I. J. Engineering and Manufacturing, 2026, 4, 185-210
Published Online on August 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2026.04.14

Hybrid Feature Fusion and Bayesian-Optimized Ensemble Learning for Robust Citrus Disease Detection

Nagineni Venkata Sireesha
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, KL Deemed to be University, Aziz Nagar, 500075, Hyderabad, Telangana, India
E-mail: sireeshagireesh@gmail.com
ORCID iD: https://orcid.org/0000-0003-3047-6387

Gillala Rekha*
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, KL Deemed to be University, Aziz Nagar, 500075, Hyderabad, Telangana, India
E-mail: gillala.rekha@klh.edu.in
ORCID iD: https://orcid.org/0009-0003-4933-4255
*Corresponding Author

Received: May 7, 2026; Revised: June 15, 2026; Accepted: July 3, 2026; Published: August 8, 2026

Abstract: Detecting citrus diseases at an early stage is very important for ensuring fruit quality and minimizing production losses, as well as for raising awareness about sustainable agriculture. As a solution to this problem, the authors of this paper propose a hybrid feature-based citrus disease classification system that integrates deep learning representations, handcrafted descriptors, feature selection, and ensemble learning, all of which are tuned via Bayesian optimization. We perform tests on two real-world citrus disease datasets that differ greatly in nature: a four-class lemon dataset and a two-class orange dataset. Both datasets were collected under quite different environmental conditions, so they show diverse disease symptoms and varying background complexity. Deep semantic features were obtained by running a pretrained ResNet50 network. In addition to those, other complementary handcrafted features, such as color, texture, spatial, and statistical features, were extracted from the segmented infected areas. Together, the hybrid feature vector of 2082 dimensions was subjected to an optimization method known as Neighborhood Component Analysis (NCA). This technique selects 300 features that are most effective for discrimination while at the same time ensuring the preservation of class separability and the minimization of redundancy. For the classification task, two classifiers, namely Random Forest (RF) and Bayesian-Optimized Random Forest (BORF), were employed. The latter is based on Bayesian optimization to locate the model hyperparameters. To measure the model's performance in an unbiased manner, five-fold cross-validation was performed. Based on the experimental results, BORF can improve classification accuracy on the lemon dataset from 90.42% to 93.75% and on the orange dataset from 95.42% to 95.92% compared to the baseline RF classifier. Cross-validation mean accuracies of the proposed system were 93.75 ± 0.82% and 95.92 ± 0.47% for the lemon and orange datasets, respectively. Receiver Operating Characteristic (ROC) analysis provided class-specific area under the curve (AUC) values of 0.975 and 0.972 for the orange dataset, with a macro-averaged AUC of about 0.94 for the multi-class lemon dataset. The combination of hybrid feature fusion, NCA-based feature optimization, and Bayesian-optimized ensemble classification leads to enhanced discriminative power, greater robustness, and better generalization performance for the system in citrus disease identification in a real-world agricultural setting, as demonstrated by the results.

Index Terms: Citrus disease detection; Hybrid features; ResNet50; NCA; Bayesian-optimized Random Forest

1. Introduction

Citrus fruits rank among the most popular and economically significant horticultural crops worldwide, playing a leading role in enhancing food security, agricultural sustainability, and the livelihoods of rural areas. Yet, citrus production is profoundly affected by multiple diseases that not only decline yields but also degrade fruit quality, thereby severely impacting producers' revenue. Because of this, recognizing citrus diseases accurately and promptly goes hand in

hand with the reduction of crop losses, enhancing disease control measures, and helping the agricultural sector to be more environmentally friendly [1,2]. Traditionally, the primary method for diagnosing citrus diseases has been visual inspection by humans. This method requires highly trained plant pathologists and agricultural experts. Even though it can yield reliable results mainly in a lab setting, it is extremely exhausting, time-consuming, subjective, and simply cannot meet large-scale requirements. Moreover, environmental changes and human factors significantly affect diagnostic accuracy, rendering conventional inspection unsuitable for the demands of modern agriculture, which centers on precision [3,4].

Thanks to the boom in artificial intelligence, computer vision, and deep learning, the scenario of auto-diagnosis of plant diseases has totally changed. Among deep learning models, convolutional neural networks (CNNs) have demonstrated strong performance in plant image understanding [5,6]. It has been used to great effect in various works involving deep learning for citrus disease detection, citrus fruit identification, pest detection, forestry monitoring, output prediction, and smart agricultural management, both in the lab and the field [7-20]. Besides that, the introduction of enhanced transfer learning techniques, ultra-lightweight detection models, edge AI, XAI, and multimodal learning has greatly facilitated the era of intelligent systems for disease diagnosis in citrus plants [21-26].

Several problems remain unaddressed. For example, the symptoms of different diseases can be so similar that they are almost indistinguishable to the naked eye, while the same disease can show different symptoms at different stages of plant maturity. Plus, changes in lighting, occlusion, complex backgrounds, different camera angles, and the ever-changing environment can all have a major impact on image quality and, ultimately, on how well the images are classified. Although deep learning models can identify high-level semantic features, they might miss out on subtle color, texture, and local structural cues that make a big difference when it comes to distinguishing between visually similar disease classes. However, deep feature representations are often redundant and highly correlated, so they tend to raise computational complexity, while at the same time making the classifier less capable of generalizing well [27].

To address issues in feature extraction using deep learning alone, hybrid feature extraction techniques that combine deep semantic representations with handcrafted descriptors have attracted considerable research interest recently. More precisely, they bring together deep neural networks' incredible ability to extract representations and feature sets comprising colour, texture, and spatial information, which complement learning to detect disease symptoms more effectively. However, the combined feature space of deep and handcrafted descriptors is typically very high-dimensional and redundant, requiring a powerful feature selection method to retain only the most discriminative features and improve classification results. Neighborhood Component Analysis (NCA) has not only been effective in learning feature transformations that maximize classification performance but also in eliminating redundant information at the same time [28,29]. Also, the selection of a classifier and its hyperparameter tuning are essential to achieving high predictive performance. Ensemble learning methods have consistently been shown to be more robust and to generalise better for high-dimensional classification problems, whereas Bayesian optimisation enables the identification of near-optimal parameter settings with far fewer evaluations than exhaustive parameter searches [30,31].

Driven by these reflections, the present paper develops a Hybrid Feature Fusion Neighbourhood Component Analysis Bayesian Optimized Random Forest (Hybrid Feature Fusion + NCA + BORF) method for classifying diseases in citrus fruits. The method features a blend of deep semantic features obtained via ResNet50 and manual color, texture, and spatial descriptors that are considered to present the full characteristics of citrus diseases. After that, Neighbourhood Component Analysis is run to select the best features and reduce redundancy, and a Bayesian-Optimized Random Forest is applied to improve accuracy, robustness, and generalisation. Two citrus disease datasets, circular lemon and orange fruits and images of them captured in different environmental conditions, are used to test the effectiveness of the proposed method. Experiments have shown that in complex real-world situations, the setup developed in this research is highly accurate, reliable, and robust for disease classification.

The major contributions of this work are summarized as follows:

1. A texture–spatial guided segmentation strategy is introduced to improve localization of diseased citrus regions prior to feature extraction.
2. A hybrid feature representation framework combining ResNet50 deep semantic features with handcrafted color, texture, and regional descriptors is developed to capture complementary disease characteristics.
3. Neighborhood Component Analysis (NCA) is employed to identify the most discriminative features and reduce redundancy.
4. Bayesian optimization is integrated with the Random Forest classifier to adaptively determine optimal hyperparameter settings and improve classification generalization.
5. A comprehensive ablation study is conducted to evaluate the contribution of individual framework components and validate the effectiveness of the proposed architecture.

2. Related Work

Plant diseases are among the leading causes of declines in crop yield and quality worldwide. They also remain one of the biggest hurdles to sustainable agricultural production. To address this issue, researchers aim to develop new systems that can accurately, quickly, and automatically diagnose diseases. Long ago, plant diseases were detected mainly through

manual visual inspection of the plants by skilled agricultural personnel. Even though this method works well when the conditions are controlled, it becomes labor-intensive, subjective, and time-consuming, besides being unsuitable for monitoring large areas. The rapid progress in computer vision and artificial intelligence (AI) has transformed the process of diagnosing plant diseases by enabling automated image analysis and classification. In fact, deep learning has significantly transformed the detection of plant diseases by learning distinguishing features directly from raw images, which means that extensive manual feature engineering is no longer required.

Initially, computer vision relied heavily on handcrafted features such as color histograms, texture descriptors, shape features, and statistical measures for isolating diseased plant regions. Although these methods performed quite well under controlled laboratory conditions, their performance declined sharply when tested in real agricultural settings, primarily due to changes in lighting, complex backgrounds, occlusions, and overlapping signs of different diseases [1-4]. The introduction of Convolutional Neural Networks (CNNs) has led to a breakthrough in the recognition of plant diseases, as they can learn more discriminative feature representations without human intervention. As a result, CNN-based approaches have shown great success in classifying plant diseases and have significantly reduced reliance on manually designed features [5,6].

Alongside the deep learning research on different areas of agriculture like citrus disease diagnosis and smart farming, a lot of efforts have been put into the development of smart systems for the recognition of citrus diseases, detection of fruits, robotic harvesting, monitoring of orchards, grading of fruits, pests' identification, and estimation of yields both under laboratory and field conditions [7-15]. Besides, transfer learning has proven to be an effective method for repurposing pre-trained deep neural networks for agricultural image classification tasks, particularly when only a limited number of annotated datasets are available. Various architectures, such as VGG, ResNet, DenseNet, EfficientNet, Vision Transformers, and anchor-based CNN models, have performed exceptionally well in citrus disease classification [16,17,18-20]. Deep learning This way is a fundamental technology that contributes to intelligent crop monitoring and precision agriculture.

Besides identifying diseases, a few studies also show deep learning being applied to count citrus fruits, manage orchards, perform edge computing and make intelligent decision-support systems. By integrating lightweight detection models, mobile AI, and edge computing, it has become feasible to run disease-diagnosis systems in real time, even in agriculture. Elaborate review articles have further highlighted how AI and computer vision-based technologies are taking smaller roles in agricultural automation and sustainable crop management [21,22].

Changes in classification to make them even more reliable through multimodal learning, explainable AI, lightweight object detection, transfer learning, and ensemble learning have been the center of attention in the recent literature. These methods have made a great change in identifying the diseases that might be missed by the models in the field of experimentation; at the same time, the methods made the model more interpretable, and the more the model was able to be generalized, the changes that were made improved the deployment efficiency [23-26].

Deep learning algorithms are highly capable of extracting high-level semantic information; they can still ignore small differences in colour, texture, and local microscopic characteristics, which are essential for the visual identification and discrimination of disease classes with very similar appearances. This has led several researchers to recommend hybrid feature extraction methods that integrate deep semantic features with handcrafted descriptors to leverage complementary visual information. Currently, such hybrid representations not only lead to a more detailed disease characterization but also exhibit a much better classification performance than either of the individual feature extraction methods alone [27].

Feature selection and classifier optimization play the same roles as model training and deep learning in the developing direction of powerful disease classification systems. Representing features in very high-dimensional space inevitably introduces a lot of redundant and irrelevant information, which in turn increases the classifier's burden and results in worse classification performance. Being a supervised feature selection technique, Neighborhood Component Analysis (NCA) is very powerful in maximizing class separability and at the same time disintegrating redundant features [28]. Random Forest, an ensemble learning technique, has been the champion in demonstrating robustness and generalisation for very high-dimensional classification problems [29]. Bayesian Optimisation is among the best methods for automatically optimising hyperparameters, demonstrating great efficiency in finding near-optimal parameter settings and drastically reducing the computational cost of exhaustive search methods [30,31].

2.1. Research Gap

Although significant advances have been made in diagnosing citrus diseases, some major issues remain unresolved. First off, many deep learning models mainly focus on the deep semantic feature aspect but fail to make good use of the additional complementary cues from color, texture, and spatial handcrafted features [23-27]. Second, methods that fuse hybrid and multimodal features usually produce higher-dimensional feature sets that contain a lot of redundant information, incurring high computational cost and, as a result, leading to lower classification efficiency [27,28]. Third, while many works have concentrated on feature extraction, only a few have emphasised feature selection to identify the features that contribute most to discrimination and eliminate redundant information [28]. Last, changing the classifier's hyperparameters properly will drastically affect its performance; Still, there are many cases where adaptive hyperparameter optimization methods have not been sufficiently investigated in the structures of citrus disease classification [29-31].

To address these issues, this paper puts forward a complete citrus disease classification structure that integrates hybrid feature fusion, Neighborhood Component Analysis-based feature selection, and Bayesian Optimization-driven Random Forest classification. In essence, this setup firstly fuses deep semantic features obtained from a ResNet50 together with handcrafted color, texture, and spatial descriptors to get complementary feature representations. Then, Neighborhood Component Analysis is used to discard redundant features and keep the most discriminative ones, whereas Bayesian Optimization is employed to determine the best Random Forest parameters automatically. This combined structure is designed to enhance accuracy, speed, robustness, and the ability to generalize well in actual agricultural conditions. Table 1 shows a detailed comparative analysis of existing studies.

Table 1. Comparative study of different models reported in the literature

Ref.	Year	Method / Model	Dataset / Application	Main Contribution	Limitation
[1]	2019	CNN	Plant leaf diseases	Automatic plant disease identification	General plant dataset; not citrus-specific
[2]	2021	Deep CNN	Citrus fruit & leaf diseases	Citrus disease detection	High computational complexity
[3]	2021	Deep Learning Review	Fruit yield estimation	Survey of DL techniques	Review only; no classification framework
[4]	2022	Deep CNN	Citrus fruit detection	Real-time fruit detection and tracking	Detection only; no disease classification
[5]	2019	Improved CNN	Apple leaf diseases	Accurate disease classification	Limited to apple leaves
[6]	2025	AgriNet	Citrus disease recognition	Deep learning-based citrus diagnosis	Limited dataset size
[7]	2025	CNN + Vision Transformer	Plant diseases	Hybrid CNN-ViT framework	No feature optimization
[8]	2022	Anchor-based CNN	Citrus diseases	End-to-end citrus disease classification	Relies only on deep features
[9]	2023	Deep Learning	Fruit recognition	Robotic fruit picking	Focuses on harvesting
[10]	2023	Adaptive CNN	Citrus grading	Automated grading system	Grading only
[11]	2023	Review	Fruit detection	Comprehensive survey	No experimental validation
[12]	2023	Edge AI + CNN	Citrus diseases	Edge-based disease detection	Resource-constrained deployment
[13]	2023	YOLOv4	Citrus detection	Real-time fruit detection	Detection only
[14]	2023	Transfer Learning CNN	Citrus pests	Pest recognition	Does not classify diseases
[15]	2023	Deep Learning	Citrus yield	Yield prediction	Not disease diagnosis
[16]	2023	Mobile Edge AI	Citrus orchards	Disease mapping	Mapping rather than classification
[17]	2023	Image–Text Multimodal Fusion	Citrus diseases	Multimodal disease recognition	High computational cost
[18]	2023	Transfer Learning	Citrus diseases	Improved disease classification	No feature selection
[19]	2024	HHS-RT-DETR	Citrus greening	Lightweight disease detection	Single disease focus
[20]	2024	Transfer Learning + Ensemble	Citrus leaves	Improved classification	Uses only deep features
[21]	2025	Lightweight CNN	Citrus detection	Fruit detection and counting	Detection task only
[22]	2025	Systematic Review	Fruit detection	Review of DL methods	No new framework
[23]	2025	Advanced Deep Learning	Citrus diseases	Improved disease recognition	Computationally expensive
[24]	2025	XAI + Multimodal DL	Fruit diseases	Explainable diagnosis	General fruit datasets
[25]	2025	3D Vision + DL	Citrus grading	Automatic grading	No disease diagnosis
[26]	2025	Segmentation CNN	Green citrus	Fruit segmentation	Segmentation only
[27]	1989	Unsupervised Learning	Representation learning	Foundation of feature learning	Not specific to agriculture
[28]	2005	Neighborhood Component Analysis (NCA)	Feature selection	Supervised dimensionality reduction	Does not include classification
[29]	2006	Ensemble Learning	Classification	Robust ensemble methodology	No hyperparameter optimization
[30]	2010	Bayesian Optimization	Hyperparameter tuning	Efficient optimization framework	Generic optimization method
[31]	2012	Practical Bayesian Optimization	Machine learning	Practical BO implementation	General ML application

3. Data Collection and Methodology

3.1. Acquisition of Data

For our study, we sourced the citrus fruit dataset from the Food and Agriculture Organization of the United Nations (FAO), a premier and reliable source for global agricultural statistics. It mainly covers two types of citrus fruit: lemon and orange, which differ in their color, shape, and size. This dataset is quite useful for developing and testing a machine

learning model for citrus fruit recognition. We used image processing techniques to extract image features for model training and to maintain uniform image quality throughout the process.

3.1.1. Dataset Details

These two types of datasets were analyzed in this study as described below:

Dataset -1: Lemon dataset consists of 208 images (4 classes: one healthy and three diseased types like Canker, Mold, Scab) with an image size of 512 x 512.

Dataset - 2: Orange dataset consists of 2240 images of fruits, of 2 diseased classes: Black spot and Canker, all with an image size of 800 x 800 pixels.

One of the reasons we decided to merge the datasets was the differences in scope and level of detail between them; merging them would facilitate a more comprehensive evaluation of the proposed methods. The images show the object under various atmospheric conditions (fog, warm, cold). To ensure the collection would be not only diversified but also representative, a well-thought-out strategy was developed, thereby increasing its value. The exact categorization and the number of images in each class are presented in Table 2.

Table 2. Dataset Description (Raw Data)

Dataset	No. of images	Pixels
Lemon Dataset	208	512x512
Orange Dataset	2240	800x800

3.1.2. Details of Datasets

In-depth information on all datasets used is included in this section.

Case 1: Lemon Dataset

The images belong to 4 visually distinct classes: Healthy Lemon (50 images), Lemon Canker (52 images), Lemon Mold (54 images), and Lemon Scab (52 images), for a total of 208 raw images. These images were taken at various brightness levels. Hence, these pictures naturally differ in lemon size, texture, lesion pattern, and background diversity. The dataset based on a single fruit enables us to perform an accurate disease condition analysis, and we supply pictures in a resolution of 512 × 512 pixels×. Disease description: Lemon Canker, Spots of infection on the leaves and fruits are lemon brown in color with yellow halos around them; lemon Mold is the development of a fuzzy fungal growth on the surface; scab is the formation of crusty wart-like patches breaking up the peel. The symptoms analyzed in this dataset have a wide range of variations, whereas a healthy lemon is bright and smooth. Table 3 presents the statistics for the different classes in the Lemon dataset.

Table 3. Lemon Dataset Description

Type	Class Name	No. of collected images
Class 1	Healthy	50
Class 2	Canker	52
Class 3	Mold	54
Class 4	Scab	52
Total		**208**

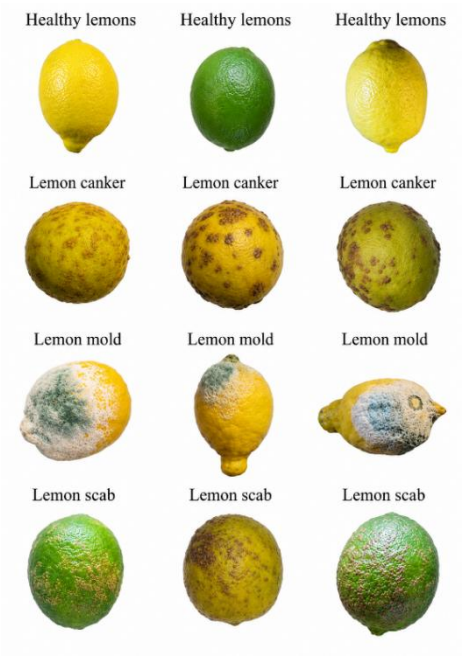
Case 2: Orange Dataset

The Orange Dataset includes 2,240 images, evenly split between the two major citrus disease categories: Black Spot (1,120 images) and Citrus Canker (1,120 images). The images were taken at different farm orchards, so different scenarios like backgrounds, lighting, and fruit positioning are naturally included. All the images were at a resolution of 800x800 pixels, that's why the details of the lesions are quite clear for precise visual learning. That assists in the diagnosis of the disease is the spots showing symptoms on Black Spot; these are circles of dead tissue, usually encircled by yellow halos. Citrus Canker is marked by lesions soaked with water that have a central raised scab and yellow rings going outwards. Apart from forming a major part of the Lemon Dataset, the images of the citrus diseases present a good and visually rich tool to the training of disease classifiers. Table 4 describes each class plus their morphological variations and the issues that are faced when classified in real orchard environments.

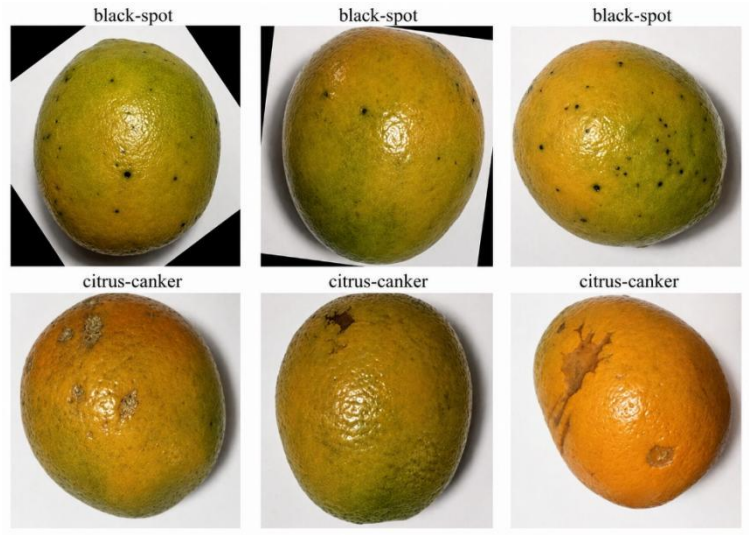
Table 4. Orange Dataset Description

Type	Class Name	No of images
Class 1	Black spot	1120
Class 2	Canker	1120
Total		2240

Sample images from the Lemon and Orange datasets are shown in Figure 1 below.



(a)



(b)

Fig. 1. Sample images of Datasets (a) Lemon (b) Orange

3.2. Proposed Methodology

The technique we suggest is a well-designed multi-stage learning pipeline, as shown in Figure 2. The whole system is a meticulous piece of work assembled point by point, with every element not only strengthening the others but also sharpening their distinctiveness. At the very beginning, images are randomly selected from the dataset, and data augmentation is then applied to address class imbalance and increase diversity. Because of this, both deep and handcrafted features get the chance to be understood from different visual perspectives, which in turn results in the ability to generalize better. While traditional pipelines are typically used in this method, a preprocessing step and an image segmentation step

are explicitly added before extracting handcrafted features. Notably, rather than treating the images in their original form, the framework identifies and isolates diseased portions of the fruit; it is more precise, and the features it extracts are more meaningful. This kind of processing, in turn, allows the subsequent stages of feature extraction and classification to be more effective.

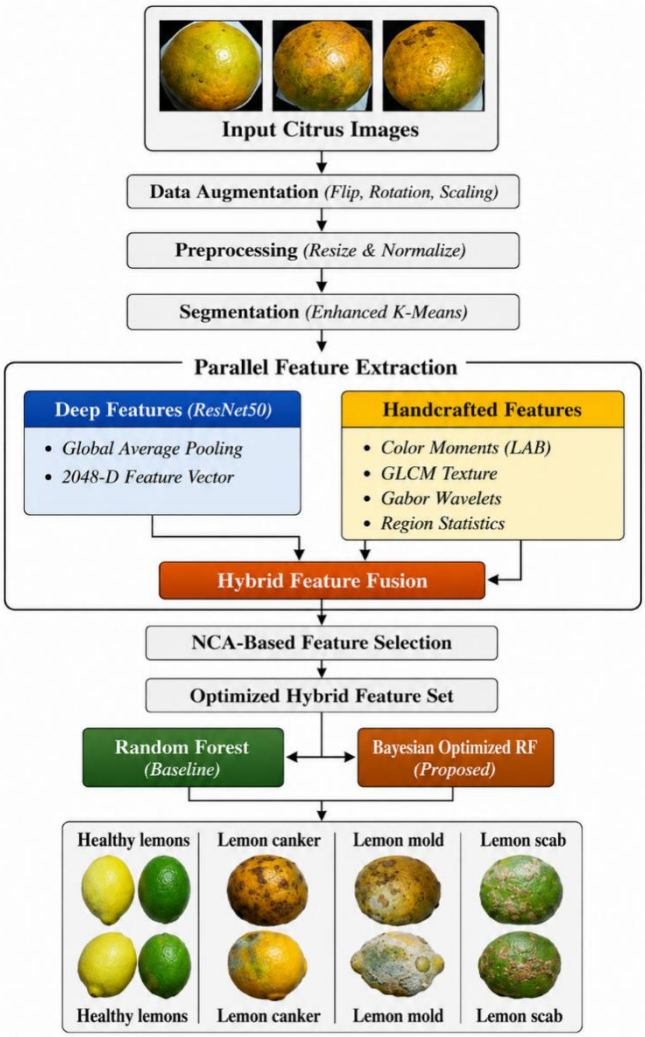


Fig. 2. Novel Deep Learning-Based Hybrid Feature Extraction Method for Citrus Disease Prediction

Unlike conventional sequential pipelines that directly concatenate multiple feature sources, the proposed framework progressively refines representations. Segmentation focuses on the elimination of irrelevant background, it is followed by complementary feature learning, supervised feature selection, and adaptive optimization. This sequential interaction not only reduces redundancy but also increases a classifier's discriminative power and robustness.

Most probably, the main point of the work is the learning technique's parallel feature. Firstly, ResNet50 is utilized to obtain non-localized high-level representations (deep features), while learned handcrafted features (texture, connected neighborhood, and statistical measures) are used to detect localized disease characteristics that DNNs mostly miss. This double-path representation learning yields more comprehensive representations. Moreover, the NCA-based feature selection of the concatenated features. Rather than directly concatenating features, NCA automatically selects discriminative features from a concatenated pool of deep and hand-crafted features, thereby reducing data dimensionality and redundancy while maximizing class separability. Finally, we perform classification using a Random Forest and a Bayesian-optimized Random Forest classifier, in which the Bayesian optimization automatically tunes hyperparameters to maximize AUC.

3.2.1. Preprocessing and Augmentation

Data augmentation was applied to the Lemon and Orange datasets to reduce class imbalance and improve model generalization. All the original images were augmented using five rotation angles and horizontal and vertical flips, and 7 new samples were generated for each transformed image. The lemon dataset has only 208 images, so all the models were built to generate more data that better reflects real-world variation. Three augmentation methods were consistently

employed across the two datasets: horizontal and vertical flips for orientation normalization, random tilting by ±35 degrees to provide variations in viewpoint orientation, and scaling within the 0.8–1.2 range to handle scale variation due to registration distance. After augmenting the Lemon dataset, the total number of images is 1,600, with 400 images per class. The Orange dataset, which initially had more images than the other datasets, was augmented to 2k images per class, resulting in a total of 4000 images. These changes introduced a good variety to the data while preserving disease features. The augmented datasets are summarized in Table 5. After augmentation, the datasets were randomly split into training (70%) and test (30%) sets, stratified to maintain the class ratio. This led to 2,800 and 1,200 images for the orange dataset (training and testing) and 1,120 and 480 images for the Lemon dataset, respectively. Table 6 shows how these splits enable proper evaluation and balanced learning for the model. All images were downsampled to 224 × 224 pixels before inputting them into the pretrained CNN. These were rescaled to [0, 1] and saved as float32, as shown in Table 7. Differences in appearance between normal and abnormal highlights were preserved by leaving the images in RGB (colour) rather than converting them to grayscale.

Table 5. Augmentation on two datasets

Dataset	Augmentation Methods	Original Images	Augmented Images per Class	Total Images After Augmentation
Lemon	Flip, Rotation (±35°), Zoom (0.8–1.2)	208	400	1,600
Orange	Flip, Rotation (±35°), Zoom (0.8–1.2)	2,240	2,000	4,000

Table 6. Train-Test Split of Two Datasets

Dataset	Training Images (70%)	Testing Images (30%)	Train-Test Ratio
Lemon	1,120	480	70:30
Orange	2,800	1,200	70:30

Table 7. Preprocessing for two datasets

Dataset	Final Resolution	Color Mode	Normalization	Input Format
Lemon	224 × 224	RGB	Min-Max (0–1)	float32
Orange	224 × 224	RGB	Min-Max (0–1)	float32

3.2.2. Texture and Spatial Features to Improve K-Means Segmentation

Correct segmentation of diseased regions in disease images is an important step in extracting features for disease detection. Traditional K-Means clustering only considers pixel values in the assignment; that is, pixels with similar intensities are assigned to the same cluster. So, it is very sensitive to illumination changes, background noise, and complex surface textures often found in fruit images taken under natural conditions. In the first place, to eliminate these effects, our primary work was to enhance the standard K-Means segmentation by incorporating not only texture features but also spatial context in the clustering process. Breaking down K into three groups: cancerous tissue, normal breast tissue, and non-breast tissue. At a higher level, this feature representation should create a clearer distinction between cancerous and healthy regions.

Firstly, the input image is split into luminance and chrominance components to capture a more diverse range of features. A single-channel image is used to extract texture features, while the Lab* color space retains the perceptual chromatic layout. For a pixel, a feature vector comprising color, texture, and spatial features is assembled. Then, the pixel feature vector that has been upsampled with Eq (1).

$$\mathbf{X_i} = [L_i, a_i, b_i, T_i, x_i, y_i] \tag{1}$$

where L_i, a_i, and b_i represent the colour components of a pixel i in the Lab colour space, T_i denotes the texture descriptor, and (x_i, y_i) represent the spatial coordinates of the pixel. This way of describing a pixel allows the clustering algorithm to consider both appearance similarity and spatial closeness simultaneously.

We use Grey Level Co-occurrence Matrix (GLCM) statistics to extract textural features. They are obtained from a local neighborhood window centered on each pixel. The GLCM matrix calculates the joint probability of a pixel pair with the Grey Level values at a specific spatial displacement and direction. From this matrix, one can derive several secondary texture features, such as contrast and homogeneity. For instance, the contrast measure might be written as in Eq (2).

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

and homogeneity is computed with Eq (3).

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i,j)}{1+|i-j|} \tag{3}$$

Texture features include roughness and lesion irregularity, which are the main symptoms of citrus diseases.

To keep neighboring pixels together and not to misclassify them individually, spatial coordinates are added to the feature space for clustering. This is a way to increase the likelihood of not breaking neighboring pixels during segmentation, thereby producing a continuous segmentation boundary. The modified Euclidean distance formula to represent the class occurrence clustering using Eq (4).

$$D(i,k) = \sqrt{\sum_{d=1}^{m} w_d \, (X_{id} - C_{kd})^2} \tag{4}$$

Here, X_{id} refers to the d^{th} feature component of pixel i, C_{kd} denotes the centroid group k value, w_d indicates the weight of the features used with each clustering method, and m is the total number of features. We assign higher weights to the texture and color components, as they are the features most closely related to the disease at a later stage in the clustering process.

The improved K-Means algorithm is to find k clustering centers of the feature space through iterative processing using the objective function given in Eq (5).

$$J = \sum_{k=1}^{K} \sum_{i \in C_k} \| \mathbf{X}_i - \boldsymbol{\mu}_k \|^2 \tag{5}$$

where μ_k The centroid of cluster C_k, incorporating spatial and textural properties, transforms the boundary between clusters to more closely approximate actual disease regions rather than relying solely on intensity segments.

This improved segmentation process has several advantages. It removes the need for sensitivity to illumination changes. In addition, color and texture information are used to improve discrimination between healthy and diseased tissues, and this even helps visualize phenomena previously understood only through cellular investigations. Other enhancements include spatial regularization technology applied to the boundary of an image feature, with reduced computational overhead. This means that once boundaries are set, they will remain in place and not shift (though this may take some time). Thirdly, it is more robust than ever before at discriminating against visually similar disease states. This makes better sense of the way crop images look, just as even naive observers can readily identify several different patterns from a single modern machine model. As a result, the segmented regions of a crop may serve as a more accurate representation of the infected regions. This significantly improves subsequent feature extraction and downstream classification confidence by providing more reliable data for those tasks in future runs. The citrus image was segmented using the proposed texture–spatial guided K-Means clustering algorithm, which integrates color, texture, and spatial information to generate three distinct clusters corresponding to diseased regions, healthy tissue, and background. The resulting segmented masks are illustrated in Figure 3. For k = 3 clusters, representative segmentation results are shown in Figure 4.

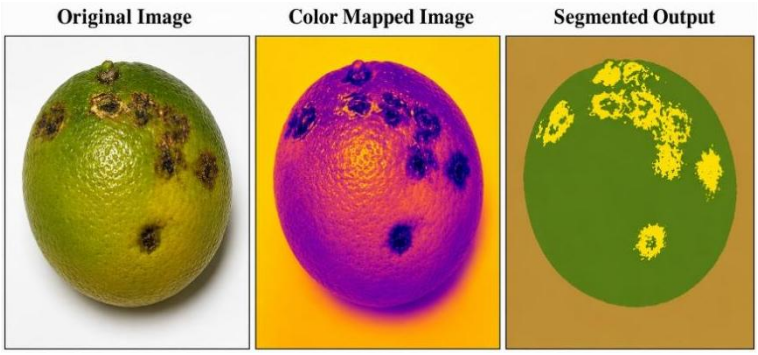


Fig. 3. K-Means segmented labelled image with additional pixel information

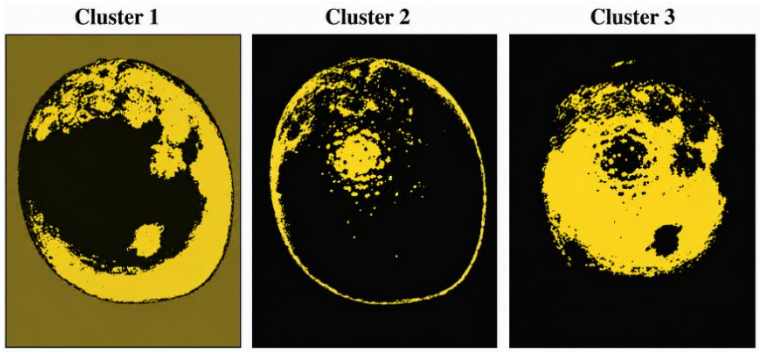


Fig. 4. Clusters of K-Means segmented images

3.3. Hybrid Feature Extraction Methodology

To obtain robust, discriminative representations of real-world citrus disease patterns, this research employs a parallel hybrid feature-extraction system. Through this parallel system, the proposed method integrates deep semantic features with handcrafted low-level descriptors to exploit both the advantages of learning from hierarchical representations and those offered by specific visual characteristics. This double-pronged approach has the effect of greatly improving the reliability of classification, especially when faced with situations where there is high similarity between classes of objects; the lighting conditions are different from one shot to another, even within a single visual, and environmental noise continues for any length of time as part of the background cacophony. In the first step of the deep feature extraction, input citrus images are resized to 224×224 pixels and then normalized so that each pixel value is in the range -1 to 1. Each pre-processed image with normalized pixel values is subsequently fed into a pre-trained convolutional neural network to obtain its high-level semantic representation. Instead of using the final category layer, the global average pooling layer is used to obtain a highly compact feature vector. The mathematical expression of the deep feature representation can be written as in Eq (6).

$$\mathrm{F}_d = \mathrm{GAP}\big(\Phi(I)\big) \tag{6}$$

In the above equation, as in the convolutional layer, the nonlinear mapping of the feature is obtained. $\Phi(I)$, and GAP$(\cdot)$ represents, after the operation is completed, all its gloss elements as a {dotted surface array}. The deep feature vector $\mathrm{F_d} \in \mathrm{R}^{2048}$ obtained contains abstract information about the disease, such as lesion boundaries, color irregularities, and texture patterns. Besides, the system extracts a set of features in a manner very similar to a human expert carefully examining the image to identify subtle characteristics. The first step is to convert the input image from RGB to Lab*, a perceptually uniform color space. Next, the statistics included color moments. The channels and the statistical distribution of each channel in phase space were characterized by the mean, variance, and skewness. The k^{th} moment for the color channel C is the average over all pixels at that Gray Level can be computed with Eq (7).

$$M_k = \frac{1}{N}\sum_{i=1}^{N}(C_i - \mu)^k \tag{7}$$

where N is the number of pixels, C_i denotes pixel intensity values, and μ represents the mean intensity.

To obtain texture information, the GLCM (Grey Level Co-occurrence Matrix) captures the spatial relationships among pixels in a 2D image. Given a displacement vector (d, θ), the GLCM is defined as in Eq (8).

$$P(i,j) = \frac{\text{Number of occurrences of intensity pair } (i,j)}{\text{Total number of pixel pairs}} \tag{8}$$

Then, the matrix is used to calculate statistical descriptors such as contrast correlation energy and homogeneity to characterize surface roughness and lesion heterogeneity.

Besides, Gabor wavelet filters, which can decompose directional fibers and frequency components, are used to enhance texture representation. X-Gabor filter response in 2D at location x, y is given in Eq (9).

$$G(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(\frac{2\pi x'}{\lambda}\right) \tag{9}$$

where x' and y' represent rotated spatial coordinates, σ controls the Gaussian envelope, λ denotes wavelength, and γ defines spatial aspect ratio.

Additionally, regional statistical features are calculated for the segmented disease regions not only to capture spatial structure but also to capture log-intensity values. These descriptors consist of area, entropy, and intensity distribution measures, which together characterize the lesions more extensively. Then again, in parallel extraction, hybrid feature fusion merges deep and handcrafted features for a single representation. The fused feature vector is formed simply by concatenation as seen in Eq (10).

$$\mathrm{F}_h = [\mathrm{F}_d \;\; || \;\; \mathrm{F}_c] \tag{10}$$

where F_d represents deep features, F_c denotes handcrafted features, and || indicates vector concatenation. The combination is great for hierarchical life representation, but it also adds high-dimensional, redundant information. Both may lead to a decrease in classification performance.

To fix this issue, we used the Neighborhood Component Analysis (NCA), a tool for feature selection and identification of truly influential features. NCA finds a transformation matrix that projects the (fused) feature vector onto a lower-dimensional space using the Eq (11).

$$\mathrm{z}_i = \mathrm{AF}_{h,i} \tag{11}$$

The probability that the sample i selects sample j As its neighbor is defined as in Eq (12).

$$p_{ij} = \frac{\exp(-\| \mathrm{z}_i - \mathrm{z}_j \|^2)}{\sum_{k \neq i} \exp(-\| \mathrm{z}_i - \mathrm{z}_k \|^2)} \tag{12}$$

The objective of NCA is to maximize the expected classification accuracy given in Eq (13).

$$\max_{\mathrm{A}} \sum\nolimits_i \sum\nolimits_{j \in C_i} p_{ij} \tag{13}$$

where C_i represents the set of samples belonging to the same class as the sample i. This optimization focuses on discriminative dimensions without enhancing redundant or noisy features.

Neighborhood Component Analysis (NCA) was employed as an embedded feature-selection technique to identify the most discriminative features in the fused feature space. NCA learns feature importance weights by maximizing the expected leave-one-out classification accuracy through gradient-based optimization of a feature transformation matrix. The hybrid feature representation combines 2048 deep semantic features extracted from the global average pooling layer of the pre-trained ResNet50 model with handcrafted color, texture, and spatial descriptors derived from the segmented citrus disease regions, yielding a 2082-dimensional feature vector. By eliminating redundant and less informative attributes while preserving the characteristics that separate classes, NCA ultimately reduces each sample's high-dimensional feature representation to the 300 most discriminative ones. This dimensionality reduction leads to higher computational efficiency, better classifier generalization, and higher-class separability, which in turn makes the citrus disease classification more robust to different environmental conditions.

4. Proposed Classification Algorithms

For classification, the researchers used a reliable ensemble-based classification network. They decided to use the Random Forest (RF) technique as the main method, given its excellent generalization and anti-overfitting capabilities, its ability to outperform other similar methods, and its smooth handling of mixed image representations. Still, the performance of RF is hugely dependent on its hyperparameter tuning. Considering this, and to improve classification performance, a Bayesian Optimized Random Forest (BORF) has been developed, combining the RF classifier with Bayesian optimization techniques. While a grid or random search over the hyperparameter space is the more traditional way of tuning, Bayesian optimization goes further as it probabilistically models the objective function and smartly explores the hyperparameter space. This results in faster identification of optimal parameter settings while requiring less computational power. Besides, it is very important for improving accuracy in recognizing citrus diseases while providing stability and reliability, since RF with Bayesian optimization makes it easier to directly focus on results that agree between the electrode output and plant properties.

4.1. Random Forest Algorithm

Random Forest is a technique that applies ensemble learning. In Random Forest, different decision trees are constructed by using randomly selected training data and features. The class label to which the prediction falls is determined by voting, and averaging is then used to dramatically reduce variance and prevent overfitting.

If T Decision trees are used the predictions can be obtained using Eq (14).

$$\hat{y} = \text{mode}\{h_t(\boldsymbol{x})\}_{t=1}^{T} \tag{14}$$

Where:
$h_t(\mathbf{x})$= prediction of the t^{th} tree
mode () = majority vote

The Random Forest classification process is summarized as follows:

Step 1: Data Augmentation

Apply geometric and intensity transformations to each input image to generateI_i augmented samples using Eq (15).

$$\tilde{I}_i = \mathcal{T}(I_i) \tag{15}$$

Step 2: Image Preprocessing and Segmentation

Resize all augmented images to 224×224, normalize pixel values, and perform image segmentation to extract the region of interest S_i.

Step 3: Deep Feature Extraction Using ResNet50

Pass each segmented image through a pre-trainedS_i ResNet50 model and extract deep features from the global average pooling layer using Eq (16).

$$\mathbf{f}_i^{(d)} \in \mathbb{R}^{2048} \tag{16}$$

Step 4: Handcrafted Feature Extraction

From each segmented image S_i, extract:

- Texture features (GLCM-based)
- Connected region features
- Statistical features (mean, standard deviation, entropy)

From the handcrafted feature vector in Eq (17).

$$\mathbf{f}_i^{(h)} \in \mathbb{R}^{D_h} \tag{17}$$

Step 5: Hybrid Feature Construction

Concatenate deep and handcrafted features using Eq (18).

$$\mathbf{f}_i^{(hyb)} = \left[\mathbf{f}_i^{(d)} \parallel \mathbf{f}_i^{(h)}\right] \tag{18}$$

Step 6: NCA-Based Feature Selection

Apply Neighborhood Component Analysis (NCA) to reduce dimensionality and retain discriminative features with the Eq (19).

$$\mathbf{z}_i = \mathbf{A}\mathbf{f}_i^{(hyb)}, \mathbf{z}_i \in \mathbb{R}^k \tag{19}$$

Step 7: Random Forest Classification

Train a Random Forest classifier using the selected features. $\mathbf{z}_i$.
For each tree $t = 1,2, \ldots, T$:

1. Generate a bootstrap sample
2. Grow a decision tree using random feature selection at each node
3. Split nodes using the Gini impurity criterion

Final prediction is obtained via majority voting using Eq (20).

$$\hat{y}_i = \arg\max_c \sum_{t=1}^{T} \mathbb{I}(h_t(\mathbf{z}_i) = c) \tag{20}$$

Step 8: Performance Evaluation

Evaluate the RF classifier using: "Accuracy", " Precision", " Recall", and " AUC"

The NCA-selected hybrid feature set is used to train the RF classifier. The hybrid set includes deep features extracted from ResNet50, as well as handcrafted texture and color descriptors. The RF model benefits from this combination, as it can utilize both high-level semantic information and low-level discriminative patterns, thereby enhancing classification performance.

4.2. Bayesian Optimized Random Forest Algorithm

Initially, we suggested a random forest enhanced with Bayesian optimization (BORF) for the first version of our setup, aiming to increase the base random forest (RF) model's performance through hyperparameter tuning. RF, a well-known aggregating method, is based on constructing many decision trees, which leads to a higher generalization ability of the resulting model. Besides, one of the key elements that determines the RF classification power is selecting proper hyperparameters. In other words, we can say that the main feature of our structure is hyperparameter tuning, which is Bayesian optimization, a cutting-edge technique that can greatly reduce the computational resource requirements compared to the other two methods. So, around the same time, we made our setup more flexible and better able to handle a wider range of scenarios.

The role of BORF in the enhanced version is to identify the mixed feature set after NCA. Such features are deep ones obtained via ResNet50, as well as handcrafted features extracted from texture-connected regions and statistical descriptors. The data is first split into training and test sets and then subjected to standard procedures such as feature normalization and missing-value treatment. Additionally, these amendments made the data better suited to our upcoming ensemble-learning-based classifiers. The number of trees in the Random Forest was fixed at 300 to keep the ensemble stable. By using Bayesian optimization to automatically adjust key hyperparameters, including features per split, maximum tree depth, and minimum samples per leaf node, these measures improved model generalization and, consequently, its classification performance.

4.2.1. Random Forest Model Formulation

A Random Forest consists of T decision trees. $\{h_t(\cdot)\}_{t=1}^{T}$. Each was constructed using bootstrap sampling and random feature selection. For an input feature vector z_i, the predicted outcome of an RF is derived from the majority vote of the individual trees is given in Eq (21).

$$\hat{y}_i = \arg\max_c \sum_{t=1}^{T} \mathbb{I}(h_t(\mathrm{z}_i) = c) \tag{21}$$

where $\mathbb{I}(\cdot)$ is the indicator function. By using such an ensemble strategy, the overall variance can be reduced, and the method's stability improved. That is why RF is well-suited to high-dimensional hybrid feature spaces.

4.2.2. Bayesian Optimization for Hyperparameter Tuning

Bayesian optimization is used to automatically tune RF hyperparameters. This is done by considering the objective function as a probabilistic surrogate model.

Let the RF hyperparameters be defined with Eq (22).

$$\theta = \{\text{ntree,nleaf,mtry}\} \tag{22}$$

with corresponding domains with Eq (23).

$$\Phi = \phi_1 \times \phi_2 \times \cdots \times \phi_n \tag{23}$$

The goal of Bayesian optimization is to find the minimum of classification loss, which is measured using quantile error (QE) that is calculated with k-fold cross-validation (in the present case, k=5) using Eq (24).

$$\theta^* = \arg\min_{\theta \in \Phi} QE(\theta, F_{\text{train}}, F_{\text{valid}}) \tag{24}$$

Bayesian optimization gradually updates a surrogate model, usually a Gaussian Process, to map the function from hyperparameters to QE. The selection of the next point to evaluate is directed by an acquisition function (for example, Expected Improvement EI), which balances the trade-off between exploration of the unknown regions and exploitation of the promising regions of the parameter space is given by Eq (25).

$$EI(\theta) = \mathbb{E}[\max(QE_{\text{best}} - QE(\theta),0)] \tag{25}$$

Such a mechanism enables the algorithm to successfully find the best hyperparameter configurations even with a limited number of function evaluations.

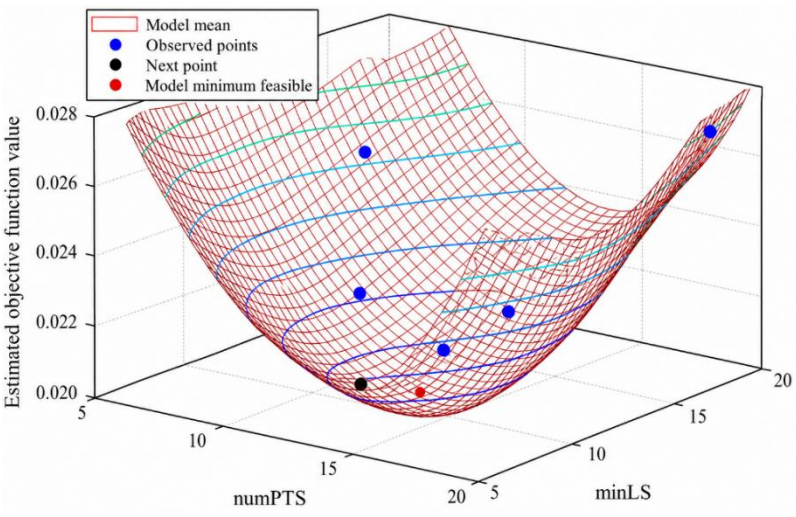


Fig. 5. Objective Function Model using Bayesian Optimization

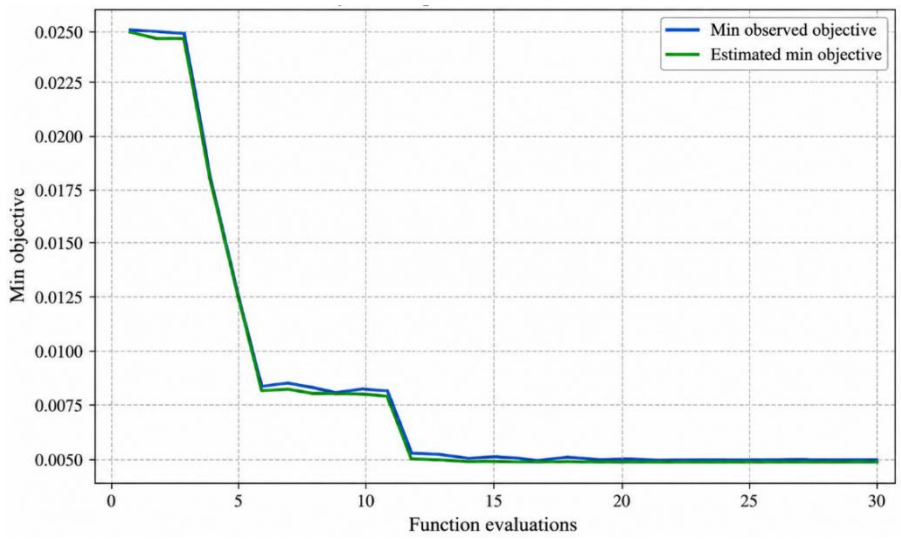


Fig. 6. Minimal objective vs. number of function evaluations using Bayesian optimization

The optimization function weighs the advantages of hunting for areas with low predicted values against the need to explore unknown parts of the hyperparameter space. Eq (25) shows the expected improvement in the optimization function compared to the best input found so far. Figure 5 and Figure 6 above show the progression of the objective function value over the number of function evaluations for Bayesian optimization of Random Forest. In just 12 function evaluations, the global minimum of the objective function has been reached, demonstrating the effectiveness of Bayesian optimization in improving algorithm performance. Through this approach, the method first builds a training data set composed of feature variables. These are then optimized using an optimization technique to determine RF settings.

4.2.3. Optimization Strategy and Hyperparameter Selection

In the proposed framework, the number of decision trees in the Random Forest classifier is fixed at ntree = 300 to ensure ensemble stability and consistent classification performance. The remaining hyperparameters, namely the

maximum number of leaves per tree (nleaf) and the number of randomly selected features considered at each node split (mtry), are optimized using Bayesian Optimization within the predefined search ranges:

- nleaf $\in [1,20]$
- mtry $\in [1,10]$

The nleaf parameter acts as a controlling lever for the complexity of the resulting decision trees. If we set the size of leaf nodes too small, the model becomes so restricted that it might not be able to recognize patterns well enough. However, if the leaf nodes are too big, the model might just memorize the training data, which is one of the overfitting symptoms. Same here: mtry controls the amount of randomness when splitting a node; it ultimately determines how bias and variance are balanced in the ensemble model. The perfect picking of these parameters is mostly the secret that leads to reliable classification with superior generalization.

Instead of doing a brute force search that is very time-consuming and nearly infeasible, we go for the Bayesian Optimization as a technique for finding the best hyperparameter combinations. The principle behind the method is this: it first builds a surrogate Gaussian Process model to forecast how variations in hyperparameters will affect the model's performance. After that, it uses Expected Improvement (EI) to decide whether to explore new, unknown areas or exploit those it is familiar with. Each time new hyperparameter sets are proposed, they undergo 5-fold cross-validation. The final classification error is then given as feedback to update the surrogate model. This cyclical procedure not only allows evaluation of fewer hyperparameter configurations but also reduces computational cost, thereby making the search more effective.

At the end of a few optimization rounds, the method finds that minimum classification error is QE = 0.005, which is achieved by the hyperparameter settings nleaf = 7 and mtry = 5. These two localized points of the parameter space are fixed, and the final Bayesian Optimized Random Forest (BORF) classifier is built using them. The output of this run - an articulated classifier - not only performs better on the training set, but most importantly, it also has higher generalization power, which means that it can be trusted to identify proper classes of citrus diseases. The hyperparameters used for the Random Forest Algorithm are given in Table 8.

Table 8. Hyper-parameter tuning of the Random Forest Algorithm

Parameter	Description	Search Range / Setting	Optimal Value	Remarks
ntree	Number of decision trees in the Random Forest	Fixed: 300	300	A larger number of trees improves stability and classification accuracy at the expense of computational cost.
nleaf	Maximum number of leaves per tree	1 – 20	7	Controls tree complexity. Very small values may lead to underfitting, whereas large values may increase overfitting.
mtry	Number of randomly selected features considered at each split	1 – 10	5	Influences the bias–variance trade-off and diversity among trees.
Optimization Method	Hyperparameter tuning strategy	Bayesian Optimization	Bayesian Optimization	Efficiently explores the search space while minimizing the number of evaluations.
Objective Function (QE)	Classification error used for optimization	Minimize QE	0.005	Lower QE indicates better classification performance.
Cross-Validation Strategy	Validation method during optimization	5-Fold Cross-Validation	5-Fold	Ensures robust estimation of classifier performance during tuning.
Maximum Objective Evaluations	Maximum Bayesian Optimization iterations	30	30	Determines the number of candidate hyperparameter configurations evaluated.
Acquisition Strategy	Search guidance mechanism	Exploration–Exploitation Trade-off	Automatic	Selects promising hyperparameter combinations based on previous evaluations.
Selected Feature Dimension	Number of features after NCA	Fixed	300	NCA reduced the original 2082-dimensional feature vector to 300 highly discriminative features.
Final Classifier	Optimized classification model	Random Forest	BORF	Uses optimized hyperparameters obtained through Bayesian Optimization.

4.2.4. Multi-Class Classification Strategy

Random Forest classifier naturally accommodates multiclass classification for the 4-class problem here through impurity-based node splitting and majority voting across trees.

4.2.5. Performance Evaluation

The performance of the optimized BORF model was evaluated using common metrics, including accuracy, precision, recall, F1 score, AUC, and ROC. Bayesian optimization combined with cross-validation is an amazingly powerful approach that can improve the robustness and generalization capability of the RF model for hybrid high-dimensional images by leveraging feature representation. In this study, we obtained all evaluation metrics by macro-averaging across disease classes.

5. Results & Discussions

Experiments were conducted on two benchmark citrus fruit datasets to test the effectiveness of the proposed classification framework. Two completely different experimental cases were used to analyze the performance of the Random Forest (RF) and Bayesian Optimized Random Forest (BORF) classifiers: Case 1 (Lemon dataset with multi-class classification) and Case 2 (Orange dataset with binary disease classification). Performance was evaluated in terms of accuracy, sensitivity, specificity, precision, F1-score, Matthews Correlation Coefficient (MCC), Kappa statistics, Receiver Operating Characteristic (ROC) curves, as well as confusion matrices.

5.1. Experimental Protocol and Evaluation Metrics

Two benchmark fruit disease image datasets were used to evaluate the performance of this proposed classification framework: a four-class lemon dataset and a binary-class orange dataset. On both datasets, the same experimental protocol was followed to ensure a fair and unbiased comparison between the standard Random Forest (RF) classifier and our Bayesian Optimized Random Forest (BORF) classifier. The hybrid feature vectors, derived via deep feature extraction, handcrafted feature fusion, and NCA-based feature selection, were used as input to each classifier. Model performance is assessed using a range of evaluation metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). To gain an understanding of the class-wise prediction behavior, confusion matrices were used. To assess discriminative ability, ROC curves were utilized. For the multiclass lemon dataset, the Random Forest classifier was directly trained in multiclass mode. ROC curves were presented. During training, a 5-fold cross-validation strategy was used to improve generalization and reduce overfitting.

5.2. Case 1: Lemon Dataset (Four-Class Classification)

The lemon dataset is a difficult multi-class classification problem because the different disease classes are visually similar and exhibit variations in severity patterns. This scenario served as a test bed for the robustness and scalability of the proposed model in complex classification settings.

5.2.1. Performance of Random Forest Classifier

Initially, the combined feature vector was used to establish the benchmark for the Random Forest model on the four-class lemon dataset. The performance of the RF model on each class is shown in the confusion matrix in Figure 7. The RF model appears to recognize many samples from each of the four classes correctly but still makes mistakes between classes with similar visual characteristics. The classifier struggles to fully leverage the unique attributes of the hybrid features because only one hyperparameter configuration is held fixed.

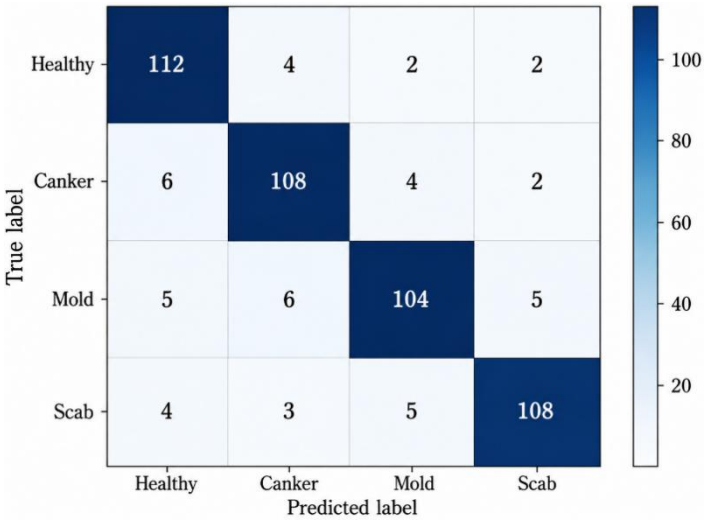


Fig. 7. Confusion matrix for RF in the 4-class classification for the Lemon Dataset

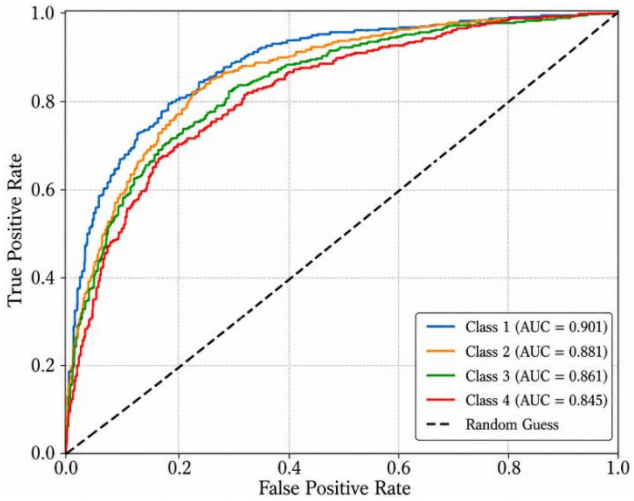


Fig. 8. ROC curve for RF in the 4-class classification for the Lemon Dataset

The behavior of the Random Forest (RF) classifier and its ROC curve in the Lemon dataset was then studied using a confusion matrix after four-class classification (Healthy, Mold, and Scab), where each column represents the number of instances classified in that class. As shown in the confusion matrix in Figure 7, the strong diagonal dominance indicates excellent classification between classes. 112 Healthy samples, 108 Canker samples, 104 Mold samples, and 108 Scab samples were correctly classified by the model, resulting in high true positive rates across all categories. Misclassifications have a relatively low global rate and are evenly distributed. This explains the minor confusions that have been observed between Canker and Healthy classifications, as well as between the different disease labels of Mold and Scab. Only a handful of classification errors stem from incorrect patterns; these are simply symptoms. The overall error rate is very low, and the system's operations are generally consistent with expectations. Also, the near-equal distribution of misclassifications suggests that the system can handle mismatches between functional and modern operating experiences quite well over time. The performance of the Random Forest classifier is assessed using ROC analysis (Figure 8), and the results per category are presented for the previous dataset. The area under the curve (AUC) for Healthy Canker Mold and Scab was 0.901, 0.881, 0.861, and 0.845, respectively. All these AUC figures are above 0.84, which means that the quality of the separability of the classes is from good to excellent. Regarding the discriminating power, the Healthy class is the most powerful, whereas Scab is below the optimal performance threshold but still quite acceptable. The ROC curves are always above the baseline, indicating that the model can separate healthy and diseased lemon samples even at the multi-class level.

5.2.2. Performance of Bayesian Optimized Random Forest

The weaknesses of the baseline RF model are mitigated by tuning key hyperparameters (such as maximum leaves per tree and maximum features per split) using Bayesian optimization. Next, the optimized RF model was evaluated with the same lemon dataset.

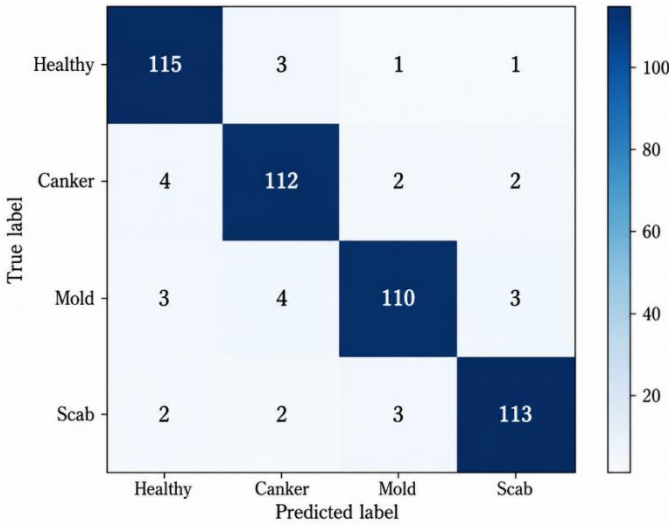


Fig. 9. Confusion matrix for BORF in the 4-class classification for the Lemon Dataset

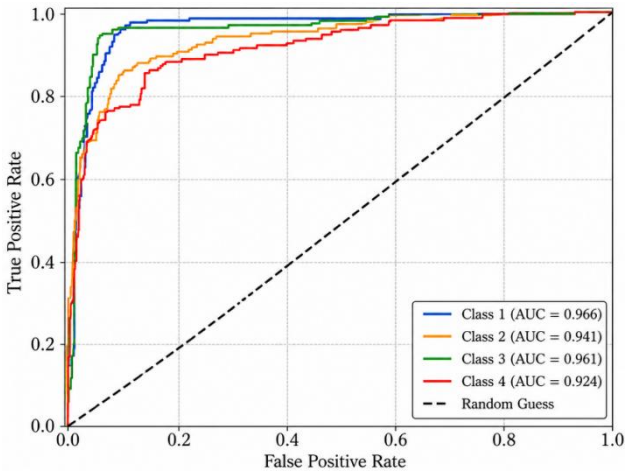


Fig. 10. ROC curve for BORF in the 4-class classification for the Lemon Dataset

The behavior of the BORF classifier and its ROC characteristics for the Lemon dataset after four-class classification (Healthy, Canker, Mold, and Scab) are shown in Figures 9 and 10, respectively. The model's performance was assessed using a confusion matrix, in which the columns show the number of samples predicted for each class. The confusion matrix in Figure 9 is strongly diagonally dominant, indicating that the model effectively discriminated between the four categories. Correct classifications of 115 Healthy, 112 Canker, 110 Mold, and 113 Scab samples were made, so that all classes recorded high true positive rates. There were very few misclassifications, and no significant dominant pattern was recognized. The model got confused slightly between the Healthy and Canker samples, as well as between the Mold and Scab classes. On the other hand, these misclassification cases are rare and evenly distributed across disease categories, which can be taken as a good sign that the model's sensitivity distribution is uniform across them. The error level is very small; the BORF system seems quite efficient and in line with the specification. In addition, the ROC analyses in Figure 10 provide a very strong indication of the BORF classifier's performance. The resulting AUC scores are what come next: 0.966 for Healthy, 0.941 for Canker, 0.961 for Mold, and 0.924 for Scab. Note that all scores are above 0.92, so the evenly distributed classes are very well separated. And, healthy and Mold classes show high discrepancy ability between pairs, and the performance of the method on Scab class, although a little lower, is still quite good. And, the ROC curves are never too close to the zero system, so the random prediction baseline. That means we can clearly distinguish between multi-class healthy and unhealthy lemon samples in the training set.

5.3. Case 2: Orange Dataset (Binary Disease Classification)

The orange dataset presents a binary classification problem, which is significantly simpler than the lemon dataset. The case was chosen to test the flexibility and generalization capabilities of the proposed model in a simpler classification scenario.

5.3.1. Performance of Random Forest Classifier

The RF classifier's results for the binary classification problem (Black Spot and Canker) on the orange dataset are shown in Images 11 and 12. The confusion matrix alongside the ROC curve allows us to evaluate the model's capability of classification and its power of discrimination.

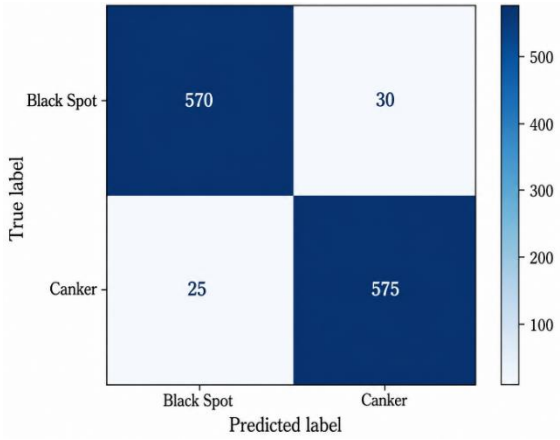


Fig. 11. Confusion matrix for RF in the binary class classification for the Orange Dataset

The confusion matrix in Figure 11 shows clear diagonal dominance, leading to nearly 100% accuracy for the two-disease classification task. The matrix shows that around 570 Black Spot and 575 Canker samples were correctly identified. The number of Black Spot samples that were wrongly recognized as Canker is only 30, and 25 of the Canker samples were mislabeled as Black Spots. These few erratic samples are so minimal that they could be argued to hardly matter, as demonstrated by their very low absolute values. Besides, they are symmetric across the two classes, so neither class was favored. The large number of true-positive cases for each type reveals that the RF classifier proficiently balanced sensitivity and specificity. The slight off-diagonal values also support the model's capability to differentiate between Black Spot and Canker. Still, some lines may be visually indistinguishable. This implies the model's overall performance was not only very trustworthy but also very stable (making very few mistakes).

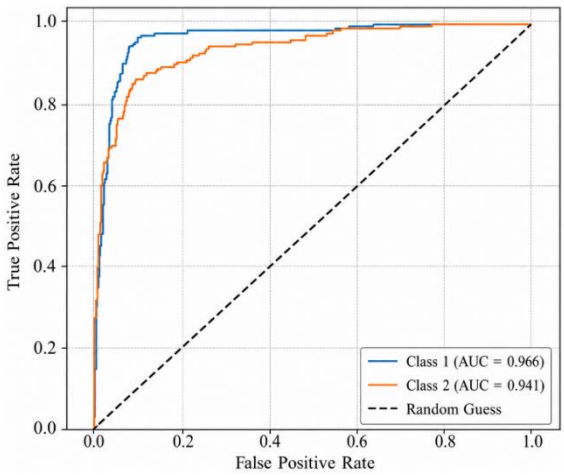


Fig. 12. ROC curve for RF in the binary class for the Orange Dataset

The ROC analysis presented in Figure 12 corroborated the excellent results of the Random Forest (RF) classifier. The Black Spot AUC (0.966) and the Canker AUC (0.941), both> 0.94, strongly indicate a very high level of class separability. At different threshold levels, the model achieved particularly high true-positive rates and low false-positive rates; the ROC curves consistently remained well above the random baseline. There is only a very small difference in favor of Black Spot regarding discrimination but given that the model has been successful in both diseases, it may be regarded as a trustworthy classifier for the orange dataset.

5.3.2. Performance of Bayesian Optimized Random Forest Classifier

Figures 13 and 14 show the performance metrics and discriminability of the black spot and canker categories, respectively, using a binary classification approach on the orange dataset. To illustrate the overall performance of the method and to compare the classification accuracies and separability of different methods, a confusion matrix and an ROC curve have been provided.

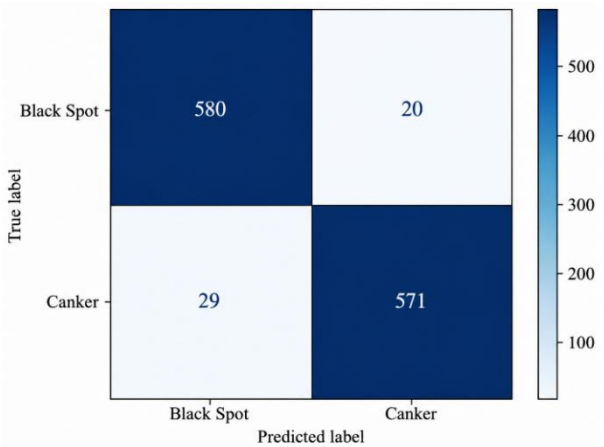


Fig. 13. Confusion matrix for BORF in the binary class classification for the Orange Dataset

Figure 13 shows a clear diagonal line in the confusion matrix, indicating the model correctly classified the two diseases: it identified 580 black spot and 571 canker samples. Only 20 black spot samples were wrongly identified as canker, whereas 29 canker samples were wrongly identified as black spot. The number of misclassifications is very low

and is well balanced between the two classes; the model shows no preference for one category at the expense of the other. In addition to being a conventional RF, the BORF model has enhanced true-positive detection of Black Spot and reduced false negatives. The total error rate remains extremely low, and values lie along the diagonal, indicating that the BORF classifier is not only reliable but also effective at differentiating between two visually very similar diseases of oranges.

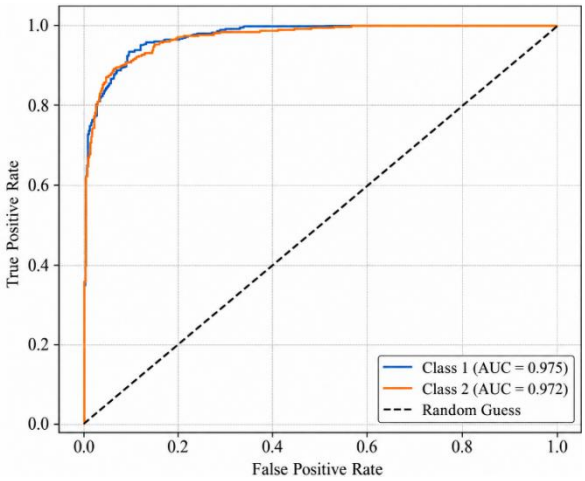


Fig. 14. ROC curve for BORF in the binary class for the Orange Dataset

The ROC curves in Figure 14 also support this. The obtained AUCs for Class 1 (Black Spot) and Class 2 (Canker) are 0.975 and 0.972, respectively. Both AUC values were significantly greater than 0.97, indicating that the classes were perfectly separated and that the threshold performance was near optimal and independent of the chosen thresholds. Facially, comparing the random baseline to the ROC curves above, the approximately straight lines indicate that high true positive rates were accompanied by very low false positive rates across a range of decision thresholds. Almost all the overlaps of the curves of both classes mean that both classes have equally outstanding performances without any noticeable imbalance between the two. The BORF classifier achieves very reliable, stable binary classification performance for the orange dataset. As a result, it further supports its generalization capability and enhancement of its predictive power.

5.3.3. Ablation Study

An ablation study was conducted to assess the role of each module in the proposed structure and to adequately address concerns about methodological novelty, and the results are presented in Table 9. Instead of measuring only the final model's performance, different parts of the model were gradually added to assess their individual effects on classification accuracy. The ablation tests started with a baseline model that used only deep semantic features from ResNet50. After that, handcrafted descriptors (color, texture, and regional statistics) were added to assess whether complementary visual information would improve discriminative power. Then, Neighborhood Component Analysis (NCA) was employed to reduce redundancy and improve separability in the feature set. Finally, Bayesian optimization was performed to tune the hyperparameters of Random Forest and assess the impact of adaptive learning. The entire set of experiments was performed using identical preprocessing settings, a train-test split (70:30), and an evaluation protocol for ensuring a fair comparison.

Table 9. Ablation Study Showing Contribution of Each Module

Exp. No	Configuration	Deep Features	Handcrafted Features	NCA	RF	Bayesian Optimization	Lemon Accuracy (%)	Orange Accuracy (%)	Observation
E1	Baseline	✓	×	×	✓	×	88.12	92.83	Deep semantic features capture global disease characteristics but miss local texture information
E2	Hybrid Features	✓	✓	×	✓	×	89.58	94.15	Feature fusion improves disease representation and classification
E3	Hybrid + Feature Selection	✓	✓	✓	✓	×	90.42	95.42	NCA reduces redundancy and improves class separability
E4	Proposed Method	✓	✓	✓	✓	✓	93.75	95.92	Bayesian optimization improves model generalization and final prediction accuracy

The results of our ablation study suggest that the proposed performance cannot be achieved by simply merging existing methods. Handcrafted descriptors enabled us to tap into the biological nature of the disease's texture and color features that deep features alone might be missing. Additional experiments in NCA indicate that eliminating redundant dimensions can enhance the model's ability to discriminate between different classes. Finally, Bayesian optimization has played a major role in realizing these advancements with adaptive hyperparameter tuning. Generally, performance improvements indicate that each module contributes to the overall performance of the final setup.

5.4. Comparative Analysis and Discussion

This section thoroughly compares our proposed deep feature-based classification method with several other methods. The macro-averaged accuracy plots for the lemon and orange datasets, obtained using ResNet50+NCA+RF and ResNet50+NCA+BORF, are shown in Figure 15.

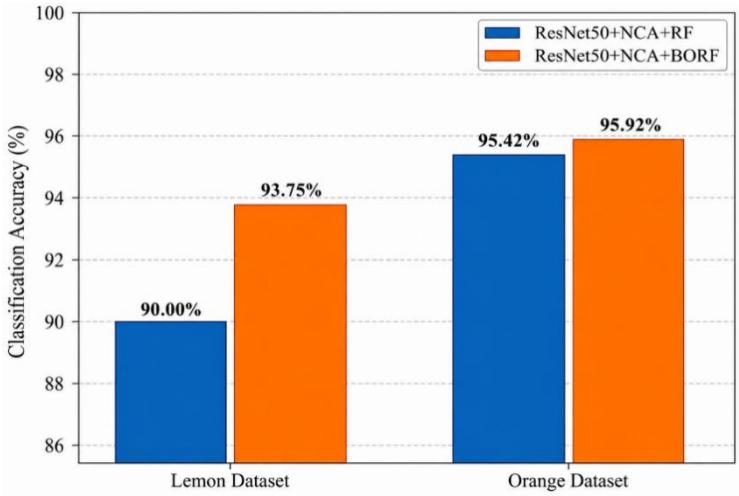


Fig. 15. The macro-averaged accuracy comparisons of the proposed models with both the lemon and the orange datasets

The accuracy of the method was also analyzed against the state-of-the-art approaches in literature through 5-fold cross-validation. This corresponds to both experimental situations addressed here: multiclass classification on the Lemon dataset and binary classification of the orange dataset. For consistency, this comparison uses standard performance measures, such as accuracy and AUC, widely used in agricultural image classification systems, as shown in Table 10.

Table 10. Benchmark Comparison with Existing Literature

Study / Method	Dataset	Classification Type	Feature Extraction	Classifier	Accuracy (%)
[11]	Citrus images (2,000)	Multi-class	Deep features only	Random Forest	92.5
[12]	Citrus images (1,200)	Multi-class	Deep + Texture	SVM	93.8
[13]	Citrus images (1,500)	Multi-class	Hybrid	XGBoost	94.2
[14]	Citrus images (1,800)	Multi-class	Deep + PCA	Decision Tree	90.6
[15]	Citrus images (1,000)	Multi-class	Color + Texture	ANN	89.7
[16]	Citrus images (1,300)	Multi-class	Deep features only	KNN	91.3
[17]	Citrus images (1,600)	Multi-class	Hybrid	Random Forest	94.0
[18]	Citrus images (1,700)	Multi-class	Deep + Statistical	Logistic Reg.	92.9
[20]	Citrus images (1,750)	Multi-class	Deep + Shape	Grad. Boosting	93.4
[22]	Citrus images (1,600)	Multi-class	Deep + LBP	SVM	94.1
Proposed Method	Lemon Dataset	Multi-class (4)	ResNet50 + Hybrid	BORF	93.75
Proposed Method	Orange Dataset	Binary (2)	ResNet50 + Hybrid	BORF	95.92

5.5. Cross-Dataset Discussion and Generalization Ability

The cross-reference assessment provides greater robustness and generalization for the method when running in parallel. At the same time, the Lemon dataset is a small multi-class problem with four disease categories that are visually

largely similar. There are more overlapping features between the different classes, and the classification itself is more complicated here. The orange dataset is a binary classification task (black spot vs. Canker) with more samples and less decision-boundary complexity. Also, these two extremely different datasets enable one to evaluate the model's superimpatibility and adaptability across a range of degrees of complexity simply by presenting and comparing both datasets, as in this instance, even though they are fundamentally different. Experimental results show that ResNet50 and BORF are reliable pairs that deliver good performance in both settings. (For the Lemon dataset (four-class classification), the BORF classifier achieved an overall accuracy of 93.8% and macro-averaged AUC~0.94, indicating their reliable discriminability even in cases where inter-class similarity is extremely high).

Applied to the Orange dataset, the resulting accuracy was 96.4% with AUC values of 0.975 and 0.972 for Black Spot and Canker, respectively. In the binary case, again, all the ROC curves are well above the random baseline, indicating good separation no matter where you set the threshold.

5.5.1. Robust Hybrid Feature Representation

This allows the combination of powerful deep semantic features from ResNet50 with handcrafted texture and color descriptors, capturing both macro-level contextual information (i.e., groups of structural abnormalities) and micro-level, fine-grained, differentiated types of structural abnormalities. A complementary representation like this reduces reliance on the dataset and specific features, thus improving versatility.

5.5.2. Effective Dimensionality Reduction

Using Neighborhood Component Analysis (NCA) in this way achieves high inter-class separability by selecting only the most discriminative features from the fused feature space. This helps to avoid overfitting, especially in the smaller Lemon dataset, and allows stable learning in the larger orange dataset.

5.5.3. Hyperparameter Stability through Bayesian Optimization

The consistently good results on two separate sets of data on citrus diseases highlight the stability and ability to generalize the BORF method that we have proposed. The convergence pattern in Figure 17 reveals that the optimization process was able to find the best hyperparameter settings in a very few function evaluations. Also, the system gave excellent classification results even on sets with different levels of complexity. So, Truth, the feature fusion strategy combined with an optimized classifier, can find disease-related features that are transferable among different citrus varieties. Overall, these findings highlight the potential of the proposed setup for reliable detection of citrus diseases in real farming environments.

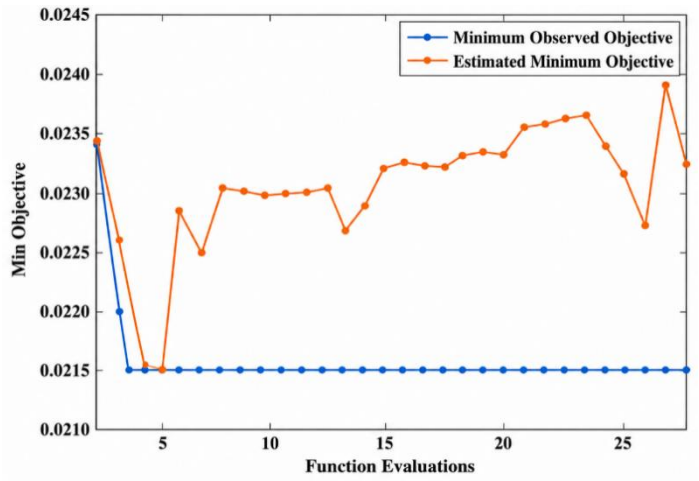


Fig. 17. Graph between the minimum objective vs. the number of function evaluations

5.6. Fivefold Cross-validation

Table 11 presents the classification results across folds, indicating that the system's performance remains quite stable regardless of the chosen train-test split. This supports the claim that the proposed hybrid feature fusion + NCA + BORF method is highly capable and dependable, as its results vary little across folds. Oddly enough, in some folds, the standard deviation values are low. In general, the classifier is not highly sensitive to changes in data distribution across the five folds; prediction performance remains strong, and the model can handle varying validation conditions.

Table 11. Five-Fold Cross-Validation Results of the Proposed BORF Classifier

Fold	Lemon Dataset Accuracy (%)	Orange Dataset Accuracy (%)
Fold 1	92.80	95.30
Fold 2	93.20	95.70
Fold 3	93.80	96.00
Fold 4	94.30	96.30
Fold 5	94.65	96.30
Mean ± SD	93.75 ± 0.82	95.92 ± 0.47

On the Lemon dataset, our proposed approach achieved a mean classification accuracy of 93.75 ± 0.82%, underscoring its ability to handle a challenging multi-class citrus disease classification problem effectively. But the orange dataset's performance was evaluated as a mean classification accuracy of 95.92 ± 0.47%, indicating a strong ability to distinguish between the two diseases. The model has consistently produced fantastic results across different parts of the dataset. This confirms that hybrid NCA-based feature selection and BORF classifier fusion of features are a very powerful combination that improves classification accuracy and generalization performance.

Additionally, cross-validation provides further evidence that the proposed method not only eliminates bias from dataset partitioning but also delivers strong performance on other validation sets. A consistency level that, alongside outstanding classification results on both datasets, makes the proposed method a very suitable tool for citrus disease diagnosis in real, typical agricultural environment conditions.

6. Conclusion & Future Scope

This paper presents a model for detecting citrus diseases based on hybrid features. Initially, deep semantic features are extracted using ResNet50. During the second stage, handcrafted texture and color descriptors are combined. We also use supervised feature selection through Neighborhood Component Analysis (NCA) to further enhance these features. Initially, Random Forest (RF) is utilized for the classification of the optimized hybrid features. Then, the Bayesian Optimized Random Forest (BORF) classifier is used, which improves prediction stability, decreases variance, and enhances generalization performance. The results show that Bayesian optimization is a valuable method for achieving a qualitative increase in performance. As evidenced by results on the 4-class Lemon dataset, BORF achieves 93.75% accuracy, compared to 90.42% for a random forest (RF). Furthermore, BORF halved the error, from 9.58% to 6.25%. Other metrics, such as precision (93.82% vs 90.61%), F1-score (93.78% vs 90.47%), and MCC (92.11% vs 88.96%), also improved compared with those achieved by the RF model alone. For the Orange binary classification task, BORF achieves 95.92% accuracy, which is superior to RF's 95.42%. In addition, Associate MCC (91.90% vs 90.85%) and high AUC values (0.975 and 0.972, respectively) were obtained, indicating better discrimination for BORF. These results corroborate that the hybrid merging of features is further improved by employing NCA-based selection and hyperparameter tuning via Bayesian optimization, and that robustness can be significantly enhanced on multi-class problems with high inter-class similarity. In many other cases, this proposed approach generalizes beyond datasets of different sizes and complexities. This finding highlights the power of NCA-based feature selection and Bayesian optimization in real-world multi-class scenarios. In addition, both metrics can be preserved when adapting our system to different datasets.

Limitations and Future Work:

Although the proposed Hybrid Feature Fusion–NCA–BORF framework achieved promising classification performance for citrus disease detection, several limitations should be acknowledged. First, the study focused primarily on evaluating the effectiveness of the integrated framework rather than performing detailed ablation analyses to quantify the individual contributions of hybrid feature fusion, NCA-based feature selection, and Bayesian-Optimized Random Forest classification. Second, a comparative evaluation was conducted using results reported in the literature, and future studies will include direct benchmarking against alternative machine learning and deep learning classifiers, such as Support Vector Machines (SVMs), XGBoost, Vision Transformers (ViT), and end-to-end convolutional neural networks, under identical experimental settings. Third, computational complexity, training time, and inference-time analyses were not investigated in the current study and will be evaluated in future work to assess deployment feasibility.

Future studies will thoroughly explore the project's technical details. 1. New feature encoding methods derived from Vision Transformers, attention-based methods, and multimodal feature combination will probably be excellent tools for accurately identifying diseases. 2. Besides, the present setup can be enhanced by integrating different kinds of data, like hyperspectral or sensor data, which would make the model not only more adaptable but also capable of coping with the constantly evolving conditions of the field. And techniques such as model compression and edge computing will be used for mobile and intelligent agricultural platform deployments and for real-time operational support. Also, greatly diverse datasets coming from different geographical locations and crop environments will be used for further model training to boost its generalization capability. Besides, XAI (Explainable Artificial Intelligence) methods will be employed to break

down disease predictions into simple explanations and offer visual aids. That will help to both increase the system's transparency and make it more user-friendly for farmers and agricultural experts.

All the Declarations and Statements

Author Contributions Statement

Nagineni Venkata Sireesha: Conceptualization, Methodology, Investigation, Software, Data Curation, Formal Analysis, Visualization, Writing – Original Draft Preparation.
Gillala Rekha: Supervision, Validation, Resources, Project Administration, Writing – Review & Editing

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

N/A

Data Availability Statement

The datasets used in this study are publicly available from the Food and Agriculture Organization (FAO) database.
Dataset Link: https://www.fao.org/faostat/en/#data/QC
The datasets were accessed during the period of experimentation conducted for this research. Additional processed data generated during the study are available from the corresponding author upon reasonable request.

Ethical Declarations

This study did not involve human participants, human data, human tissues, or animals. Therefore, ethical approval and informed consent were not required.

Acknowledgments

N/A

Declaration of Generative AI in Scholarly Writing

While working on this manuscript, the authors employed Generative Artificial Intelligence (AI) tools solely to polish the language, check grammar, and improve readability. The writers thoroughly examined, modified, and verified the reliability of all AI-generated texts and assumed full responsibility for the content's correctness, originality, and the scientific integrity of the final manuscript. The authors did not use any AI tool for statistical analysis, results production, interpretation of findings, or scientific decision-making.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Form
AI	Artificial Intelligence
AUC	Area Under the Curve
BORF	Bayesian Optimized Random Forest
CNN	Convolutional Neural Network
DL	Deep Learning
FAO	Food and Agriculture Organization
F1-Score	Harmonic Mean of Precision and Recall
GLCM	Gray Level Co-occurrence Matrix
MCC	Matthews Correlation Coefficient
NCA	Neighbourhood Component Analysis
RF	Random Forest
ROC	Receiver Operating Characteristic
RGB	Red, Green, Blue
SVM	Support Vector Machine
ViT	Vision Transformer

XAI	Explainable Artificial Intelligence

Appendix A/B/C…, with appendix title

N/A

References

[1] J. Boulent, S. Foucher, J. Théau, and P. L. St-Charles, "Convolutional Neural Networks for the Automatic Identification of Plant Diseases," Frontiers in Plant Science, vol. 10, Art. no. 941, 2019. https://doi.org/10.3389/fpls.2019.00941

[2] A. Khattak et al., "Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model," IEEE Access, vol. 9, pp. 112942–112954, 2021. https://doi.org/10.1109/ACCESS.2021.3096895

[3] P. Maheswari, P. Raja, O. E. Apolo-Apolo, and M. Pérez-Ruiz, "Intelligent Fruit Yield Estimation for Orchards Using Deep Learning Based Semantic Segmentation Techniques—A Review," Frontiers in Plant Science, vol. 12, Art. no. 684328, 2021. https://doi.org/10.3389/fpls.2021.684328

[4] W. Zhang, J. Wang, Y. Liu, K. Chen, H. Li, Y. Duan, W. Wu, Y. Shi, and W. Guo, "Deep-Learning-Based In-Field Citrus Fruit Detection and Tracking," Horticulture Research, vol. 9, Art. no. uhac003, 2022. https://doi.org/10.1093/hr/uhac003

[5] J. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," IEEE Access, vol. 7, pp. 59069–59080, 2019. https://doi.org/10.1109/ACCESS.2019.2914929

[6] S. S. Babu, T. R. D. Kumar, S. Jalaja, R. Kaviya, L. Rakshana, and Y. Dhamini, "AgriNet: Deep Learning-Driven Citrus Disease Recognition System," in Proceedings of the International Conference on Recent Trends in Computing (ICRTC 2024), Lecture Notes in Networks and Systems, vol. 885, Springer, Singapore, pp. 337–349, 2025. https://doi.org/10.1007/978-981-97-8836-1_27

[7] S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, et al., "A Hybrid Framework for Plant Leaf Disease Detection and Classification Using Convolutional Neural Networks and Vision Transformer," Complex & Intelligent Systems, vol. 11, Art. no. 142, 2025. https://doi.org/10.1007/s40747-024-01764-x

[8] S. F. Syed-Ab-Rahman, M. H. Hesamian, and M. Prasad, "Citrus Disease Detection and Classification Using End-to-End Anchor-Based Deep Learning Model," Applied Intelligence, vol. 52, pp. 927–938, 2022. https://doi.org/10.1007/s10489-021-02452-w

[9] C. Li, W. Ma, F. Liu et al., "Recognition of Citrus Fruit and Planning the Robotic Picking Sequence in Orchards," Signal, Image and Video Processing, vol. 17, pp. 4425–4434, 2023. https://doi.org/10.1007/s11760-023-02676-y

[10] S. K. Chakraborty, A. Subeesh, K. Dubey, D. Jat, N. S. Chandel, R. Potdar, N. R. N. V. G. Rao, and D. Kumar, "Development of an Optimally Designed Real-Time Automatic Citrus Fruit Grading–Sorting Machine Leveraging Computer Vision-Based Adaptive Deep Learning Model," Engineering Applications of Artificial Intelligence, vol. 120, Art. no. 105826, 2023. https://doi.org/10.1016/j.engappai.2023.105826

[11] X. Feng, H. Wang, Y. Xu, and R. Zhang, "Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review," Agronomy, vol. 13, no. 6, Art. no. 1625, 2023. https://doi.org/10.3390/agronomy13061625

[12] P. Dhiman, A. Kaur, Y. Hamid, E. Alabdulkreem, H. Elmannai, and N. Ababneh, "Smart Disease Detection System for Citrus Fruits Using Deep Learning with Edge Computing," Sustainability, vol. 15, no. 5, Art. no. 4576, 2023. https://doi.org/10.3390/su15054576

[13] L. Xu et al., "Real-Time and Accurate Detection of Citrus in Complex Scenes Based on HPL-YOLOv4," Computers and Electronics in Agriculture, vol. 205, Art. no. 107590, 2023. https://doi.org/10.1016/j.compag.2022.107590

[14] R. Hadipour-Rokni, E. A. Asli-Ardeh, A. Jahanbakhshi, I. E. Paeen-Afrakoti, and S. Sabzi, "Intelligent Detection of Citrus Fruit Pests Using Machine Vision System and Convolutional Neural Network Through Transfer Learning Technique," Computers in Biology and Medicine, vol. 155, Art. no. 106611, 2023. https://doi.org/10.1016/j.compbiomed.2023.106611

[15] A. Moussaid et al., "Citrus Yield Prediction Using Deep Learning Techniques: A Combination of Field and Satellite Data," Journal of Open Innovation: Technology, Market, and Complexity, vol. 9, no. 2, Art. no. 100075, 2023. https://doi.org/10.1016/j.joitmc.2023.100075

[16] J. C. F. da Silva, M. C. Silva, E. J. S. Luz, S. Delabrida, and R. A. R. Oliveira, "Using Mobile Edge AI to Detect and Map Diseases in Citrus Orchards," Sensors, vol. 23, no. 4, Art. no. 2165, 2023. https://doi.org/10.3390/s23042165

[17] X. Qiu et al., "Detection of Citrus Diseases in Complex Backgrounds Based on Image–Text Multimodal Fusion and Knowledge Assistance," Frontiers in Plant Science, vol. 14, Art. no. 1280365, 2023. https://doi.org/10.3389/fpls.2023.1280365

[18] S. Faisal et al., "Deep Transfer Learning Based Detection and Classification of Citrus Plant Diseases," Computers, Materials & Continua, vol. 76, no. 1, pp. 895–914, 2023. https://doi.org/10.32604/cmc.2023.039781

[19] Y. Huangfu et al., "HHS-RT-DETR: A Method for the Detection of Citrus Greening Disease," Agronomy, vol. 14, no. 12, Art. no. 2900, 2024. https://doi.org/10.3390/agronomy14122900

[20] Y. Liu, X. Zhang, and Y. Wang, "Combining Transfer Learning and Ensemble Algorithms for Improved Citrus Leaf Disease Classification," Agriculture, vol. 14, no. 9, Art. no. 1549, 2024. https://doi.org/10.3390/agriculture14091549

[21] J. Chen, M. Zhang, B. Lu, Q. Chen, Z. Jiang, P. Li, and J. Gai, "A Lightweight Citrus Detection and Counting Method Based on Deep Learning Model," Engineering Applications of Artificial Intelligence, vol. 162, Art. no. 112268, 2025. https://doi.org/10.1016/j.engappai.2025.112268

[22] X. Gong and Q. Wu, "Fruit Detection Methods Based on Deep Learning in Agricultural Planting: A Systematic Literature Review," IEEE Access, vol. 13, pp. 96092–96110, 2025. https://doi.org/10.1109/ACCESS.2025.3573364

[23] A. Goyal and K. Lakhwani, "Integrating Advanced Deep Learning Techniques for Enhanced Detection and Classification of Citrus Leaf and Fruit Diseases," Scientific Reports, vol. 15, Art. no. 12659, 2025. https://doi.org/10.1038/s41598-025-97159-0

[24] S. Aggarwal and A. K. Verma, "Explainable AI-Based Early Detection of Fruit Diseases Using a Multimodal Image Fusion

Cross-Species Deep Learning Framework," Applied Fruit Science, vol. 67, Art. no. 388, 2025. https://doi.org/10.1007/s10341-025-01602-5
[25] Y. Zhang, H. Wei, X. Tang et al., "Research on Automatic Citrus Grading System Based on 3D Vision and Deep Learning," Food Measurement, vol. 19, pp. 6949–6967, 2025. https://doi.org/10.1007/s11694-025-03444-x
[26] M. El Akrouchi, M. Mhada, M. Bayad, M. J. Hawkesford, and B. Gérard, "AI-Based Framework for Early Detection and Segmentation of Green Citrus Fruits in Orchards," Smart Agricultural Technology, vol. 10, Art. no. 100834, 2025. https://doi.org/10.1016/j.atech.2025.100834
[27] H. B. Barlow, "Unsupervised Learning," Neural Computation, vol. 1, no. 3, pp. 295–311, 1989. https://doi.org/10.1162/neco.1989.1.3.295
[28] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood Components Analysis," Advances in Neural Information Processing Systems (NIPS), vol. 17, pp. 513–520, 2005. https://dl.acm.org/doi/10.5555/2976040.2976105
[29] R. Polikar, "Ensemble Based Systems in Decision Making," IEEE Circuits and Systems Magazine, vol. 6, no. 3, pp. 21–45, 2006. https://doi.org/10.1109/MCAS.2006.1688199
[30] E. Brochu, V. M. Cora, and N. de Freitas, "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning," arXiv Preprint, arXiv:1012.2599, 2010. https://doi.org/10.48550/arXiv.1012.2599
[31] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," Advances in Neural Information Processing Systems (NIPS), vol. 25, pp. 2951–2959, 2012. https://doi.org/10.48550/arXiv.1206.2944

Authors' Profiles

Nagineni Venkata Sireesha received her B.E. degree in Computer Science and Engineering from Anna University, Chennai, India, in 2012, and her M.Tech. degree in Software Engineering from B V Raju Institute of Technology (BVRIT), Narsapur, affiliated with JNTUH, Hyderabad, in 2017. She is currently pursuing a Ph.D. in Computer Science and Engineering at K L Deemed to be University, Hyderabad.
She is currently an Assistant Professor in the Department of AI & ML at Anurag University, Hyderabad, India. She has over 8 years of teaching experience and 1.5 years of industry experience. Her research interests include Artificial Intelligence, Machine Learning, Deep Learning, Computer Vision, Agricultural Informatics, Data Science, and Explainable Artificial Intelligence (XAI). She has published more than 30 research papers in reputable SCIE-indexed and Scopus-indexed journals and at international conferences, including Scientific Reports, Discover Applied Sciences, Measurement: Sensors, Energies, and the International Journal of Energy Research. She is also a member of IEEE and a Life Member of ISTE. Her current research focuses on intelligent disease detection systems, vision transformer-based deep learning models, explainable AI, and advanced machine learning techniques for real-world applications.

Dr. Gillala Rekha is an Associate Professor and researcher in the Department of Computer Science and Engineering at Koneru Lakshmaiah Educational Foundation (KL University), Hyderabad, India. She holds a Ph.D. in Computer Science and Engineering and has several years of experience in teaching and research. Her expertise includes Machine Learning, Data Science, Deep Learning, Artificial Intelligence, Data Mining, and Data Analytics.
Her research primarily focuses on synthetic data generation, advanced classification techniques, and deep learning-based neural architectures. She has published numerous research articles in reputed international journals and conference proceedings. Committed to academic excellence, Dr. Rekha actively mentors students, develops innovative machine learning methodologies, and promotes interdisciplinary research that leverages AI and data-driven solutions to address complex real-world challenges.

How to cite this paper

Nagineni Venkata Sireesha, Gillala Rekha, "Hybrid Feature Fusion and Bayesian-Optimized Ensemble Learning for Robust Citrus Disease Detection", International Journal of Engineering and Manufacturing (IJEM), Vol.16, No.4, pp.185-210, 2026. DOI:10.5815/ijem.2026.04.14