International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Deep Convolutional Networks for Crop Disease Diagnosis: A Survey on Architectures, Feature Engineering, and Explainability for Real-Time Agricultural Deployment

Yashwardhan Bhosale[1] , Prof. S.N.Shelke[2] , Shravan Sutar[3] , Shravani Chaskar[4]

[1] *First Author: Department of Computer Engineering, Sinhgad Academy of Engineering, Pune*
[2] *Second Author: HOD-Department of Computer Engineering, Sinhgad Academy of Engineering, Pune*
[3] *Third Author: Department of Computer Engineering, Sinhgad Academy of Engineering, Pune*
[4] *Fourth Author: Department of Computer Engineering, Sinhgad Academy of Engineering, Pune*

---***---

Abstract –
This research explores the intersection of Artificial Intelligence and modern farming techniques, specifically focusing on Precision Agriculture and its potential to transform traditional crop management practices for 21st-century food security challenges. Within this broad domain, our work concentrates on leveraging deep neural networks and computer vision to diagnose plant diseases from images, bringing together expertise from machine learning, digital imaging, plant pathology, and agricultural engineering to develop practical diagnostic tools for field use. The urgency of this research is underscored by the devastating reality that plant diseases cause approximately 30-33% of global crop losses annually, while traditional expert-based visual inspection methods remain slow, expensive, subjective, and often detect problems too late to prevent significant damage.

Current deep learning approaches, despite achieving impressive accuracy, face critical limitations including high computational costs, demanding hardware requirements, lack of interpretability in decision-making processes, insufficient labeled data for rare diseases, high variability in field conditions, and the inability to provide real-time diagnostics without cloud connectivity. To address these challenges, this survey advocates for compact convolutional architectures such as MobileNetV2 and EfficientNetB0 that can be deployed on low-power devices while maintaining diagnostic accuracy..

Key Words: Precision Agriculture, Deep Learning (DL), Convolutional Neural Networks (CNNs), Lightweight Architectures (MobileNetV2, EfficientNetB0), Explainable AI (XAI), LIME, Transfer Learning, Feature Engineering, Image Segmentation, Edge Computing, Mobile Deployment.

1.INTRODUCTION

Our world faces an increasingly urgent food crisis. With the global population expected to exceed 9.7 billion by mid-century, combined with climatic volatility creating unpredictable growing conditions, keeping crops healthy has never been more critical (Jafar et al., 2024). Plant infections caused by bacteria, viruses, and fungi don't just hurt individual farmers—they ripple through entire supply chains, affecting food prices and threatening the livelihoods of millions of smallholder farmers, especially in developing nations.

The old-school approach of having agricultural experts walk through fields checking plants simply doesn't scale. Expert plant pathologists are scarce, particularly in regions where they're needed most. Even trained professionals struggle with early-stage diagnosis because many diseases look similar when symptoms first appear. Manual inspection is also inherently subjective—what one expert sees as early blight, another might classify differently. This is where AI and Deep Learning become game-changers, offering the ability to rapidly analyze subtle visual changes on plant surfaces and provide consistent, objective diagnoses (Jafar et al., 2024). These automated tools can democratize access to expert-level diagnostics, bringing sophisticated disease identification to any farmer with a smartphone.

Understanding Plant Health Issues: Biotic vs. Abiotic

When plants show signs of distress, the causes generally fall into two categories. Biotic diseases come from living organisms—the pathogens attacking the plant. These include fungi like Powdery Mildew and Anthracnose that spread across leaves, bacteria such as Bacterial Blight that cause water-soaked lesions and wilting, and viruses like Mosaic Virus that create distinctive patterns on leaves. Each type of pathogen operates differently and progresses at different rates.

On the other hand, abiotic stresses aren't caused by infections at all—they're environmental problems. Extreme temperatures, pollution, nutrient deficiencies, and water stress can all cause visible symptoms on plants. You might see yellowing (what scientists call chlorosis), dead tissue (necrosis), curled leaves, or stunted growth. Here's the tricky part: environmental stress often mimics infection symptoms visually. This is why accurate diagnostic tools need to distinguish between these causes—treating a nutrient deficiency with fungicide won't help, and vice versa (Javidan et al., 2024).

How Automated Disease Diagnosis Works

The process of automated plant disease diagnosis (PDD) follows what's called an Image Vision Recognition (IVR) pipeline. Think of it as a series of steps: first, you capture the image; then you clean it up and prepare it; next, you identify the relevant parts; after that, you extract important features; and finally, you use AI to classify what disease (if any) is present. Each step plays a crucial role in turning a simple photo into actionable diagnostic information. The modular design is actually beneficial—it means we can improve individual components as better techniques emerge without rebuilding the entire system.

2.) Related Work: How Current Systems Approach Plant Disease Detection

2.1 The Complete Pipeline from Field to Diagnosis

Capturing and Preparing Images

Getting good training data starts with image collection, typically from smartphones, high-resolution cameras, or even drones flying over fields. Many researchers use established datasets like PlantVillage.Modern systems are exploring beyond standard RGB photography—multispectral and hyperspectral imaging can capture disease signatures invisible to the human eye. The widespread availability of smartphones has made image capture accessible globally, though this convenience introduces challenges with inconsistent image quality.

Once images are captured, preprocessing becomes essential. Raw field photos need standardization—they're resized to consistent dimensions (typically 224×224 or 256×256 pixels), cleaned of noise using filters like Gaussian or median filtering, and often converted to different color spaces (Ahmed et al., 2022). The HSV color space, for example, separates color information from lighting variations, making disease symptoms stand out more clearly. The Hue component is particularly valuable since it captures the color characteristics associated with disease manifestation.

The CLAHE technique enhances contrast locally by adjusting pixel intensity within small patches, followed by smoothing across tile boundaries to maintain image consistency (Gonzalez & Woods, 2018). This is especially crucial for field images where lighting is unpredictable—think shadows from leaves or passing clouds. Without this preprocessing, subtle disease spots might be invisible to the model.

Isolating the Problem Areas

Segmentation is about separating what matters (the infected leaf tissue) from what doesn't (soil, shadows, healthy parts). Getting this right can boost classification accuracy by 10-15% simply by eliminating confusing background information.

Traditional segmentation approaches include thresholding (where Otsu's method automatically finds the best cutoff point), K-means clustering (grouping similar-colored pixels), and edge detection using Canny or Sobel operators. These methods are computationally cheap and fast, but they struggle with real-world complexity. When diseased regions have similar colors to backgrounds, or when you have multiple disease stages in one image, these simple mathematical rules often fail.

Deep learning-based segmentation offers a more robust solution. Architectures like U-Net, Mask R-CNN, and DeepLabV3+ can classify every single pixel (semantic segmentation) or identify individual disease lesions as separate objects (instance segmentation). These models achieve impressive Intersection over Union (IoU) scores exceeding 0.90, meaning they're very precise about where disease boundaries actually are.

2.2 Extracting Meaningful Information: Features That Matter

Traditional Hand-Crafted Features

Before deep learning dominated the field, researchers manually designed features based on domain knowledge. This required understanding both plant pathology and image processing.

Color-based features capture statistical properties—mean, maximum, median, and standard deviation—from different color spaces (RGB, HSV, HSI). These quantify visual changes like yellowing or tissue death. More advanced approaches use color moments, which describe the distribution's skewness and kurtosis, or color correlograms that capture how colors relate spatially (Javidan et al., 2024).

Texture properties such as contrast or entropy can be quantified through statistical matrices like GLCM, which capture spatial correlations among pixel intensities (Javidan et al., 2024). Local Binary Patterns (LBP) and Gabor filters offer alternative ways to represent texture that work well for capturing repetitive patterns caused by pathogen growth.

Shape descriptors measure geometric attributes—area, perimeter, circularity, aspect ratio. Hu moments provide shape descriptions that don't change with rotation, while Fourier descriptors characterize complex lesion boundaries. The compactness ratio (perimeter squared divided by area) helps distinguish regular fungal spots from irregular bacterial lesions (Javidan et al., 2024).

Deep Learning's Automatic Feature Discovery

Convolutional Neural Networks changed the game by learning features automatically from raw pixels. Instead of manually designing features, CNNs learn hierarchical representations through their layers—early layers detect simple edges and textures, while deeper layers recognize complex semantic concepts (Shelke et al., 2024). This automatic learning has proven superior to hand-crafted features, especially for complex real-world scenarios where defining optimal features manually is nearly impossible.

2.3 Classification Approaches: From Traditional ML to Modern Deep Learning

Classical Machine Learning Methods

Support Vector Machines, Random Forests, Decision Trees, K-Nearest Neighbors, and Artificial Neural Networks represent the traditional toolkit. SVMs with radial basis function kernels work particularly well for binary classification problems. Random Forest's ensemble approach provides some protection against overfitting and handles multi-class scenarios reasonably well, though performance typically plateaus around 85-92% accuracy. These methods shine with smaller datasets (500-2000 images) but depend heavily on the quality of hand-crafted features you feed them.

Convolutional Neural Networks: The Modern Standard

CNNs dominate image classification because they excel at capturing spatial patterns and generalizing to new data.

Heavy-duty models like VGG19 and ResNet achieve outstanding accuracy—VGG19 can hit 99.48% for tomato disease classification. However, there's a catch: VGG19's model size reaches 76.48MB and requires 40.05 million floating-point operations, making it impractical for mobile devices (Ahmed et al., 2022). ResNet variants with 50-152 layers need substantial GPU resources and take over 100 milliseconds to process a single image on smartphones. These architectures were designed for data centers, not farm fields.

Efficient architectures like MobileNetV2 and EfficientNetB0 change the equation entirely. MobileNetV2 achieves comparable accuracy (99.30%) while drastically cutting computational overhead to 4.87 million floating-point operations (Ahmed et al., 2022). Its 9.60MB size means it runs on devices with just 2GB of RAM. EfficientNetB0 takes a different approach, using compound scaling to uniformly adjust depth, width, and resolution, achieving peak accuracies up to 99.69% (Nigar et al., 2024). These models process images in under 50 milliseconds on modern smartphones, making real-time diagnosis feasible.

2.4 What Makes Lightweight Models Efficient

Understanding why models like MobileNetV2 work so well on mobile devices requires looking at their core innovations.

Depthwise Separable Convolution is the key breakthrough. This method isolates spatial and channel-wise filtering, substantially decreasing computation compared to standard convolutions (Sandler et al., 2018). Instead of processing all channels together, it first applies a separate filter to each input channel (depthwise), then uses 1×1 convolutions to combine the results (pointwise). For a 3×3 kernel, this reduces computational cost by roughly 9 times.

ReLU6 activation functions—defined as min(max(0, x), 6)—provide bounded outputs that improve numerical stability, especially important when quantizing models to 8-bit or lower precision for deployment on resource-constrained hardware (Ahmed et al., 2022).

Inverted residual blocks in MobileNetV2 expand features in intermediate layers while keeping inputs and outputs narrow, improving gradient flow and representational capacity without sacrificing efficiency (Sandler et al., 2018).

2.5 System Architecture in Practice

Practical deployment architectures follow modular designs to support real-time farm use. The data flow moves sequentially: sensors or users capture data, which flows to image processing modules (applying CLAHE and segmentation), then to the deep learning model for classification, followed by XAI generation (using LIME), and finally to output interfaces like mobile applications (Jafar et al., 2024). This clear separation allows independent optimization of each component.

For deployment, the architecture loads a compressed and optimized DL model (like MobileNetV3-small, with parameters reduced to approximately 0.93 million through quantization) directly onto mobile devices for fast, on-device inference (Damaševičius & Maskeliūnas, 2024). Supporting both online and offline operation modes is crucial for rural areas with limited connectivity.

3.) Advantages, Limitations, and Where We're Heading

3.1 Comparing Different Approaches

Deep learning models deliver superior performance through automated feature learning and strong generalization capabilities, often achieving accuracies above 99%. However, they're data-hungry—typically requiring 10,000+ labeled images—and computationally demanding. Training these models needs substantial GPU resources, taking anywhere from several hours to days depending on dataset size and model complexity. The labeling process itself creates a bottleneck, requiring expert knowledge and significant time investment. Plus, their "black-box" nature makes it hard to understand why they make specific predictions (Shelke et al., 2024).

Traditional machine learning models offer simplicity—they train faster on smaller datasets (500-2000 images suffice), completing training in minutes on standard CPUs. But they top out around 85-92% accuracy and depend entirely on the quality of hand-crafted features. The feature engineering process demands domain expertise and iterative experimentation, making it time-consuming and potentially suboptimal (Javidan et al., 2024).

The efficiency gap between heavy and lightweight deep learning models is striking. MobileNetV2 matches VGG19's accuracy (99.30% vs. 99.48%) while requiring approximately 8 times less memory (9.60MB vs. 76.48MB) and 8 times fewer computations (4.87 vs. 40.05 million floating-point operations). This efficiency enables all-day battery operation and reduces thermal constraints on mobile devices (Ahmed et al., 2022).

3.2 Future Research Directions

Several persistent challenges need addressing for widespread deployment:

Making Models More Efficient: Ongoing work on quantization (reducing precision from 32-bit to 8-bit or lower) and compression techniques like pruning and knowledge distillation can further minimize model sizes. Neural Architecture Search techniques could automatically discover optimal lightweight architectures tailored to specific deployment constraints (Damaševičius & Maskeliūnas, 2024).

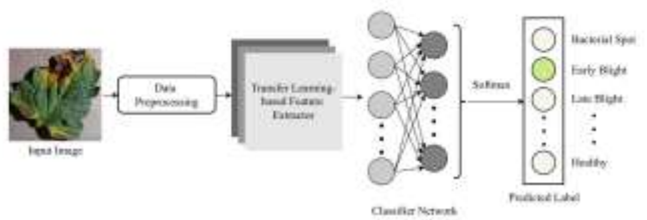
Handling Multiple Problems Simultaneously: Current systems typically identify one disease at a time, but real-world plants often face multiple challenges concurrently—diseases, pests, and nutrient deficiencies don't occur in isolation. Future systems need multi-label prediction capabilities and quantitative severity assessment through pixel-level segmentation. Disease progression modeling using temporal analysis of image sequences could enable predictive diagnostics, catching infections before visible symptoms appear (Javidan et al., 2024).

Building User Trust: Explainable AI methods like LIME and gradient-based visualization techniques such as Grad-CAM provide transparency into model decision-making, fostering necessary user trust for agricultural adoption (Nigar et al., 2024). However, explanations must be presented in agronomically meaningful terms rather than technical jargon to be useful for farmers.

4.) System Architecture Overview

A practical PDD system connects multiple components working together:

Component	Purpose	Technologies Used
Data Capture	Get images from the field	Smartphones, IoT sensors, drones, high-resolution cameras
Image Enhancement	Improve quality and focus on problem areas	CLAHE, color space conversion, segmentation, normalization
Disease Classification	Identify the specific disease	Lightweight CNNs (MobileNetV2, EfficientNetB0)
Explanation Generation	Show why the model made its decision	XAI modules (LIME, Grad-CAM)
User Interface	Deliver results and recommendations	Mobile apps, cloud reporting, SMS alerts



5.) Our Proposed Approach: Combining Efficiency with Transparency

Based on our analysis of current research, we propose a hybrid architecture that balances maximum accuracy with practical deployability on resource-constrained platforms. The system uses MobileNetV2 or EfficientNetB0 as the feature extraction backbone.

How the System Works

Step 1: Preprocessing and Data Augmentation

Input images first undergo illumination normalization using CLAHE. During training, we apply continuous data augmentation—rotating images by ±20 degrees, shifting by ±10%, horizontally flipping, and zooming by ±15%. This significantly improves the model's ability to handle real-world field variability while preventing data leakage (Ahmed et al., 2022). Augmentation effectively multiplies training dataset size by 10-20 times, crucial when labeled data is limited.

Step 2: Transfer Learning Foundation

Rather than training from scratch, we use pre-trained convolutional layers from the lightweight backbone as our starting point. These layers already know how to recognize generic visual patterns from being trained on massive datasets like ImageNet (1.2 million images). Empirical evidence suggests transfer learning can markedly shorten training duration and enhance accuracy, particularly when disease-specific datasets are limited (Pan & Yang, 2010). This approach typically reduces training time by 60-70% and improves accuracy by 5-10% compared to training from scratch.

Step 3: Specialized Classification Head

The robust feature vector (typically 1280-dimensional for MobileNetV2) feeds into a smaller, application-specific classification head with shallow dense layers (for example, 512 → 256 → number of disease classes). This head trains from scratch to specialize the general features for our specific plant disease classes—say, 38 distinct conditions. We include dropout layers (rate 0.3-0.5) to prevent overfitting to limited disease-specific training data (Ahmed et al., 2022)

Step 4: Explainable Predictions

Each final prediction pairs with a LIME-generated explanation. This module visually highlights which image regions—like specific lesion spots—most influenced the model's decision, providing essential diagnostic context that builds user confidence (Nigar et al., 2024). LIME works by perturbing the input image and observing how classifications change, identifying critical regions. The superpixel-based visualization makes it intuitive even for non-technical users to understand.

This integrated framework combining efficiency-optimized features with transparent decision-making yields a robust, low-latency system suitable for real-time mobile deployment. Applications like 'PlantCare' demonstrate that end-to-end inference (from image capture to explained diagnosis) can complete in under 2 seconds on mid-range smartphones, enabling practical field usage (Nigar et al., 2024).

6.) Conclusion:

Shifting from traditional visual inspection to AI-driven diagnosis is essential for sustainable agricultural production, especially given that plant diseases threaten one-third of global crop yields annually. Our survey reveals that achieving truly deployable plant disease detection solutions requires prioritizing lightweight deep learning models—specifically MobileNetV2 and EfficientNetB0—which deliver near-state-of-the-art accuracy (approaching 99.69%) while maintaining computational footprints small enough for mobile deployment (Ahmed et al., 2022; Nigar et al., 2024).

Incorporating Explainable AI, exemplified by LIME, is not optional—it's essential for overcoming the "black-box" problem and providing users with prediction traceability and validation. This transparency isn't just a nice feature; it's fundamental for farmer adoption and trust. Looking forward, research must concentrate on multi-disease detection, quantitative severity assessment, and improving model robustness to environmental noise for seamless translation from laboratory success to autonomous field operation. Combining lightweight architectures, explainability, and edge computing represents our most promising path toward democratizing advanced plant disease diagnostics globally.

7.) Future Research Opportunities

The persistent challenge of bridging the gap between laboratory performance and field deployment needs urgent attention. Models trained exclusively on controlled laboratory images often struggle with complex field conditions—occlusion by other leaves, variable lighting throughout the day, and cluttered backgrounds all hurt performance (Javidan et al., 2024). Here's where future work should focus:

Improving Training Data Quality and Quantity

Generative techniques like GANs (Generative Adversarial Networks) and Diffusion Models could synthetically expand limited datasets, particularly valuable for rare diseases affecting less common crops (Shelke et al., 2024). These approaches generate photorealistic images with controlled disease characteristics, addressing class imbalance problems. Self-supervised learning approaches could leverage vast amounts of unlabeled agricultural imagery to improve feature representations without requiring expensive expert labeling.

Optimizing for Edge Devices

Continuing post-training model compression and quantization work (reducing MobileNetV3-small parameters to approximately 0.93 million through 8-bit quantization) enables high-accuracy performance on low-power IoT and mobile devices (Damaševičius & Maskeliūnas, 2024). This minimizes bandwidth and computational demands for portable field solutions. Hardware-aware neural architecture search could design models specifically optimized for agricultural edge devices. Federated learning approaches might enable collaborative model improvement across farms while preserving data privacy.

Advanced Diagnostic Capabilities

Moving beyond simple disease classification to support multi-label predictions (diagnosing several co-occurring disorders, pests, or deficiencies simultaneously) represents a critical advancement. Integrating pest detection, weed identification, and crop growth monitoring would create comprehensive agricultural intelligence systems. Temporal modeling using recurrent architectures or temporal convolutions could track disease progression and predict future spread patterns, enabling proactive rather than reactive interventions (Javidan et al., 2024).

Making Explanations Actionable

Evolving XAI outputs from simple visual maps to integrated decision-support systems that provide specific, non-technical treatment recommendations and dosage information directly to farmers would greatly increase practical value (Nigar et al., 2024). Natural language generation could convert visual explanations into plain-language diagnostic reports. Integrating with agricultural knowledge bases and weather data could provide context-aware treatment recommendations optimized for local conditions, crop varieties, and regulatory constraints.

8.) References

1. Jafar A, Bibi N, Naqvi RA, Sadeghi-Niaraki A and Jeong D (2024) Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. Front. Plant Sci. 15:1356260.
2. Ahmed S, Hasan MB, Ahmed T, Sony MRK, And Kabir MH (2022) Less Is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification. IEEE Access, 10, 68868-68884.
3. Nigar N, Faisal HM, Umer M, Okid O, And Lukose JM (2024) Improving Plant Disease Classification With Deep-Learning-Based Prediction Model Using Explainable Artificial Intelligence. IEEE Access, 12, 100005-100014.
4. Shelke SN, Yadav MP, & Salve S (2024) A Systematic Review of AI Driven Crop Disease Detection using Deep Learning: Analysis, Research Challenges and Future Trends.
5. Javidan SM, Banakar A, Rahnama K, Vakilian KA, Ampatzidis Y (2024) Feature engineering to identify plant diseases using image processing and artificial intelligence: A comprehensive review. Smart Agricultural Technology, 8, 100480.
6. Damaševičius, R., & Maskeliūnas, R. (2024). High-accuracy plant disease classification on edge computing devices using MobileNetV3-small.
7. Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing (4th ed.). Pearson