverage the covariance structure of the feature space.
2. **Probabilistic Methods:** OpenMax and Weibull-based tail modeling.
3. **Foundation Model Strategies:** The specific utility of DINOv2 and the trade-off between linear probing and non-parametric nearest-neighbor evaluations.
4. **Adaptation-Based Detection:** The novel exploitation of LoRA parameters as uncertainty sensors, including unmerged embeddings, selective low-rank approximation (SeTAR), and boxed abstraction monitors (LoRA-BAM).
5. **Validation:** The use of attention mechanisms and CAM to visually verify OOD decisions.

---

2. Geometric Approaches: Mahalanobis Distance and Its Evolutions

The geometric interpretation of neural feature spaces provides the most mathematically grounded framework for OOD detection. The core hypothesis is that a well-trained network maps ID data to a union of compact, low-dimensional manifolds (approximated as Gaussians), and OOD data falls into the low-density regions between or outside these manifolds.

2.1 Theoretical Foundations of Mahalanobis Distance (MD)

The Mahalanobis Distance (MD) is a generalized distance metric that accounts for the correlations between variables in a dataset. Unlike the Euclidean distance, which assumes that features are uncorrelated and have unit variance (isotropic), MD "whitens" the data based on the empirical covariance matrix.

Formally, consider a pre-trained feature extractor $f(x)$ that maps an input image x to a feature vector $z \in \mathbb{R}^d$ in the penultimate layer. We model the distribution of features for each class $c \in \{1, \dots, C\}$ as a multivariate Gaussian distribution $\mathcal{N}(\mu_c, \Sigma)$.

The class-conditional mean is estimated as:

\hat{\mu}_c = \frac{1}{N_c} \sum_{i: y_i = c} f(x_i)$$To ensure robust estimation, especially in high-dimensional spaces where the number of samples per class (N_c) might be small relative to the dimensionality (d), we typically assume a **tied covariance matrix** Σ shared across all classes. This is calculated by pooling the sample covariances:$$\hat{\Sigma} = \frac{1}{N} \sum_{c=1}^{C} \sum_{i: y_i = c} (f(x_i) - \hat{\mu}_c)(f(x_i) - \hat{\mu}_c)^T

The OOD score for a test sample x is defined as the minimum squared distance to any class centroid, scaled by the inverse covariance:

$$M(x) = \min_c (f(x) - \hat{\mu}_c)^T \hat{\Sigma}^{-1} (f(x) - \hat{\mu}_c)$$

2.1.1 Why MD Outperforms Softmax

The Softmax function in the final layer is a projection that compresses the high-dimensional feature vector into a probability simplex. This compression inevitably results in information loss. The Softmax logits effectively represent the distance to the decision boundary (hyperplane), not the distance to the class prototype. A sample can be extremely far from the centroid (an outlier) but still be far from the decision boundary (high confidence), leading to the "high-confidence fool" problem. MD, by operating in the feature space, captures the distance to the density mode, providing a direct measure of "typicality."

2.2 The "Simple Fix": Relative Mahalanobis Distance (RMD)

Despite the theoretical elegance of MD, it suffers from a critical failure mode in "Near-OOD" scenarios—situations where the OOD samples share significant semantic overlap with the ID data. This is precisely the case in plant disease detection, where a "Tomato Yellow Leaf Curl" image (OOD) looks structurally identical to a "Tomato Mosaic Virus" image (ID) except for specific coloration patterns.

2.2.1 The Background Confounding Problem

In deep neural networks, the magnitude (norm) of the feature vector carries significant information about the "background" content of the image. For instance, the presence of a leaf shape and green pixels triggers a baseline level of activation across the network filters. This "background signal" contributes to the total distance $M(x)$. When both ID and Near-OOD samples share this background, their $M(x)$ scores become indistinguishable, as the shared background distance dominates the subtle class-specific distance.

2.2.2 The RMD Methodology

Ren et al. proposed the Relative Mahalanobis Distance (RMD) to isolate the class-specific signal. The method assumes that the feature vector is composed of a class-specific component and a class-agnostic (background) component.

To estimate the background contribution, RMD fits a second Gaussian distribution $\mathcal{N}(\mu_0, \Sigma_0)$ to the **entire training set**, ignoring class labels.

The RMD score is the difference between the class-specific distance and the background distance:

$$RMD(x) = M(x) - M_0(x)$$

where $M_0(x) = (f(x) - \hat{\mu}_0)^T \hat{\Sigma}_0^{-1} (f(x) - \hat{\mu}_0)$.

This formulation is mathematically equivalent to a log-likelihood ratio test between a class-conditional model and a background model.

$$RMD(x) \approx -\log P(x|\text{Class}_k) + \log P(x|\text{Background})$$

By subtracting the background term, RMD cancels out the common factors (e.g., "it is a leaf"). A high RMD score implies that the sample is close to the class centroid *relative* to how close it is to the generic background. If an image is just a generic leaf (Near-OOD), it will be close to μ_0 and moderately close to μ_c, resulting in a small difference. If it is a specific disease

instance (ID), it will be very close to μ_c but generic relative to μ_0, maximizing the score.

2.2.3 Empirical Validation in Fine-Grained Tasks

The efficacy of RMD is most pronounced in fine-grained tasks. On the Genomics OOD benchmark—which involves distinguishing between DNA sequences from different bacterial classes (highly similar)—RMD improved the Area Under the Receiver Operating Characteristic (AUROC) by nearly **15.8 points** compared to standard MD. In the context of plant diseases, this suggests that RMD is essential for distinguishing between visually similar pathologies.

2.3 Mahalanobis++: Addressing Feature Norm Instability

While RMD addresses the semantic overlap, recent research has highlighted a structural instability in MD related to the statistical properties of ViT features. The study "Mahalanobis++" identifies that the L2 norms of feature vectors in modern pre-trained models are often heavy-tailed and not normally distributed.

2.3.1 The Norm-Variance Correlation

In many models, there is a strong correlation between the norm of the feature vector $\|z\|_2$ and the prediction confidence. OOD samples often result in feature vectors with significantly smaller (or larger) norms than ID samples. However, the standard MD calculation allows the covariance matrix to be dominated by the directions of high variance, which are often just the directions of magnitude change. This creates "blind spots" where OOD samples with typical norms but atypical angles are not detected.

2.3.2 The Normalization Solution

Mahalanobis++ proposes a preprocessing step: **L2-normalization** of the features before computing the Gaussian statistics.

$$z_{norm} = \frac{z}{\|z\|_2}$$

By projecting all features onto the unit hypersphere, the method eliminates the variance due to magnitude. The MD then becomes purely a measure of angular distance (cosine similarity) scaled by the angular spread of the class cluster. Experimental results on 44 different models (including ViTs and ConvNeXts) showed that this normalization consistently improved OOD detection performance, reducing the False Positive Rate at 95% True Positive Rate (FPR95) by an average of **9.6%** compared to the conventional Mahalanobis score. [5]

For Vision Transformers in agriculture, which often process images with variable background clutter (leading to variable feature energy), Mahalanobis++ provides a crucial stabilization mechanism, ensuring that the detection is based on the *pattern* of the disease, not the

contrast or brightness of the image.

---

3. Extreme Value Theory and the Legacy of OpenMax

Before the dominance of geometric distance methods, the field of Open Set Recognition (OSR) was defined by **OpenMax**, an algorithm grounded in Extreme Value Theory (EVT).

3.1 The Statistical Basis: Extreme Value Theory (EVT)

EVT is a branch of statistics used to model the risk of extreme deviations from the median of probability distributions. The central theorem (Fisher-Tippett-Gnedenko) states that the maximum of a sequence of independent and identically distributed random variables converges to one of three distributions: Gumbel, Fréchet, or **Weibull**.

In the context of neural networks, the "scores" (activations) of correct classes are considered the "normal" distribution. The "scores" of outliers or unknown classes are expected to fall into the "tail" of this distribution. By modeling this tail, one can estimate the probability that a given input belongs to the set of "knowns" or if it is an "extreme" value belonging to the "unknown."

3.2 The OpenMax Algorithm Deep Dive

OpenMax was designed as a drop-in replacement for the Softmax layer. Its operation involves two phases: training (calibration) and inference.

3.2.1 Phase 1: Calibration (Weibull Fitting)

1. **Feature Extraction:** For each training sample x_i correctly classified as class c, the activation vector $v(x_i)$ from the penultimate layer is extracted.
2. **Centroid Calculation:** The Mean Activation Vector (MAV) μ_c is computed for each class.
3. **Distance Calculation:** The Euclidean distance between each sample and its class MAV is computed: $d_i = \|v(x_i) - \mu_c\|_2$.
4. **Tail Modeling:** For each class, the largest distances (e.g., the top 20 outliers within the class) are selected. A **Weibull distribution** is fitted to these distances. The Weibull PDF provides the probability of a distance being "too large" for that class.

3.2.2 Phase 2: Inference

1. **Prediction:** Given a test sample x, the distances to the top-α predicted classes are computed.

2. **Rejection Probability:** The fitted Weibull models are used to calculate the probability ω_c that the sample is an outlier for class c.
3. **Score Revision:** The activation (logit) for class c is recalibrated: $\hat{v}_c = v_c(1 - \omega_c) + \cdots$. Essentially, if the outlier probability is high, the activation is dampened.
4. **Unknown Class:** The "subtracted" probability mass is assigned to a new pseudo-class y_{K+1} (the "Unknown" class).
5. **Softmax:** A Softmax is applied over the $K + 1$ classes.

3.3 The Failure of OpenMax in Modern Architectures

While OpenMax represented a significant conceptual leap, empirical evidence suggests it struggles in the feature spaces of modern Deep Neural Networks (DNNs), particularly ViTs.

1. **The Curse of Dimensionality:** EVT relies on the concept of distance. In high-dimensional spaces (ViT-Base has 768 dimensions), the distribution of distances becomes concentrated, and the distinction between the "bulk" and the "tail" blurs.
2. **Hyperparameter Sensitivity:** The performance of OpenMax is notoriously sensitive to the "tail size" parameter (how many samples are used to fit the Weibull). [1]
3. **Logit vs. Latent:** OpenMax modifies logits based on latent distances. However, Mahalanobis Distance methods have shown that operating *purely* in the latent space (pre-logit) is more effective.

---

4. The Foundation Model Era: DINOv2 and the OOD Landscape

The paradigm of training models from scratch (supervised learning) is being superseded by the use of **Foundation Models**. **DINOv2** (Discriminative Self-supervised Learning) represents the state-of-the-art in this domain, offering feature spaces that are remarkably robust for OOD detection.

4.1 DINOv2: Architecture and Pre-training Signals

DINOv2 employs a student-teacher architecture trained with a combination of **DINO** (self-distillation) and **iBOT** (Masked Image Modeling) losses.

- **DINO Loss:** Encourages the student network to match the teacher's output on different augmented views of the same image (global consistency).
- **iBOT Loss:** Forces the student to reconstruct masked patches of the image based on the visible context (local consistency).

This dual objective is critical for fine-grained plant disease tasks. Unlike CLIP, which aligns images to text captions (often losing fine-grained visual details), DINOv2 is purely visual and learns features that capture subtle spatial patterns required for depth and fine-grained segmentation.

4.2 The Linear Probe vs. Nearest Neighbor (k-NN) Debate

When adapting a foundation model like DINOv2, a critical choice arises between using a Linear Probe or Non-Parametric Nearest Neighbor (k-NN) evaluation. [6]

4.2.1 Linear Probing: The Bottleneck Effect

Linear probing involves freezing the DINOv2 backbone and training a linear layer on the labeled ID dataset.

- **OOD Weakness:** Linear probing can degrade OOD detection as the linear projection may map OOD samples far from the ID manifold into "high confidence" regions of the decision space. [3]

4.2.2 Nearest Neighbor (k-NN): Manifold Preservation

The k-NN approach stores feature embeddings of the training set. For a test sample, the OOD score is the distance to the k-th nearest neighbor.

- **OOD Strength:** k-NN operates directly on the native DINOv2 manifold. Because DINOv2 is trained to cluster visually similar images, ID samples naturally cluster tightly, while OOD samples fall into sparse regions.
- **Performance:** DINOv2 + k-NN consistently outperforms all other foundation models and linear probes by a large margin on challenging benchmarks like iNaturalist and NINCO.

---

5. State-of-the-Art LoRA Implementations for OOD Detection

While DINOv2 provides general features, specialized crop disease tasks benefit from **Low-Rank Adaptation (LoRA)**. Modern research has moved beyond simple merging to exploit LoRA modules as active uncertainty sensors.

5.1 Unmerged LoRA Embeddings (LoRA-MD)

A significant breakthrough involves using the **activations of the LoRA branch (BAx)** before they are merged into the pre-trained weights. [7]

- **Mechanism:** The LoRA embedding $E_{LORA}(x)$ is defined as the concatenation of

intermediate activations A_{ij} for all L layers, averaged across input tokens. [7]
- **Near-OOD Superiority:** In fine-grained scenarios where standard MD on last-layer activations fails (AUROC ~0.4), MD using LoRA embeddings can reach AUROCs as high as **0.890**. [7] This is because the adapter specifically captures the "delta" required to process ID diseases, while ignoring general background features. [7]

5.2 Selective Low-Rank Approximation (SeTAR)

SeTAR is a novel, training-free OOD detection method that leverages selective low-rank approximation of weight matrices.

- **Selective Rank Reduction:** Instead of a uniform rank for all adapters, SeTAR uses a greedy search algorithm to identify the most impactful layers and singular values to retain.
- **Stability Enhancements:** By approximating weight matrices with optimal low-rank configurations, SeTAR filters out "minor" singular components that often contain noise and compromise OOD stability.
- **SeTAR+FT:** An extension that fine-tunes the "minor" components of weight matrices while freezing major ones, achieving state-of-the-art results on ImageNet1K by reducing false positive rates by up to **18.95%**.

5.3 LoRA-BAM: Boxed Abstraction Monitors

LoRA-BAM (Boxed Abstraction Monitors) implements lightweight monitors directly over the LoRA layers to filter queries beyond the model's specialized competence.

- **Implementation:** It applies k-means clustering and **boxed abstraction** (defining conservative geometric bounds) to the LoRA feature vectors. [8]
- **Interpretability:** Unlike black-box detectors, LoRA-BAM provides interpretable OOD rejection by determining if a sample falls within the "box" of learned task-specific features.
- **Performance:** It has shown an 88% reduction in hallucination errors in open-world detection benchmarks and rejects up to 95% of far-OOD queries while retaining nearly all legitimate ID samples.

---

6. Fine-Grained Attention Mechanisms

The ability of a ViT to detect OOD samples is linked to multi-scale features and attention focus.

6.1 Multi-Scale Attention Architectures

In FGVC, attention must capture both global structure and minute pathological details.

- **Cascaded Multi-Scale Attention (CMSA):** Tailored for hybrid architectures to integration features across scales without downsampling, capturing both coarse and fine-grained details simultaneously.
- **MAHL Framework:** A multi-scale attention-driven approach where spatial regions are adaptively selected based on discriminative contribution scores from hierarchical classifiers. [9]

6.2 Attention Entropy as an Uncertainty Metric

The entropy of the attention map provides a training-free OOD signal.

- **ID vs. OOD:** Known diseases produce "peaked" attention (Low Entropy). Unknown objects result in diffuse attention (High Entropy). [10]
- **VIPAMIN:** A novel prompt tuning initialization method that specializes prompt attention to minimize entropy on ID data, thereby maximizing the gap for OOD samples. [10]

---

7. Interpretability and Validation: Class Activation Mapping (CAM)

Trust in agricultural AI is maintained through visual verification.

- **Grad-CAM for ViTs:** Highlights specific regions influencing class prediction, ensuring the model focuses on the lesion rather than background noise. [13]
- **MECAM-OOD:** A recent approach integrating CAM with multi-exit networks to enhance the spatial awareness of OOD detection.
- **Validation:** CAM serves as a final "sanity check." If a model is confident but the CAM shows focus on soil or image corners, the prediction is flagged as unreliable. [13]

---

8. Synthesis and Recommendation for Agricultural Deployment

For the specific challenge of **Fine-Grained Plant Disease Detection**, a single detection method is rarely sufficient. The most robust state-of-the-art framework follows a layered approach:

Component	Implementation Recommendation	Primary Benefit

Backbone	**DINOv2 (ViT-L/14)**	Superior visual texture modeling for fine-grained lesions.
Adaptation	**SeTAR+FT**	Selective low-rank fine-tuning to stabilize the OOD-robust manifold.
Detection	**LoRA-MD (Unmerged)**	Concatenated $A_i a$ activations for maximum sensitivity to disease patterns. [7]
Normalization	**Mahalanobis++**	L2-normalization of features to stabilize ViT angular distance. [5]
Refinement	**LoRA-BAM**	Interpretable "boxed" filtering to reject semantically far outliers.
Visual Check	**Grad-CAM**	Verification that the OOD rejection or ID score is based on leaf pathology. [13]

Conclusion

The current "State-of-the-Art" moves beyond treating OOD detection as a post-processing step on the final output. By integrating detection into the **PEFT architecture itself** (via unmerged adapters, selective approximation, and boxed monitors), practitioners can achieve high-fidelity reliability even in the most challenging near-OOD agricultural scenarios.

References (Integrated)

- **** Li, Y., et al. (2024). *SeTAR: Out-of-Distribution Detection with Selective Low-Rank Approximation*. NeurIPS.
- [7] Salimbeni, E., et al. (2024). *Beyond Fine-Tuning: LoRA Modules Boost Near-OOD Detection*. DLSP.
- **** *LoRA-BAM: Input Filtering for Fine-tuned LLMs via Boxed Abstraction Monitors over LoRA Layers*.

- **** Mueller, M. & Hein, M. (2025). *Mahalanobis++: Improving OOD Detection via Feature Normalization*. ICML.
- **** Ren, J., et al. (2021). *A Simple Fix to Mahalanobis Distance for Improving Near-OOD Detection*.
- **** Oquab, M., et al. (2023). *DINOv2: Learning Robust Visual Features without Supervision*.
- [10] *VIPAMIN: Novel initialization method for visual prompt tuning for OOD*.
- [13] Various authors on Grad-CAM for plant disease interpretation in Transformers.

Alıntılanan çalışmalar

1. OpenMax with Clustering for Open-Set Classification - IFI UZH, erişim tarihi Şubat 2, 2026, https://www.ifi.uzh.ch/dam/jcr:85dfa060-839a-4d4c-83a0-58d334ee064c/huber2024bachelor.pdf
2. Combining pre-trained Vision Transformers and CIDER for Out Of Domain Detection - arXiv, erişim tarihi Şubat 2, 2026, https://arxiv.org/pdf/2309.03047
3. Unsupervised Image Classification with Adaptive Nearest Neighbor Selection and Cluster Ensembles - arXiv, erişim tarihi Şubat 2, 2026, https://arxiv.org/html/2511.16213v1
4. A Closer Look at Benchmarking Self-supervised Pre-training with Image Classification, erişim tarihi Şubat 2, 2026, https://www.researchgate.net/publication/391218098_A_Closer_Look_at_Benchmarking_Self-supervised_Pre-training_with_Image_Classification
5. Mahalanobis++: Improving OOD Detection via Feature Normalization - arXiv, erişim tarihi Şubat 2, 2026, https://arxiv.org/html/2505.18032v1
6. OpenOOD v1.5: Enhanced Benchmark for Out-of-Distribution Detection - arXiv, erişim tarihi Şubat 2, 2026, https://arxiv.org/html/2306.09301v5
7. Beyond fine-tuning: LoRA modules boost near-OOD ... - DLSP 2024, erişim tarihi Şubat 2, 2026, https://dlsp2024.ieee-security.org/papers/dls2024-final19.pdf
8. LoRA-BAM: Input Filtering for Fine-tuned LLMs via Boxed Abstraction Monitors over LoRA Layers - ResearchGate, erişim tarihi Şubat 2, 2026, https://www.researchgate.net/publication/392334388_LoRA-BAM_Input_Filtering_for_Fine-tuned_LLMs_via_Boxed_Abstraction_Monitors_over_LoRA_Layers
9. Multi-Scale Attention-Driven Hierarchical Learning for Fine-Grained Visual Categorization, erişim tarihi Şubat 2, 2026, https://www.mdpi.com/2079-9292/14/14/2869
10. VIPAMIN: Visual Prompt Initialization via Embedding Selection and Subspace Expansion, erişim tarihi Şubat 2, 2026, https://arxiv.org/html/2510.16446v1
11. Bridging Attribution and Open-Set Detection using Graph-Augmented Instance Learning in Synthetic Speech - arXiv, erişim tarihi Şubat 2, 2026, https://arxiv.org/html/2601.07064v1

12. Can Pre-trained Networks Detect Familiar Out-of-Distribution Data? - OpenReview, erişim tarihi Şubat 2, 2026, https://openreview.net/forum?id=Pb9PIECnNF
13. Enhancing multiclass plant disease classification using GAN-boosted vision transformer with XAI insights - NIH, erişim tarihi Şubat 2, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12827657/
14. [Project] Recent Class Activation Map Methods for CNNs and Vision Transformers - Reddit, erişim tarihi Şubat 2, 2026, https://www.reddit.com/r/MachineLearning/comments/myenmh/project_recent_class_activation_map_methods_for/
15. Implementation of Explainable AI in Deep Learning Methods for Multiclass Classification of Plant Diseases in Mango Leaves - ddd-UAB, erişim tarihi Şubat 2, 2026, https://ddd.uab.cat/pub/elcvia/elcvia_a2025v24n1/elcvia_a2025v24n1p104.pdf