Physical Education, Health and Social Sciences

https://journal-of-social-education.org E-ISSN: 2958-5996

P-ISSN: 2958-5988

Deep Learning for Early Detection and Prediction of Emerging Plant Diseases

Dr. Nadia Jabeen *[1], Muhammad Shahid[2], Hafiza Muqaddas Qavi Mehmood[3], Osama Akbar[4], Abdulrehman Niazi[5]

1 Department of Agriculture, Hazara University Mansehra, KPK. nadia_khan909090@yahoo.com nadia.agri@hu.edu.pk ORCID ID: 0000-0001-6617-8301

2 Department of Plant Pathology, University of Agriculture Faisalabad. shahidallahrakha7176@gmail.com

3 Department of Plant Pathology, University of Agriculture Faisalabad. muqaddasqavi@gmail.com

4 Department of Plant Breeding and genetics, University of Agriculture Faisalabad. osamashahwani0@gmail.com

5 Department of Plant Breeding and Genetics, University of Agriculture Faisalabad. abdulrehmanniazi493@gmail.com

DOI: https://doi.org/10.63163/jpehss.v3i4.1697

Abstract

Emerging plant diseases and invasive agricultural pests pose escalating threats to global food security, crop productivity, and economic stability, exacerbated by climate change-driven shifts in pathogen transmission dynamics. Traditional visual inspection methods are inherently reactive, subjective, and inadequate for pre-symptomatic detection, often resulting in delayed interventions and excessive pesticide application. This comprehensive review examines the transformative role of deep learning architectures in advancing early detection and prediction of plant diseases. We systematically analyze the evolution from foundational Convolutional Neural Networks (CNNs) to sophisticated Vision Transformers (ViTs) and hybrid CNN-Transformer architectures, evaluating their diagnostic performance, architectural innovations, and deployment suitability across agricultural contexts. Critically, we explore multimodal sensing modalities beyond conventional RGB imaging, including hyperspectral imaging for pre-visual biomarker detection, infrared thermography for vascular stress assessment, and environmental microclimate integration for predictive forecasting. The review synthesizes current advances in multimodal fusion architectures, vision-language foundation models, and few-shot learning paradigms that address persistent challenges of data scarcity, domain shift, and the pervasive "lab-to-field" performance degradation. We examine edge deployment strategies, model optimization techniques including INT8 quantization, and explainable AI frameworks essential for practical agricultural adoption. Benchmark datasets are critically assessed for their limitations and generalization dynamics. Key findings demonstrate that integrated multimodal deep learning systems incorporating spatial-spectral-temporal data fusion can achieve proactive detection 5-7 days pre-symptomatically, enabling targeted interventions that reduce prophylactic fungicide application by up to 41%. Despite remarkable progress, significant challenges remain in bridging generalization gaps, developing robust field-deployable systems, and establishing standardized evaluation protocols for real-world agricultural environments.

Keywords: Deep Learning, Plant Disease Detection, Multimodal Fusion, Hyperspectral Imaging, Precision Agriculture, Computer Vision, Early Warning Systems, Transfer Learning, Few-Shot Learning, Vision Transformers

1. Introduction

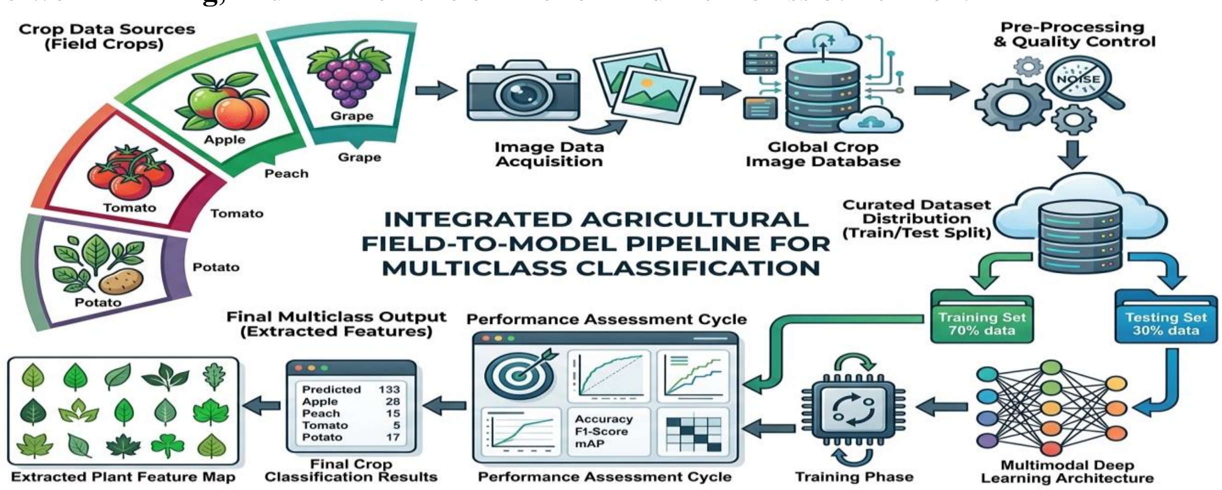
Emerging plant diseases and invasive agricultural pests pose a continuous risk to global food security, threatening crop yields, destabilizing agricultural supply chains, and causing severe economic destabilization. The intensifying impacts of global climate change manifested through rising ambient temperatures and erratic precipitation patterns have fundamentally altered pathogen transmission dynamics (Garrett et al., 2006). Warmer microclimates accelerate the reproduction rates of fungal, bacterial, and viral vectors, while shifting precipitation causes root damage and moisture stress, thereby lowering natural host plant resistance (Velásquez et al., 2018). Consequently, plant pathogens are expanding into previously non-endemic geographic regions at unprecedented rates, presenting novel diagnostic challenges to modern agriculture (Singh et al., 2023).

Historically, disease identification relied on visual inspection by agricultural experts or localized manual sampling by farmers. These traditional paradigms exhibit profound limitations. Visual diagnosis requires specialized phytopathological expertise that is rarely accessible at scale in resource-constrained or remote farming communities. Furthermore, human visual assessment is inherently subjective, prone to error, and inherently reactive (Barbedo, 2016). By the time pathological symptoms such as necrotic lesions, severe chlorosis, or leaf curling become visually conspicuous to the naked eye, the infection has typically advanced into systemic stages. This late-stage diagnosis frequently leads to the over-application of broad-spectrum chemical pesticides, driving environmental pollution, raising production costs, and hastening pesticide resistance without effectively salvaging crop yields (Rani & Khetarpal, 2015).

To address these systemic vulnerabilities, the agricultural domain has undergone a paradigm shift toward precision agriculture powered by artificial intelligence (AI), computer vision, and deep learning (DL) (Liu & Wang, 2021). Deep learning architectures eliminate the need for manual feature engineering by automatically learning complex hierarchical visual representations directly from raw multi-spectral and spatial data. Over the past decade, deep learning models have evolved from simple convolutional classifiers evaluated on curated laboratory datasets to sophisticated multimodal architectures capable of synthesizing high-resolution imagery, hyperspectral reflectance, thermal signatures, and temporal environmental data (Too et al., 2019).

The primary challenge in modern digital phytopathology is shifting from post-symptomatic recognition to pre-symptomatic prediction. Achieving true early warning capability requires understanding the complex interactions between high-dimensional deep learning architectures, physiological host-pathogen dynamics, non-visible sensing modalities, and real-world field variability (Singh et al., 2023).

Figure 1. Architectural workflow of an integrated field-to-model pipeline for crop disease data acquisition, database management, quality pre-processing, 70/30 data splitting, deep neural network training, and final feature extraction and multiclass evaluation.



2. Deep Learning Architectures in Phytopathology

2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) serve as the foundational architecture for automated plant disease identification due to their ability to process grid-structured spatial data. Standard CNNs operate via learnable convolutional kernels that slide across input images to capture local visual patterns, such as leaf margins, textural anomalies, color gradients, and lesion geometry, at escalating levels of abstraction. These convolutional layers are paired with spatial pooling operations to reduce dimensional complexity while preserving invariant spatial features, culminating in dense fully connected layers for multi-class pathology classification (Juroszek & von Tiedemann, 2013).

The evolution of CNN backbones reflects a continuous trade-off between diagnostic accuracy, network depth, and edge-device deployability. Early network implementations, such as AlexNet, introduced Rectified Linear Unit (ReLU) activations and dropout regularization to achieve automated leaf disease classification (Chakraborty & Newton, 2011). Subsequent deep architectures like VGG19 demonstrated that stacking uniform 3×3 convolutional filters allow the network to learn rich fine-grained representations. For instance, combining a pre-trained VGG19 backbone with downstream k-Nearest Neighbors (k-NN) classification achieved a 99.1% accuracy on specialized crop datasets. However, the extreme parameterization of VGG architectures (exceeding 138 million parameters) creates significant memory bottlenecks that impede deployment on resource-constrained mobile hardware or unmanned aerial vehicles (UAVs) (Bebber et al., 2014).

To overcome parameter inflation and vanishing gradient dynamics in ultra-deep networks, Residual Networks (ResNet) introduced identity skip connections that allow gradients to flow directly across layers. Compact variants like ResNet-9 and deeper structures like ResNet-50 enable stable training across deep feature hierarchies, maintaining high performance under varied parameter constraints (Savary et al., 2019). Dense Convolutional Networks (DenseNet) extended this concept by establishing direct feed-forward connections between all sequential layers. DenseNet-121 achieves near-optimal classification accuracies (up to 99.75%) on standard benchmark sets, leveraging dense feature reuse to resist overfitting even when trained on small agricultural sample populations (Mahlein, 2016).

EfficientNet introduced compound scaling, an architectural breakthrough that uniformly scales network depth, width, and input image resolution using a fixed set of scaling coefficients. Comparative benchmarks demonstrate that EfficientNet-B5 achieves top-tier accuracies (99.91% on unaugmented data and 99.97% on augmented datasets) while requiring significantly fewer floating-point operations (FLOPs) than traditional VGG or ResNet networks, making it a primary backbone for deployment in smart agriculture systems (Bebber et al., 2013).

2.2 Vision Transformers and Attention Mechanisms

While CNNs excel at extracting local visual features, their reliance on localized receptive fields limits their capacity to model long-range spatial relationships across an entire plant canopy or leaf surface. Vision Transformers (ViTs) reframe image processing by tokenizing spatial image patches and passing them through multi-head self-attention mechanisms. This structural design allows the network to evaluate global contextual relationships across disparate regions of an image simultaneously (Pimentel, 2005).

The Swin Transformer improves upon standard ViT architectures by constructing hierarchical feature maps through Shifted Windows. By restricting self-attention computations to localized non-overlapping windows while permitting cross-window connections in sequential layers, the Swin Transformer achieves linear computational complexity relative to image size. In agricultural computer vision, Swin Transformer backbones outperform conventional CNNs when handling complex multi-lesion images, overlapping foliage, and subtle background variations (Kamilaris & Prenafeta-Boldú, 2018).

2.3 Hybrid CNN-Transformer Architectures

To merge the localized feature extraction strengths of CNNs with the global contextual modeling of Transformers, hybrid architectures have emerged as a dominant paradigm in phytopathological AI. Models such as CTPlantNet and MobilePlantViT embed lightweight convolutional blocks within transformer self-attention pipelines. The initial convolutional layers extract low-level spatial features (such as edge transitions and micro-lesions), while the subsequent transformer blocks model structural canopy relationships and global disease distribution patterns. Hybrid networks achieve diagnostic accuracy competitive with large ViTs while reducing parameter counts to levels suitable for edge integration (Mohanty et al., 2016).

Table 1: Performance Comparison and Architectural Innovations of Deep Learning Backbones in Plant Pathology

Model Architecture	Structural Innovation	Benchmarked Dataset	Reported Accuracy / Performance	Deployment Suitability
AlexNet	ReLU activations, dropout regularization, GPU acceleration	PlantVillage	95.00% – 99.00%	Low (High memory footprint)
VGG19 (+ k-NN)	Deep stacking of uniform 3×3 convolutions	Custard Apple Dataset	99.10%	Low (138M parameters)
ResNet-50 / ResNet-9	Residual identity skip connections	PlantVillage / Crop Targets	99.20% – 99.53%	Moderate (Balanced compute)
DenseNet-121	Direct layer-to-layer feature propagation	PlantVillage	99.75%	Moderate (High accuracy on small data)
EfficientNet-B5	Compound scaling of depth, width, and resolution	PlantVillage (Augmented)	99.97%	High (Optimized FLOP efficiency)
CTPlantNet (Hybrid)	CNN local feature extraction + Transformer attention	Multi-crop Benchmarks	>95.00%	High (Balanced local-global modeling)
MobilePlantViT	Lightweight convolutional transformer hybrid	Resource-Constrained Sets	Competitive with ViT-B	Extreme High (Edge/Mobile optimized)

3. Beyond Visual RGB: Sensing Modalities for Pre-Symptomatic Detection

3.1 RGB Imaging Constraints and Background Bias

Standard RGB sensors (operating within the 400–700 nm visible light spectrum) form the basis of most public disease detection models. Despite their low cost and accessibility, RGB-based vision systems operate under fundamental physical limitations. RGB imagery captures surface color changes, morphological alterations, and structural necrosis. Consequently, RGB-trained models are inherently reactive; they can only register a disease after physical cellular destruction has manifested visually on the leaf lamina (LeCun et al., 2015).

Furthermore, deep models trained exclusively on RGB leaf imagery are susceptible to background bias. Controlled laboratory datasets often feature leaves excised from the plant and placed against uniform, dark posterboards. Neural networks trained on these images frequently overfit to non-pathological background features, studio illumination angles, or shadow boundaries rather than learning true physiological disease markers. When these models are evaluated in real-world field environments where variable sunlight, soil clutter, overlapping canopies, and weed species are present their diagnostic accuracy degrades sharply (Terentev et al., 2022).

3.2 Hyperspectral Imaging for Pre-Visual Biomarkers

Hyperspectral Imaging (HSI) addresses the limitations of visible-spectrum sensors by collecting narrow, contiguous spectral bands across the visible, near-infrared (NIR), and short-wave infrared (SWIR) spectra (typically 350–2500 nm). When a pathogen infects host tissue, it induces physiological alterations such as stomatal disruption, changes in internal leaf moisture, breakdown of photosynthetic pigments (chlorophyll a and b, carotenoids), and cell wall degradation days before visible symptoms appear (Mahlein, 2016).

These biochemical shifts alter the plant's spectral reflectance signature. For example, fungal pathogens disrupt leaf mesophyll structure, causing measurable shifts in NIR reflectance (700–1100 nm). Simultaneously, moisture stress induced by vascular bacterial pathogens alters specific absorption troughs in the SWIR spectrum (1400–1900 nm). Deep learning architectures designed for HSI data such as 3D Convolutional Neural Networks that process spatial-spectral hypercubes can detect pre-symptomatic infections up to 5 to 7 days before physical lesions emerge (Lowe et al., 2017).

However, HSI sensors encounter operational challenges. Point spectrometers often average spectral signatures over a localized field-of-view, which can dilute early sub-millimeter disease foci. Additionally, hyperspectral hypercubes generate high-dimensional data streams that require significant computational power for noise reduction, dimensionality reduction, and real-time spatial-spectral decoding (Laughney, 2012).

3.3 Infrared Thermography and Vascular Stress Detection

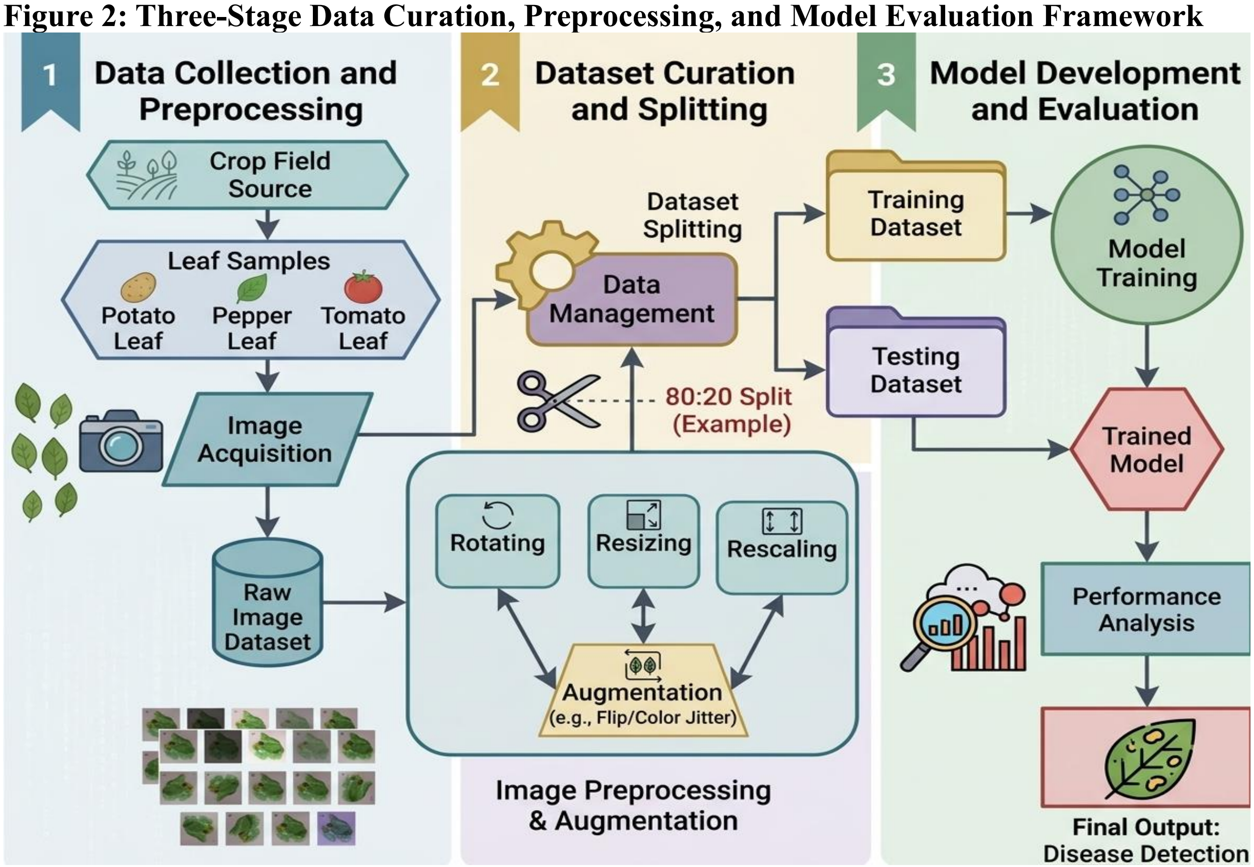
Thermal infrared imaging (operating in the 8–14 μm long-wave infrared band) measures surface skin temperature dynamics dictated by plant transpiration rates. One of the earliest host responses to pathogen attack particularly root-rot pathogens, soil-borne fungi, or vascular wilt bacteria such as *Xylella fastidiosa* is stomatal closure triggered by synthesized abscisic acid. Stomatal closure reduces evaporative cooling, resulting in a localized rise in leaf surface temperature relative to ambient air temperature (Jones, 2018).

Deep learning models integrated with thermal sensors analyze these transpiration anomalies to assess stress severity. When combined with RGB or HSI pipelines, thermal imagery provides an early indicator of vascular stress, helping distinguish between nutrient-induced chlorosis and active biotic pathogen attacks (Shakeel et al., 2022).

Table 2: Operational and Diagnostic Profile of Non-Destructive Sensing Modalities

Imaging Modality	Spectral / Thermal Range	Target Physiological / Biochemical Marker	Diagnostic Lead Time	Operational Limitations
Visible (RGB)	400 – 700 nm	Surface chlorosis, necrotic lesions, structural atrophy	Reactive (post-symptomatic visible lesions)	Zero pre-visual lead time; sensitive to field lighting and background noise
Hyperspectral (HSI)	350 – 2500 nm	Chlorophyll degradation, leaf moisture, cell wall breakdown	Proactive (5–7 days prior to visual symptoms)	High sensor cost; large data volumes; spatial averaging dilution
Infrared Thermal	8 – 14 μm	Stomatal closure, transpiration rate, leaf surface temperature	Proactive (Early physiological stress response)	Highly sensitive to ambient wind and transient weather shifts

Multimodal Sensing	Combined RGB + HSI + Thermal + Microclimate	Integrated biochemical, structural, and thermodynamic stress profiles	Maximum Proactive (Up to 7 days pre-symptomatic prediction)	Complex multi-sensor alignment, high processing overhead

4. Multimodal Data Fusion and Environmental Predictive Modeling
4.1 Architectures for Multimodal Fusion
Relying on a single sensing modality is often insufficient for robust field diagnosis under variable agricultural conditions. Multimodal deep learning architectures address this limitation by synthesizing diverse data streams such as multi-spectrum imagery, environmental sensor telemetry, and agronomic text descriptors into unified predictive frameworks. Within these systems, modality fusion typically follows four primary architectural paradigms: early, intermediate, late, and dynamic fusion (Baltrušaitis et al., 2019). Early fusion concatenates raw multi-sensory inputs or low-level feature tensors at the initial input layer, whereas intermediate fusion extracts independent feature representations from each modality branch prior to projecting them into a shared latent space (Guarrasi et al., 2025). Late fusion operates at the decision level by combining prediction probabilities from independent unimodal networks through ensemble weighting or averaging. In contrast, dynamic fusion employs adaptive gating mechanisms and attention channels to dynamically re-weight individual modalities based on real-time environmental context and signal fidelity (Zhang et al., 2026).
An exemplar of dynamic integration is the Dynamic Multi-Modal Fusion Network (DMMF-Net), which utilizes an Adaptive Fusion Module (AFM) with learned gating mechanisms to dynamically re-weight visual, textual, and microclimate inputs. DMMF-Net achieves 99% accuracy on multimodal benchmarks with low computational complexity (4.2 GFLOPs), outperforming single-modality baselines while maintaining diagnostic stability if an individual sensor modality experiences transient failure (Boulahia et al., 2021).
Figure 2: Three-Stage Data Curation, Preprocessing, and Model Evaluation Framework


4.2 Integrating Microclimate Dynamics for Disease Forecasting

Pathogen proliferation is strongly conditioned by environmental microclimates. High relative humidity, extended leaf wetness duration, and favorable temperature windows are often required for fungal spore germination and bacterial colonization. Visual models alone cannot account for these epidemiological drivers; a plant leaf may appear visually healthy despite being exposed to environmental conditions that guarantee pathogen outbreak within days (Zhao et al., 2024).

To enable true early warning capabilities, advanced deep learning frameworks combine visual sensing with real-time temporal microclimate data. These dual-branch architectures process high-resolution image patches alongside continuous streams of environmental data (ambient temperature, canopy humidity, leaf wetness, precipitation, and soil moisture). Within these dual-branch systems, temporal sequences of microclimate metrics are modeled using Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) units, or temporal Transformers.

Environment-guided visual attention modules use this microclimate embedding to dynamically adjust the attention weights of the vision backbone (Sangeetha et al., 2026). When environmental risk indices are elevated, the visual network increases its sensitivity to subtle visual anomalies. Integrating 3D-CNNs (for hyperspectral spatial-spectral analysis) with temporal Transformer weather modules yields high diagnostic accuracy (up to 94.5%) and enables disease detection 5 to 7 days before physical symptom onset. In field simulations, these predictive systems enabled targeted, site-specific chemical applications, reducing prophylactic fungicide use by up to 41% (Zhang et al., 2024).

4.3 Multimodal Vision-Language Models in Phytopathology

The adaptation of vision-language foundation models (such as CLIP, Segment Anything Model [SAM], and DINOv2) offers new possibilities for agricultural diagnostics. These models align visual domain features with textual domain representations through large-scale contrastive learning (Iqbal, 2025).

Domain-specific Vision-Language Models (VLMs), such as SCOLD, achieve high performance in disease identification (99.15%) and species classification (94.92%). Evaluation on complex datasets like LeafBench which contains 186,000 images paired with 13,950 expert-annotated visual-textual question-answer pairs demonstrates that fine-tuned VLMs outperform vision-only models by up to 27.76% on diagnostic reasoning tasks. Furthermore, multimodal few-shot frameworks (such as MMFSL) utilize language-derived text templates to guide visual Transformers, maintaining classification accuracy even when presented with extremely limited training samples (Manda et al., 2026).

5. Benchmark Datasets, Domain Shift, and Generalization Dynamics

5.1 Overview of Standard Benchmark Corpuses

The development and evaluation of deep learning models in digital phytopathology rely heavily on structured benchmark datasets, which vary substantially in scale, acquisition environment, class taxonomy, and annotation depth. The PlantVillage dataset remains the most widely cited benchmark, containing 54,303 excised single-leaf images spanning 38 class labels across 14 crop species. Captured under controlled laboratory conditions using uniform dark backgrounds and standardized illumination, PlantVillage provided an essential foundation for early CNN benchmarking (Rai et al., 2025).

To address the domain constraints of studio datasets, the PlantDoc corpus was constructed with 2,598 images covering 13 crop species and 17 disease classes, collected directly from unconstrained field environments to evaluate in-the-wild visual recognition (Saini et al., 2025). Expanding upon this scale, the PlantWild dataset encompasses 18,542 expert-validated images across 89 classes, incorporating a broad spectrum of healthy and diseased foliage captured under complex field lighting. Similarly, the Plant Disease Dataset 271 (PDD271) provides 220,592 leaf images across 271 disease categories, establishing a platform for evaluating fine-grained classification under long-tailed class distributions. More recently, the LeafBench dataset introduced a multimodal benchmark comprising 186,000 images across 97 disease classes and 22 host species, paired with 13,950 expert-curated visual-text question-answer pairs to evaluate advanced vision-language architectures (Zulfiqar et al., 2026).

5.2 The "Lab-to-Field" Performance Drop

A pervasive hurdle in operationalizing deep learning for plant pathology is the sharp drop in diagnostic accuracy observed when transitioning from lab-trained models to open-field applications. Modern neural backbones like DenseNet-121 and EfficientNet-B5 frequently achieve near-saturated accuracies between 99.7% and 99.8% on laboratory benchmarks such as PlantVillage. However, when these identical model checkpoints are evaluated on field-collected leaf samples, classification accuracy degrades precipitously to approximately 61.97% (Sperschneider, 2020).

This performance degradation is driven by several interrelated factors inherent to field environments. First, models trained on studio imagery tend to exploit background artifacts such as posterboard textures, mounting pins, or consistent shadow boundaries as shortcut features rather than learning genuine pathological symptoms. Second, natural field environments introduce high levels of environmental and illumination variance, including direct sunlight glare, dynamic canopy shadows, and color shifts caused by diurnal weather changes (Verma et al., 2020). Finally, field foliage presents structural canopy complexity, featuring overlapping leaves, soil backgrounds, non-pathological mechanical damage, pest feeding marks, and the frequent co-occurrence of multiple distinct diseases on a single leaf lamina. Bridging this generalization gap requires training frameworks that incorporate field-collected datasets, advanced data augmentation, domain adaptation techniques, and patch-based feature extraction (Assimakopoulos et al., 2024).

Table 3: Comparison of Standard Benchmark Datasets for Deep Learning in Plant Pathology

Dataset Name	Sample Size	Category / Class Coverage	Imaging Environment	Primary Limitation / Challenge
PlantVillage	54,303 images	38 classes (14 crop species)	Controlled laboratory (excised leaves, uniform background)	Severe lab-to-field domain shift; background shortcut learning
PlantDoc	2,598 images	17 disease classes (13 species)	In-the-wild (field-captured, unconstrained)	Small dataset scale; noisy web labels; class imbalance
PlantWild	18,542 images	89 classes (33 healthy, 56 diseased)	In-the-wild (multi-engine web search, validated)	High visual variance; non-standardized field framing
PDD271	220,592 images	271 categories	Mixed controlled and field settings	Extreme class imbalance; long-tailed distribution challenge
LeafBench	186,000 images	97 disease classes (22 species)	In-the-wild visual-text paired corpus	High architecture complexity needed for text-vision evaluation

6. Advanced Learning Paradigms for Data Scarcity and Emerging Diseases

6.1 Transfer Learning and Fine-Tuning Dynamics

Collecting expert-annotated phytopathological datasets for novel or rare crop diseases is labor-intensive and costly. Transfer learning addresses this data scarcity by repurposing deep networks pre-trained on large-scale datasets like ImageNet. Initial layers which capture general visual primitives such as edges, textures, and color gradients are retained, while higher-level layers are fine-tuned on crop-specific image sets. Transfer learning accelerates model convergence and enables stable feature learning on small agricultural datasets, achieving diagnostic accuracies between 93% and 99.5% across various disease classes (Hu et al., 2026).

6.2 Few-Shot and Metric-Based Learning

When an emerging pathogen enters a new geographic region, only a few physical leaf samples may be available for model training. Standard deep learning models fine-tuned on limited samples often suffer from catastrophic overfitting. Few-shot learning (FSL) addresses this challenge by learning a metric space where samples are categorized based on similarity rather than explicit class boundaries (Ghimire et al., 2026).

Metric-based FSL frameworks using Siamese Networks or Prototypical Networks trained with triplet loss functions learn to evaluate relative feature distances. Under low-data regimes on the PlantVillage benchmark, Siamese few-shot models achieve median diagnostic accuracies of 55.5% with 1 sample per class (1-shot), 80.0% with 15 samples per class, and 90.0% with 80 samples per class. This represents an 89.1% reduction in required training data compared to standard supervised fine-tuning, offering a viable strategy for rapidly identifying emerging quarantine pathogens (Prashanthi et al., 2025).

Multimodal Few-Shot Learning (MMFSL) further improves low-sample generalization by incorporating text-branch guidance. By combining a Swin-Tiny visual encoder with class-label text embeddings via bilinear metric space alignment and Model-Agnostic Meta-Learning (MAML), MMFSL achieves 86.97% accuracy under 5-way 1-shot settings and 96.33% under 5-way 5-shot settings. When transferred directly to field conditions, MMFSL maintains 74.49% accuracy in 5-way 5-shot tasks, outperforming unimodal few-shot architectures (Rahman et al., 2024).

6.3 Self-Supervised Learning

Self-Supervised Learning (SSL) reduces reliance on manual human annotations by pre-training models on unlabeled image corpuses. Unified SSL frameworks combine Bootstrap Your Own Latent (BYOL), Masked Image Modeling (MIM), and contrastive learning representations within backbones like ResNet-101 (Garrett et al., 2006).

By learning invariant visual representations from unlabeled field images, SSL-pre-trained models demonstrate high zero-shot and fine-tuned generalization. An SSL ResNet-101 model fine-tuned on the PlantDoc dataset achieved 77.82% accuracy, outperforming standard supervised pre-trained baselines. When subsequently fine-tuned on PlantVillage, accuracy reached 99.85%, demonstrating that self-supervised pre-training yields feature extractors that adapt well across both studio and in-the-wild agricultural domains (Velásquez et al., 2018).

6.4 Localized Patch Segmentation for Cross-Crop Generalization

Standard image-level classification models often struggle to generalize across different plant species infected by the same pathogen family. A model trained to recognize powdery mildew on wheat leaves may fail when presented with powdery mildew on a cucumber leaf, as the network prioritizes overall leaf morphology over lesion features (Barbedo, 2016).

To decouple pathogen detection from host-crop morphology, localized patch-based segmentation divides leaf images into dense grid patches. The network evaluates individual patches, separating healthy tissue from diseased micro-regions. Diagnostic models using lightweight Inception backbones trained on local lesion patches can calculate total leaf disease severity while achieving high diagnostic performance (up to 99.75% accuracy) on novel crop species omitted entirely from the training dataset (Saini et al., 2025).

7. Field Deployment, Edge Computing, and Real-Time Infrastructure

7.1 Model Optimization and INT8 Quantization

Deploying deep learning models on agricultural machinery, edge sensors, or mobile phones requires optimizing network parameters for real-time inference under energy and compute constraints. Full-precision 32-bit floating-point (FP32) models can be converted to 8-bit integer (INT8) precision using linear quantization. The mathematical transformation for linear parameter mapping is expressed as:

$$Q(x) = \text{round}\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \cdot (2^b - 1)\right)$$

where x represents the original floating-point tensor value, $x_{\min}$ and $x_{\max}$ define the dynamic scalar bounds of the tensor, b denotes the target integer bit-width ($b = 8$), and $Q(x)$ yields the quantized integer value (Zhao et al., 2024).

Applying INT8 quantization reduces network memory footprints by roughly 75% and lowers compute demands, allowing complex multimodal frameworks to execute inference runs in under 20 milliseconds per frame. This performance enables real-time pathogen identification on resource-constrained embedded platforms (Mahlein, 2016).

7.2 Edge Deployment on Autonomous Hardware

To support continuous field monitoring, lightened object detection architectures such as Single Shot MultiBox Detectors (SSD) and optimized YOLO variants are integrated directly into edge hardware, embedded compute modules (e.g., NVIDIA Jetson platforms), and IoT agricultural equipment (Shakeel et al., 2022).

Mounting edge computer hardware onto Unmanned Aerial Vehicles (UAVs) or tractor booms enables automated, high-throughput crop monitoring. As the autonomous vehicle navigates crop rows, the edge device captures visual and thermal imagery, executes localized patch segmentation, and flags disease hot-spots in real time. These systems feed geo-tagged disease maps directly into precision spraying systems, enabling variable-rate pesticide application that targets infected foliage while sparing healthy crop zones (Rani & Khetarpal, 2015).

7.3 Explainable AI (XAI) for Agronomic Decision Support

A major barrier to the adoption of deep learning in industrial agriculture is the "black-box" nature of deep neural networks. Agronomists and farming operators are hesitant to apply costly or localized chemical interventions based solely on an uninterpretable model output (Singh et al., 2023).

Explainable AI (XAI) techniques address this issue by visualizing the visual features driving a network's diagnostic output. Gradient-weighted Class Activation Mapping (Grad-CAM), attention-map visualization, and layer-wise relevance propagation highlight the spatial regions of a leaf image that contributed most to a classification score. By confirming that the model is making decisions based on actual pathological lesions rather than non-relevant background clutter, XAI builds trust among agronomic users and provides actionable insights for crop management (Manda et al., 2026).

Conclusion

This review has comprehensively examined the transformative potential of deep learning for early detection and prediction of emerging plant diseases, highlighting both significant advancements and persistent challenges in the field. The evolution from conventional CNN architectures to sophisticated Vision Transformers and hybrid multimodal systems has dramatically enhanced diagnostic capabilities, enabling automated, high-throughput disease identification with unprecedented accuracy. Critically, the integration of non-visible sensing modalities hyperspectral imaging for pre-visual biomarker detection and infrared thermography for vascular stress assessment has shifted the diagnostic paradigm from reactive post-symptomatic recognition to proactive pre-symptomatic prediction, offering lead times of 5-7 days before physical symptom manifestation. Multimodal data fusion architectures that synthesize visual, spectral, thermal, and environmental microclimate data have demonstrated superior diagnostic stability and predictive forecasting capabilities, with field simulations indicating up to 41% reduction in prophylactic fungicide application through targeted interventions. However, the operationalization of these technologies faces formidable hurdles. The pervasive "lab-to-field" performance drop, where near-perfect laboratory accuracies degrade to approximately 62% in real-world conditions, underscores the critical need for field-collected datasets, domain adaptation techniques, and robust augmentation strategies. Data scarcity for emerging pathogens necessitates advanced learning paradigms including few-shot learning, self-supervised pre-training, and vision-language foundation models that leverage multimodal supervision. Edge deployment challenges require continued optimization through model quantization, lightweight architectural design, and integration with autonomous hardware platforms. Furthermore, the adoption of deep learning in agricultural practice demands transparent, explainable AI frameworks that build stakeholder trust and provide actionable decision support. Future research directions should prioritize standardized benchmarking protocols for field conditions, development of generalized models capable of cross-crop and cross-pathogen transfer, and integration of predictive modeling with decision support systems for precision agriculture. Ultimately, realizing the full potential of deep learning for plant disease management requires synergistic collaboration among computer scientists, plant pathologists, agronomists, and farmers to translate technological innovations into accessible, reliable, and sustainable agricultural solutions that safeguard global food security against evolving pathogen threats.

References

Assimakopoulos, F., Vassilakis, C., Margaris, D., Kotis, K., & Spiliotopoulos, D. (2024). Artificial intelligence tools for the agriculture value chain: Status and prospects. *Electronics*, *13*(22), 4362.

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering, 144*, 52–60. https://doi.org/10.1016/j.biosystemseng.2016.01.017

Bebber, D. P., Holmes, T., Smith, D., & Gurr, S. J. (2014). Economic and physical determinants of the global distributions of crop pests and pathogens. *New Phytologist, 202*(3), 901–910. https://doi.org/10.1111/nph.12704

Bebber, D. P., Ramotowski, M. A. T., & Gurr, S. J. (2013). Crop pests and pathogens move polewards in a warming world. *Nature Climate Change, 3*, 985–988. https://doi.org/10.1038/nclimate1990

Boulahia, S. Y., Amamra, A., Madi, M. R., & Daikh, S. (2021). Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Machine Vision and Applications, 32*, Article 121. https://doi.org/10.1007/s00138-021-01249-8

Chakraborty, S., & Newton, A. C. (2011). Climate change, plant diseases and food security: An overview. *Plant Pathology, 60*(1), 2–14. https://doi.org/10.1111/j.1365-3059.2010.02411.x

Dong, J., Fuentes, A., Zhou, H., Jeong, Y., Yoon, S., & Park, D. S. (2024). The impact of fine-tuning paradigms on unknown plant diseases recognition. *Scientific Reports*, *14*(1), 17900.

Garrett, K. A., Dendy, S. P., Frank, E. E., Rouse, M. N., & Travers, S. E. (2006). Climate change effects on plant disease: Genomes to ecosystems. *Annual Review of Phytopathology, 44*, 489–509. https://doi.org/10.1146/annurev.phyto.44.070505.143420

Ghimire, S., Lamsal, R., Sankuratri, A., Lakkarsu, P., & Vutla, S. (2026). Artificial intelligence-driven plant disease detection and diagnosis: A comprehensive review of deep learning approaches, multimodal sensing technologies, and future perspectives in precision agriculture. *The Plant Pathology Journal*, *42*(2), 121.

Guarrasi, V., Aksu, F., Caruso, C. M., Di Feola, F., Rofena, A., Ruffini, F., & Soda, P. (2025). A systematic review of intermediate fusion in multimodal deep learning for biomedical applications. *Image and Vision Computing, 158*, 105509. https://doi.org/10.1016/j.imavis.2025.105509

Hu, Y., Li, H., Yang, C., Chen, N., Pan, Z., & Ke, W. (2026). Challenges and opportunities in tomato leaf disease detection with limited and multimodal data: A review. *Mathematics*, *14*(3), 422.

Iqbal, A. (2025). Comprehensive AI framework for automated classification, detection, segmentation, and severity estimation of date palm diseases using vision-language models and generative AI. *Frontiers in Plant Science*, *16*, 1710188.

Jones, H. G. (2018). Thermal imaging and infrared sensing in plant ecophysiology. In *Advances in plant ecophysiology techniques* (pp. 135-151). Cham: Springer International Publishing.

Juroszek, P., & von Tiedemann, A. (2013). Climate change and potential future risks through wheat diseases in Europe. *European Journal of Plant Pathology, 136*, 21–33. https://doi.org/10.1007/s10658-012-0147-7

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Laughney, A. M. (2012). *Localized spectroscopic imaging for the detection and discrimination of surgical breast tissue pathologies*. Dartmouth College.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: A review. *Plant Methods, 17*, 22. https://doi.org/10.1186/s13007-021-00722-9

Lowe, A., Harrison, N., & French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods, 13*, 80. https://doi.org/10.1186/s13007-017-0233-z

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors Parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease, 100*(2), 241–251. https://doi.org/10.1094/PDIS-03-15-0340-FE

Manda, T., Huang, T., Ding, Y., Dai, S., Yang, L., & Dai, T. (2026). Multimodal Deep Learning and Foundation Models for Early Detection and Forecasting of Plant Diseases. *Plants*, *15*(17), 2564.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. https://doi.org/10.3389/fpls.2016.01419

Pimentel, D. (2005). Environmental and economic costs of the application of pesticides primarily in the United States. *Environment, Development and Sustainability, 7*, 229–252. https://doi.org/10.1007/s10668-005-0709-4

Prashanthi, B., Krishna, A. V. P., & Rao, C. M. (2025). A comparative study of fine-tuning deep learning models for leaf disease identification and classification. *Engineering, Technology & Applied Science Research*, *15*(1), 19661-19669.

Rahman, Z. U., Asaari, M. S. M., Ibrahim, H., Abidin, I. S. Z., & Ishak, M. K. (2024). Generative adversarial networks (GANs) for image augmentation in farming: A review. *IEEE Access*, *12*, 179912-179943.

Rai, N., Boyd, N. S., & Schumann, A. W. (2025). Advancing site-specific disease and pest management in precision agriculture: From reasoning-driven foundation models to adaptive, feedback-based learning. *arXiv preprint arXiv:2510.24650*.

Rani, P., & Khetarpal, R. K. (2015). Pesticide resistance and environmental impacts: A global concern. *International Journal of Current Research, 7*(12), 238–242.

Saini, A., Madaan, P., Kumari, P., Panwar, S., & Kaushik, S. (2025, December). Multi-Crop Plant Disease Detection using PlantDoc. In *2025 4th International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 1657-1664). IEEE.

Sangeetha, S. K. B., Royappan, L., Deeba, K., Benjula Anbu Malar, M. B., Shivahare, B. D., & Mathivanan, S. K. (2026). Adaptive fuzzy deep learning with multimodal sensor fusion for enhanced plant disease detection. *Scientific Reports, 16*, 22302. https://doi.org/10.1038/s41598-026-50281-z

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution, 3*, 430–439. https://doi.org/10.1038/s41559-018-0793-y

Shakeel, Q., Bajwa, R. T., Rashid, I., Aslam, H. M. U., Iftikhar, Y., Mubeen, M., ... & Wu, M. (2022). Concept and application of infrared thermography for plant disease measurement. *Trends in Plant Disease Assessment*, 109-125.

Singh, B. K., Delgado-Baquerizo, M., Egidi, E., Guirado, E., Leach, J. E., Liu, H., & Trivedi, P. (2023). Climate change impacts on plant pathogens, food security and paths forward. *Nature Reviews Microbiology, 21*, 640–656. https://doi.org/10.1038/s41579-023-00900-7

Sperschneider, J. (2020). Machine learning in plant–pathogen interactions: empowering biological predictions from field scale to genome scale. *New Phytologist*, *228*(1), 35-41.

Terentev, A., Dolzhenko, V., Fedotov, A., & Eremenko, D. (2022). Current state of hyperspectral remote sensing for early plant disease detection: A review. *Sensors, 22*(3), 757. https://doi.org/10.3390/s22030757

Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture, 161*, 272–279. https://doi.org/10.1016/j.compag.2018.03.032

Velásquez, A. C., Castroverde, C. D. M., & He, S. Y. (2018). Plant–pathogen warfare under changing climate conditions. *Current Biology, 28*(10), R619–R634. https://doi.org/10.1016/j.cub.2018.03.054

Verma, N. V., Shukla, M., Kulkarni, R., Srivastava, K., Claudic, B., Savara, J., ... & Pandya, A. (2022). Emerging extraction and diagnostic tools for detection of plant pathogens: Recent trends, challenges, and future scope. *ACS Agricultural Science & Technology*, *2*(5), 858-881.

Zhang, Q., Wei, Y., Han, Z., Fu, H., Peng, X., Hu, Q., Deng, C., Xu, C., Wen, J., Hu, D., & Zhang, C. (2026). Multimodal fusion on low-quality data: A comprehensive survey. *Information Fusion*. https://doi.org/10.1016/j.inffus.2026.104437

Zhang, Y., Yang, J., & others. (2024). A comprehensive survey on deep learning multi-modal fusion: Methods, technologies and applications. *Computers, Materials & Continua, 80*(1), 1–35. https://doi.org/10.32604/cmc.2024.053204

Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. *ACM Computing Surveys, 56*(9), Article 216. https://doi.org/10.1145/3649447

Zulfiqar, B., Chaudhary, S., Dastgheer, F., & Shabbir, T. (2026). Artificial Intelligence and Machine Learning for Early Detection and Diagnosis of Plant Diseases. *Research Consortium Archive*, *4*(2), 3358-3373.