

Image Processing and Machine Learning Approaches for Leaf Disease Identification: A Survey

Arshdeep Singh¹, Meenakshi Arora², Rohini Sharma³

¹ P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana

² Assistant Professor, of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

³ Assistant Professor, CS, GPGCW, Rohtak, Haryana, India

Abstract:

Identification of leaf diseases is essential for maintaining plant health and raising agricultural output. Plant leaf diseases can greatly lower crop losses and improve yield quality when detected early and accurately. Combining machine learning algorithms with image processing techniques has become a potent tool for automated disease diagnosis and tracking in recent years. This survey provides a thorough summary of the current approaches that use machine learning and image processing to detect leaf diseases. The accuracy, computational efficiency, and scalability of several supervised and unsupervised learning methods, and deep learning models are examined and contrasted. In addition to discussing potential future paths for improving resilience and real-time application, the research addresses important issues such complicated backgrounds and fluctuating illumination conditions. Researchers and practitioners attempting to create intelligent plant disease diagnosis systems might use this survey as a reference.

Keywords: Leaf Disease Detection, Image Processing, Machine Learning, Deep Learning.

INTRODUCTION: Numerous economies rely heavily on the agricultural sector, particularly in developing nations where a sizable section of the populace makes their living from farming. However, crop health and food security are seriously threatened by plant diseases, especially those that harm leaves. In order to minimize financial losses and guarantee sustainable agriculture, it is imperative that these diseases be identified promptly and accurately [1]. Agricultural specialists do manual inspection as part of traditional plant disease detection techniques, which is labor-intensive,

time-consuming, and susceptible to human error. As a result, there has been a lot of interest lately in the use of automated methods that make use of image processing and machine learning (ML).

The first step in the automation pipeline is image processing, which makes it possible to extract pertinent visual characteristics from digital photos of leaves. These characteristics—like color, texture, and shape—are crucial for identifying and differentiating between different kinds of disease [2]. To improve image quality and identify regions of interest, methods such as segmentation, noise filtering, histogram equalization, and grayscale conversion are frequently used [3].

These characteristics are subsequently analyzed by machine learning models, which categorize the leaf photos into several illness groups. Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests are examples of classical machine learning techniques that have been effectively used for illness classification [4]. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in automatically learning hierarchical features from raw picture data, frequently surpassing conventional methods, with the advent of deep learning [5]. By offering substantial collections of tagged leaf photos from a variety of crops and illnesses, several publicly accessible datasets, like the PlantVillage dataset, have sped up research [6]. Notwithstanding these developments, issues including lighting differences, backdrop complexity, intra-class variability, and restricted field deployment continue to exist. Research is still being done to address these problems.

Research Background:

Over the past 20 years, automated leaf disease detection research has advanced dramatically, moving from basic image analysis techniques to extremely sophisticated, learning-based systems. In order to detect disease symptoms on leaves, rule-based and thresholding approaches were the main focus of early research in this field [7]. Color and texture thresholds were manually defined. These techniques were computationally cheap, but they weren't reliable for different plant species and lighting circumstances. Researchers created custom feature-based machine learning models to get around the drawbacks of rule-based systems. These methods extracted disease-related texture and color patterns using descriptors including color histograms, Local Binary Patterns (LBP), and Gray-Level Co-occurrence Matrix (GLCM) [8]. Then, using conventional machine learning techniques like SVM, Random Forest, and k-NN, the retrieved characteristics were categorized. These models were more adaptable, but they frequently needed a lot of feature engineering and were sensitive to background noise.

With the advent of deep learning, and in particular Convolutional Neural Networks (CNNs), which greatly improved feature learning and model generalization, the paradigm changed. By directly learning hierarchical representations from images, CNNs removed the requirement for human feature extraction. CNNs could attain over 99% accuracy on benchmark datasets like PlantVillage, according to studies like Ferentinos (2018) [9]. These outcomes did not always translate well to actual field situations, either, as they were frequently obtained in controlled settings with clear, well-lit photographs. Recent work has explored transfer learning, data augmentation, and ensemble learning to improve performance in noisy environments [10]. Simultaneously, lightweight CNN architectures and mobile deployment frameworks (e.g., TensorFlow Lite, PyTorch Mobile) are being developed to make disease detection tools accessible to farmers via smartphones [11]. The current research is focused on improving robustness, explainability, and real-world applicability of detection systems. Hybrid approaches combining image processing, deep learning, and expert knowledge are also being explored to bridge the gap between lab performance and field deployment.

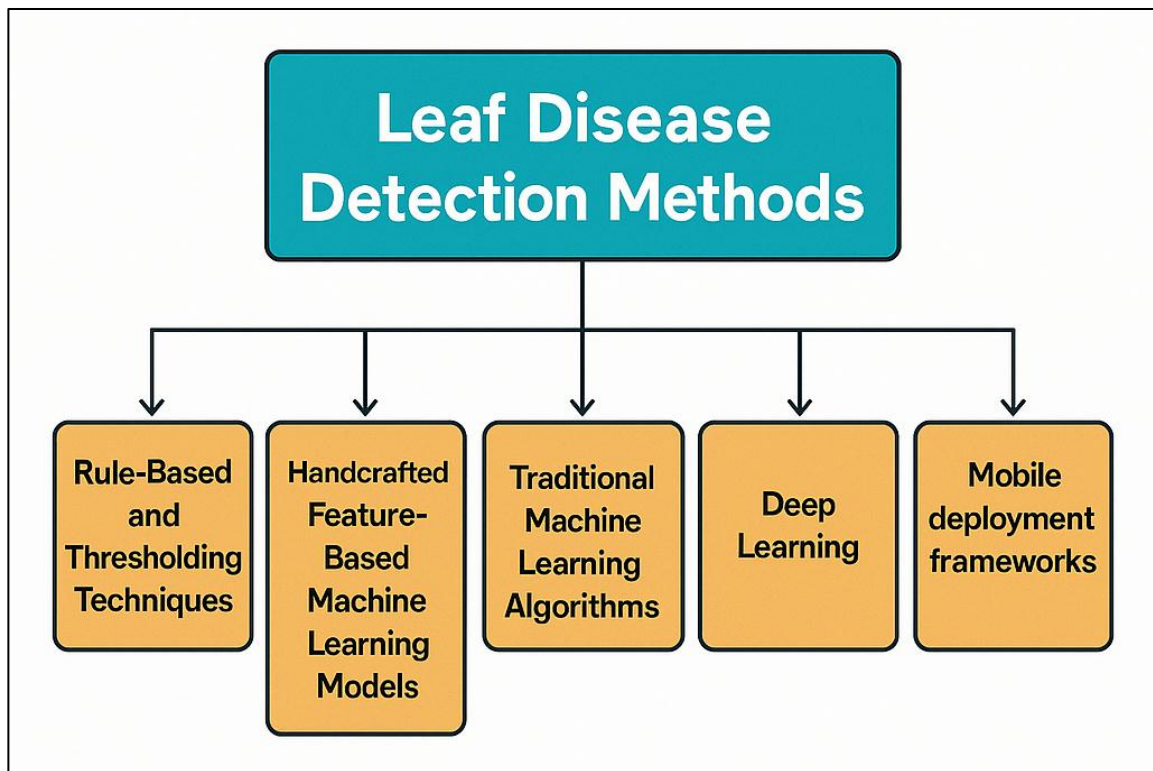


Figure 1: Classification of Leaf Disease Detection Methods

CLASSIFICATION OF LEAF DISEASE DETECTION METHODS

Among the oldest and most basic approaches utilized in image-based leaf disease detection systems are rule-based and thresholding strategies. These methods separate disease-affected areas from healthy ones by applying preset guidelines or fixed thresholds to particular picture characteristics like color, intensity, or texture.

1. Rule-Based Techniques: Rule-based methods use heuristic if-then rules defined by domain experts to detect symptoms of disease. These rules may depend on observable patterns in leaf images, such as: Color changes (e.g., brown or yellow patches), Shape irregularities (e.g., circular spots or edges) and Texture patterns (e.g., rough or dry areas). If the pixel color falls within a certain RGB or HSV range typical of a disease symptom, it may be marked as infected.

2. Thresholding Techniques: Thresholding is a fundamental image segmentation method used to separate the object of interest (e.g., diseased region) from the background. It involves setting a fixed or adaptive threshold to divide pixel values into two or more categories (e.g., healthy vs diseased). There are various types of thresholding techniques: Global Thresholding: A single threshold value is used across the entire image. Adaptive Thresholding: Different thresholds are applied to different parts of the image, useful for uneven lighting. Otsu's Method: An automatic technique that computes an optimal threshold by minimizing intra-class variance.

Rule-based and thresholding techniques laid the foundation for automated plant disease detection. While they are no longer state-of-the-art, they are still valuable for simple applications, low-resource environments, and educational purposes. Their main limitation is the lack of adaptability and robustness, which led to the evolution of more sophisticated machine learning and deep learning-based methods.

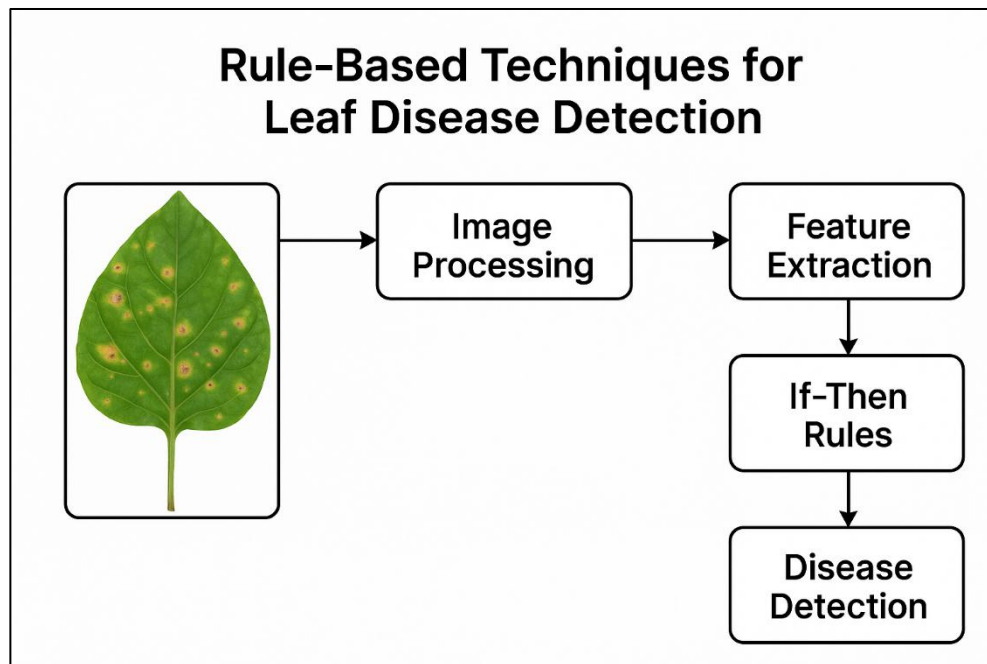


Figure 2: Rule Based Disease Detection

3. Handcrafted Feature-Based Machine Learning Models for Leaf Disease Detection:

Handcrafted feature-based machine learning refers to traditional image classification approaches where domain-specific features are manually designed and extracted from images before being fed into machine learning (ML) classifiers. In the context of leaf disease detection, this method involves identifying visual patterns related to diseases—such as color, shape, texture, and edge information—and using them to train ML models to distinguish between healthy and diseased leaves. High-resolution images of healthy and diseased leaves are captured using cameras or smartphones. Noise reduction, contrast enhancement, and segmentation techniques (e.g., thresholding, edge detection) are applied to isolate the leaf and enhance features. Feature Extraction (Handcrafted): This is the core step where specific features are manually designed or chosen based on domain knowledge. Dimensionality reduction techniques like PCA (Principal Component Analysis) or feature ranking may be used to eliminate redundant or irrelevant features. Selected features are used to train traditional ML classifiers, such as: SVM, RF, KNN and Naive Bayes.

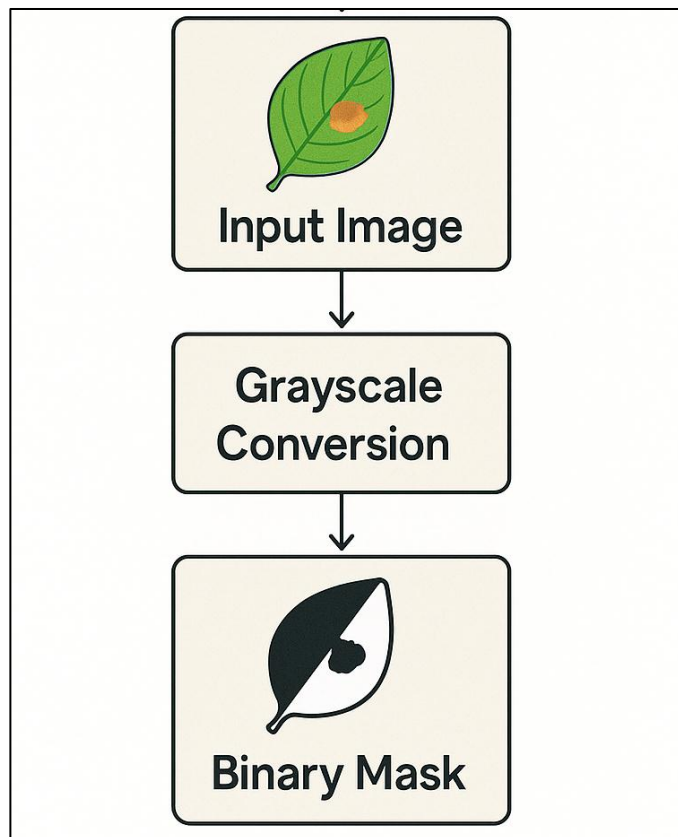


Figure 3: Thresholding Based Leaf Disease Detection

Handcrafted feature-based machine learning approaches were essential precursors to deep learning in plant pathology. While they offer interpretability and lower computational cost, their reliance on manual design and limited adaptability makes them less suitable for large-scale, real-time agricultural applications today. Nonetheless, they remain valuable for controlled environments, educational use, or when computational resources are limited.

4. Traditional Machine Learning Models for Leaf Disease Detection: Traditional (ML) models have played a crucial role in early automated leaf disease detection systems. These approaches focus on using labeled datasets where feature vectors (extracted manually or using basic algorithms) are used to train classifiers capable of recognizing patterns associated with different types of plant diseases. In these methods: Digital images of healthy and diseased leaves are collected using cameras or mobile devices. It involves noise removal, color normalization, background elimination, and segmentation to isolate the leaf region and enhance image quality. Algorithms such as PCA or LDA (Linear Discriminant Analysis) may be used to select the most relevant features. Then classification models are used such as: the SVM maximizes the margin between classes using hyperplanes, the Random Forest (RF) ensemble of decision trees that improves accuracy and robustness, the k-NN classifies based on majority vote of nearest data points, Naive Bayes which is a probabilistic model based on Bayes' theorem, assuming feature independence. The Artificial Neural Networks (ANN) shallows networks capable of modeling complex relationships.

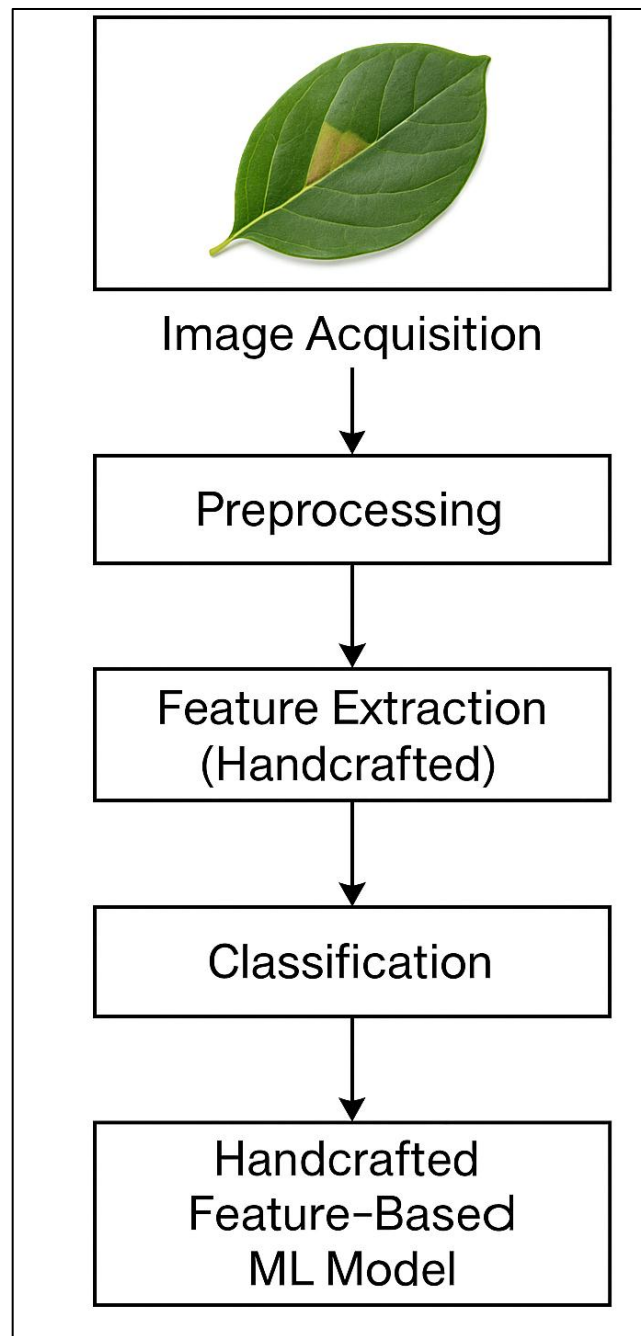


Figure 4: Handcraft Feature Based ML Model

Earlier automated leaf disease detection systems have benefited greatly from the use of traditional machine learning models. Although they work well in controlled conditions and for simpler challenges, they frequently perform poorly in a variety of real-world agricultural contexts. Deep learning techniques have been popular due to the requirement for manual feature engineering and sensitivity to changes in background, lighting, and disease severity. These models can be used as baselines or as lightweight substitutes for deep models, although they are nonetheless useful in settings with limited resources.

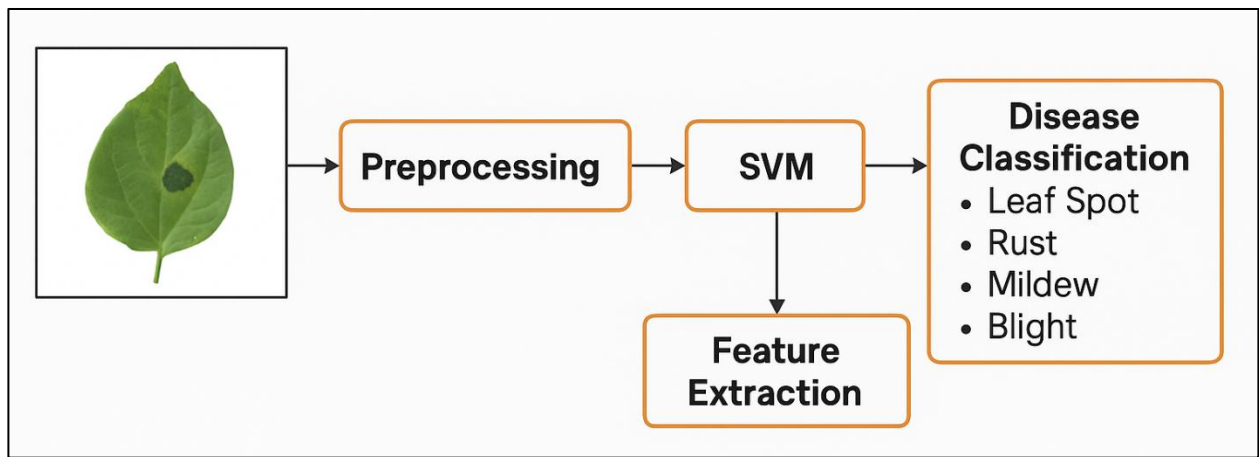


Figure 5: SVM Classify and decide the leaf disease

The Support Vector Machine (SVM) model is effective in classifying plant leaf diseases based on carefully extracted features such as color, texture, and shape. SVM works well when: The feature space is well-defined and linearly separable (or kernel-transformed), The dataset is not too large, making training computationally efficient. SVM provides high classification accuracy in controlled environments with handcrafted features, however it requires precise preprocessing and feature engineering, making it less scalable than deep learning for complex or variable real-world scenarios.

5. Deep Learning Models for Leaf Disease Detection: Deep learning has revolutionized leaf disease detection by enabling models to automatically learn complex, hierarchical features directly from raw images eliminating the need for manual feature engineering. The most commonly used deep learning architecture in this field is the Convolutional Neural Network (CNN). Traditional machine learning relies on manually selected features (color, texture, shape), which may not generalize well across different lighting conditions or disease stages. Deep learning models: Automatically learn features from data, performs better on large and diverse datasets. Deep learning especially CNNs and transfer learning has dramatically improved the accuracy, speed, and scalability of leaf disease detection. It is now considered the state-of-the-art approach, especially for real-world deployment in smart agriculture and mobile diagnostics. Common Deep Learning Architectures are as follows:

Convolutional Neural Networks: Layers include convolution, pooling, and fully connected layers. It learn spatial hierarchies of features (edges → patterns → object).

Transfer Learning Models: Use pre-trained CNNs (trained on ImageNet) and fine-tune them for leaf disease detection, e.g. VGG16/VGG19, ResNet50/101 and InceptionV3.

Region-Based CNNs (R-CNN, Faster R-CNN): Detect and classify multiple diseases or infected regions within the same leaf. Useful when localization of disease symptoms is needed.

GANs (Generative Adversarial Networks): Used for data augmentation: generate synthetic diseased leaf images to expand dataset size and variety.

6. Mobile Deployment Frameworks for Leaf Disease Detection:

With the growing need for real-time, on-field leaf disease detection, mobile deployment frameworks have become crucial for translating deep learning models from lab environments to practical agricultural tools. These frameworks allow trained models to be integrated into mobile or embedded devices, making plant disease diagnosis accessible to farmers, agronomists, and agricultural extension workers without requiring internet access or high-end computing hardware. Farmers in rural areas often lack access to desktop computers or cloud-based systems. Mobile phones are ubiquitous, even in low-resource settings. Real-time disease detection on mobile devices

enables instant decision-making for treatment [12]. There are some key mobile deployment frameworks such as:

TensorFlow Lite (TFLite): It is a Lightweight version of TensorFlow optimized for mobile and embedded devices (Android/iOS). A CNN model trained on a leaf dataset can be converted to TFLite and deployed in a farming app to classify diseases offline.

PyTorch Mobile: It enables deployment of PyTorch models on mobile platforms. It supports Android and iOS with TorchScript (serialized models). It is useful for apps that require dynamic computation graphs.

ML Kit by Google: It provides on-device ML capabilities for Android and iOS apps. It can integrate custom TensorFlow Lite models. It offers APIs for image labeling and object detection.

Core ML (Apple): It is Apple's native framework for running machine learning models on iOS devices. It can convert models from Keras, TensorFlow, or ONNX to .mlmodel format.

ONNX Runtime Mobile: Open Neural Network Exchange (ONNX) model format support for cross-platform deployment.

Mobile deployment frameworks bridge the gap between deep learning research and practical, user-friendly solutions for agriculture. With tools like TensorFlow Lite, PyTorch Mobile, and ML Kit, it's now possible to deliver real-time, offline plant disease diagnosis directly to farmers' smartphones—empowering them with actionable insights and improving crop health at scale.

Table 1: Comparative analysis of Leaf Disease Detection Methods:

Author(s)	Method / Technique	Dataset Used	Accuracy / Performance	Deployment / Application
Mohanty et al. [12]	Deep learning using CNN (AlexNet, GoogLeNet)	PlantVillage	~99.35%	Suggested real-time mobile deployment for farmers
Ferentinos, K. P. [5]	CNN architectures trained on PlantVillage	PlantVillage (87,848 images)	>99%	Emphasized accurate diagnosis; potential for smart agriculture
Pujari et al. [3]	Color and texture feature extraction + SVM classification	Own dataset (fungal infections)	~95% (depending on crop type)	Lab-based diagnosis; not real-time
Arivazhagan et al. [2]	GLCM texture features + SVM	Own image dataset (leaf images)	~92–94%	Offline analysis of unhealthy leaf regions
Camargo & Smith [4]	Color and texture analysis + neural networks	Multiple crop leaf images	~93%	Focused on early symptom detection, no real-time system
Patil & Kumar [7]	Survey of image processing techniques (color, shape, edge, and texture)	Review-based (no specific dataset)	Not applicable (review paper)	Overview of trends; emphasized automated analysis pipelines

Table 2: Comparative Analysis of Reviewed Leaf Disease Detection Methods:

Criteria	Mobile Deployment Frameworks	Deep Learning Models	Traditional ML Models	Handcrafted Feature-Based ML	Thresholding Techniques	Rule-Based Techniques
Definition	Platforms or tools for deploying leaf disease models on smartphones or edge devices	End-to-end models like CNNs, often requiring large datasets for learning features and classification	Models like SVM, RF, k-NN trained on structured features	Uses expert-designed features (e.g., color, texture) fed to ML classifiers	Pixel-based segmentation using intensity thresholds	Hard-coded logic based on if-then rules
Feature Extraction	Usually done on-device or pre-trained	Automatic from raw images	Requires structured input	Manual (e.g., GLCM, color histograms)	None – intensity based	Manually defined conditions
Accuracy	Dependent on embedded model size	High (85–99% in many cases)	Moderate to High	Moderate	Low to Moderate	Low
Data Requirement	Optimized for small, compressed models	High – needs large annotated datasets	Medium – depends on feature quality	Medium	Low	Very Low
Computational Cost	Optimized for edge computing	High (needs GPUs for training)	Moderate	Low to Moderate	Very Low	Very Low
Interpretability	Moderate	Low (black-box models)	High	High	High	Very High
Deployment Complexity	Needs framework like TensorFlow Lite, CoreML, etc.	Requires resource-intensive environment	Simple to moderate deployment	Relatively easy	Very simple	Very simple
Adaptability to New Diseases	Moderate to High (via model updates)	High (with retraining)	Moderate	Low to Moderate	Very Low	Very Low
Examples	TensorFlow Lite,	CNNs, ResNet,	SVM, Random	Color/Texture features	Otsu, Fixed thresholding	IF leaf color is brown,

	PyTorch Mobile, ML Kit	MobileNet, EfficientNet	Forest, k- NN	+ SVM	g	THEN disease = Blight
Advantages	Real-time, portable, user- friendly	High accuracy, automatic feature learning	Faster training, interpretable	No deep model needed, easier tuning	Simple, fast, unsupervised	Human- understandable rules
Disadvantages	Hardware constraints, battery use	Training complexity , needs large data	Depends on features, not as robust	Requires expert domain knowledge	Sensitive to noise, lighting	Not scalable, lacks flexibility

CONCLUSION: From straightforward rule-based methods to sophisticated deep learning strategies and mobile deployment frameworks, leaf disease detection has undergone substantial development. Although they are simple to use and understand, traditional thresholding and rule-based approaches are not as reliable or scalable as what is required in actual agricultural settings. The practical use of these techniques is limited by their extreme sensitivity to changes in background, lighting, and leaf appearance. While each method has its merits, deep learning combined with mobile deployment stands out as the most promising solution for real-time, accurate, and scalable leaf disease detection. Future developments should focus on making these models more lightweight, interpretable, and adaptable to new plant species and diseases with minimal retraining effort.

REFERENCES

1. 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.
2. Arivazhagan, S., Newlin Shebiah, R., Ananthi, S., & Vishnu Varthini, S. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal*, 15(1), 211–217.
3. Pujari, J. D., Yakkundimath, R., & Byadgi, A. S. (2016). Image processing based detection of fungal diseases in plants. *Procedia Computer Science*, 46, 1802–1808.
4. Camargo, A., & Smith, J. S. (2009). An image-processing based algorithm to automatically identify plant disease visual symptoms. *Biosystems Engineering*, 102(1), 9–21.
5. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.
6. Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*.
7. Patil, J., & Kumar, R. (2011). Advances in image processing for detection of plant diseases. *Journal of Advanced Bioinformatics Applications and Research*, 2(2), 135–141.
8. Arivazhagan, S., Shebiah, R. N., Ananthi, S., & Varthini, S. V. (2013). Detection of unhealthy leaf spots using texture features. *Agricultural Engineering International: CIGR Journal*, 15(1), 211–217.
9. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.

10. 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.
11. Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence*, 31(4), 299–315.
12. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419.