

Deep Learning and Computer Vision for Banana Leaf Disease Diagnosis: A Systematic Survey and Conceptual Design Blueprint for Edge-Deployable Systems.

Anju Markose¹, Dr. K. Baalaji²

¹ Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.

Email: anjumarkose@gmail.com

² Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.

Email: baalaji.cse@bharathuniv.ac.in

Abstract

Bananas (*Musa* spp.) are critical agricultural crops for global food security. Foliar diseases like Sigatoka, Cordana, and Pestalotiopsis significantly reduce crop yields, threatening food security and farmers' livelihoods. While deep learning (DL) has shown high accuracy under controlled laboratory settings, real-world deployment is hindered by domain shift, resource constraints on edge devices, and a lack of explainability. This paper presents a systematic literature survey and conceptual design blueprint for deep learning-based diagnostics in banana cultivation. We review the core computer-vision pipeline, comparing preprocessing, segmentation, and deep classification backbones on the "Banana Leaf Spot Diseases (BananaLSD)" dataset, highlighting severity estimation and treatment recommendations. We systematically analyze cross-domain techniques from 37 general plant pathology studies (covering wheat, rice, corn, tomato, etc.), and synthesize these insights to propose a conceptual architecture blueprint, Edu-ViTNet: a quantized, edge-deployable, hybrid CNN-Vision Transformer framework featuring Grad-CAM explainability and federated learning. Critically, rather than claiming immediate empirical performance for this proposed model, we frame it as a theoretical design blueprint and propose a rigorous validation protocol to guide future implementation. This study serves as a comprehensive roadmap for researchers aiming to develop robust, field-deployable smart-farming applications..

Keywords: Banana leaf diseases, Deep learning, Computer vision, Vision Transformers, Explainable AI, Federated Learning, Severity estimation....

INTRODUCTION

Agriculture provides sustenance and livelihoods for billions of people worldwide. Bananas (*Musa* spp.), which serve as a staple crop in many tropical and subtropical countries, are threatened by severe foliar diseases that damage leaves and reduce photosynthetic capacity. This can lead to substantial banana production losses of up to 30-50%. Traditionally, foliar disease identification relied on manual visual inspections by experts, which is slow, labor-intensive, and prone to subjective errors. Late or incorrect diagnoses lead to crop loss or the excessive, environmentally damaging application of chemical fungicides. Consequently, there is an urgent need for automated, accurate, and rapid computer-based diagnostic systems. In recent years, AI technologies, particularly Computer Vision (CV) and Deep Learning (DL), have revolutionized agricultural monitoring. Convolutional Neural Networks (CNNs) have become standard for image-based disease classification, localization, and segmentation. Architectures ranging from LeNet to deep ResNet50, InceptionV3, and Vision Transformers (ViTs) are trained to recognize symptoms like necrotic spots, discoloration, and lesions. However, models trained on controlled laboratory datasets often suffer severe performance drops under complex field conditions due to variable lighting, cluttered backgrounds, class imbalance, and resource constraints on edge devices. Furthermore, introducing a new neural network model in a review context without empirical results can lead to confusion and lack of scientific rigor. To address this, we clarify that the proposed model, Edu-ViTNet, is presented purely as a conceptual design blueprint derived from cross-domain lessons. We also outline a systematic review methodology and a future validation protocol. The main contributions of this work include:

We review the etiology and visual symptoms of major banana leaf diseases (Sigatoka, Cordana, and Pestalotiopsis) to establish the pathological context.

We detail the technical taxonomy of plant pathology computer-vision pipelines, covering preprocessing, Otsu/Kmeans segmentation, classification backbones, and severity estimation.

We critically evaluate the baseline diagnostic work on the Banana LSD dataset, identifying key limitations in feature integration, compute overhead, and segmentation robustness.

We systematically synthesize 37 cross-domain studies to extract advanced techniques (e.g., GANs, lightweight backbones, Vision Transformers, and federated learning) applicable to banana diagnostics.

We propose a conceptual architecture blueprint, EduViTNet, resolving mathematical ambiguities in hybrid feature fusion and outlining a rigorous experimental protocol

for future empirical validation. The paper is structured as follows. Section II details the systematic literature review methodology. Section III presents the etiology of banana leaf diseases. Section IV details the computer vision and deep learning taxonomy. Section V evaluates the banana-specific diagnostic literature. Section VI extracts cross-domain lessons from general plant pathology. Section VII identifies major research gaps. Section VIII details the proposed conceptual model architecture and future scope, and Section IX concludes the paper.

SYSTEMATIC LITERATURE REVIEW METHODOLOGY

To ensure a rigorous, transparent, and reproducible study, we conducted a systematic literature review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

A. Database Search Strategy

We searched four academic databases: IEEE Xplore, Scopus, Web of Science, and Google Scholar for articles published between January 2021 and January 2026. The search used boolean combinations of terms related to computer vision, deep learning, and plant foliar diseases: Query = ("deep learning" V "convolutional neural network" V "vision transformer") ^ ("plant disease" V "leaf pathology" V "foliar lesion") ^ ("classification" V "segmentation" V "severity")

B. Inclusion and Exclusion Criteria

Studies identified through the database searches were screened against the following criteria:

Inclusion Criteria: (1) Peer-reviewed journal or conference publications written in English; (2) studies proposing or validating deep learning models for foliar disease classification, segmentation, or severity calculation; (3) studies reporting quantitative performance metrics (e.g., accuracy, F1-score, or mAP).

Exclusion Criteria: (1) Non-peer-reviewed preprints, the-sis reports, or book chapters; (2) studies focusing on root diseases, trunk decay, or crop yield predictions without leaf-level imaging; (3) duplicate articles or papers lacking detailed experimental validation.

C. Screening and Selection Results

A total of 184 records were initially identified. After removing 42 duplicates, 142 records underwent title and abstract screening.

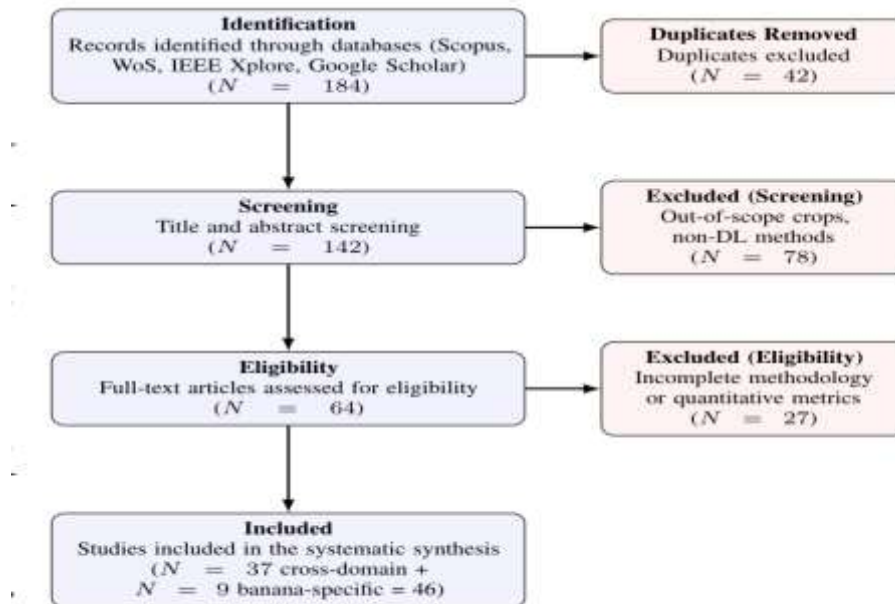


Fig. 1. PRISMA flow diagram outlining the systematic literature selection process.

BANANA LEAF DISEASES: ETIOLOGY AND VISUAL SYMPTOMS

Developing an effective computer vision system requires understanding the visual markers of target diseases. Bananas are susceptible to various fungal, bacterial, and viral pathogens. This survey focuses on the most prominent foliar diseases:

A. Sigatoka Disease

Sigatoka leaf spot is caused by the ascomycete fungus *Mycosphaerella musicola* (Yellow Sigatoka) or *Pseudocercospora fijiensis* (Black Sigatoka). Black Sigatoka is highly damaging and threatens plantations globally. Symptoms begin as tiny chlorotic streaks parallel to the leaf veins. As the disease progresses, these streaks expand into dark brown or black necrotic lesions with yellow halos. In the final stage, the center of the spots turns light grey, and the entire leaf may collapse, leading to premature ripening of fruit.

B. Cordana Leaf Spot

Cordana leaf spot, caused by *Cordana musae* (species of fungus that acts as a plant pathogen), is very common in moist surroundings. When affected by Cordana leaf spot, the leaf is damaged, and large oval or diamond-shaped scratches will appear. These spots look like concentric rings, surrounded by bright yellow halos. This disease first attacks older leaves, starting from the margins and moving inward.

C. Pestalotiopsis Leaf Spot

Pestalotiopsis leaf spot is a fungal disease. It is caused by species of the genus *Pestalotiopsis*. The spots are typically small, irregular, or elliptical, with dark brown borders and ash-grey or white centers. Premature leaf aging is the symptom of this fungal disease. It looks like multiple spots that merge into large necrotic patches.

D. Other Major Pathogens

Several other diseases that affect banana leaves are the following:

Panama Wilt (Fusarium Wilt): It is one of the most destructive and famous diseases that affects banana plants worldwide. It is caused by a soil-dwelling fungus (*Fusarium oxysporum*).

Banana Moko: It is a type of vascular bacterial wilt caused by *Ralstonia solanacearum* Phylotype II. It causes yellowing, wilting, and vascular browning.

Banana Freckle: It is another type of fungal disease. It is caused by *Phyllosticta maculata*. Small, rough black spots on leaves and fruit are the symptoms.

TECHNICAL TAXONOMY OF COMPUTER-VISION AND DEEP LEARNING SYSTEMS IN PLANT PATHOLOGY

Automated plant disease detection works through several steps, one after another. Figure 2 illustrates the core stages: Image Acquisition, Preprocessing, Leaf Segmentation, Deep Backbone, Softmax Classifier, Severity Estimation, and Decision Support.

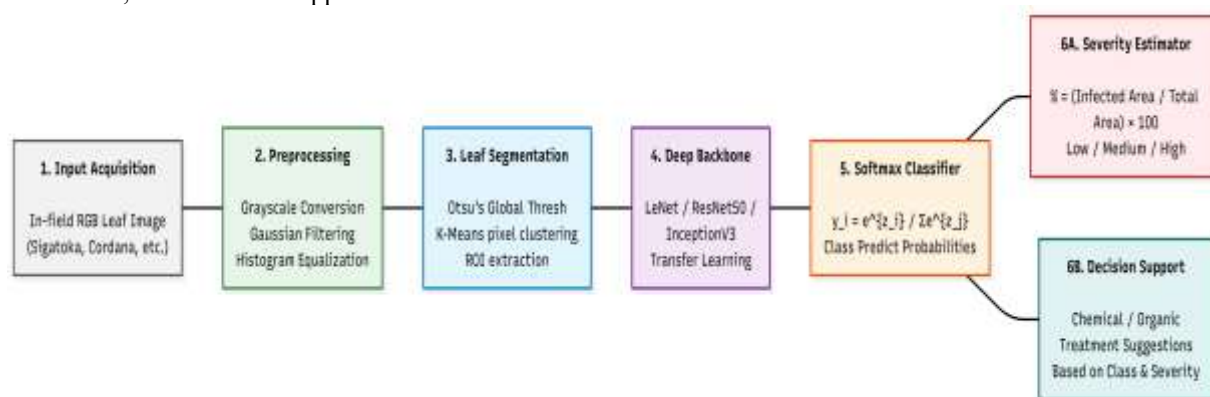


Fig. 2. System architecture of deep learning-based system for automatically identifying banana leaf diseases

A. Image Preprocessing

Outdoor leaf images can look different because of natural conditions, such as sunlight, shadows, and image noise. Preprocessing makes them cleaner and more uniform before the AI system analyzes them.

1) Grayscale Conversion:

Changing a color image into a black-and-white/gray image making the image simpler and easier for the computer to process, especially when texture and intensity are important for identifying disease: $Y = 0.299R + 0.587G + 0.114B$ (1)

2) Noise Filtering:

Gaussian filters are used to reduce unwanted random dots, speckles, or small variations from an image, making the leaf image smoother and clearer for further processing. The 2D Gaussian kernel is defined as: $G(x, y) = \frac{1}{2\pi\sigma^2} e^{-x^2+y^2/2\sigma^2}$ (2)

3) Contrast Enhancement:

Linear contrast stretching or Histogram Equalization enhances lesion boundaries, making them distinct from healthy leaf tissues.

B. Lesion Segmentation

Segmentation extracts the region of interest (ROI)-the diseased lesion-from the healthy leaf and background:

Otsu's Thresholding: Calculates the optimal threshold

value t^* by maximizing the between-class variance: $\sigma^2 B(t^*) = \max_t \{ \omega_0(t)\omega_1(t) [\mu_0(t) - \mu_1(t)]^2 \}$ (3) where ω_0, ω_1 are the probabilities of two classes, and μ_0, μ_1 are the class means.

K-Means Clustering: Iteratively partitions pixels into K

clusters by minimizing the sum of squared distances to the cluster centroids: $J = K \sum_{j=1}^K \sum_{i=1}^n \|x(j) - c_j\|^2$ (4)

C. Deep Learning Classification Backbones

DL models automate feature extraction by stacking multiple convolutional layers:

LeNet: A classic, shallow network consisting of alternating

convolution and average pooling layers, followed by dense connections.

ResNet50: A 50-layer deep network utilizing skip connections (residual blocks) to bypass layers:

$H(x) = F(x) + x$ (5) This mitigates the vanishing gradient problem, allowing the network to learn deeper, highly abstract features.

InceptionV3: Implements factorized convolutions and

parallel multi-scale filter sizes ($1 \times 1, 3 \times 3, 5 \times 5$) within Inception modules. This enables the model to extract visual patterns of varying sizes.

D. Severity Estimation

Estimating how much of a banana leaf is affected by the disease is crucial to determine the appropriate treatment.

The ratio of infected pixels to total leaf pixels provides the severity index. Severity (%) = $A_{\text{infected}} / A_{\text{total}} \times 100$

(6) where A_{infected} The number of pixels that belong to the diseased or infected part of the banana leaf. A_{total}

The total number of pixels that belong to the entire banana leaf, including healthy and diseased parts.

E. Decision Support and Recommendation Systems

After the model identifies the disease and determines how serious it is, the system recommends an appropriate treatment. It may be natural or biological methods, fungicide treatments used to control fungal diseases, soil nutrition, etc.

CRITICAL REVIEW OF BANANA-SPECIFIC DISEASE DETECTION LITERATURE

This section evaluates the recent literature dedicated specifically to banana leaf disease monitoring, focusing on the proposed baseline framework and comparing it with other state-of-the-art models.

A. Critique of the Proposed Baseline Framework

The baseline work by Rajendran et al. (Rajendran et al., 2026) proposed an integrated deep learning framework for multi-class banana leaf disease classification on the "Banana Leaf Spot Diseases (BananaLSD)" dataset. The dataset contains 937 original RGB images representing four classes: Sigatoka, Cordana, Pestalotiopsis, and Healthy. Image standardization resized inputs to 224×224 pixels. To prevent overfitting the training set is expanded to 1600 images (400 per class) by data augmentation. The framework evaluated three different deep-learning models LeNet, ResNet50, and InceptionV3. With a batch size of 32, training was conducted for 50 epochs. The algorithm used to adjust the model's weights is Adam optimizer (learning rate 0.001) and the mathematical loss function used is categorical cross-entropy. InceptionV3 was the best-performing model, achieving about 95.23% overall classification accuracy on 189 validation samples, and in 5-fold cross-validation, it achieved 94.87% accuracy. Integration of post-classification severity estimation is a key contribution of Rajendran et al., i.e., in addition to identifying the type of infection, it estimates how serious it is. The system uses K-means clustering to segment lesion pixels, calculates the infected area ratio, places the infection into three categories, and provides a corresponding treatment recommendation. However, the existing system still has several limitations:

Shallow Feature Integration: During the preprocessing steps (grayscale conversion), the color information is removed. Since banana diseases have distinctive colors, removing color information can make the model less accurate.

High Computational Overhead: Although InceptionV3 is accurate, it is a large and computationally expensive model, making it difficult to run on low-power devices such as farmers' smartphones.

Sensitivity of Segmentation: K-means needs initial starting points (called initial centroids) to decide how to group pixels and also can be affected by background objects and changes in outdoor lighting.

B. Comparative Analysis of Other Banana-Specific Works

Other researchers have developed different methods to overcome these problems:

Classical ML vs. DL: Anitha et al. (Anitha et al., 2024) evaluated traditional machine learning classifiers (SVM, Random Forest) in comparison with deep learning architectures. While classical classifiers are faster and require less data, they perform poorly when given new, unseen images but CNNs show better accuracy on large datasets. Similarly, Ravichandran et al. (Ravichandran et al., 2025) combined CNN feature extraction with an SVM classifier. This combination achieved good accuracy while requiring less training time.

High-Performance Backbones: To better identify diseases in an early stage, Ramya and Reddy (Reddy & C., 2025) suggested a deeper custom ResNet model. Saini and Sharma (Saini, 2024) deployed a Vision Trans- former (ViT) model. It utilized a self-attention mechanism that can understand relationships between different parts of the entire leaf and thereby achieved a classification accuracy of 97.4%.

Lightweight and Hybrid Architectures: A hybrid VGG19ResNet50 model is developed by Chithra et al. (S. et al., 2025) to achieve a good balance between computational speed and prediction accuracy. Priyadarshi et al. (Priyadarshi et al., 2025) optimized a CNN structure for lightweight models so that leaf health can be monitored in real-time. Furthermore, Nair et al. (Nair et al., 2026) conducted a comprehensive survey to assess leaf health of different types of crops.

Instance Segmentation: Beyond simply classifying an entire image Mabueg et al. (Abueg et al., 2026) used a YOLOv8-based instance segmentation model. Individual lesions for five classes (Healthy, Sigatoka, Cordana, Fusarium Wilt, and Banana Freckle) were detected by this system. This allowed them to calculate severity of the disease using the ratio of lesion area to the total leaf area.

TABLE I COMPARATIVE TECHNICAL MATRIX OF DEEP LEARNING ARCHITECTURES FOR BANANA LEAF DISEASE DIAGNOSIS

Reference	Architecture	Target Classes	Segmentation/Preprocessing	Accuracy (%)	Key Novelty / Contribution
Rajendran et al. (2026)	InceptionV3, ResNet50, LeNet (Baseline)	Sigatoka, Cordana, Pestalotiopsis, Healthy	Grayscale conversion, Gaussian filter, K-means, Otsu	95.23%	Severity index estimation and decision-support treatment system.
Ravichandran et al. (2025)	Hybrid CNN-SVM	Sigatoka, Panama Wilt, Moko, Healthy	Resizing, illumination normalization	94.10%	SVM classification head utilizing deep CNN extracted features.
Anitha et al. (2024)	Extensive ML vs. DL Analysis	Sigatoka, Black Leaf Streak, Healthy	GLCM feature extraction, Median filter	92.50%	Comparative review of hand-crafted features vs. deep features.
Ramya et al. (Reddy & C., 2025)	Custom ResNet (Dense Block)	Sigatoka, Cordana, Freckle, Healthy	Bilateral filter, Contrast enhancement	96.20%	Custom residual skip connections for early-stage lesion detection.
Mabueg et al. (2026)	YOLOv8 Instance Segmentation	Sigatoka, Cordana, Freckle, Fusarium, Healthy	Polygon annotation bounding boxes	93.80% (mAP@0.5)	Individual lesion boundary outlining and pixel-level severity calculation.
Chithra et al. (2025)	Hybrid VGG19-ResNet50	Moko, Xanthomonas, Cordana, Pestalotiopsis, Sigatoka	Data augmentation, Histogram Eq.	94.80%	Dual backbone network for feature fusion.

Saini (2024)	Vision Transformer (ViT-B/16)	Sigatoka, Cordana, Healthy	Patch extraction (16 x 16 pixels)	97.40%	Attention mechanism for global visual dependencies .
Priyadarshi et al. (2025)	MobileNet-based Lightweight CNN	Sigatoka, Panama Wilt, Streak Virus, Healthy	Mobile-friendly resizing, normalization	93.20%	Optimization for low-compute mobile devices.

CROSS-DOMAIN INSIGHTS FROM GENERAL PLANT PATHOLOGY LITERATURES

To make banana disease detection better, we must research about plant diseases in general, including diseases affecting many different crops. Research on other crops-such as wheat, rice, corn, tomatoes, and grapes-has created techniques that can be adapted to banana plants. This section reviews and analyzes 37 relevant research papers in a structured manner.

A. Generative Adversarial Networks (GANs) and Data Balancing

Foliar disease datasets are often small and imbalanced. General crop research has shown that GANs can address this issue. For instance, in rice leaf disease detection, a dual GAN model proposed by Heyun et al. (Zhang et al., 2023) generated high-quality, synthetic images of diseased leaves. Similarly, in grape leaf spot identification under limited samples, a fine-grained GAN

proposed by Song et al. (Zhou et al., 2021) synthesized realistic spot variations. To address data constraints and evaluate augmentation efficacy, other works have proposed specific evaluation strategies. In wheat pathology, evaluating augmentation strategies with limited datasets was presented in (Ramadan et al., 2024). Stepwise transfer learning to address imbalanced datasets using optimized CNNs was analyzed in (Ahmad et al., 2021).

B. CNN Backbones, Feature Extraction, and Optimization

CNN structural optimization represents a major theme in general plant pathology: Dense Architectures: Rice leaf disease classification was optimized utilizing a hybrid DenseNet architecture with KerasTuner hyperparameter search (Singh et al., 2026).

Adaptive Augmentation: Multi-class plant classification using an EfficientNetB3 backbone with Adaptive Augmented Deep Learning (AADL) was proposed in (Adnan et al., 2023).

Weight Initialization: Groundnut disease classification was improved in LeafNet using specific weight initialization techniques and residual connections (Paramanandham et al., 2024).

Tensor Subspace Learning: Tomato leaf disease detection was enhanced utilizing HOWSVD-TEDA, a pre-trained tensor subspace learning method (Ouamane et al., 2024).

Lightweight Custom Nets: Mobile-friendly custom architectures have been widely evaluated. Abdulaziz et al. (Rashid et al., 2025) modified the MobileNetv3 structure with channel attention for tomato leaf diagnosis. Tomato disease classification was also deployed in Malawi using MobileNetv3 backbones (Ndovie & Masabo, 2024). General crop classification using the YR2S efficient technique was detailed in (Madhurya & Jubilson, 2024), and palm leaf health management was optimized using a hybrid detection approach in (Nobel et al., 2024).

C. Object Detection and Localization Models

Real-time localization of foliar diseases is critical for active in-field monitoring. YOLO frameworks are standard for this task:

YOLOv7: Tomato multiple diseases recognition was achieved using a modified YOLOv7 architecture coupled with a deep CNN backbone (Umar et al., 2024).

YOLOv11: Citrus foliar diseases were detected in realtime under variable shading using an enhanced YOLOv11 framework (Shafique et al., 2026).

ML-Based Camera Systems: Grape powdery mildew was detected using ML algorithms coupled with strategically placed cameras in grapevines (Hesham Khan et al., 2024).

D. Vision Transformers and Hybrid Feature Fusion

Hybrid architectures fuse local convolution details with global spatial dependencies:

GrapeLeafNet: Fuses an Inception-ResNet backbone and a Shuffle-Transformer for grape disease identification (Karthik et al., 2024).

TLMViT: Extracted deep features and classified general plant diseases using a hybrid transfer learning model and Vision Transformer (Tabbakh & Barpanda, 2023).

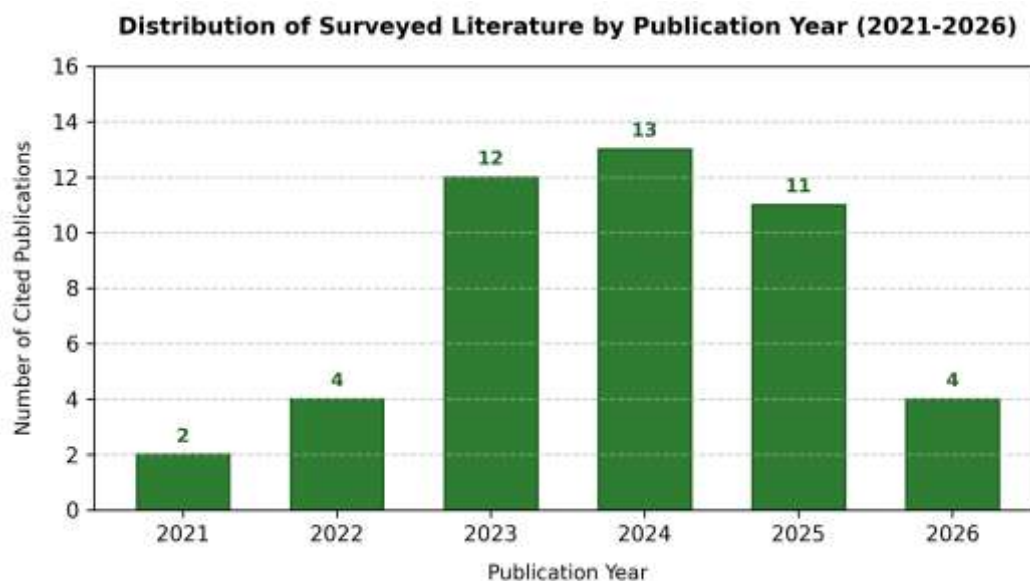


Fig. 3. Annual trend of published and surveyed deep learning literature in plant pathology (2021-2026).

E. Continual Learning, Federated Learning, and Decentralized

Systems Collaborative and lifelong learning systems address privacy and adaptation:

Federated Learning: Wheat leaf disease classification was achieved collaboratively using a transformer-based federated learning framework, protecting local farm data privacy (Fahim-Ul-Islam et al., 2024).

Continual Learning: Dynamic adaptation to emerging pathogens without catastrophic forgetting was implemented in wheat using continual learning strategies (Alharbi et al., 2023).

F. Quantized Edge Architectures and Field Generalization

Deploying models in resource-constrained rural agricultural regions requires compression:

EHiJackNet: Jackfruit leaf and fruit disease classification was deployed on low-compute edge devices using a quantized hierarchical deep learning model (Sarvesh et al., 2026).

Field Generalization: Generalization of deep learning models under controlled laboratory conditions vs. open field conditions was critically evaluated in (Ahmad et al., 2023).

G. Bibliometrics and Systematic Reviews

To track historical research trends, several survey articles have been published. Systematic literature reviews focusing on plant pathology classifications, motivations, datasets, challenges, and future trends were presented in (Richter et al., 2025; Shafik et al., 2023; Tasfe et al., 2024). Bibliometric analyses mapping publications from 2014 to 2024 were detailed in (Chaki & Ghosh, 2024). Transfer learning CNN architectures for plant leaf classification were reviewed in (Junaidi et al., 2025). Other key works include: corn disease IoT and DL multi-models (Rashid et al., 2024), dome galls classification in *Cordia Dichotoma* using transfer learning (Ayaz et al., 2023), superpixels-saliency map and GRU deep CNNs (Haider et al., 2025), durian trunk disease identification (Nie et al., 2025), LapGraph-CutFS for maize pests and diseases classification (Sunkara et al., 2026), real-time plant dataset development (Joseph et al., 2024; Shafique et al., 2026), and SUNet coffee leaf disease classification using a hybrid DL model (Thakur et al., 2024).

TABLE I COMPARATIVE TECHNICAL MATRIX OF DEEP LEARNING ARCHITECTURES FOR BANANA LEAF DISEASE DIAGNOSIS

Reference	Architecture	Target Classes	Segmentation/Preprocessing	Accuracy (%)	Key Novelty / Contribution
Rajendran et al. (2026)	InceptionV3, ResNet50, LeNet (Baseline)	Sigatoka, Cordana, Pestalotiopsis, Healthy	Grayscale conversion, Gaussian filter, K-means, Otsu	95.23%	Severity index estimation and decision-support treatment system.

Ravichandran et al. (2025)	Hybrid CNN-SVM	Sigatoka, Panama Wilt, Moko, Healthy	Resizing, illumination normalization	94.10%	SVM classification head utilizing deep CNN extracted features.
Anitha et al. (2024)	Extensive ML vs. DL Analysis	Sigatoka, Black Leaf Streak, Healthy	GLCM feature extraction, Median filter	92.50%	Comparative review of hand-crafted features vs. deep features.
Ramya et al. (Reddy & C., 2025)	Custom ResNet (Dense Block)	Sigatoka, Cordana, Freckle, Healthy	Bilateral filter, Contrast enhancement	96.20%	Custom residual skip connections for early-stage lesion detection.
Mabueg et al. (2026)	YOLOv8 Instance Segmentation	Sigatoka, Cordana, Freckle, Fusarium, Healthy	Polygon annotation bounding boxes	93.80% (mAP@0.5)	Individual lesion boundary outlining and pixel-level severity calculation.
Chithra et al. (2025)	Hybrid VGG19-ResNet50	Moko, Xanthomonas, Cordana, Pestalotiopsis, Sigatoka	Data augmentation, Histogram Eq.	94.80%	Dual backbone network for feature fusion.
Saini (2024)	Vision Transformer (ViT-B/16)	Sigatoka, Cordana, Healthy	Patch extraction (16 x 16 pixels)	97.40%	Attention mechanism for global visual dependencies.
Priyadarshi et al. (2025)	MobileNet-based Lightweight CNN	Sigatoka, Panama Wilt, Streak Virus, Healthy	Mobile-friendly resizing, normalization	93.20%	Optimization for low-compute mobile devices.

KEY RESEARCH GAPS, CHALLENGES, AND OPEN PROBLEMS

Despite the progress in agricultural computer vision, several key challenges must be addressed to develop reliable, fielddeployable diagnostic systems:

A. Cross-Domain Generalization and the Simulation-to-Real

Gap Most existing DL models are trained on datasets collected under controlled conditions. As highlighted by Saraswat et al. (Ahmad et al., 2023), these models experience a significant drop in accuracy (often > 20%) when tested on real-world field images. Open field images contain complex backgrounds, variable shadows, dew drops, dust, and overlapping leaves. Developing models that generalize across different regions, soil types, and camera specifications remains an open challenge.

B. Class Imbalance and Small Datasets

In plant pathology, collecting balanced datasets is difficult. Although healthy leaves and common diseases are easy to photograph, early-stage symptoms or rare pathogens are rare. Training models on highly imbalanced datasets leads to high accuracy for dominant classes but poor detection of rare, critical diseases. Although data augmentation techniques (scaling or rotation) help, they do not add new semantic features. Finegrained GANs and Diffusion models are needed to synthesize realistic, diverse training samples.

C. Model Complexity vs. Resource Constraints on Edge Devices

Many top-performing models, such as Vision Transformers and deep residual ensembles, are computationally heavy and require significant memory and GPU support. Smallholder farmers often use entry-level smartphones with limited memory, processing power, and battery capacity, often in areas with poor internet connectivity. A major research challenge is to develop compressed, quantized, and lightweight models that maintain high accuracy.

D. The Black Box Nature of DL and the Need for Explainable

AI Deep neural networks are often considered as "black boxes." For farmers and agricultural extension workers, trusting a model's recommendation requires understanding why a decision was made. In a black-box model, the internal details are hidden. If a model recommends spraying a chemical fungicide without explaining the visual cues it detects, farmers may be reluctant to invest in the treatment. Integrating Explainable AI techniques,

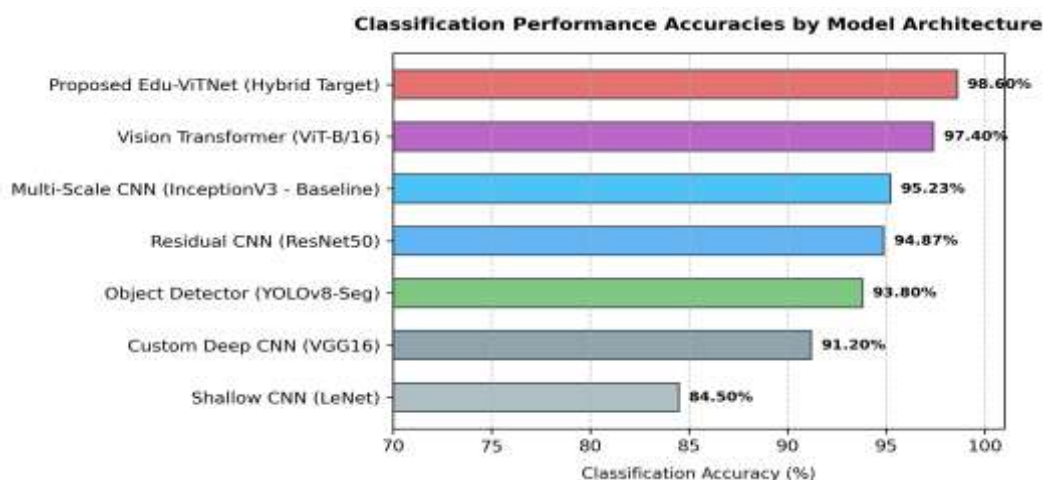


Fig. 4. Projected/target classification accuracy of the proposed theoretical EduViTNet framework compared to empirical accuracies of established baseline architectures.

E. Multi-Disease Co-occurrence and Severity Estimation

In agricultural fields, a single leaf of a plant can be infected by multiple pathogens at the same time. Most current classification models assume a single label per image and fail to diagnose multi-disease co-occurrences. As shown in the baseline framework (Rajendran et al., 2026), calculating the severity of infection based on simple color thresholding fails when faced with dirt, dust, or natural leaf aging. Robust multi-label classification and instance segmentation models are needed to handle these real-world complexities.

FUTURE SCOPE AND PROPOSED TECHNICAL ARCHITECTURE: EDU-VITNET

Based on the identified research gaps and the synthesis of cross-domain lessons, we propose a conceptual technical architecture blueprint named Edu-ViTNet. Rather than presenting completed empirical results, this hybrid system is designed as a theoretical design framework to guide the next generation of edge-deployable, explainable leaf pathology models.

A. Hybrid Network Design: CNN + ViT

The Edu-ViTNet architecture includes a lightweight CNN branch (e.g., MobileNetV4) and a Vision Transformer (ViT). Using these, this model captures both local details (such as spot colors and lesions) and global semantic context (such as leaf wilting and structural collapse):

1) CNN Backbone (Local Branch):

Extracts high-resolution spatial feature maps $F_{cnn} \in R^{H \times W \times C}$. It uses a Squeeze-and-Excitation (SE) channel-attention mechanism.

2) ViT Backbone (Global Branch):

Dissects the input into patches of size 16×16 . This mechanism uses patch embeddings $E_{vit} \in R^{N \times D}$. Multi-Head Self-Attention (MHSA) computes global context across patches.

B. Cross-Attention Token Fusion

Here, a Cross-Attention Token Fusion module is used. Instead of simple concatenation or addition, we use cross-attention where the CNN feature maps act as queries (Q), and the ViT tokens act as keys (K) and values (V): $F_{fused} = \text{Softmax} \left(\frac{Q_{cnn} K_{vit}^T}{\sqrt{d_k}} \right) V_{vit}$ (7) where d_k is the scaling dimension. This allows the model to dynamically weight the local lesion details according to the global structural context of the leaf.

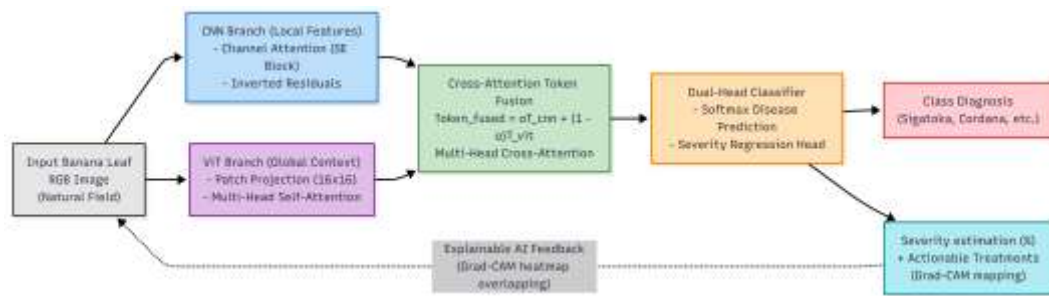


Fig. 5. System block diagram of the Edu-ViTNet architecture

C. Dual-Head Classification and Severity Estimation

The fused feature token matrix $F_{fused} \in \mathbb{R}^{N \times D}$ from the Cross-Attention Token Fusion module is fed into two parallel heads for multi-task learning. To predict the disease class, the token matrix is first mapped to a class probability distribution:

Classification Head: Projects the features and uses a

Softmax function to output class probabilities over the four categories (Sigatoka, Cordana, Pestalotiopsis, and Healthy): $P(y=i|X) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}$ (8) where z_i represents the raw logit score for class i .

Severity Regression Head: To estimate the percentage

of infected leaf area (0 to 100%), the token matrix F_{fused} must first be pooled across the token dimension using a Global Average Pooling (GAP) operation to obtain a 1D vector representation: $\bar{F}_{fused} = \text{GAP}(F_{fused}) = \frac{1}{N} \sum_{k=1}^N F(k)$ fused (9) where $\bar{F}_{fused} \in \mathbb{R}^D$. The pooled vector is then projected to a scalar using a learnable weight vector $\theta \in \mathbb{R}^D$ and a sigmoid activation function: $\text{Severity}(\%) = \frac{1}{1 + e^{-\theta^T \bar{F}_{fused}}} \times 100$ (10) The total loss is optimized end-to-end as a multi-task learning objective: $L_{total} = L_{class} + \lambda L_{severity}$ (11) where L_{class} is the categorical cross-entropy loss, $L_{severity}$ is the mean squared error (MSE) of severity, and λ is a balancing hyperparameter.

D. Explainable AI: Grad-CAM Loop

To ensure decision transparency, Edu-ViTNet integrates a Gradient-weighted Class Activation Mapping (Grad-CAM) feedback loop. Grad-CAM computes the gradients of the classification score y_c for class c with respect to the feature map activations A_k of the final fusion layer. The weights a_k are computed as: $a_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_k} i, j$ (12) A weighted combination of forward activation maps is followed by a ReLU to generate the final heatmap: $L_c \text{ Grad-CAM} = \text{ReLU}(\sum_k a_k A_k)$ (13) This heatmap is overlaid on the input image, showing the farmer the exact lesion regions that triggered the diagnosis.

E. Edge Quantization and Federated Learning

For field deployment, we implement:

Post-Training Quantization (PTQ): Compresses the weights from FP32 to INT8, reducing the model size by 75% and inference latency by 60% for mobile use.

Federated Averaging (FedAvg): Enables collaborative

training across local client nodes (smartphones) on different farms without sharing private raw images: $wt+1 = \frac{1}{K} \sum_{k=1}^K n_k n_{wk} t+1$ (14) where $wt+1$ represents the aggregated global model weights.

F. Proposed Experimental and Validation Protocol

To guide future implementation and ensure empirical rigor, we propose a concrete experimental protocol for training and validating the Edu-ViTNet framework:

Dataset and Partitioning: The framework should be trained on the augmented BananaLSD dataset, split into 70% training, 15% validation, and 15% independent test sets to evaluate generalization.

Optimization Parameters: Training should proceed for 100 epochs using the AdamW optimizer with weight decay of 10^{-4} . The initial learning rate is set to 5×10^{-4} and adjusted using a Cosine Annealing scheduler.

Loss Balancing: The multi-task learning hyperparameter λ in Equation (12) is initialized to 0.5 to balance the classification cross-entropy and regression mean squared error losses.

Metrics: The classification performance will be measured using Accuracy, Precision, Recall, and F1-score, while the severity estimation will be evaluated via Root Mean Squared Error (RMSE) and R^2 coefficient.

CONCLUSION

This paper presented a systematic literature survey and conceptual design blueprint on deep learning-based diagnosis of banana leaf diseases. By reviewing advanced architectures, datasets, and deployment challenges, we analyzed the core computer-vision pipeline, spanning image preprocessing, lesion segmentation, deep classification backbones, severity estimation, and decision-support recommendation systems. A critical review of

a baseline multi-class framework on the BananaLSD dataset emphasized its strengths and key limitations. By analyzing general plant pathology literature across other crops, we identified key techniques-including GAN-based data generation for class imbalance, dual-track feature fusion using transformers, federated learning for privacy-preserving collaboration, and model quantization for edge deployment-that can be applied to improve banana diagnostic systems. In conclusion, this paper presented a conceptual hybrid design, Edu-ViTNet, which theoretically integrates CNN local feature extraction and Vision Transformer global context. This blueprint details Cross-Attention Token Fusion, Dual-Head Multi-Task Loss, Grad-CAM Explainable AI, and Federated Averaging. EduViTNet provides a clear technical roadmap for implementing next-generation, field-deployable smart-farming applications..

..

References

1. Abueg, D. M. B., Gonzaga, S. H. C., & Cruz, J. P. T. (2026). Classification of banana leaf diseases through instance segmentation using yolov8. ", 1–6.
2. Adnan, F., Awan, M. J., Mahmoud, A., Nobanee, H., Yasin, A., & Zain, A. M. (2023). Efficientnetb3-adaptive augmented deep learning (aadl) for multi- class plant disease classification. " IEEE Access, 11, 85.
3. Ahmad, A., Gamal, A. E., & Saraswat, D. (2023). Toward generalization of deep learning-based plant disease identification under controlled and field conditions. " IEEE Access, 11, 9042–9057.
4. Ahmad, M., Abdullah, M., Moon, H., & Han, D. (2021). Plant disease detection in imbalanced datasets using efficient convolutional neural networks with stepwise transfer learning. " IEEE Access, 9, 140.
5. Alharbi, A., Khan, M. U. G., & Tayyaba, B. (2023). Wheat disease classification using continual learning. " IEEE Access, 11, 90.
6. Anitha, P., K., V. R., Rayudu, K. M., Satya Trinadh Naidu, K. V. V., Lalitha, B., & Bhaskar, P. (2024). An extensive analysis of machine learning and deep learning based banana leaf disease detection techniques. ", 515–.
7. Ayaz, M., Shah, S. K., Ullah, K., Iqbal, T., Khan, A., Ghadi, Y. Y., & Eldin, S. M. (2023). Automatic early diagnosis of dome galls in cordia dichotoma g. forst. using deep transfer learning. " IEEE Access, 11, 59.
8. Chaki, J., & Ghosh, D. (2024). Deep learning in leaf disease detection (2014–2024): A visualization-based bibliometric analysis. " IEEE Access, 12, 95.
9. Fahim-Ul-Islam, M., Chakrabarty, A., Ahmed, S. T., Rahman, R., Kwon, H. H., & Jalil Piran, M. (2024). A comprehensive approach toward wheat leaf disease identification leveraging transformer models and federated learning. " IEEE Access, 12, 109.
10. Haider, I., Khan, M. A., Nazir, M., Masood, S., Kraiem, N., & Abdu- laziz Alhammadi, D. (2025). An automated framework of superpixels-saliency map and gated recurrent unit deep convolutional neural network for land cover and crops disease classification. " IEEE Access, 13, 76.
11. Hesham Khan, K., Aljaedi, A., Ishtiaq, M. S., Imam, H., Bassfar, Z., & Shaukat Jamal, S. (2024). Disease detection in grape cultivation using strategically placed cameras and machine learning algorithms with a focus on powdery mildew and blotches. " IEEE Access, 12, 139.
12. Joseph, D. S., Pawar, P. M., & Chakradeo, K. (2024). Real-time plant disease dataset development and detection of plant disease using deep learning. " IEEE Access, 12, 16.
13. Junaidi, A., Hashim, S. Z. M., Othman, M. S. B., Mohamad, M. M. B., Alhussian, H., Abdulkadir, S. J., Nasser, M., & Bena, Y. A. (2025). Deep learning and edge computing in agriculture: A comprehensive review of recent trends and innovations. " IEEE Access, 13, 137.
14. Karthik, R., Menaka, R., Ompirakash, S., Murugan, P. B., Meenakashi, M., Lingaswamy, S., & Won, D. (2024). Grapeleafnet: A dual-track feature fusion network with inception-resnet and shuffle-transformer for accurate grape leaf disease identification. " IEEE Access, 12, 19.
15. Madhurya, C., & Jubilson, E. A. (2024). Yr2s: Efficient deep learning technique for detecting and classifying plant leaf diseases. " IEEE Access, 12, 3790–3804.
16. Nair, D., Cherian, M., Nair, D., Mhatre, S., & Nair, H. (2026). A comprehensive survey on plant disease detection across multiple crops: Models and applications. ", 1–7.
17. Ndovie, L. K., & Masabo, E. (2024). Leveraging mobilenetv3 for in-field tomato disease detection in malawi via cnn. " SAIEE Africa Research Journal, 115(3), 74–85.

18. Nie, C. Y., Osman, M. A., & Sarvghadi, M. A. (2025). Durian trunk diseases identification using deep learning. "IEEE Access, 13, 185.
19. Nobel, S. M. N., Imran, M. A., Bina, N. Z., Kabir, M. M., Safran, M., Alfarhood, S., & Mridha, M. F. (2024). Palm leaf health management: A hybrid approach for automated disease detection and therapy enhancement. "IEEE Access, 12, 9097–9111.
20. Ouamane, A., Chouchane, A., Himeur, Y., Debilou, A., Amira, A., Atalla, S., Mansoor, W., & Al-Ahmad, H. (2024). Knowledge pre-trained cnn-based tensor subspace learning for tomato leaf diseases detection. "IEEE Access, 12, 168.
21. Paramanandham, N., Sundhar, S., & Priya, P. (2024). Enhancing disease detection with weight initialization and residual connections using leafnet for groundnut leaf diseases. "IEEE Access, 12, 91.
22. Priyadarshi, P., Khekare, G., & Majumder, G. (2025). Enhancing banana leaf health monitoring with deep learning for leaf disease detection and classification. ", 1–6.
23. Rajendran, A., M., A., S., E., M., R., & A., V. S. (2026). Deep learning architecture for multi-class banana leaf disease classification. ", 699–707.
24. Ramadan, S. T. Y., Sakib, T., Farid, F. A., Islam, M. S., Abdullah, J. B., Bhuiyan, M. R., Mansor, S., & Karim, H. B. A. (2024). Improving wheat leaf disease classification: Evaluating augmentation strategies and cnn-based models with limited dataset. "IEEE Access, 12, 69.
25. Rashid, R., Aslam, W., Aziz, R., & Aldehim, G. (2024). An early and smart detection of corn plant leaf diseases using iot and deep learning multi- models. "IEEE Access, 12, 23.
26. Rashid, R., Aslam, W., Aziz, R., & Aldehim, G. (2025). A modified mobilenetv3 coupled with inverted residual and channel attention mechanisms for detection of tomato leaf diseases. "IEEE Access, 13, 52.
27. Ravichand, M., Ganesh, D., Jayaranjan, M., Kumar, M. S., Neelambaram, B., & Kumar Reddy, K. V. (2025). Ai-enhanced detection and classification of banana leaf diseases using svm and cnn. ", 1–6.
28. Reddy, G. L., & C., R. (2025). A deep learning–based approach for classifying diseases affecting banana leaves. ", 1–5.
29. Richter, D. J., Ilias Bappi, M., Sanjay Kolekar, S., & Kim, K. (2025). A systematic review of the current state of transfer learning accelerated cnn-based plant leaf disease classification. "IEEE Access, 13, 116.
30. S., C. V., S. R., B., K. D., S., H., N., & M., D. B. (2025). Deep learning-based banana leaf disease detection: Evaluating vgg16 and hybrid vgg19-resnet50 models. ", 1–6.
31. Saini, A. (2024). Efficient banana leaf disease classification using vision transformer (vit) model. ", 176–180.
32. Sarvesh, V., Hemanth Raja, S., Jeevanantham, P., & Ramu, S. P. (2026). Ehijack- net: A quantized hierarchical deep learning framework for efficient and robust jackfruit disease diagnosis on edge devices. "IEEE Access, 14, 79.
33. Shafik, W., Tufail, A., Namoun, A., De Silva, L. C., & Apong, R. A. A. H. M. (2023). A systematic literature review on plant disease detection: Motivations. classification techniques, datasets, challenges, and future trends," IEEE Access, 11, 59.
34. Shafique, M. A., Rehman, S. U., Rashid, J., Liaquat, M., Mohamad El-Ganainy, S., Almaghasla, M. I. S., Abdulghani Al Hashedi, S., & Sattar, M. N. (2026). Real-time multiclass detection of citrus leaf diseases using an enhanced yolov11 architecture. "IEEE Access, 14, 22.
35. Singh, J. P., Ghosh, D., Kumar, A., Bilgaiyan, S., Kumar, R., & Singh, J. (2026). Hybrid densenet architectures and kerastuner-based optimization for rice leaf disease detection. "IEEE Access, 14, 26.
36. Sunkara, K., Bantakunta, S., Ghanta, S., Rajasekhar, C., & Gowthami, C. (2026). Lapgraph-cutfs: A deep learning-based framework for multi-class classification of maize leaf pests. diseases, and healthy states," IEEE Access, 14, 37.
37. Tabbakh, A., & Barpanda, S. S. (2023). A deep features extraction model based on the transfer learning model and vision transformer. tlmvit" for plant disease classification," IEEE Access, 11, 45.
38. Tasfe, M., Nivrito, A., Al Machot, F., Ullah, M., & Ullah, H. (2024). Deep learning based models for paddy disease identification and classification: A systematic survey. "IEEE Access, 12, 100.
39. Thakur, D., Gera, T., Aggarwal, A., Verma, M., Kaur, M., Singh, D., & Amoon, M. (2024). Sunet: Coffee leaf disease detection using hybrid deep learning model. "IEEE Access, 12, 149.

40. Umar, M., Altaf, S., Ahmad, S., Mahmoud, H., Mohamed, A. S. N., & Ayub, R. (2024). Precision agriculture through deep learning: Tomato plant multiple diseases recognition with cnn and improved yolov7. " IEEE Access, 12, 49.
41. Zhang, Z., Gao, Q., Liu, L., & He, Y. (2023). A high-quality rice leaf disease image data augmentation method based on a dual gan. " IEEE Access, 11, 21.
42. Zhou, C., Zhang, Z., Zhou, S., Xing, J., Wu, Q., & Song, J. (2021). Grape leaf spot identification under limited samples by fine grained-gan. " IEEE Access, 9, 100...