

# Artificial Intelligence in Pathogen Based Plant Disease Detection: Advances, Trends, and Challenges

Adeeb P H

Department of Computer Science and Engineering  
College of Engineering, Trivandrum  
Kerala, India  
adeebharis@gmail.com

Thasni T

Department of Computer Science and Engineering  
College of Engineering, Trivandrum  
Kerala, India  
thasnit@cet.ac.in

**Abstract**—Plant disease outbreaks caused by fungi, bacteria, and viruses continue to threaten crop yields, food security, and the long-term health of agricultural systems across the globe. Detecting these diseases early, accurately, and at scale remains a difficult task, especially as climate patterns shift and food supply chains stretch across continents. Conventional diagnostic methods such as visual inspection, lab-based tests, and molecular assays have long been the standard, yet they are often slow, expensive, and tied to controlled laboratory settings. In recent years, advances in artificial intelligence (AI) have begun to transform this landscape. Modern approaches powered by convolutional neural networks (CNNs), transformers, and multimodal AI now take on the challenge of interpreting complex agricultural data, whether from images, sensors, or text. This review explores the current ecosystem of AI-driven techniques for detecting plant diseases caused by pathogens, comparing how different methods perform and where each one fits best. Special focus is given to the shift from traditional image-based deep learning toward newer frontiers such as explainable AI (XAI), graph neural networks (GNNs), few-shot and foundation models, and fully multimodal systems. We structure our comparison not around accuracy rankings alone, but around what each paradigm can and cannot do: the conditions under which it was tested, what it requires to operate, what it reports about its own decisions, and how close it has come to real agricultural deployment. Along the way, the review identifies persistent challenges including dataset over-reliance on controlled benchmarks, poor domain transfer to real-world conditions, constraints in real-time deployment, and the absence of standardised explainability evaluation. Drawing from the latest research, it outlines the major gaps that still remain and offers practical directions for future work.

**Keywords**—Plant disease detection; Transfer learning; Explainable AI; Few-shot learning; Graph neural networks; Multimodal learning; Vision transformers

## I. INTRODUCTION

Agricultural productivity around the world is increasingly threatened by plant diseases caused by fungi, bacteria, and viruses, with millions of hectares of crops affected each year and epidemic outbreaks especially common in regions already facing food insecurity [1]. The impact is severe: pests and diseases account for up to 40% of global crop losses annually, resulting in over \$220 billion in economic damage [2], with yield reductions ranging from 21.5% in wheat to 40% in

potatoes [3]. Despite this urgency, current diagnostic systems remain inadequate for large-scale use; although laboratory-based methods such as PCR, ELISA, and microscopy provide high accuracy, they are slow, resource-intensive, and often inaccessible to farmers in field or low-resource environments [4], [5].

Advances in AI are transforming plant disease detection, with deep learning models—including Convolutional Neural Networks (CNNs), Transformers, and Graph Convolutional Networks (GCNs)—achieving state-of-the-art performance. Ensemble frameworks now surpass 99% accuracy on benchmark datasets [6], [7], and compact hybrid Vision Transformer (ViT) architectures achieve 92.6–98.3% accuracy with only 0.8M parameters [8]. Adversarial training further enhances robustness; Fast Gradient Sign Method (FGSM)-trained ViT models reach 99.4% accuracy on more than 100,000 bean disease samples [9].

However, a critical limitation pervades this literature: performance is evaluated almost exclusively through benchmark accuracy on controlled datasets such as PlantVillage, which features clean backgrounds and laboratory-quality imaging. This single-metric framing obscures wide variation in what methods actually require, what they report about their own decisions, and how closely they have been tested under real agricultural conditions. A model achieving 99.89% accuracy on PlantVillage may behave very differently under field lighting, background clutter, or when encountering disease classes absent from its training set.

Several reviews have examined AI approaches to plant disease detection, many focusing on CNN accuracy benchmarks on public datasets [10], [11]. More recent work has examined large language models (LLMs) and foundation models in broader agricultural contexts [12]. This review differs from its predecessors in two respects: it covers a wider range of paradigms—including spectral methods, graph neural networks (GNNs), multimodal LLMs, and few-shot and zero-shot approaches alongside standard CNNs—and it organises its comparison around structural capabilities and deployment attributes rather than accuracy rankings, enabling practitioners

to assess method suitability for their specific context rather than identifying a universal winner.

This review addresses the evaluation gap directly. We synthesise 18 studies published between 2019 and 2025, covering CNNs, hybrid transformers, hyperspectral and spectroscopic models, explainable AI (XAI) techniques, graph-based approaches, few-shot and foundation models, and multimodal architectures. Rather than imposing subjective quality scores, we organise our comparison around documented, factual attributes of each study: the dataset and imaging conditions used, the number of crops and disease classes covered, whether inference time and model size were reported, what explainability mechanisms were included, and what form of deployment (if any) was demonstrated. This allows readers to assess suitability for their own context rather than accepting a ranked leaderboard.

## II. RELATED WORK

This review was conducted through a structured literature search across IEEE Xplore, Scopus, Web of Science, and Google Scholar. Search terms included combinations of “plant disease detection,” “crop disease classification,” “deep learning,” “convolutional neural network,” “vision transformer,” “graph neural network,” “explainable AI,” “few-shot learning,” “hyperspectral,” and “multimodal learning,” restricted to publications from January 2019 to June 2025. This window was selected to capture the transition from CNN-dominant approaches toward more recent paradigms including foundation models, multimodal architectures, and few-shot methods.

Studies were included if they: (1) addressed detection or classification of plant diseases caused by pathogens (fungal, bacterial, or viral), or in one case a broader plant health framework also encompassing nutritional deficiency alongside pathogen-driven disease [13]; (2) employed a machine learning or deep learning method as a core component; (3) reported quantitative evaluation results; and (4) were published in peer-reviewed journals or indexed conference proceedings. Studies were excluded if they addressed only post-harvest disease, plant phenotyping without a disease focus, or non-plant biological systems.

An initial pool of over 200 candidate papers was reduced to the 18 studies included in this review by prioritising methodological diversity across six paradigms (CNN/ensemble, hybrid CNN-ViT, spectral, GNN, multimodal LLM, and few-shot/zero-shot), recency within the 2019–2025 window, and relevance to real-world deployment considerations. Where multiple studies employed nearly identical architectures on the same benchmark dataset, the most recent or the one reporting the broadest attribute set was preferred. No formal PRISMA protocol was followed, but the selection rationale described above is sufficient to reconstruct the corpus from the stated search sources and criteria. Fig. 1 provides a visual overview of all six paradigms and their structural capabilities before the detailed sections that follow.

## III. DEEP LEARNING ARCHITECTURES FOR PLANT DISEASE DETECTION

### A. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) continue to be the most widely used architecture for plant disease detection because of their strong ability to learn hierarchical spatial features from images. Meena et al. [14] applied Mask R-CNN to the PlantVillage dataset of 22,500 images covering 14 crop species, achieving 98% classification accuracy and 94% instance segmentation accuracy. The framework employs a pre-trained backbone (ResNet or VGG), a Region Proposal Network (RPN) to generate bounding boxes, and an RoI Align layer for precise spatial alignment. Region-based CNNs not only classify diseases but also localise lesions, offering insights for assessing severity in precision agriculture.

Rahman et al. [7] developed a real-time plant leaf disease monitoring system using a deep CNN pipeline with image preprocessing and augmentation stages. Evaluated on multi-crop benchmark data, the system achieved over 99% accuracy and was designed with continuous monitoring integration in mind, though inference latency and hardware configuration were not reported.

Despite these strong benchmark results, PlantVillage images are captured under controlled conditions, making direct comparison with field performance difficult. None of the CNN-based studies reviewed here provide systematic evaluation on independently collected field data—a gap that is rarely acknowledged within the papers themselves.

### B. 3D Convolutional Networks for Hyperspectral Data

Nagasubramanian et al. [15] introduced a 3D Deep Convolutional Neural Network (3D-DCNN) to detect charcoal rot in soybean stems using hyperspectral images across 240 wavelength bands (400–1000 nm). The 3D architecture processes both spatial and spectral information simultaneously, achieving 95.73% accuracy and an F1-score of 0.87 for infected samples. Saliency map analysis revealed that near-infrared wavelengths were the most informative for identifying disease, enabling pre-symptomatic detection before visible lesions appear. This is a qualitatively different capability from RGB-based methods: spectral approaches can detect infection at a stage when treatment is most effective. The requirement for specialised hyperspectral hardware limits field scalability.

### C. Vision Transformers and Hybrid Architectures

Thakur et al. [8] introduced PlantXViT, a hybrid architecture integrating CNN-based feature extraction with Vision Transformer (ViT) attention mechanisms. With only 0.8 million parameters, it achieves 92.6–98.3% accuracy and is designed for mobile deployment, incorporating GradCAM and LIME visualisations for interpretability. It is the only architecture in this review explicitly benchmarked for parameter efficiency.

Njau et al. [9] strengthened transformer robustness through Fast Gradient Sign Method (FGSM) adversarial training for common bean disease detection. Their adversarially trained ViT attained 99.4% accuracy on a dataset of over 100,000

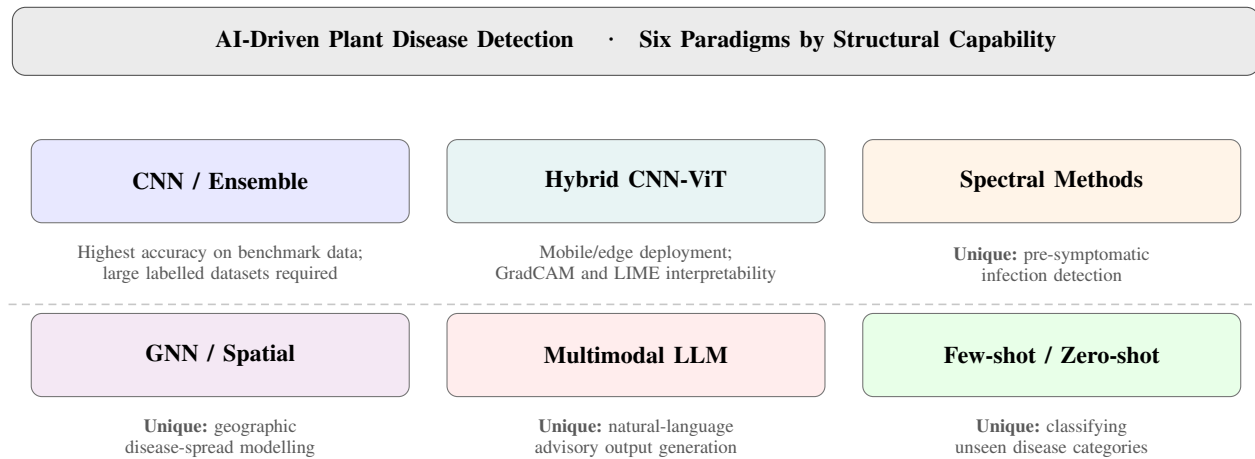


Fig. 1: Taxonomy of the six AI paradigms covered in this review, organised by structural capability. Entries marked **Unique:** identify a capability that is categorically unavailable to all other paradigms regardless of implementation quality; entries without this label indicate relative rather than exclusive strengths.

annotated images, demonstrating resilience to lighting, background, and orientation variation—conditions closer to field variability than most benchmark evaluations test for. Among the reviewed studies, adversarial training is the only technique that directly targets the domain gap at training time.

#### D. Multi-Stage Detection and Ensemble Approaches

Assaduzzaman et al. [16] introduced XSE-TomatoNet, enhancing EfficientNetB0 with squeeze-and-excitation blocks and multi-scale feature fusion. Trained on 32,532 tomato images, it achieved 99.11% accuracy with the broadest XAI suite reviewed—combining SHAP, LIME, GradCAM, and GradCAM++—and was deployed as both a web and Android application.

Ali et al. [6] combined DenseNet201, EfficientNetB0, EfficientNetB3, InceptionResNetV2, and ResNet50v2 with CLAHE and Adaptive Median Filter preprocessing, achieving 99.89% accuracy on 87,000 PlantVillage images—the highest benchmark result in this review—without reporting XAI capabilities or inference speed.

Polly and Devi [17] proposed a six-stage pipeline using drone imagery: YOLOv8-based leaf detection, background removal via DeepLabV3+, disease classification (96.97% training, 92.89% validation), U-Net semantic segmentation (99% accuracy), severity estimation, and treatment recommendations. Among the reviewed studies, it is the only one integrating drone-based field acquisition with end-to-end severity quantification.

### IV. CLASSICAL AND HYBRID MACHINE LEARNING APPROACHES

#### A. Spectroscopy with Optimised Machine Learning

Wang et al. [18] applied the Information Feedback Optimization (INFO) algorithm to fine-tune Support Vector Machine (SVM) parameters for detecting Verticillium wilt in cotton using Raman spectroscopy. The optimised model

achieved 97.5% training accuracy (F1-score: 0.974) and 90.0% validation accuracy (F1-score: 0.867), detecting infections before visible symptoms appeared. The INFO algorithm's feature importance outputs provide interpretability absent from most black-box deep learning classifiers, though deployment requires laboratory-grade Raman instrumentation.

#### B. Ensemble Methods for High-Resolution RGB Imaging

Almadhor et al. [19] integrated colour and texture feature extraction with ensemble tree classifiers for guava disease recognition using DSLR imagery, achieving 99% accuracy on 3,651 images. This highlights the continued relevance of classical ensemble learning for high-quality, small-scale datasets, though DSLR-quality image requirements constrain practical scalability.

### V. MULTIMODAL LEARNING AND INTEGRATION

Wang et al. [20] developed LLMI-CDP, a multimodal large language model (LLM) built on VisualGLM, fine-tuned using LoRA and Q-Former for visual-text alignment. Trained on 2,498 Chinese agricultural images covering 141 categories, LLMI-CDP achieved 78.8% disease recognition and 86.7% pest detection accuracy. Among the reviewed studies, it is the only system that generates prevention strategy text alongside predictions. High inference latency is explicitly noted as limiting real-time deployment.

Zeng et al. [21] introduced DIC-Transformer, integrating Faster R-CNN with a Swin Transformer backbone for detection and caption generation (BLEU-1: 0.756, CIDEr-D: 450.51, ROUGE: 0.721) alongside disease classification (accuracy: 0.854) across 3,971 images in 18 categories. Caption generation provides non-expert users with textual disease descriptions alongside class labels—an output type absent from all other reviewed architectures.

## VI. GRAPH NEURAL NETWORKS AND SPATIAL MODELLING

Bera et al. [13] presented PND-Net, integrating CNNs and Graph Convolutional Networks (GCNs) through spatial pyramid pooling. It is noted that this work addresses both pathogen-driven disease and nutritional deficiency—the latter being an abiotic condition outside the strict pathogen scope of this review; it is included here given its methodological contribution to multi-dataset GNN evaluation. Tested across four datasets—banana, coffee, potato, and PlantDoc—it achieved 90.00%, 90.54%, 96.18%, and 84.30% accuracy respectively. The PlantDoc evaluation is notable because PlantDoc contains field-collected images with complex, realistic backgrounds, making it a more credible test of generalisation than PlantVillage.

Hou et al. [22] focused on detecting wheat soil-borne mosaic virus using Bayesian GCNs for spatial trust propagation, spectral clustering for regional heterogeneity, and multi-indicator feedback for interpretability. Unlike any other reviewed study, this system models the spatial dynamics of disease spread across geographic plots rather than classifying individual leaf images—a fundamentally different task suited to regional monitoring at scale.

Meenalochini and Amudha [23] built a machine vision-based expert system for cauliflower disease detection integrating deep learning with IoT capabilities. Trained on 750 images across four diseases, the GNN achieved 89% accuracy, while IoT integration enabled real-time monitoring and mobile alert notifications for farmers—the closest any reviewed study comes to a functioning field deployment, though this claim rests on a single study published in a limited-indexing venue and has not been independently corroborated.

## VII. FEW-SHOT LEARNING AND FOUNDATION MODELS

Jiang et al. [24] proposed PlantCaFo, leveraging foundation models (CLIP, DINO, DALL-E, GPT-3) with a Dilated Contextual Adapter (DCon-Adapter) and weight decomposition matrix. PlantCaFo achieved 93.53% accuracy in a 16-shot setting, a 7.74% improvement over baselines, and was additionally evaluated on a new out-of-distribution (OOD) dataset—rare among reviewed studies. The OOD evaluation quantifies accuracy decline under domain shift rather than simply noting it as a limitation.

Rezaei et al. [25] proposed PMFFA, combining pre-training, meta-learning, and feature attention. Using ViT, it achieves 90.12% accuracy on PlantDoc with 1.11 ms inference time—combining near-field-realistic benchmark performance with edge-viable inference speed. The feature attention module improved accuracy by 3.10% by emphasising discriminative regions and suppressing background noise.

Farooq et al. [26] extended zero-shot learning for potato early blight using GloVe embeddings and cosine similarity projection, achieving 98.50% on seen and 99.91% on unseen disease categories. Zero-shot generalisation to unseen classes is qualitatively important for emerging diseases where training data is unavailable.

## VIII. EXPLAINABILITY AND INTERPRETABILITY

Integrating explainable AI (XAI) techniques has become more common across the reviewed literature, but the evaluation of those explanations remains shallow. Saliency mapping [15] identified near-infrared wavelengths (700–1000 nm) as the most discriminative for charcoal rot. Grad-CAM [8], [16], [22] produced class-specific heatmaps emphasising lesion regions. LIME [8], [16] provided localised interpretability by perturbing inputs to confirm predictions relied on disease symptoms. Vision Transformers [8], [9] inherently supported interpretability through attention visualisation.

A critical gap persists: no reviewed paper evaluates explanation quality using objective metrics such as pointing games, deletion-insertion curves, or localisation accuracy tests. The presence of a Grad-CAM visualisation is treated as evidence of interpretability, but it is not evidence that the explanation is accurate, complete, or useful to agronomists. No study reports a user study with farmers or domain experts to assess whether explanations support better decision-making in practice.

## IX. CROSS-PARADIGM SYNTHESIS

A cross-paradigm observation concerns the tradeoff between data requirements and generalisation capability. CNN and ensemble models achieve the highest absolute accuracies but require thousands of labelled images per disease class and are structurally confined to categories present in their training set. Few-shot and zero-shot approaches accept lower accuracy in exchange for the ability to operate with five to sixteen labelled examples per class, or none at all for unseen categories. Foundation-model approaches such as PlantCaFo [24] occupy an intermediate position: they leverage large pretraining corpora but adapt with modest domain-specific data.

A second observation concerns the distinction between demonstrated and theoretical capability. Several claims in the reviewed literature conflate architectural potential with empirical performance. Zero-shot methods can in principle classify any disease representable in a semantic embedding space; in practice, the reviewed implementations demonstrate this on a limited set of potato disease categories [26]. The structured comparison that follows attributes capabilities only where they are empirically demonstrated in the reviewed corpus.

## X. METHODOLOGICAL COMPARISON

### A. Grouped Narrative Comparison

Table I organises the reviewed studies by paradigm across five factual attributes—dataset and imaging conditions, crops and disease classes covered, explainability reported, inference speed and model size, and structural limitations—drawn directly from what each study reports, with no quality judgement imposed. These five attributes were chosen because they jointly determine real-world deployability—evaluation realism, breadth of testing, transparency, edge feasibility, and categorical limits—rather than being an arbitrary subset of what each study reports.

TABLE I: GROUPED NARRATIVE COMPARISON OF AI PARADIGMS FOR PLANT DISEASE DETECTION.

| Paradigm                                                   | Studies                                                                         | Datasets & imaging conditions                                                                                                                                             | Crops / disease classes                                                                                | Explainability reported                                                                                              | Inference & model speed size reported                                                                   | What this group cannot do                                                                                                                          |
|------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>CNN and ensemble models</b>                             |                                                                                 |                                                                                                                                                                           |                                                                                                        |                                                                                                                      |                                                                                                         |                                                                                                                                                    |
| CNN / ensemble                                             | Ali et al. [6]; Almadhor et al. [19]; Rahman et al. [7]; Meena et al. [14]      | PlantVillage (controlled studio); Plant Pathology 2020; guava DSLR imagery (Almadhor et al.); RGB images only                                                             | Up to 38 disease classes; up to 14 species                                                             | None reported                                                                                                        | Inference time not reported; model sizes not reported                                                   | Detect disease before visible symptoms appear; operate with fewer than hundreds of labelled images per class; generalise to unseen disease classes |
| <b>Lightweight and explainable CNN-transformer hybrids</b> |                                                                                 |                                                                                                                                                                           |                                                                                                        |                                                                                                                      |                                                                                                         |                                                                                                                                                    |
| Hybrid CNN-ViT with XAI                                    | Thakur et al. [8]; Assaduzzaman et al. [16]; Polly & Devi [17]; Njau et al. [9] | Multi-crop public datasets; PlantVillage (drone imagery in Polly et al.); 100,000+ annotated bean images (Njau et al.); RGB only                                          | 11 tomato diseases (Assaduzzaman); multi-crop (Thakur); 3 crops from drone (Polly); common bean (Njau) | GradCAM, LIME, SHAP, GradCAM++ (Assaduzzaman); GradCAM + LIME (Thakur); severity maps (Polly); attention maps (Njau) | 0.8M parameters reported (Thakur); web + Android deployment (Assaduzzaman); inference time not reported | Generate natural language descriptions of disease; operate with very limited labelled data; detect pre-symptomatic infection                       |
| <b>Spectral and non-RGB methods</b>                        |                                                                                 |                                                                                                                                                                           |                                                                                                        |                                                                                                                      |                                                                                                         |                                                                                                                                                    |
| Hyperspectral & Raman                                      | Nagasubramanian et al. [15]; Wang et al. [18]                                   | 240-band hyperspectral (soybean); Raman spectra (cotton); both laboratory instruments                                                                                     | 1 crop each; charcoal rot and Verticillium wilt respectively                                           | Saliency maps (Nagasubramanian); feature importance via INFO (Wang)                                                  | Not reported; requires specialised hardware                                                             | Operate without specialist spectral equipment; scale to field conditions without infrastructure investment                                         |
| <b>Graph neural networks and spatial modelling</b>         |                                                                                 |                                                                                                                                                                           |                                                                                                        |                                                                                                                      |                                                                                                         |                                                                                                                                                    |
| GNN / graph-based                                          | Bera et al. [13]; Hou et al. [22]; Meenalochini & Amudha [23]                   | PlantDoc (field images); WSBM wheat dataset (plot-level geographic data); 750 cauliflower images with IoT capture                                                         | 4 datasets / crop types (Bera); regional wheat plots (Hou); 4 cauliflower diseases (Meenalochini)      | Spatial pyramid visualisation (Bera); multi-indicator feedback (Hou); none (Meenalochini)                            | Real-time reported in Meenalochini; model details not reported                                          | Detect disease before visible symptoms; generate explanatory text for non-experts                                                                  |
| <b>Multimodal large language models</b>                    |                                                                                 |                                                                                                                                                                           |                                                                                                        |                                                                                                                      |                                                                                                         |                                                                                                                                                    |
| Multimodal LLM                                             | Wang et al. [20]; Zeng et al. [21]                                              | Chinese agricultural images; ADCG-18 (purpose-built, 18 disease categories); RGB only                                                                                     | 141 crop/pest categories (Wang); 18 disease categories across 10 plant species (Zeng)                  | Natural language captions and prevention strategies; GradCAM in Zeng et al.                                          | High inference latency noted as limitation (Wang); model sizes not reported                             | Match CNN-level accuracy; operate in real time on edge hardware; function without large language model infrastructure                              |
| <b>Few-shot and zero-shot learning</b>                     |                                                                                 |                                                                                                                                                                           |                                                                                                        |                                                                                                                      |                                                                                                         |                                                                                                                                                    |
| Few-shot / zero-shot                                       | Jiang et al. [24]; Rezaei et al. [25]; Farooq et al. [26]                       | PlantVillage; PlantDoc (field images); Cassava; OOD dataset (Jiang); 5–16 labelled examples per class (Rezaei, Jiang); zero labelled examples for unseen classes (Farooq) | 27–38 disease classes; multi-crop; unseen categories in Farooq et al.                                  | Attention maps (Jiang, Rezaei); cosine similarity scores (Farooq)                                                    | 1.11 ms per image reported (Rezaei ViT); 0.57 ms (Rezaei ResNet50)                                      | Match large-dataset CNN accuracy under data-rich conditions; generate natural language explanations                                                |

Note: All attributes are drawn directly from reported experimental conditions. “—” indicates the attribute was not reported in any study within the group.

## B. Quantitative Overview

Across the 18 reviewed studies, average reported accuracy varies substantially by paradigm: CNN and ensemble models average 98.7% on their respective benchmarks; hybrid CNN-ViT models average 96.8%; spectral methods average 93.9%; GNN-based studies average 88.5%; multimodal LLMs average 82.8%; and few-shot and zero-shot methods average 94.0%. These figures must be interpreted with caution, as they aggregate results from different datasets, imaging conditions, and evaluation protocols, rendering cross-paradigm comparisons unreliable in isolation.

Of the 18 studies, 10 were published in 2024–2025, reflecting rapid acceleration in the field. Seven studies use PlantVillage as a primary or secondary benchmark, four incorporate PlantDoc (field-collected images), and the remainder use domain-specific or purpose-built datasets. Dataset sizes range from 750 images (Meenalochini et al.) to over 100,000 (Njau et al.), with the majority falling in the 2,000–90,000 range. Five studies report some form of model deployment (web, Android, or IoT), and only two (Rezaei et al. and Meenalochini et al.) report inference speed.

## C. What the Grouped Comparison Reveals

The structured grouping in Table I makes several cross-cutting patterns visible that accuracy-only comparisons conceal.

**Inference speed and model size are rarely reported.** Only two studies (Rezaei et al. and Meenalochini et al.) provide inference speed figures, and only one (Thakur et al.) reports parameter count. Since deployment suitability for mobile and edge devices depends directly on these values, the field as a whole lacks the data needed to make principled deployment recommendations. Standardising the reporting of inference time, memory footprint, and hardware configuration should be a baseline expectation.

**Dataset conditions differ fundamentally across groups.** CNN and ensemble models are validated almost entirely on PlantVillage, which provides controlled studio-quality images. GNN-based studies include PlantDoc and IoT-captured field images. Few-shot methods include OOD evaluation in at least one case. These are not equivalent evaluation contexts, and cross-group accuracy comparisons are therefore misleading. The apparent accuracy advantage of CNN ensembles over GNNs (99.89% vs. 84.30%) partly reflects more favourable evaluation conditions rather than genuine superiority under field conditions. Notably, no CNN or ensemble study in this corpus was evaluated on PlantDoc, making a controlled same-dataset comparison unavailable—which is itself a finding.

**Each paradigm has a structurally unique capability.** Spectral methods are the only group capable of pre-symptomatic detection; this is not a performance difference but a categorical one—RGB-based methods cannot operate on data that does not yet exist visually. Multimodal LLMs are the only group producing natural language explanations and actionable prevention guidance; this cannot be achieved by adding an XAI tool to a CNN. Zero-shot methods are the only group that

can classify disease classes entirely absent from training; this is structurally unavailable to fixed-class classifiers regardless of their accuracy on known classes. Recognising these categorical distinctions is important: selecting a method because it has the highest benchmark score on PlantVillage, when the deployment task requires pre-symptomatic detection or zero-shot generalisation, reflects a mismatch between evaluation and need.

**Explainability is present in more studies but evaluated in none.** Every paradigm includes at least one study with some form of visualisation or attribution mechanism, which represents genuine progress. However, no study in any group evaluates whether those explanations are accurate using quantitative XAI metrics, or whether they are useful to farmers and agronomists through user studies. The inclusion of a Grad-CAM figure does not constitute explainability validation.

## D. Deployment Context Guidance

Based on the factual attributes documented in Table I, different deployment contexts call for different paradigm choices, independent of benchmark rankings:

- **Large labelled dataset available, maximum classification accuracy needed:** CNN/ensemble approaches (e.g., Ali et al. [6]) deliver the highest reported accuracy, though field generalisation should be independently verified.
- **Mobile or edge deployment required:** Hybrid CNN-ViT architectures (PlantXViT [8]) and few-shot ViT pipelines (PMFFA [25], reporting 1.11 ms inference) are the only groups with documented lightweight operation among the reviewed studies.
- **Rare or emerging disease with few available images:** Few-shot and zero-shot methods are the appropriate choice; large CNN ensembles require data volumes unavailable in these scenarios.
- **Pre-symptomatic monitoring required:** Spectral methods are the only option among the reviewed studies; no image-based approach reviewed here detects disease before visual symptoms appear.
- **Explanation for non-expert farmers required:** Multimodal LLMs (LLMI-CDP [20], DIC-Transformer [21]) are the only systems generating natural language outputs among the reviewed studies; XAI tools on CNN models produce heatmaps requiring technical interpretation.
- **Regional spread monitoring across geographic areas:** Bayesian GCN approaches (Hou et al. [22]) are designed for this spatial task; leaf-level classifiers are not.
- **Real-time IoT field integration:** The GNN-IoT system of Meenalochini et al. [23] is the only reviewed study demonstrating this integration, though on a small dataset and reported in a single, limited-indexing source.

## XI. LIMITATIONS AND CHALLENGES

### A. Benchmark Conditions Do Not Represent Field Conditions

Approximately ten of the eighteen reviewed studies validate primarily on PlantVillage or similarly controlled datasets—

the latter category explicitly including laboratory-instrument spectral studies (Nagasubramanian et al. [15]; Wang et al. [18]). The remaining eight incorporate field-collected images, purpose-built datasets, or out-of-distribution evaluation sets. Studies that evaluated on field-collected data—PlantDoc in Bera et al. [13] and Rezaei et al. [25]; drone imagery in Polly and Devi [17]—show accuracy figures 5–15 percentage points lower than PlantVillage results from comparable architectures. The magnitude of this gap is rarely acknowledged or quantified by studies that test only on controlled benchmarks. The field requires standardised evaluation protocols under realistic field conditions to enable honest cross-study comparison.

#### *B. Explainability Is Reported But Not Validated*

Every paradigm includes at least one XAI mechanism, but no study evaluates the accuracy or utility of its explanations using objective metrics. Pointing game accuracy, deletion-insertion curves, and segmentation-based localisation tests are established XAI evaluation methods entirely absent from this literature. Additionally, no study reports a user study with farmers or agronomists to assess whether provided explanations translate into better decision-making. Claiming that a model is “explainable” on the basis of a visualisation figure alone overstates what has been demonstrated.

#### *C. Inference Speed and Model Size Are Rarely Documented*

Only two of eighteen studies report inference speed, and only one reports model parameter count. Without this information, it is impossible to assess deployment feasibility on mobile or edge hardware. Future work should treat inference time, memory footprint, and the hardware used for evaluation as mandatory reporting requirements, equivalent in importance to accuracy metrics.

#### *D. Spectral Detection Remains Hardware-Constrained*

Pre-symptomatic detection via hyperspectral or Raman spectroscopy is the only reviewed approach that identifies infection before visible symptoms appear—the window of highest intervention value. However, both reviewed spectral studies operate exclusively with laboratory-grade instruments that are impractical for field deployment at scale. Bridging this gap through lower-cost multispectral sensors or sensor-camera fusion represents a high-priority direction.

#### *E. Dataset Scale and Crop Coverage Are Narrow*

Several studies use datasets too small for reliable generalisation (Meenalochini et al.: 750 images; LLMI-CDP: 2,498 images). The majority of studies cover a small fraction of commercially important crop-disease combinations. Rare diseases and crops prevalent in lower-income agricultural regions are particularly underrepresented.

#### *F. Robustness, Uncertainty, and Reproducibility*

None of the reviewed studies report uncertainty estimates alongside predictions, limiting the ability of end-users to calibrate trust in model outputs. Adversarial robustness is addressed in only one study (Njau et al. [9]). Reproducibility

is further constrained by the frequent absence of released code, model weights, or dataset access details; only a subset of reviewed studies make their implementation publicly available, limiting independent replication and comparative evaluation.

#### *G. No End-User Validation Exists*

No reviewed study evaluates its system with farmers or agronomists as participants, or measures socio-economic impact. Systems described as “field-deployable” have not demonstrated utility within real farming workflows. Participatory design and user-centred evaluation are absent from this literature as a whole.

## XII. FUTURE DIRECTIONS

### *A. Standardised Field-Condition Benchmarks*

The community requires agreed-upon field-condition evaluation datasets that include variable lighting, background clutter, geographic diversity, and realistic imaging equipment. These should complement PlantVillage rather than replace it, enabling fair comparison across paradigms evaluated under equivalent conditions.

### *B. Mandatory Reporting of Deployment Attributes*

Publication standards should require reporting of inference time, memory footprint, hardware configuration, code availability, and—where XAI is claimed—quantitative evaluation of explanation quality. These attributes are necessary for the field to move beyond accuracy-only comparison.

### *C. Closing the Spectral-Field Gap*

Developing affordable multispectral sensing hardware compatible with mobile platforms, or integrating spectral bands into standard agricultural camera systems, could extend pre-symptomatic detection beyond laboratory settings. AI models that fuse RGB and spectral signals represent a high-impact integration opportunity.

### *D. Participatory Design and Field Validation*

Broader field trials, participatory development involving farmers and extension agents, and socio-economic impact evaluation are necessary for translating research advances into agricultural practice. Co-design with end-users is a prerequisite for ensuring that model outputs—including both classifications and explanations—are actionable within real farming workflows.

### *E. Multimodal Advisory Systems*

Next-generation systems will extend beyond disease identification to provide advisory outputs combining image analysis with weather, soil, and historical data. Multimodal LLMs are promising for this role but require accuracy improvements and latency reductions before they can serve as primary diagnostic tools.

### XIII. CONCLUSION

This review has examined 18 AI-based approaches to pathogen-driven plant disease detection across six methodological paradigms. Rather than ranking them by benchmark accuracy alone, we have shown that each paradigm possesses structurally distinct capabilities and limitations: spectral methods enable pre-symptomatic detection no image-based approach can replicate; multimodal LLMs uniquely generate natural language guidance; zero-shot methods operate on disease classes absent from training; GNNs model spatial spread dynamics; and lightweight CNN-ViT hybrids operate under edge hardware constraints. These differences are categorical, not quantitative, and selecting a method by its PlantVillage accuracy without accounting for deployment context risks choosing the wrong tool for the task.

The field's most urgent needs are not architectural but methodological: standardised field-condition benchmarks, mandatory reporting of inference speed and model size, quantitative evaluation of explainability claims, and user studies with farming communities. Addressing these gaps will allow AI plant disease detection research to produce not just high-accuracy systems, but systems that are genuinely useful under the conditions agriculture actually presents.

### REFERENCES

- [1] J. B. Ristaino, P. K. Anderson, D. P. Bebber, K. A. Brauman, N. J. Cunniffe, N. V. Fedoroff, C. Finegold, K. A. Garrett, C. A. Gilligan, C. M. Jones, M. D. Martin, G. K. MacDonald, P. Neenan, A. Ospina, S. Savary, and J. E. Yuen, "The persistent threat of emerging plant disease pandemics to global food security," *Proceedings of the National Academy of Sciences*, vol. 118, no. 23, p. e2022239118, 2021.
- [2] Food and Agriculture Organization of the United Nations, "About fao's work on plant production and protection," <https://www.fao.org/plant-production-protection/en/>, 2021, accessed: 2025-11-13.
- [3] E.-C. Oerke, "Crop losses to pests," *The Journal of Agricultural Science*, vol. 144, no. 1, pp. 31–43, 2006.
- [4] N. Boonham, J. Kreuze, S. Winter, R. van der Vlugt, J. Bergervoet, J. Tomlinson, and R. Mumford, "Methods in virus diagnostics: from elisa to next generation sequencing," *Virus Research*, vol. 186, pp. 20–31, 2014.
- [5] S. Chakraborty, M. Chakraborty, Y. Muhovsky, G. Niebala, K. Kara, and A. Carlier, "Portable solutions for plant pathogen diagnostics," *Frontiers in Plant Science*, vol. 12, p. 778742, 2021.
- [6] A. H. Ali, A. Youssef, M. Abdelal, and M. A. Raja, "An ensemble of deep learning architectures for accurate plant disease classification," *Ecological Informatics*, vol. 79, p. 102507, 2024. [Online]. Available: <https://doi.org/10.1016/j.ecoinf.2024.102507>
- [7] K. N. Rahman, S. C. Banik, R. Islam, and A. A. Fahim, "A real-time monitoring system for accurate plant leaves disease detection using deep learning," *Crop Design*, vol. 4, p. 100092, February 2025, open Access. [Online]. Available: <https://doi.org/10.1016/j.crdes.2024.100092>
- [8] P. S. Thakur, T. Ahmad, A. Jain, D. Singh, and H.-N. Lee, "Vision transformer meets convolutional neural network for plant disease classification," *Ecological Informatics*, vol. 77, p. 102245, November 2023.
- [9] E. T. Njau, M. A. Dida, D. G. Nyambo, and C. M. Marwa, "Enhancing detection of common bean diseases using fast gradient sign method adversarial training in vision transformers," *Frontiers in Plant Science*, vol. 16, p. 1496308, August 2025.
- [10] S. P. Mohanty, D. P. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.
- [11] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, 2018.
- [12] H. Li, H. Wu, Q. Li, and C. Zhao, "A review on enhancing agricultural intelligence with large language models," *Artificial Intelligence in Agriculture*, vol. 15, no. 4, pp. 671–685, December 2025, open Access. [Online]. Available: <https://doi.org/10.1016/j.aiia.2025.100185>
- [13] A. Bera, D. Bhattacharjee, and O. Krejcar, "Pnd-net: Plant nutrition deficiency and disease classification using graph convolutional network," *Scientific Reports*, vol. 14, no. 18084, July 2024, open Access. [Online]. Available: <https://doi.org/10.1038/s41598-024-66543-7>
- [14] S. D. Meena, H. Katragadda, P. B. Sanny, K. Dande, and J. Sheela, "Using region-based deep learning algorithm mask rcnn algorithm on image-based plant disease detection," in *2024 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2024, pp. 1–6, conference Location: Coimbatore, India.
- [15] K. Nagasubramanian, S. Jones, A. K. Singh, S. Sarkar, A. Singh, and B. Ganapathysubramanian, "Plant disease identification using explainable 3d deep learning on hyperspectral images," *Plant Methods*, vol. 15, p. 98, August 2019, open Access.
- [16] M. Assaduzzaman, P. Bishshash, M. A. S. Nirob, A. A. Marouf, J. G. Rokne, and R. Alhajj, "Xse-tomatonet: An explainable ai based tomato leaf disease classification method using efficientnetb0 with squeeze-and-excitation blocks and multi-scale feature fusion," *MethodsX*, vol. 14, p. 103159, 2025, eCollection 2025 Jun. [Online]. Available: <https://doi.org/10.1016/j.mex.2025.103159>
- [17] R. Polly and E. A. Devi, "Semantic segmentation for plant leaf disease classification and damage detection: A deep learning approach," *Smart Agricultural Technology*, vol. 9, p. 100456, 2024. [Online]. Available: <https://doi.org/10.1016/j.atech.2024.100456>
- [18] L. Wang, J. Chi, F. Zhang, Y. Lu, T. Li, K. Wang, X. Wang, and X. Zhang, "Early detection and severity classification of verticillium wilt in cotton stems using raman spectroscopy and optimized machine learning," *Frontiers in Plant Science*, vol. 16, p. 1496621, September 2025.
- [19] A. S. Almadhor, H. T. Rauf, M. I. U. Lali, R. Damaševičius, B. Alouffi, and A. S. Alharbi, "Ai-driven framework for recognition of guava plant diseases through machine learning from dslr camera sensor based high resolution imagery," *Sensors*, vol. 21, no. 11, p. 3830, 2021.
- [20] Y. Wang, F. Wang, W. Chen, B. Lv, M. Liu, X. Kong, C. Zhao, and Z. Pan, "A large language model for multimodal identification of crop diseases and pests," *Scientific Reports*, vol. 15, no. 1, p. 21959, July 2025, open Access. [Online]. Available: <https://doi.org/10.1038/s41598-025-01908-0>
- [21] Q. Zeng, J. Sun, and S. Wang, "Dic-transformer: interpretation of plant disease classification results using image caption generation technology," *Frontiers in Plant Science*, vol. 15, p. 1273029, 2024. [Online]. Available: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1273029/full>
- [22] X. Hou, C. Zhang, Y. Song, T. Alghamdi, M. Aborokbah, H. Zhang, H. La, and Y. Wang, "Wheat soil-borne mosaic virus disease detection: A perspective of agricultural decision-making via spectral clustering and multi-indicator feedback," *Plants*, vol. 14, no. 15, p. 2260, July 2025, open Access.
- [23] M. Meenalochini and P. Amudha, "Gnn based cauliflower plant disease prediction using deep learning techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, pp. 859–869, 2024, ISSN: 2147-6799. [Online]. Available: <https://www.ijisae.org>
- [24] X. Jiang, J. Wang, K. Xie, C. Cui, A. Du, X. Shi, W. Yang, and R. Zhai, "Plantcafo: An efficient few-shot plant disease recognition method based on foundation models," *Plant Phenomics*, vol. 7, p. 100024, 2025, open Access. [Online]. Available: <https://doi.org/10.1016/j.plaphe.2025.100024>
- [25] M. Rezaei, D. Diepeveen, H. Laga, M. G. K. Jones, and F. Sohel, "Plant disease recognition in a low data scenario using few-shot learning," *Computers and Electronics in Agriculture*, vol. 219, p. 108812, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169924002035>
- [26] M. S. Farooq, A. Kamran, S. A. Raza, M. F. Wasiq, B. Hassan, and N. J. Herzog, "Advancing early blight detection in potato leaves through zeroshot learning," *Journal of Imaging*, vol. 11, no. 8, p. 256, July 2025, open Access.