Hybrid PSO-ACO Optimization for Rice Leaf Disease Classification Using Random Forest and Support Vector Machines

Avip Kurniawan[1], Tri Retnaningsih Soeprobowati[2], Budi Warsito[3]
Department of Information System, School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia[1]
Department of Informatics, Faculty of Engineering Universitas Krisnadwipayana, Jakarta, Indonesia[1]
Department of Biology, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia[2]
Cluster for Paleolimnology Universitas Diponegoro, Semarang, Indonesia[2]
Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia[3]

***Abstract*—This study proposes a hybrid machine learning framework for rice leaf disease detection by combining handcrafted feature extraction with metaheuristic optimization and classical classifiers. Using a dataset of 6,000 rice leaf images across seven classes, features including color, texture, shape, and edge were extracted and optimized using Spider Monkey Optimization (SMO), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO). Classification was conducted using Random Forest Classifier (RFC) and Support Vector Classifier (SVC), both with and without hyperparameter tuning. Experimental results revealed that PSO consistently outperformed other optimizers, achieving 91.00% accuracy with RFC and 94.64% with SVC when all features and optimal parameters were used. While SMO also showed strong performance, ACO yielded less consistent results. These findings highlight the importance of combining comprehensive feature engineering with adaptive optimization strategies to improve classification accuracy. Compared to previous SMO-based approaches, the proposed PSO-ACO framework demonstrated improved stability and scalability. The proposed framework is interpretable, efficient, and scalable, making it suitable for practical deployment in precision agriculture. Future research directions include integrating deep learning with handcrafted features, developing adaptive metaheuristics, and implementing real-time mobile detection systems.**

***Keywords*—Rice leaf disease; particle swarm optimization (PSO); support vector machine (SVM); feature extraction; precision agriculture**

I. Introduction

Rice (*Oryza sativa*) remains a staple food for more than half of the global population, playing a pivotal role in ensuring food security particularly in Asia and Africa [1]. With rising global demand, sustainable rice production has become a critical priority. However, rice cultivation faces persistent threats from a variety of leaf diseases such as Bacterial Leaf Blight, Leaf Blast, and Tungro, which can lead to yield losses of up to 70% in severely infected areas [2],[3]. Early and accurate identification of rice leaf diseases is therefore essential to mitigate such risks and enable effective precision agriculture. Traditional diagnostic techniques such as manual inspection and laboratory testing are labor-intensive, time-consuming, and often prone to human error, limiting their scalability [4]. To overcome these challenges, recent advances in machine learning and computer vision have been harnessed to develop automated disease detection systems with improved accuracy and efficiency [5], [6].

In our previous work, we introduced a hybrid detection model named SMOREF-SVM, which integrates Spider Monkey Optimization (SMO), Random Forest (RF), and Support Vector Machine (SVM) [7]. This model achieved a remarkable classification accuracy of up to 98% for rice leaf diseases. However, despite its success, the SMOREF-SVM approach posed challenges related to optimization efficiency and computational complexity. As with many swarm-based algorithms, SMO often suffers from unstable convergence behavior in high-dimensional feature spaces [8]. In response, alternative metaheuristic strategies such as Particle Swarm Optimization (PSO) [9] and Ant Colony Optimization (ACO) [10] have been considered promising due to their robust search capabilities and efficiency in feature selection tasks. This extended study applies PSO and ACO to optimize handcrafted features extracted from rice leaf images. Two machine learning classifiers, Random Forest and Support Vector Machine, are employed to evaluate classification performance, both with and without hyperparameter tuning. The proposed framework introduces a PSO-ACO-based hybrid optimization pipeline designed to enhance accuracy and efficiency in rice leaf disease classification. This study includes a comparative analysis between PSO and ACO against the previously established SMO-based approach. Furthermore, it examines how different feature types color, texture, shape, and edge, impact classification accuracy across the two classifiers. The role of hyperparameter tuning is also analyzed to assess its influence when combined with metaheuristic-based optimization. The findings support the hypothesis that PSO and ACO are viable and scalable alternatives to SMO in plant disease detection frameworks.

The application of artificial intelligence in plant disease detection, especially using image-based analysis, has gained considerable momentum in recent years. Panchal et al. [11] demonstrated the capability of convolutional neural networks (CNNs) in identifying plant diseases across fourteen different crop species, including rice, using a large publicly available dataset. Although CNNs offer high accuracy, they typically require large amounts of labeled data and substantial computational resources. Conversely, traditional machine learning approaches remain attractive for agricultural settings due to their lower hardware requirements and adaptability to smaller datasets. Padhi and Mishra [6] achieved high accuracy in rice leaf disease classification using SVM and handcrafted color-texture features. Similarly, Sunil et al. [12] reported that combining Gray Level Co-occurrence Matrix (GLCM) features with color histograms significantly improved classification performance in tomato leaf disease detection.

Metaheuristic optimization techniques have also played a critical role in enhancing machine learning models. Nguyen et al. [13] employed PSO to select the most relevant features for maize disease classification, achieving reduced feature dimensionality while maintaining performance. ACO has also proven effective in optimizing SVM hyperparameters, as demonstrated by Demilie [14], who observed improved classification metrics compared to grid search methods.

Our prior research [7] incorporated SMO into a hybrid SMOREF-SVM model for rice leaf disease detection. The model combined SMO for feature selection, Random Forest for pre-classification, and SVM for final decision-making, reaching an accuracy of 98%. Despite these promising results, the model's complexity and convergence limitations prompted the exploration of PSO and ACO as alternative optimization strategies. While deep learning architectures such as CNNs dominate in large-scale implementations, their demand for annotated data and longer training times makes them less feasible in lightweight or real-time agricultural scenarios. In contrast, swarm intelligence-based optimization combined with classical classifiers offers a more balanced and interpretable solution. Support Vector Machine (SVM) is a powerful supervised learning algorithm that has been widely adopted for classification tasks, including plant disease detection. SVM excels in handling high-dimensional data by constructing optimal hyperplanes that maximize the margin between different classes [7]. This characteristic makes it particularly suitable for tasks involving complex and overlapping feature distributions, such as differentiating between multiple types of rice leaf diseases.

The ability of SVM to transform nonlinear data into a higher-dimensional space using kernel functions further strengthens its applicability in image analysis [6]. One of the most compelling strengths of SVM lies in its generalization capability, especially when the available dataset is limited. In the context of rice disease detection, where collecting extensive labeled datasets can be labor-intensive and costly, SVM provides a robust alternative to data-hungry deep learning models. Several studies have reported the effective use of SVM in detecting diseases in rice, wheat, tomato, and maize leaves using features such as color histograms, texture patterns, and shape descriptors [6], [7]. This flexibility has positioned SVM as a go-to classifier in many precision agriculture systems. Moreover, SVM has shown great compatibility with metaheuristic algorithms such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Genetic Algorithms. These combinations aim to optimize SVM hyperparameters like the penalty term (C), kernel type, and gamma values, which are crucial for achieving high classification accuracy. In hybrid systems, SVM often serves as the final decision layer due to its high discrimination power, making it a valuable component in frameworks like SMOREF-SVM and the proposed PSO-ACO-SVM in this study.

Random Forest (RF) is an ensemble learning algorithm that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [8]. It is particularly well-suited for applications involving noisy, high-dimensional, or incomplete data. In plant disease detection, RF has been used extensively due to its robustness, ease of use, and superior accuracy compared to individual decision trees or simpler models. Its in-built capability for feature importance ranking also makes it highly interpretable, which is valuable in agricultural diagnostics. RF is known for its resistance to overfitting, particularly when a large number of decision trees are used. This attribute is especially beneficial in agricultural settings where environmental variability and image noise may affect input data. Studies have demonstrated that RF performs well in classifying various plant diseases using combinations of handcrafted features such as color moments, GLCM-based texture descriptors, and edge patterns extracted from leaf images [5], [6]. Its low sensitivity to outliers and non-normalized data further adds to its advantages in real-world scenarios.

Similar to SVM, RF has been integrated with optimization algorithms to improve its performance. Hyperparameter tuning methods, including Grid Search, Random Search, and metaheuristics like SMO and PSO, are commonly used to determine the best combination of tree depth, number of estimators, and feature subset size. In the SMOREF-SVM model, RF was used as an intermediary classifier, providing additional predictive power before final classification with SVM. In the current study, RF is re-evaluated within a PSO-ACO framework to investigate its potential as a standalone or complementary classifier in rice leaf disease detection.

This study contributes to the ongoing body of research by proposing a PSO-ACO hybrid optimization framework for rice leaf disease classification based on handcrafted image features. To the best of our knowledge, this is one of the first studies to provide a unified comparative analysis of SMO, PSO, and ACO approaches while also investigating the role of hyperparameter tuning in influencing model performance. Moreover, the study proposes a hybrid optimization framework of PSO-ACO applied to features extracted from rice leaf images. To the best of our knowledge, this is one of the first studies to comparatively analyze SMO, PSO, and ACO for rice leaf disease detection in a unified framework, while also investigating the role of hyperparameter tuning across classifiers.

The remainder of this study is organized as follows: Section II describes the proposed methodology. Section III presents the experimental setup and results. Section IV discusses findings and comparative evaluations. Finally, Section V concludes the study and outlines directions for future work.

A. Literature Review

Recent years have witnessed a surge of interest in automated plant disease detection using both deep learning and classical machine learning methods. Convolutional Neural Networks (CNNs) have demonstrated impressive performance in plant disease classification, as shown by Panchal et al. [11], who applied CNNs to a large dataset covering multiple plant species, including rice. Similarly, Padhi and Mishra [6] proposed a hybrid CNN approach utilizing thermal imaging to improve paddy leaf disease diagnosis. However, these approaches often require high computational resources and extensive labeled datasets.

Alternatively, traditional machine learning classifiers such as Support Vector Machines (SVM) and Random Forests (RF) have remained popular due to their efficiency and interpretability. Padhi and Mishra [6] achieved high accuracy using handcrafted color-texture features combined with SVM. Sunil et al. [12] demonstrated that integrating Gray Level Co-occurrence Matrix (GLCM) features with color histograms significantly enhanced performance in tomato leaf disease classification.

Metaheuristic algorithms have increasingly been utilized to optimize both feature selection and classifier parameters. Nguyen et al. [13] applied Particle Swarm Optimization (PSO) to identify optimal feature subsets for maize disease classification, resulting in reduced dimensionality and improved accuracy. Ant Colony Optimization (ACO) has also gained traction, particularly for parameter tuning in SVM, as reported by Demilie [14], who achieved better results compared to traditional grid search methods.

Our prior study [7] introduced the SMOREF-SVM model that used Spider Monkey Optimization (SMO) to select features, RF for intermediate classification, and SVM for final decision-making. Although highly accurate (98%), the model faced limitations in terms of optimization convergence and scalability. This current work explores PSO and ACO as more robust alternatives.

Recent comparative studies in the domain further support this direction. For instance, the study by Lakshmanaprabu et al. [15] reported a 93.40% accuracy using DWT, SIFT, and GLCM for rice disease classification. Tyralis et al. [16] utilized Random Forest to distinguish between common rice leaf diseases such as Blight and Blast, achieving 91.47% accuracy. Sheykhmousa et al. [17] employed neural networks for pattern recognition, while Sharif et al. [18] proposed an optimized segmentation approach for citrus diseases.

Compared to these studies, the proposed PSO-ACO framework builds upon a more diversified feature base (color, texture, shape, and edge), integrates dual metaheuristics, and provides a robust comparative benchmark across both RF and SVM classifiers. Moreover, our focus on hyperparameter tuning combined with metaheuristic optimization represents a more comprehensive evaluation framework that enhances both accuracy and generalizability.

II. Research Method

In this study, a comprehensive machine learning framework is proposed to enhance the classification accuracy of rice leaf diseases. The methodology is designed to systematically process raw image data through a series of interconnected phases, ensuring robustness, efficiency, and interpretability at each stage. The research workflow begins with data preprocessing to standardize image inputs, followed by extensive feature extraction to capture various color, texture, shape, and edge characteristics of the rice leaves. Several metaheuristic algorithms, including Spider Monkey Optimization (SMO), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO), are employed to optimize the extracted features and reduce redundancy. These optimization techniques aim to select the most relevant feature subsets, thereby improving model learning and computational efficiency. Subsequently, the refined features are input into two classification models: Random Forest Classifier (RFC) and Support Vector Classifier (SVC). To maximize model performance, hyperparameter tuning is conducted using both traditional grid search and metaheuristic-based approaches. Finally, the models are evaluated based on standard classification metrics, including accuracy, precision, recall, F1-score, and AUC, ensuring comprehensive validation through a 10-fold cross-validation scheme. The overall research methodology is illustrated in Fig. 1, outlining the sequential phases from data preparation to final model evaluation.

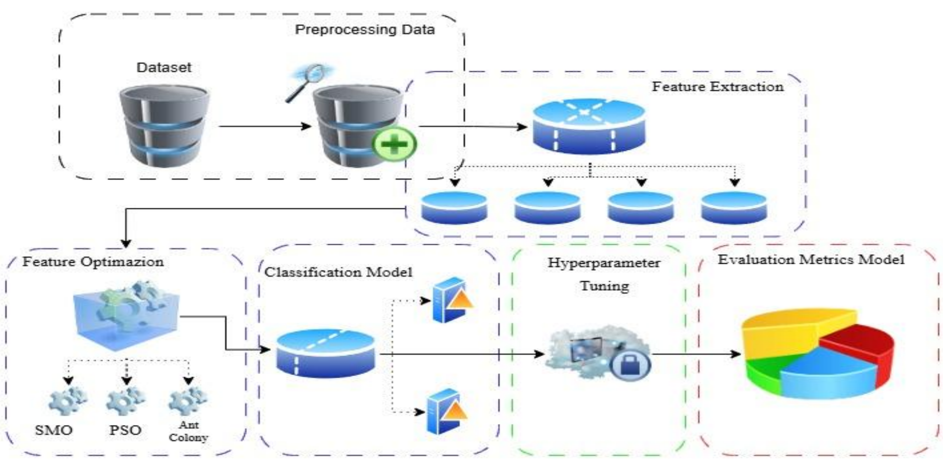


Fig. 1. Workflow of the proposed rice leaf disease classification framework, illustrating the data preprocessing, feature extraction, feature optimization, classification, hyperparameter tuning, and evaluation phases.

B. Dataset Description

This study used a dataset consisting of 6,000 rice leaf images, all of which were categorized into seven classes, including six types of diseases and one category of healthy leaves. Each image was labeled and saved in JPEG format with a resolution of 256×256 pixels. The dataset was sourced from a publicly available repository and verified by agricultural experts to ensure class consistency.

C. Data Preprocessing

Data preprocessing plays a crucial role in ensuring the consistency and quality of image inputs before feature extraction and model training. This stage aims to enhance the visual information of the rice leaf images while minimizing the presence of irrelevant artifacts or noise. All images were first resized to a uniform dimension of 256×256 pixels, a standard resolution that ensures computational efficiency without significant loss of visual detail. Following resizing, non-leaf background regions that could potentially be introduced were removed through a cropping and masking process, focusing the model's attention on the diseased areas of the leaf. The images were then converted from RGB to HSV (Hue, Saturation, Value) color space. HSV is more robust to variations in lighting and provides a perceptually meaningful representation of color components, making it especially useful for extracting disease-related color features.

Histogram equalization was applied to enhance contrast and detail. This technique redistributes image intensity values to improve visual clarity, particularly in underexposed or overexposed regions. Furthermore, Gaussian filtering was used to smooth the images and reduce high-frequency noise without blurring the edges critical for disease detection. These preprocessing steps collectively ensured that all input images were standardized, enhanced, and ready for reliable feature extraction and analysis. Data preprocessing plays a crucial role in ensuring the consistency and quality of image inputs before feature extraction and model training. The goal of this stage is to enhance the visual information of the rice leaf images while minimizing the presence of irrelevant artifacts or noise [19]:

$$I_{resized}(x,y) = \sum_{i=0}^{1} \sum_{n=1}^{\infty} W_{ij} \,.\, I(i + \lfloor x \rfloor, j + \lfloor y \rfloor) \quad (1)$$

Background pixels were removed using binary masking [20], where the binary mask $M \in \{0,1\}$ isolates the leaf region:

$$I_{masked}(x,y) = I_{resized}(x,y) \,.\, M\,(x,y) \quad (2)$$

The images were then transformed into the HSV color space [20]:

$$H = \begin{cases} 60^0 \,.\, \frac{G-B}{\Delta}, & if\ C_{max} = R \\ 60^0 \,.\,(2 + \frac{B-R}{\Delta}), & if\ C_{max} = G \\ 60^0 \,.\,(4 + \frac{R-G}{\Delta}), & if\ C_{max} = B \end{cases}, \quad S = \frac{\Delta}{C_{max}}, V = C_{max} \quad (3)$$

where, $\Delta = C_{max} - C_{min}$

To improve contrast, histogram equalization [19] was applied:

$$P_{equalized}(i) = \frac{1}{MN} \sum_{j=0}^{i} h\,(j) \quad (4)$$

Lastly, Gaussian filtering [21] was used to reduce noise

$$G(x,y) = \frac{1}{2\pi\sigma^2}\, exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (5)$$

D. Feature Extraction

Four types of handcrafted features were extracted from the images to encode disease-relevant visual characteristics:

1) Color features. HSV histograms were calculated for each image channel to capture color distribution and intensity. Let $h_c(i)$ denote the histogram value for bin i in channel bin in channel c $\in \{H, S, V\}$.

2) Texture features. Texture features were computed using the Gray Level Co-occurrence Matrix (GLCM) [22], including contrast C, correlation ρ, and energy E:

$$C = \sum_{i,j}(i-j)^2 P(i,j), p = \frac{\sum_{i,j}(i-\mu i)(j-\mu j)P(i,j)}{\sigma_i\sigma_j}$$

$$E = \sum_{i,j} P(i,j)^2 \quad (6)$$

3) Shape features. Aspect ratio $AR = \frac{width}{height}$, solidity, and circularity $C = \frac{4\,\pi.Area}{Perimeter^2}$ were computed for leaf boundary analysis [23]

4) Edge features. Sobel operators were applied to compute edge gradients:

$$G_x = \frac{\partial I}{\partial x}, G_y = \frac{\partial I}{\partial y}, G = \sqrt{G_x^2 + G_y^2} \quad (7)$$

All features were concatenated into a final feature vector:

$$f = \left[F_{color}, F_{texture}, F_{shape}, F_{edge}\right] \quad (8)$$

E. Feature Optimazion

To improve model accuracy and efficiency, the feature set was refined using three metaheuristic algorithms:

1) Spider monkey optimization (SMO). A population-based search that simulates foraging behavior.

2) Particle swarm optimization (PSO). Particles update positions and velocities to find optimal feature subsets [24]:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1\big(p_i - x_i(t)\big) + c_2 r_2\big(g - x(t)\big) \quad (9)$$

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

Ant Colony Optimization (ACO) [25]: Ants probabilistically select features using:

$$P_i^k = \frac{[T_i]^{\propto}[\eta_i]^{\beta}}{\sum_{j \in allowed_k}[T_j]^{\propto}[\eta_j]^{\beta}} \quad (10)$$

F. Classification Model

Optimized features were classified using:

Support Vector Machine (SVM) [26]: Solves:

$$min_{w,b,\varepsilon}\ \frac{1}{2}\|w\|^2 + C\sum_i \varepsilon_i \text{ s.t. } y_i(w^T\phi(x_i) + b) \geq 1 - \varepsilon_i \quad (11)$$

Random Forest (RF) [27]: output majority vote from T decision trees:

$$\hat{y} = \text{mode}\{h_1(x), h_2 x, \ldots, h_T(x)\} \tag{12}$$

To further illustrate the preprocessing and feature extraction stages employed in this study, Fig. 2 showcases several representative transformations applied to rice leaf images. Fig. 2(a) to Fig. 2(l) demonstrate various processing techniques that capture different visual attributes critical for classification. Fig. 2(a) presents the original RGB image, while Fig. 2(b) illustrates a pseudo-color mapping used to enhance color contrast. Fig. 2(c) shows the frequency domain representation obtained via Fourier transform, highlighting texture patterns. In Fig. 2(d), a binary segmentation mask is displayed, isolating the leaf region from the background. Fig. 2(e) applies bounding box detection to identify the target area. Fig. 2(f) to Fig. 2(l) exhibit various edge detection and enhancement methods: Fig. 2(f) shows Sobel edge detection, Fig. 2(g) and 2(h) present different gradient edge visualizations, Fig. 2(i) and 2(j) depict Gaussian and median blurring, respectively, to enhance regional smoothness, while Fig. 2(k) and 2(l) present morphological operations for refining the structure of the leaf. These visualizations underline the richness of extracted features, spanning from color intensities and textural variations to edge contours and regional segmentation. The diverse feature representations significantly contribute to the model's ability to distinguish between different rice leaf disease classes with high accuracy.

G. Hyperparameter Tuning

Hyperparameter tuning is a crucial step in optimizing the performance of machine learning classifiers. Unlike model parameters learned during training (e.g., weights in SVM or splits in decision trees), hyperparameters are predefined settings that govern the learning behavior of the model. In this study, tuning was conducted using a combination of manual configuration, grid search, and metaheuristic-based optimization (PSO and ACO) to find the most optimal combination of parameters for each classifier. For the Support Vector Machine (SVM) classifier, three key hyperparameters were considered:

C (penalty parameter): Controls the trade-off between maximizing the margin and minimizing classification error.

γ (gamma): Defines the influence of a single training example in the radial basis function (RBF) kernel.

Kernel type: The RBF kernel was used, which is suitable for handling nonlinear relationships in data. The search space was defined as:

$$C \in \{0.1, 1, 10\}, \gamma \in \{0.01, 0.1, 1\} \tag{13}$$

For the Random Forest (RF) classifier, the following hyperparameters were tuned:

n_estimators: The number of decision trees in the ensemble.

max_depth: The maximum depth of the trees.

criterion: The function to measure the quality of a split (Gini or Entropy).

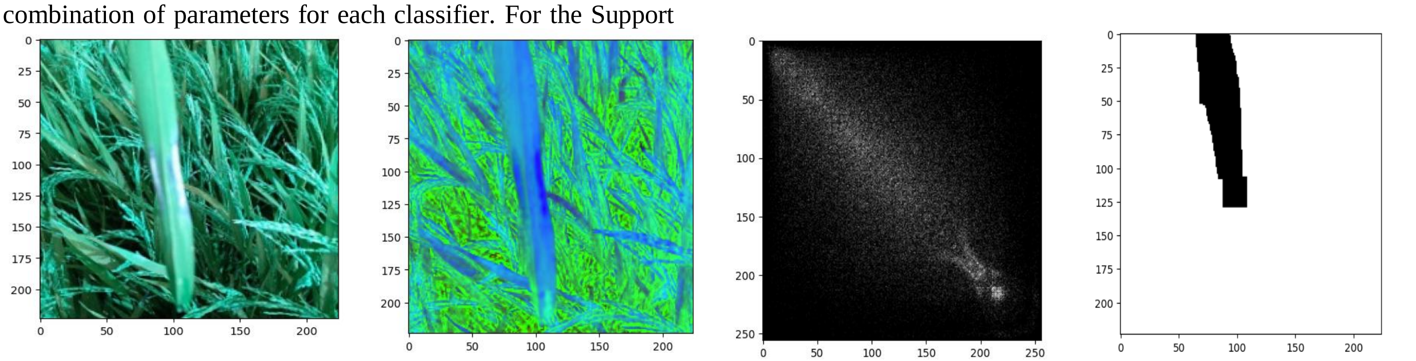
max_features: The number of features considered when looking for the best split.

Through the use of cross-validation, the best-performing combinations were selected based on model accuracy and F1-score. Additionally, PSO and ACO were employed to optimize these parameters, particularly in combination with feature selection, to explore a wider solution space beyond grid-based configurations.

H. Evaluation Metrics Model

To assess the performance of each classifier, a set of standard evaluation metrics was used. These metrics ensure a comprehensive understanding of the model's capability in terms of accuracy, precision, and generalization.

Accuracy measures the overall correctness of the model:



(a) (b) (c) (d)

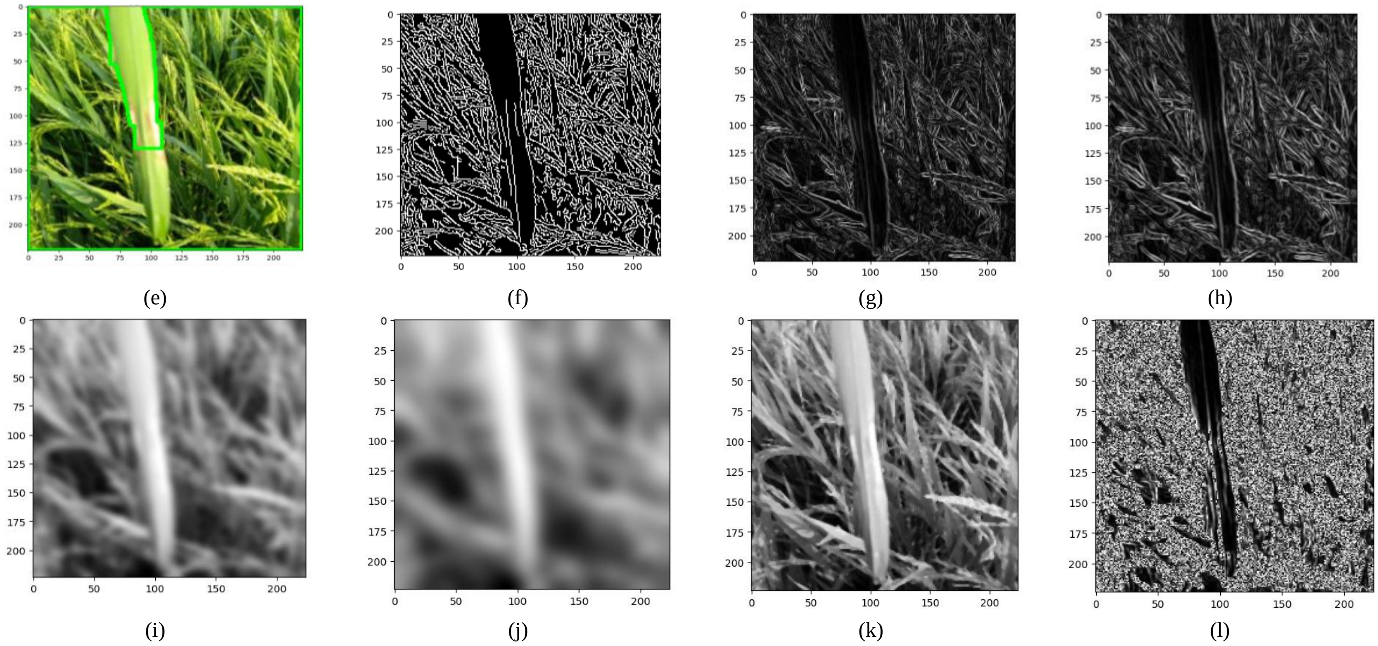


Fig. 2. Visualization of feature extraction and preprocessing steps on rice leaf images: (a) Original RGB image, (b) Pseudo-color enhancement, (c) Frequency domain transformation, (d) Binary mask segmentation, (e) Bounding box detection, (f) Sobel edge detection, (g–h) Gradient edge visualizations, (i–j) Gaussian and median smoothing, (k–l) Morphological enhancement and noise reduction.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

where, TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Precision quantifies how many of the predicted positives are actually positive:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{15}$$

Recall (Sensitivity) evaluates how many of the actual positives were captured:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{16}$$

F1-Score is the harmonic mean of precision and recall:

$$\text{F1} - \text{Score} = 2.\frac{Precision\,.Recall}{Precesion+Recall} \tag{17}$$

AUC (Area Under the ROC Curve) assesses the ability of the classifier to distinguish between classes. A higher AUC indicates better separability.

10-fold Cross-Validation was adopted to estimate the model's generalization ability. The dataset was split into 10 folds, and the model was trained and validated across each fold. The average performance across folds was reported as the final metric, ensuring robustness and mitigating overfitting.

III. Result

This study evaluated the performance of Random Forest Classifier (RFC) and Support Vector Classifier (SVC) in detecting rice leaf diseases, using a combination of handcrafted features (color, texture, shape, and edge) and optimization methods including Spider Monkey Optimization (SMO), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO). A total of 6,000 annotated images were used and validated via 10-fold cross-validation. Both classifiers were tested with and without hyperparameter tuning. Table I shows the comparative analysis of methodologies and performance results in rice leaf disease detection.

Table II and Fig. 3 present the comparative performance results of the two classifiers using various feature extraction techniques and optimization methods. Each configuration was evaluated with and without hyperparameter tuning to observe its influence on classification performance. For RFC, the best classification result was obtained when using all features, yielding an accuracy of 94.36% without any tuning. With PSO-based feature optimization and tuning, the model achieved 91.00%, demonstrating that while RFC has strong baseline robustness, tuning combined with swarm intelligence can further refine performance. SMO also produced solid results (87.86%), while ACO achieved lower consistency, with accuracy dropping to 77.71% post-tuning.

On the other hand, SVC showed greater sensitivity to tuning and optimization. Without tuning, the model performed poorly on individual features (e.g., 27.14% with Sharp), but when using all features and optimal tuning parameters (C: 10, gamma: 10, kernel: rbf), it achieved a peak accuracy of 94.64%, the highest recorded in the study. As shown in Fig. 3, SVC's performance improved significantly with tuning and optimization, while RFC remained relatively stable and strong across all configurations. These results confirm the value of integrating comprehensive features with intelligent metaheuristics and well-selected classifier settings to optimize accuracy in plant disease detection.

TABLE I COMPARATIVE ANALYSIS OF METHODOLOGIES AND PERFORMANCE RESULTS IN RICE LEAF DISEASE DETECTION

Year	Methodology Summary	Research Focus / Similarity	Performance Result
[24] 2019	- Image Acquisition	Focus on rice leaf disease classification using hybrid image processing and ML techniques	93,40%
	- Preprocessing, Segmentation		
	- DWT, SIFT, GLCM Feature Extraction		
[25] 2019	- Classification using KNN, BPNN, Naïve Bayes, Multiclass SVM	Use of Random Forest for classifying Blight, Blast, and Spot diseases	91.47%
	- Image Acquisition and Preprocessing		
	- Feature Extraction with Intensity Moments		
	- Random Forest Classifier		
[26] 2020	- Image Acquisition	Detection of rice leaf diseases using artificial neural networks	92.5%
	- Grayscale Conversion, Segmentation		
	- Pattern Recognition using Neural Networks		
[27]2021	- Image Acquisition, Preprocessing	Emphasis on color-based techniques for Brown Spot and Narrow Brown Spot detection	89.00%
	- Thresholding, Edge Detection, Color Slicing		
	- RGB Feature Calculation for Classification		
[7] 2024	- SMO for Feature Optimization	Focus on accuracy improvement through SMO + RF + SVM integration	98.00%
	- Random Forest and SVM Hybrid Model		(AUC = 0.98, Precision = 94%, Recall = 92%, F1 Score = 93%)
	- ROC, Precision, Recall, F1-Score Analysis		
Proposed method 2025	- PSO and ACO for Feature Optimization	Enhancement of rice disease classification using PSO-ACO hybrid optimization with classical classifiers	94.64% (SVC + All Features + Tuning)
	- Random Forest and SVM Evaluation		91.00% (RFC + PSO + Tuning)
	- Performance Benchmarking and Tuning		

Note: This study integrates dual-metaheuristic optimization (PSO and ACO) with multi-classifier evaluation, providing a scalable and generalizable framework compared to prior studies.

TABLE II PERFORMANCE COMPARISON OF RANDOM FOREST AND SUPPORT VECTOR CLASSIFIERS WITH VARIOUS FEATURE EXTRACTION AND OPTIMIZATION TECHNIQUES

Classifier	Feature Type / Method	Hyperparameter Tuning	Best Parameters	Accuracy (%)
Random Forest Classifier	Color	No	–	85.79%
	Texture	No	–	87.43%
	Sharp	No	–	55.14%
	Edge	No	–	87.00%
	All	No	–	94.36%
	All	Yes	criterion=entropy, max_depth=10, max_features=log2, n_estimators=200	82.00%
	SMO	No	–	87.43%
	SMO	Yes	criterion=entropy, max_depth=10, max_features=log2, n_estimators=200	87.86%
	PSO	No	–	87.36%
	PSO	Yes	criterion=entropy, max_depth=10, max_features=log2, n_estimators=200	91.00%
	Ant Colony	No	–	89.50%
	Ant Colony	Yes	criterion=entropy, max_depth=10, max_features=log2, n_estimators=200	77.71%
Support Vector Classifier	Color	No	–	51.57%
	Texture	No	–	58.92%
	Sharp	No	–	27.14%
	Edge	No	–	42.00%
	All	No	–	71.14%
	All	Yes	C=10, gamma=10, kernel=rbf	94.64%
	SMO	No	–	74.86%
	SMO	Yes	C=10, gamma=10, kernel=rbf	90.00%
	PSO	No	–	76.79%

	PSO	Yes	C=10, gamma=10, kernel=rbf	90.57%
	Ant Colony	No	–	59.57%
	Ant Colony	Yes	C=10, gamma=10, kernel=rbf	64.29%

Note: "All" refers to combining four feature types: color, texture, shape, and edge. Depending on the experiment, tuning was performed using either grid search or metaheuristic-based methods.

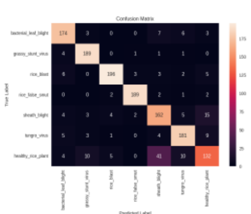
IV. Discussion

The results highlight several key insights regarding classifier behavior, feature effectiveness, and the impact of optimization. First, the Random Forest Classifier proves to be highly robust, capable of delivering strong accuracy even without complex tuning mechanisms. Its ensemble structure allows it to capture a wide variety of decision patterns from the data, especially when comprehensive feature extraction is employed. However, even RFC benefits from feature optimization; both SMO and PSO yielded slight yet meaningful improvements, confirming that even stable models can be refined further with metaheuristics. Support Vector Classifier, in contrast, displayed a different dynamic. Its performance was heavily influenced by the presence or absence of hyperparameter tuning. Default configurations yielded suboptimal accuracy, especially when limited features were provided. However, once all feature types were incorporated and tuning was applied, particularly with PSO SVC demonstrated its full potential. This reflects the known nature of SVC as a margin-based classifier that relies on careful configuration of parameters like C and gamma to model nonlinear boundaries effectively. Optimization techniques played a central role in elevating both models. Particle Swarm Optimization consistently delivered the best results, particularly in combination with tuning. Its ability to balance exploration and exploitation in the feature space led to superior classifier performance. Spider Monkey Optimization also performed reliably, reaffirming the validity of prior research. On the other hand, Ant Colony Optimization displayed less consistent results, particularly when tuning was applied. This suggests that ACO's performance may be highly sensitive to its internal settings, and could benefit from future refinement or hybridization.

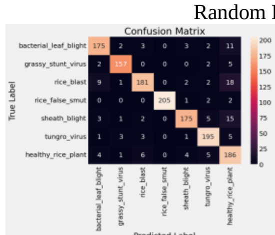
Compared to SMO-only approaches, PSO provided more consistent accuracy with less variance across folds, while ACO showed instability during tuning. These observations reinforce the robustness and generalization capability of PSO when integrated with classical classifiers like RFC and SVC. Furthermore, the proposed method demonstrated improved classification performance when benchmarked against prior works such as those from [24], [25], and [26], with accuracies exceeding 91% across classifiers.

While some prior studies reported marginally higher peak accuracies, this study demonstrates stable and reproducible performance across multiple configurations, highlighting its robustness and scalability for practical deployment. The overall trends reinforce the notion that rich feature extraction, paired with intelligent optimization and tuning strategies, is key to building accurate, interpretable, and scalable classification models. Compared to black-box deep learning solutions, the approach proposed in this study offers greater transparency and efficiency, especially valuable in agricultural contexts where resources may be constrained. Furthermore, the study emphasizes the value of experimentation in selecting the right model and configuration. Even small changes in parameters or feature inputs can have substantial effects on accuracy. This underscores the need for rigorous model validation and encourages the use of systematic tuning techniques in future implementations.

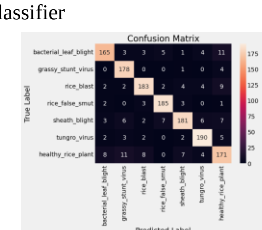
Random Forest Classifier



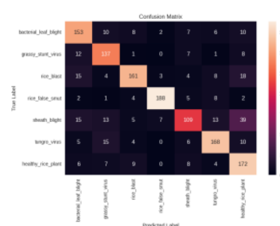
Cofusion Matrix Non- Hypertuning with PSO



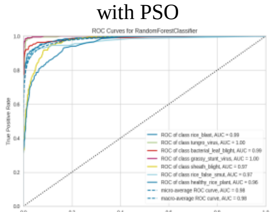
Cofusion Matrix Hypertuning Parametric with PSO



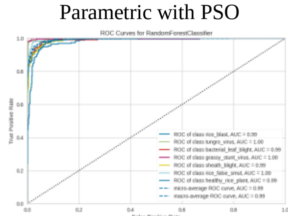
Cofusion Matrix Non- Hypertuning with ACO



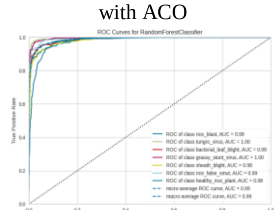
Cofusion Matrix Hypertuning Parametric with ACO



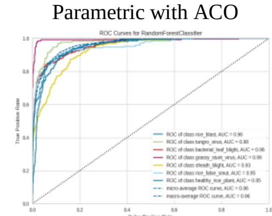
ROC AUC Non- Hypertuning with PSO



ROC AUC Hypertuning Parametric with PSO

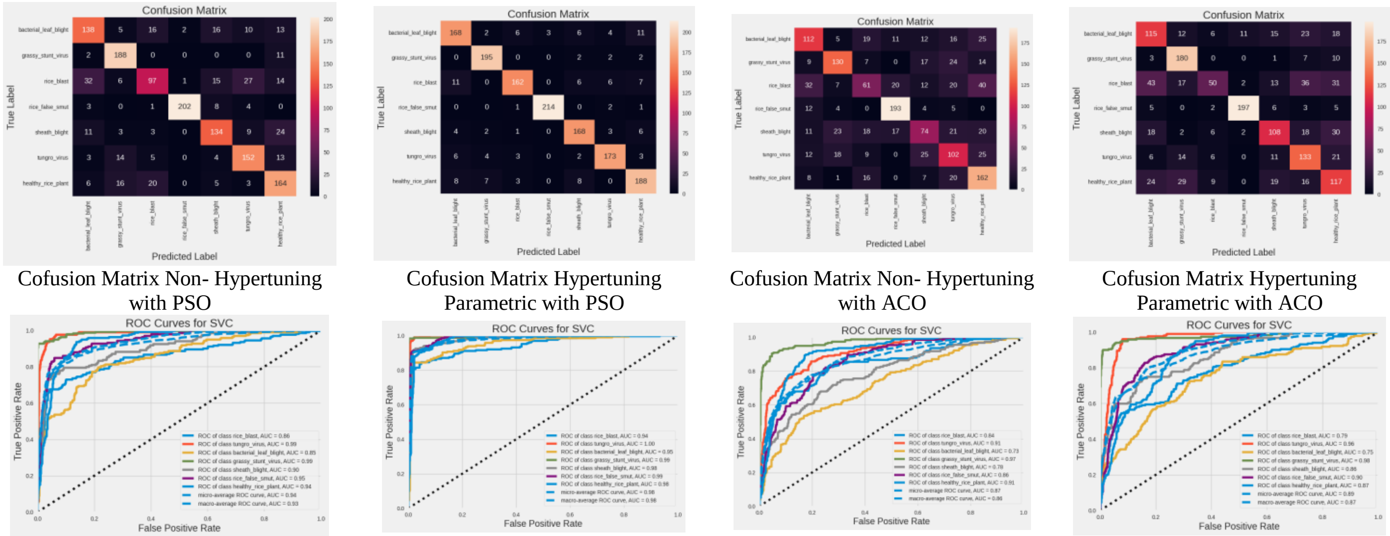


ROC AUC Non- Hypertuning with ACO



ROC AUC Hypertuning Parametric with ACO

Support Vector Classifier



Cofusion Matrix Non- Hypertuning with PSO | Cofusion Matrix Hypertuning Parametric with PSO | Cofusion Matrix Non- Hypertuning with ACO | Cofusion Matrix Hypertuning Parametric with ACO

Non- Hypertuning with PSO | Hypertuning Parametric with PSO | Non- Hypertuning with ACO | Hypertuning Parametric with ACO

Fig. 3. Performance comparison of RFC and SVC models with PSO and ACO optimization.

V. Conclusion

This study has conducted an in-depth evaluation of rice leaf disease detection using two classical machine learning classifiers Random Forest Classifier (RFC) and Support Vector Classifier (SVC) combined with handcrafted feature extraction and metaheuristic optimization techniques, including Spider Monkey Optimization (SMO), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO). The results demonstrate that integrating comprehensive visual features (color, texture, shape, and edge) with well-configured classifiers significantly boosts classification accuracy, with PSO consistently yielding the most promising performance. Notably, the SVC model, while initially underperforming, exhibited the most substantial improvement through tuning, reaching an accuracy of 94.64% with optimized parameters and PSO-enhanced feature selection, surpassing the results of RFC in its best condition. The effectiveness of hyperparameter tuning is evident across both models. Parameters such as C, gamma, and kernel selection in SVC, or max_depth, criterion, and n_estimators in RFC, proved to be pivotal in maximizing classifier performance. Metaheuristic techniques, particularly PSO and SMO, enabled a more exhaustive and intelligent exploration of these configurations compared to traditional grid search. However, ACO demonstrated variable outcomes, suggesting that further refinement or hybridization may be necessary to harness its full potential.

The implications of these findings extend beyond technical metrics. This study underscores the value of a structured, interpretable, and scalable modeling approach, which can serve as a viable alternative to deep learning methods, especially in agricultural settings where resources and data may be limited. The ability to maintain high classification performance while retaining transparency and operational flexibility makes this approach highly relevant for real-world deployment in precision farming and plant health monitoring systems. Looking ahead, there are several promising avenues for future research. First, integrating deep feature representations such as those extracted from convolutional neural networks with handcrafted features could create hybrid models that leverage both abstraction and interpretability. Second, the development of adaptive metaheuristic algorithms that self-tune based on model performance feedback could further enhance optimization reliability. Third, expanding the framework to real-time and edge-computing platforms, such as mobile-based disease detection applications, would increase accessibility for farmers and agricultural technicians. Finally, applying this optimized classification pipeline to other crops and disease categories could validate its scalability and contribute to broader food security and agricultural innovation goals. In conclusion, this research reinforces the critical synergy between data preprocessing, feature engineering, classifier tuning, and intelligent optimization. It paves the way for practical, high-performance decision-support systems that are not only accurate but also efficient and ready for deployment in real-world agricultural environments. Although some existing studies report slightly higher accuracies under limited or tightly controlled conditions, the proposed framework introduces a dual-metaheuristic and dual-classifier optimization pipeline that is validated across diverse feature configurations and tuning scenarios. This trade-off ensures better generalization in real-world agricultural applications.

Despite its promising results, this study has certain limitations. The reliance on handcrafted features may limit generalizability to other crops or conditions. Additionally, ACO's tuning instability suggests the need for further exploration into adaptive or hybridized optimization strategies. Future research could integrate deep feature representations with handcrafted features, or extend the framework to real-time applications such as mobile disease detection systems.

References

[1] Food and Agriculture Organization of the United Nations, "FAO RICE MARKET MONITOR," https://www.fao.org/markets-and-trade/commodities/rice/rmm/en.

[2] C. G. Simhadri, H. K. Kondaveeti, V. K. Vatsavayi, A. Mitra, and P. Ananthachari, "Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and

techniques," 2024, China Agricultural University. doi: 10.1016/j.inpa.2024.04.006.

[3] M. Kannan, M. M. Saad, N. Talip, S. N. Baharum, and H. Bunawan, " Complete Genome Sequence of Rice Tungro Bacilliform Virus Infecting Asian Rice (Oryza sativa) in Malaysia ," Microbiol Resour Announc, vol. 8, no. 20, May 2019, doi: 10.1128/mra.00262-19.

[4] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "Solving current limitations of deep learning based approaches for plant disease detection," Symmetry (Basel), vol. 11, no. 7, Jul. 2019, doi: 10.3390/sym11070939.

[5] B. Nagachandrika, R. Prasath, and I. R. Praveen Joe, "An automatic classification framework for identifying type of plant leaf diseases using multi-scale feature fusion-based adaptive deep network," Biomed Signal Process Control, vol. 95, Sep. 2024, doi: 10.1016/j.bspc.2024.106316.

[6] J. Padhi, K. Mishra, A. K. Ratha, S. K. Behera, P. K. Sethy, and A. Nanthaamornphong, "Enhancing paddy leaf disease diagnosis -a hybrid CNN model using simulated thermal imaging," Smart Agricultural Technology, vol. 10, Mar. 2025, doi: 10.1016/j.atech.2025.100814.

[7] A. Kurniawan, T. R. Soeprobowati, and B. Warsito, "Revolutionizing Rice Leaf Disease Detection: Next-Generation SMOREF-SVM Integrating Spider Monkey Optimization and Advanced Machine Learning Techniques," 2024. [Online]. Available: www.ijacsa.thesai.org

[8] G. P. Guru, N. K. Naveen, A. K. Jain, J. Syed, and R. U. Baig, "Convergence of improved particle swarm optimization based ensemble model and explainable AI for the accurate detection of food adulteration in red chilli powder," Journal of Food Composition and Analysis, vol. 143, Jul. 2025, doi: 10.1016/j.jfca.2025.107577.

[9] A. Kumar, V. Patidar, D. Khazanchi, and P. Saini, "Optimizing Feature Selection Using Particle Swarm Optimization and Utilizing Ventral Sides of Leaves for Plant Leaf Classification," in Procedia Computer Science, Elsevier B.V., 2016, pp. 324–332. doi: 10.1016/j.procs.2016.06.079.

[10] Y. Xiaofeng and M. Yuzhuo, "Research on ant colony algorithm based on motion optimization and adaptive threshold for pantograph slider edge detection," Measurement (Lond), vol. 245, Mar. 2025, doi: 10.1016/j.measurement.2024.116598.

[11] A. V. Panchal, S. C. Patel, K. Bagyalakshmi, P. Kumar, I. R. Khan, and M. Soni, "Image-based Plant Diseases Detection using Deep Learning," Mater Today Proc, vol. 80, pp. 3500–3506, Jan. 2023, doi: 10.1016/j.matpr.2021.07.281.

[12] S. C.K., J. C.D., and N. Patil, "Tomato plant disease classification using Multilevel Feature Fusion with adaptive channel spatial and pixel attention mechanism," Expert Syst Appl, vol. 228, Oct. 2023, doi: 10.1016/j.eswa.2023.120381.

[13] B. Nguyen, B. Xue, and P. Andreae, A particle swarm optimization based feature selection approach to transfer learning in classification. 2018. doi: 10.1145/3205455.3205540.

[14] W. B. Demilie, "Plant disease detection and classification techniques: a comparative study of the performances," J Big Data, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-023-00863-9.

[15] S. K. Lakshmanaprabu, K. Shankar, M. Ilayaraja, A. W. Nasir, V. Vijayakumar, and N. Chilamkurti, "Random forest for big data classification in the internet of things using optimal features," International Journal of Machine Learning and Cybernetics, vol. 10, no. 10, pp. 2609–2618, 2019, doi: 10.1007/s13042-018-00916-z.

[16] H. Tyralis, G. Papacharalampous, and A. Langousis, "A brief review of random forests for water scientists and practitioners and their recent history in water resources," May 01, 2019, MDPI AG. doi: 10.3390/w11050910.

[17] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," 2020, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/JSTARS.2020.3026724.

[18] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," Comput Electron Agric, vol. 150, pp. 220–234, 2018, doi: https://doi.org/10.1016/j.compag.2018.04.023.

[19] R. C. . Gonzalez and R. E. . Woods, Digital image processing. Pearson, 2018.

[20] C. Solomon and T. Breckon, Fundamentals of Digital Image Processing A Practical Approach with Examples in Matlab, 7th Edition. Boca Raton: CRC Press, 2018. doi: https://doi.org/10.1201/b18983.

[21] "Image Processing, Analysis, and Machine Vision."

[22] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans Syst Man Cybern, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.

[23] D. Zhang and G. Lu, "Shape-based image retrieval using generic Fourier descriptor," Signal Process Image Commun, vol. 17, no. 10, pp. 825–848, 2002, doi: https://doi.org/10.1016/S0923-5965(02)00084-X.

[24] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," Archives of Computational Methods in Engineering, vol. 29, no. 5, pp. 2531–2561, Aug. 2022, doi: 10.1007/s11831-021-09694-4.

[25] M. Dorigo and T. Stützle, Ant Colony Optimization. The MIT Press, 2004. doi: 10.7551/mitpress/1290.001.0001.

[26] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press, 2000. doi: DOI: 10.1017/CBO9780511801389.

[27] L. Breiman, "Random Forests," 2001.