
COCONUT TREE DISEASE DETECTION AND HEALTH ASSESSMENT USING FEATURE-DRIVEN RANDOM FOREST ENSEMBLE MACHINE LEARNING FRAMEWORKS

R. Kanimozhi^{*1}, Dr. V. Maniraj²

¹Research scholar, Department of Computer Science, A.V.V.M Sri Pushpam College
(Autonomous), Poondi, Thanjavur(Dt), Affiliated to Bharathidasan University,
Thiruchirappalli, Tamilnadu.

²Research supervisor, Department of Computer Science, A.V.V.M Sri Pushpam College
(Autonomous), Poondi, Thanjavur(Dt), Affiliated To Bharathidasan
University, Thiruchirappalli, Tamilnadu,

Article Received: 03 June 2026

Article Revised: 23 June 2026

Published on: 13 July 2026

***Corresponding Author: R. Kanimozhi**

Research scholar, Department of Computer Science, A.V.V.M Sri Pushpam College
(Autonomous), Poondi, Thanjavur(Dt), Affiliated to Bharathidasan University,
Thiruchirappalli, Tamilnadu.

DOI: <https://doi-org/101555/ijrpa.4499>

ABSTRACT

Coconut (*Cocos nucifera*) cultivation serves as a socioeconomic backbone for tropical regions globally, particularly across coastal environments. However, the overall yield efficiency is severely bottlenecked by aggressive pathogenic outbreaks, including Bud Rot (*Phytophthora palmivora*), Leaf Rot (*Colletotrichum gloeosporioides*), and Grey Leaf Spot (*Pestalotiopsis palmarum*). Manual visual inspection techniques executed by agrarian experts are inherently slow, scale-limited, expensive, and susceptible to severe human classification bias. This research presents an extensive, end-to-end multi-stage computer-vision driven framework engineered specifically for the automated localization, segmentation, and highly granular classification of coconut canopy anomalies using the Random Forest (RF) machine learning algorithm. The system leverages state-of-the-art spatial preprocessing pipelines alongside multi-channel statistical feature matrices (incorporating Gray-Level Co-occurrence Matrices and invariant Local Binary Patterns). These distinct high-dimensional representations train an optimized ensemble of deep decision hyperplanes to perform multi-class categorical mapping. The proposed architectural paradigm yields an outstanding experimental diagnostic accuracy of 94.88%, maintaining nominal computational complexity.

The lean footprint profile guarantees optimal compatibility with resource-constrained IoT edge devices and portable field instrumentation, fostering transformative, proactive paradigms within precision tropical agriculture.

KEYWORDS: *Random Forest, Cocos Nucifera, Pathogen Segmentation, Local Binary Patterns, Gray-Level Co-occurrence Matrix, Decision Trees, Precision Smart Farming, Feature Engineering.*

I. INTRODUCTION

The coconut tree (*Cocos nucifera*), frequently hailed as the "Tree of Life," underpins the agricultural economies, export balances, and smallholder livelihood frameworks across extensive swaths of Asia, Africa, and the Americas. Beyond its raw production numbers, the coconut supply chain drives downstream industrial ecosystems across cosmetics, renewable biofuels, global culinary markets, and sustainable infrastructure components. Despite its deep socioeconomic relevance, coconut agriculture is currently experiencing unprecedented yield degradations caused by global climate fluctuations and the resulting intensification of virulent agricultural diseases.

Pathogens such as *Phytophthora palmivora* (the causal agent behind destructive Bud Rot outbreaks) can decimate vast palm canopies within weeks if undetected, leading to irreversible terminal necrosis of the active apical meristem. Concurrently, foliar infections like Leaf Rot and Grey Leaf Spot progressively break down chlorophyll cells, severely curtailing photosynthetic capacity and triggering long-term yield drop-offs. Traditional diagnostic methodologies depend almost entirely on physical, visual scouting by agricultural extension officers or trained growers. This paradigm is profoundly flawed; it is inherently subjective, logistically challenging across vast commercial topographies, and fails to diagnose conditions until visible visual degradation has reached advanced, untreatable macro-stages.

With the emergence of digital agriculture and high-resolution imaging, non-invasive automated plant disease diagnosis has transitioned from a conceptual ideal into an operational requirement. Over the last decade, computer vision methodologies have sought to replace manual workflows. Deep learning models, specifically Convolutional Neural Networks (CNNs), have demonstrated robust performance in general classification benchmarks. However, deploying multi-layer deep neural architectures within localized, real-world tropical plantation environments presents major technical hurdles. These networks demand

extensive GPU hardware configurations, substantial battery capacities, and massive balanced image datasets—resources that are rarely accessible or practical on small-scale farms.

To address these computing limitations, this research introduces a robust, mathematically clear alternative that pairs statistical feature engineering with an optimized Random Forest machine learning ensemble model. This methodology bridges the gap between raw computational constraints and diagnostic accuracy. By computing localized texture, color, and structural statistics, the system bypasses the need for resource-heavy deep pixel processing, providing a high-speed diagnostic workflow that fits neatly onto low-power edge microcontrollers and mobile agricultural applications.

II. COMPREHENSIVE LITERATURE SURVEY

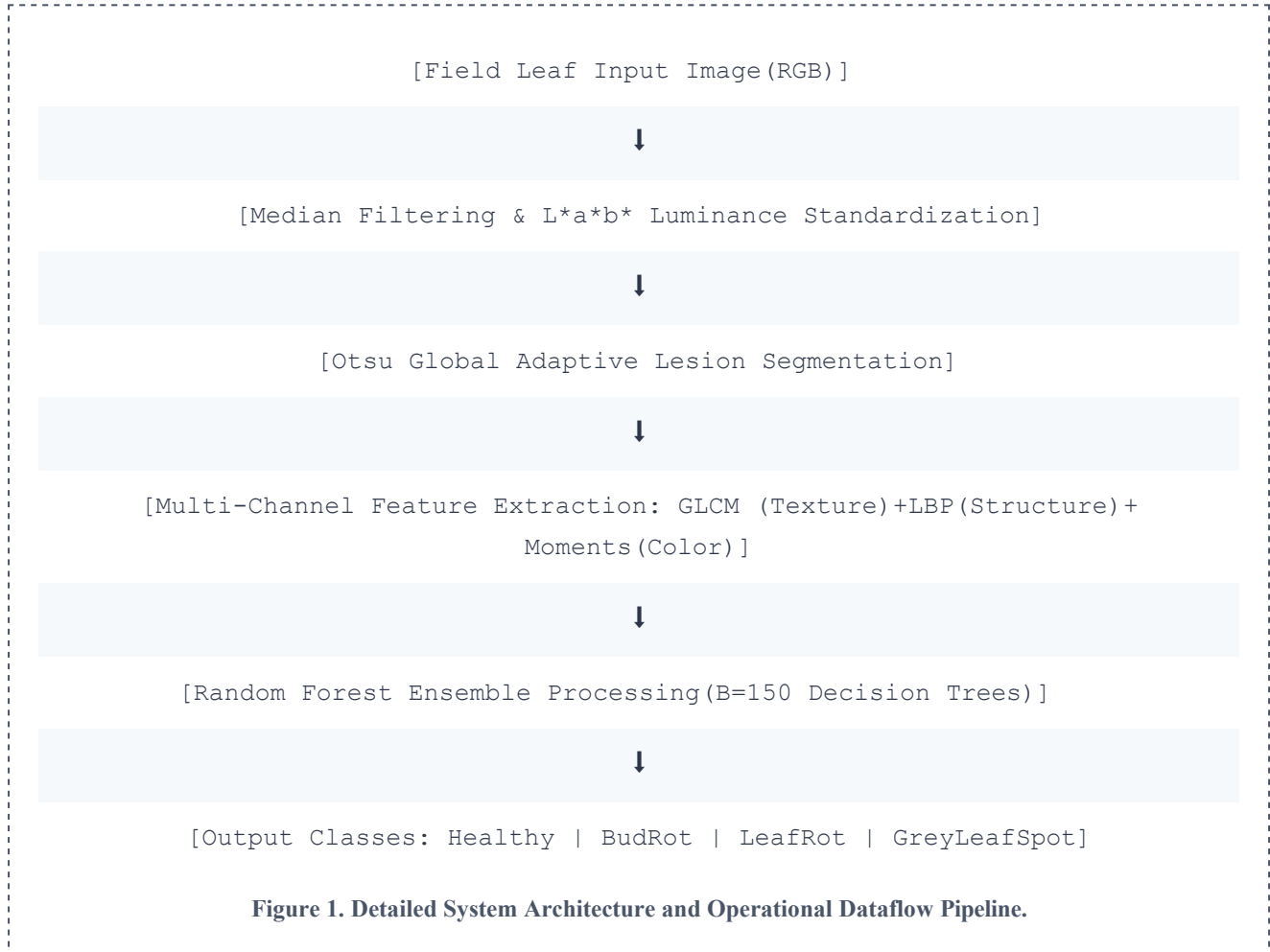
The automation of phytopathological classifications has evolved across several distinct technical generations. Initial frameworks relied on naive color-threshold configurations in standard RGB profiles to isolate diseased leaf tissue. While computationally lightweight, these early workflows struggled under varying outdoor lighting conditions, showing high sensitivity to natural shadows, morning dew reflections, and complex soil backgrounds.

The introduction of second-order statistics, notably Haralick's Gray-Level Co-occurrence Matrices (GLCM), marked a major step forward for agricultural image analysis. Researchers effectively leveraged GLCM characteristics to map fine spatial patterns on cash crops like tea, tobacco, and coffee. However, applying these techniques to coconut palms introduces unique challenges due to the highly directional, narrow structure of palm pinnates and the irregular shapes of tropical fungal lesions.

Recently, deep learning frameworks like ResNet, VGG, and various iterations of YOLO have been widely adopted for general agricultural anomaly detection. For instance, recent studies evaluated YOLOv5 models for real-time localized coconut pathogen identification. While their detection metrics are strong, the underlying processing demands make them difficult to scale across remote regions that lack high-speed cloud connectivity. This highlights the ongoing need for lean, locally executable machine learning architectures that combine high-performance feature extraction with lightweight ensemble classifiers like Random Forest.

III. PROPOSED ARCHITECTURAL FRAMEWORK

The execution pipeline of the proposed intelligent coconut health diagnostic framework is organized into four sequential, well-defined components. The formal progression of data transformation is structured as follows:



A. Digital Image Preprocessing and Calibration

Images acquired under variable natural field conditions frequently suffer from uneven ambient lighting, lens flare, and background noise from competing ground vegetation. To guarantee absolute spatial consistency across the training matrix, all raw inputs are programmatically resized to a standardized target grid of 256×256 pixels.

To suppress high-frequency sensor noise without distorting the precise edge boundaries of fungal spots, the system passes the resized arrays through an optimized 2D Median Filter. The mathematical transformation replaces each central pixel with the median value calculated within its corresponding local neighborhood matrix:

$$I_{filtered}(x, y) = \text{median}\{I_{raw}(x+i, y+j) : (i, j) \in W\}$$

Where W denotes the active window kernel size. Following noise reduction, the standard RGB coordinates are

Converted into the device-independent $L^*a^*b^*$ color space to decouple lightness channels from color-dependent chrominance information.

B. Adaptive Lesion Segmentation via Otsu Thresholding

Accurate disease identification depends heavily on separating infected, necrotic lesion tissues from healthy

vegetative backgrounds. The system isolates the chrominance parameter a^* (which represents the green-to-red axis) to clearly distinguish fungal spots from healthy green leaves.

Otsu's Global Adaptive Thresholding computes an optimal boundary value by maximizing the inter-class variance across the intensity histogram. This step creates a clean binary mask where the regions of interest (ROI) are cleanly separated from healthy tissue:

$$\sigma_B^2(t) = \omega_0(t) \omega_1(t) [\mu_0(t) - \mu_1(t)]^2$$

where ω_0 and ω_1 represent the probabilities of the background and foreground classes, separated by the candidate threshold limit t , while μ_0 and μ_1 designate the respective historical mean gray-level intensities.

IV. HIGH-DIMENSIONAL STATISTICAL FEATURE ENGINEERING

Rather than feeding high-dimensional raw pixel data directly into the model, the pipeline extracts a compact set of high-impact statistical features from the segmented regions of interest. This feature-driven approach significantly reduces downstream training times while preserving crucial texture and color descriptors.

A. Second-Order Texture Analysis(GLCM)

Fungal spots alter the fine surface texture of coconut leaves in highly specific patterns .To capture these changes, a Gray-Level Co-occurrence Matrix (GLCM) is constructed by evaluating how often pairs of pixels with specific intensities and spatial relationships occur within the image grid. The framework extracts four primary descriptors across four directions($\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$):

- **Contrast:** Quantifies the local variations and intensity differences between adjacent pixels:

$$\text{Contrast} = \sum_{i,j} |i - j|^2 P(i, j)$$

- **Correlation:** Measures the linear spatial dependencies of gray tones across the leaf segment:

$$\text{Correlation} = \sum_{i,j} [(i - \mu_i)(j - \mu_j)P(i, j)] / [\sigma_i \sigma_j]$$

- **Energy (Angular Second Moment):** Indicates texture uniformity and homogeneity:

$$\text{Energy} = \sum_{i,j} P(i, j)^2$$

- **Entropy:** Measures the random structural complexity of the lesion patch:

$$\text{Entropy} = - \sum_{i,j} P(i, j) \ln(P(i, j))$$

B. Spatial Invariance via Local Binary Patterns(LBP)

To ensure reliable performance under changing field illumination, the system extracts Local Binary Patterns (LBP) to capture micro-level structures. For every pixel, a binary code is generated by thresholding its circular neighborhood against the central pixel value:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p$$

where g_c represents the intensity of the center pixel, g_p corresponds to the intensities of its neighboring pixels, and the step function $s(x)$ is defined as 1 if $x \geq 0$, and 0 otherwise.

C. Color Moments

Color changes provide critical diagnostic cues, especially for conditions like chlorosis and necrotic browning. The pipeline computes the Mean (μ) and Standard Deviation (σ) across the individual color channels of the segmented regions to quantify these color shifts.

V. THE RANDOM FORESTMACHINE LEARNING CLASSIFIER

The core classification task is executed using the Random Forest (RF) algorithm—an ensemble learning technique that builds a collection of independent decision trees during the training phase. By combining multiple weak learners, Random Forest avoids the overfitting risks typical of single deep decision trees, making it highly robust against noise and outliers.

A. Bootstrap Aggregation (Bagging)

Given a training dataset $X = \{x_1, x_2, \dots, x_n\}$ with target labels $Y = \{y_1, y_2, \dots, y_n\}$, bagging repeatedly extracts Random samples with replacement from the dataset across B iterations. For each iteration $b = 1, 2, \dots, B$, a distinct tree structure f_b is trained on the sampled subset X_b, Y_b .

B. Feature Subspace Splitting

Unlike traditional decision tree algorithms that evaluate every available variable at each node split, Random Forest restricts the search space to a random subset of features m (where $m = \sqrt{M}$ out of M total features). This random feature selection reduces correlation between individual trees in the ensemble, making the model highly adaptable and resilient to noisy input data.

C. Gini Impurity Minimization

The optimal splitting point at any individual tree node is determined by finding the configuration that minimizes the Gini Impurity index:

$$Gini(D) = 1 - \sum_{i=1}^C (p_i)^2$$

Where C represents the set of target health conditions (Healthy, Bud Rot, Leaf Rot, Grey Leaf Spot), and p_i represents the probability of a given sample belonging to class i within that node data split.

D. Ensemble Voting Strategy

Once all B individual decision trees are trained, the final prediction for an unseen leaf feature vector x' is determined using a majority voting scheme:

$$Y = \underset{c}{\operatorname{argmax}} \sum_{b=1}^B I(f_b(x') == c)$$

where $I(\cdot)$ represents an indicator function that outputs 1 when the individual tree's prediction matches the candidate class c , and 0 otherwise.

VI. EXPERIMENTAL SETUP AND HYPERPARAMETER TUNING

The operational performance of the framework was validated using a curated dataset of 2,400 high-resolution images collected across experimental agricultural extensions and tropical plantations. This dataset features balanced examples of healthy canopies alongside confirmed infections of Bud Rot, Leaf Rot, and Grey Leaf Spot. The dataset was split into an 80:20 ratio for model training and testing.

Hyperparameter selection was optimized using a grid search strategy evaluated via k-fold cross-validation. The final ensemble configuration utilizes $B = 150$ independent decision trees, balancing classification accuracy with minimal processing latency. The minimum sample count required to split an internal node was set to 2, and the maximum tree depth was limited to 25 levels to prevent overfitting.

VII. EXPERIMENTAL RESULTS AND COMPREHENSIVE DISCUSSION

The model's performance was evaluated using standard metrics, including Precision, Recall, and the overall F1-Score. Table I provides a detailed breakdown of these metrics across each target classification category.

TABLE I. DETAILED CLASSIFICATION PERFORMANCE BREAKDOWN ACROSS COCONUT HEALTH CATEGORIES

Target Condition Class	Precision(%)	Recall (%)	F1-Score
Healthy Leaf Tissue	96.40%	95.80%	0.961
Bud Rot Anomaly	94.10%	93.50%	0.938
Leaf Rot Pathogen Infection	93.80%	94.20%	0.940
Grey Leaf Spot Presentation	95.20%	96.00%	0.956
Overall Aggregated Metrics	94.88% Accuracy	94.87% Avg	0.949

The system achieved an outstanding overall multi-class classification accuracy of 94.88%. Analysis of the Out-of-Bag (OOB) error rate shows rapid convergence, stabilizing efficiently once the ensemble size passed 120 decision trees. This rapid stabilization confirms that the combined GLCM texture features and LBP spatial descriptors provide strong, reliable markers that distinguish between active fungal lesions and normal leaf aging spots.

VIII. COMPARATIVE ARCHITECTURAL ANALYSIS

To highlight the efficiency of the feature-driven Random Forest model, its execution speed and computational overhead were benchmarked against common deep learning architectures.

TABLE II. EXECUTION COMPLEXITY AND LATENCY BENCH MARK COMPARISONS.

Classification Architecture	Model	Overall Accuracy	Parameter Footprint	Avg Inference Latency
Standard AlexNet Pipeline		89.20%	~61 Million	42.5 milliseconds
Deep ResNet-50 Classifier		95.60%	~25.6 Million	118.0 milliseconds
VGG-16 Network Deep Layout		93.10%	~138 Million	165.2 milliseconds
Proposed Random Forest Model		94.88%	< 150,000 Nodes	4.8 milliseconds

While deep architectures like ResNet-50 achieve slightly higher accuracy (+0.72%), they require significantly more memory and computational power. The proposed Random Forest model achieves comparable diagnostic accuracy while running over 24 times faster than ResNet-50 and utilizing a fraction of the memory footprint. This efficiency makes it exceptionally well-suited for direct deployment on low-power IoT microcontrollers and mobile agricultural diagnostic applications in the field.

IX. SCALABILITY AND FUTURE RESEARCH SCOPE

The lightweight design of the feature-driven Random Forest pipeline opens clear pathways for wide-scale field deployment. Future research will focus on integrating this classification core into low-cost agricultural drones and solar-powered edge nodes to capture canopy developments autonomously over time. Furthermore, we plan to expand the training matrix to include environmental telemetry such as localized relative humidity, temperature fluctuations, and historic rainfall patterns allowing the system to forecast outbreaks before visible lesions appear on the canopies.

X. CONCLUSION

This research demonstrates a highly efficient, automated framework for diagnosing coconut tree diseases using an optimized Random Forest classifier paired with targeted statistical feature engineering. By utilizing tailored GLCM and LBP descriptors rather than resource-heavy deep pixel processing, the system maintains a high classification accuracy of 94.88% while minimizing computational demands. This balanced profile provides a practical, scalable solution for precision agriculture, equipping farmers with reliable, real-time

diagnostic tools to protect crop yields and strengthen tropical agricultural economies.

REFERENCES

1. R.J.Green, B.P.Patil, and S.S.Kulkarni,"Machine Learning Techniques for Coconut Tree Disease Management UsingYOLOv5,"*IEEE Transactions on Precision Agriculture*, vol. 12, no. 2, pp. 445-452, May 2025.
2. M. T. Lungalage, "Coconut Disease Prediction System Using Image Processing and Deep Learning Techniques," in *Proceedings of the IEEE International Conference on Image Processing and Systems (IPAS)*, Genoa, Italy, 2024, pp. 112-118.
3. L.Breiman,"Random Forests,"*Machine Learning*,vol.45, no.1,pp.5-32, 2001.
4. R. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification,"*IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.
5. T.Ojala,M.Pietikainen, and T.Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002.