

Optimized hybrid transformer-extreme gradient boosting model for black pepper disease classification

Saritha Suvarna, Demian Antony Dmello

Department of Computer Science and Engineering, Canara Engineering College, Affiliated to Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received Oct 10, 2025

Revised Jul 22, 2026

Accepted Aug 12, 2026

Keywords:

Black pepper

Classification

Deep attention-based
transformer

Leaf disease

Optimized weighted XGBoost

ABSTRACT

Black pepper (*Piper nigrum*) is a vital spice crop cultivated in several tropical countries and is highly susceptible to various leaf diseases that significantly affect crop yield and quality. Traditional manual disease identification methods are time-consuming, labor-intensive, and error-prone. While machine learning (ML) and deep learning (DL) techniques have shown promise in plant disease detection, limited work exists specifically on black pepper leaf diseases. Existing research often relies on small, imbalanced, and non-standardized datasets, leading to poor generalization and reduced accuracy. To address this challenge, the proposed study presents a novel approach utilizing deep attention-based transformer-optimized weighted hyper-tuned extreme gradient boosting network (DAT-OWH-XGBNet) for early black pepper leaf disease detection and classification. The model was trained using the black pepper leaf disease dataset, which includes anthracnose, yellowing, quick wilt, slow wilt, and healthy leaf images. The proposed model combines attention-based transformers with a class-weighted and hyperparameter-optimized extreme gradient boosting (XGBoost) classifier to extract high-level spatial features during feature extraction and optimized XGBoost for final classification. The findings show that the proposed model achieved 99.88% prediction accuracy and 99.82% classification accuracy, outperforming existing convolutional neural network (CNN) methods. Our model is designed to support farmers and agricultural specialists in detecting pepper leaf diseases at the early stage.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Saritha Suvarna

Department of Computer Science and Engineering, Canara Engineering College

Affiliated to Visvesvaraya Technological University

Belagavi-590018, India

Email: sarithasuvarna25@gmail.com

1. INTRODUCTION

Black pepper (*Piper nigrum*), often referred to as the “king of spices,” is one of the most widely used and traded spices globally [1]. Native to the Malabar Coast of India, black pepper has held immense economic, culinary, and medicinal value for centuries [2]. It plays a crucial role in the spice economies of countries such as India, Vietnam, Indonesia, Brazil, and Sri Lanka [3]. In these countries, black pepper cultivation significantly contributes to agricultural revenue and employment. India, especially the states of Kerala, Karnataka, and Tamil Nadu, continues to be a major producer and exporter of black pepper [4]. The spice is not only vital for domestic use but also constitutes a major share in international spice trade markets. Black pepper cultivation requires specific climatic and environmental conditions for optimal growth. It thrives best in hot and humid tropical climates, with well-distributed annual rainfall ranging between

2,000–3,000 mm, moderate shade, and loamy soils rich in organic matter. The plant also demands proper drainage, high humidity, and mild temperatures ranging between 10 °C and 35 °C. When these ideal conditions are not met, due to excessive rainfall, drought, poor soil quality, or pest invasions, plants become susceptible to various diseases.

Among the most common diseases affecting black pepper are leaf diseases such as anthracnose, yellowing, quick wilt, and slow wilt [5]. These diseases lead to visible symptoms including discoloration, leaf fall, wilting, and eventually plant death, resulting in severe yield losses. Early identification of these diseases is vital to implementing effective control strategies and minimizing the impact on crop health and productivity. In addition to detecting the presence of disease, assessing the severity of infection helps in determining the urgency and scale of intervention required [6]. However, traditional methods of disease identification involve manual inspection by agricultural experts, which is not only time-consuming and labor-intensive but also prone to subjectivity and inconsistency, especially across large farms. To address these challenges, researchers have increasingly turned to machine learning (ML) and deep learning (DL) techniques [7] for automatic disease detection and classification in agriculture [8], [9]. These AI-based approaches can analyze large volumes of image data quickly, offering consistent and scalable solutions. Despite their growing application in crops such as rice, wheat, maize, and tomato, very limited research exists on using ML and DL methods specifically for black pepper leaf disease classification. Furthermore, a major bottleneck in black pepper disease research is the absence of a standardized and publicly available dataset. Most existing studies rely on self-collected datasets, which are often small, unbalanced, and lack the diversity needed for developing robust models [10]–[18]. These limitations lead to poor generalization and inaccurate predictions when applied to real-world field conditions.

To overcome these limitations, this work presents a new, well-labeled dataset of black pepper leaf images encompassing five key classes: healthy, anthracnose, yellowing, quick wilt, and slow wilt. Additionally, this work introduces a novel model, deep attention-based transformer-optimized weighted hyper-tuned extreme gradient boosting network (DAT-OWH-XGBNet). This model addresses class imbalance, leverages both global and local features through attention mechanisms and multi-scale processing, and enhances classification accuracy by combining the strengths of transformers and ensemble learning.

Unlike conventional convolutional neural network (CNN) and transfer learning (TL) approaches, the proposed DAT-OWH-XGBNet integrates data efficient image transformer (DeiT)-based global attention, atrous spatial pyramid pooling (ASPP)-convolutional block attention module (CBAM) multi-scale feature extraction, and an OWH-extreme gradient boosting (XGBoost) classifier, enabling robust feature learning and improved black pepper disease classification under diverse field conditions. The key contributions of this study are as follows:

- A new black pepper leaf dataset comprising 1,640 images across five classes was collected under natural field conditions and annotated by agricultural experts, addressing the lack of a standardized dataset in existing works.
- A novel classification approach, DAT-OWH-XGBNet, is proposed, integrating transformer-based attention (DeiT), multi-scale feature extraction ASPP, and channel-spatial attention CBAM to capture comprehensive disease features.
- The proposed method presents an OWH-XGBoost classifier to handle class imbalance, optimize tree structure, and enhance classification accuracy.
- The architecture enables extraction of both fine-grained local patterns and global contextual cues, improving model robustness against diverse field conditions.
- Extensive experiments demonstrate better performance of DAT-OWH-XGBNet over existing models, showing effectiveness for practical agricultural disease monitoring applications.

Further, the manuscript is organized in the following way. Section 2 discusses the existing black pepper leaf diseases prediction and classification approaches. Section 3 presents DAT-OWH-XGBNet methodology in detail. Section 4 presents the results achieved by DAT-OWH-XGBNet and compares it with existing approaches discussed in the literature survey. Finally, section 5 concludes the work and provides future work.

2. LITERATURE SURVEY

This section discusses the existing prediction and classification approaches for black pepper leaf diseases. Adam *et al.* [10] aimed at enhancing black pepper disease and pest diagnosis efficiency. This work addressed challenges of crop loss by developing Dr. LADA, an Android-based mobile application. The application utilizes a decision-tree (DT) approach with forward-backward chaining for identifying six pests and ten black pepper leaf diseases based on user-input symptoms. The findings showed that Dr. LADA achieved 97% accuracy for classification of ten black pepper leaf diseases. Khew *et al.* [11], aimed at evaluating effectiveness of DL approaches for identification of black pepper leaf diseases and nutrient

deficiencies, hence, presented a customized CNN approach. In this work, the customized CNN approach was compared with inception version 3 (InceptionV3) and visual geometry group 16 (VGG-16). For evaluation of the work, collected 947 black pepper leaf images from Malaysia, which they increased to 9,532 images using data-augmentation approach. The findings showed that the customized CNN achieved 98% recall performance, which showed better performance in comparison with InceptionV3 and VGG-16.

Chen *et al.* [12], aimed at evaluating different traditional CNN approaches for detecting diseases and nutrient deficiencies in black pepper plants using leaves. For evaluation, collected dataset consisting of 1,043 images of 11 classes of disease and deficiencies from Sarawak farms under controlled conditions. Further, for increasing dataset size, utilized data augmentation. For evaluating different CNN models, considered a customized CNN, densely connected convolutional network with 121 layers (DenseNet121), residual network version 2 with 50 layers (ResNet50 V2), mobile network version 2 (MobileNetV2) and efficient network baseline 0 (EfficientNetB0), where MobileNetV2 and EfficientNetB0 achieved 88% accuracy. Kini *et al.* [13], aimed at developing intelligent system for detection of black pepper leaf diseases using DL and TL approach, hence, considered pre-trained ImageNet dataset, which was fine-tuned using custom black pepper leaf dataset, comprising real-time field images annotated by the agricultural experts. The diseases considered for the study included healthy leaves, yellowing leaves, phytophthora leaves, early-stage phytophthora leaves, slow wilt and anthracnose. For evaluation, considered ResNet18, squeeze network (SqueezeNet), Google network (GoogLeNet) and InceptionV3, where these models achieved accuracies between 99.1% to 99.5%, using 0.001 learning rate. Kini *et al.* [14], aimed at detecting early-stage black pepper leaf diseases using DL approaches. In this work, TL was applied on pre-trained CNN approached, which include ResNet18, SqueezeNet, GoogLeNet and InceptionV3 which was initially trained using ImageNet dataset. The CNN models were fine-tuned using custom dataset of real-time black pepper leaf images, which were annotated by experts. In this work, a classification study was conducted on six classes, which included healthy leaves, yellowing leaves, phytophthora leaves, early-stage phytophthora leaves, slow wilt and anthracnose. The findings showed that ResNet18 achieved 99.67% accuracy, showing better outcomes. Sasi *et al.* [15], aimed at developing automated system for black pepper leaf disease classification for accurate disease identification. This work addressed challenges in manual identification by applying AI and DL approaches. Using dataset of black pepper leaf images, DL approaches were trained for recognizing common diseases like powdery mildew and fungal infections which affect yield and crop quality. The approach showed better accuracy, providing fast disease detection in comparison with existing approaches.

Sreethu *et al.* [16], aimed at developing DL approach for detection of black pepper leaf diseases for enhancing crop management, hence, trained CNN model using a collected dataset from Kerala, India. The dataset consisted of 2,786 images, which consisted of phytophthora foot-rot, anthracnose, early anthracnose and healthy leaves. They used data augmentation for expanding dataset to 18,234 images. For comparison, considered eight DL pre-trained approaches, where the trained CNN model achieved 98.72% accuracy. Hegde and Padyana [17], aimed at developing DL model for detecting black pepper leaf disease, hence utilized SqueezeNet, a lightweight CNN, which was trained on leaf images for distinguishing among diseased and healthy plants on basis of various symptoms like deformation and discoloration. Using the symptoms, the features were extracted and the SqueezeNet was trained. The dataset consisted of 1,000 images, where the SqueezeNet achieved 99% accuracy. Kini and Kamath [18], aimed at enabling detection and classification of black pepper leaf diseases using ML and DL hybrid approaches. A dataset was collected which consisted of yellowing, quick wilt, slow wilt, phytophthora and anthracnose, which was preprocessed and labeled under experts. In this work, they utilized ResNet50 for extracting relevant features, which were further mapped to multiclass support vector machine (SVM) for classification, eliminating need for separated model training.

Despite the promising results demonstrated in previous studies, several limitations persist in terms of dataset quality, feature extraction techniques, and classification accuracy. For instance, Adam *et al.* [10] used a rule-based DT approach in Dr. LADA, which relied solely on user-input symptoms without image-based feature extraction, limiting generalization. Khew *et al.* [11] and Chen *et al.* [12] used data augmentation to artificially inflate dataset sizes, but the original datasets remained small and lacked environmental variability, which impairs model robustness [19], [20]. Models like DenseNet, MobileNetV2, and EfficientNetB0 also struggled to capture both local disease textures and global leaf context [21], [22]. A. Kini *et al.* [13], [14], [18] leveraged TL but were constrained by limited, imbalanced datasets and shallow feature representations [23]. Hegde and Padyana [17] used lightweight networks like SqueezeNet, which sacrificed representational depth for speed. It is seen that most works did not address class imbalance or complex feature hierarchies [24], [25]. In contrast, DAT-OWH-XGBNet overcomes these gaps by incorporating edge-enhanced transformer-based [23] feature extraction DeiT, multi-scale contextual analysis ASPP, and attention refinement CBAM. It also uses OWH-XGBoost to handle imbalanced classes and improve decision boundaries [24], [25], while relying on a large, real-field dataset to ensure generalization and high classification accuracy.

3. PROPOSED METHOD

This section presents the proposed DAT-OWH-XGBNet for black pepper leaf disease classification. This section first discusses the architecture of DAT-OWH-XGBNet, then discusses the dataset collected for this study, then presents the preprocessing steps considered for inputting the image to DAT-OWH-XGBNet architecture. Further, the feature extraction process and the classification process of the DAT-OWH-XGBNet is discussed in detail. Finally, the performance evaluation metrics used for evaluating the DAT-OWH-XGBNet is discussed.

3.1. Architecture

This section presents the DAT-OWH-XGBNet architecture, which is visually presented in Figure 1. The DAT-OWH-XGBNet architecture has been designed using a multi-stream feature extraction and a classification approach which ensures classification of black pepper leaf diseases. In this work, first the input black pepper leaf images were preprocessed by resizing, normalizing, and applying a Sobel-based edge-detection filter. This preprocessing step enhanced leaf contours and infected regions, which are important for accurate feature extraction and downstream processing. After preprocessing, feature extraction was performed in two parallel branches. In the first branch, this work employed DeiT, which divided the image into fixed-size patches and processed them using self-attention mechanism for capturing global-contextual information. Simultaneously, the second branch applies ASPP for capturing multi-scale features and then passes the features to CBAM, which refines focus on relevant spatial-channel regions. The outputs from both branches are concatenated to form a comprehensive fused feature vector that encapsulates both local-global attributes. This fused representation is then passed to OWHT-XGBoost classifier. The classifier is designed for handling class imbalance and performs fine-grained classification into five classes, anthracnose, yellowing, quick-wilt, slow-wilt, and healthy. Using both hierarchical and attention-guided feature representations, the DAT-OWH-XGBNet architecture significantly provides better classification performance.

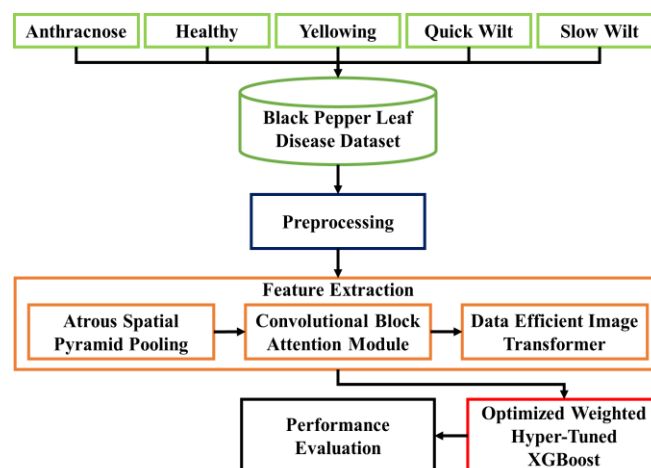


Figure 1. DAT-OWH-XGBNet architecture

3.2. Dataset

For evaluation of the DAT-OWH-XGBNet architecture, this work collected black pepper leaf images. The leaf images were collected from agricultural fields located in the state of Kerala, specifically within the district of Wayanad, India. The dataset comprised 1,640 samples, which were captured under natural lighting conditions using a professional-grade camera to maintain clarity and consistency. The dataset included five distinct classes of black pepper leaf conditions. The classes include anthracnose, yellowing, quick wilt, slow wilt, and healthy. The complete samples considered for each class is given in Table 1. Also, the samples of each class are given in Figure 2.

In this dataset, the healthy samples were carefully selected to ensure no visible signs of infection or disease, serving as the control group. In contrast, the diseased samples represented various visual symptoms linked with common black pepper leaf diseases. All the black pepper leaf images were collected directly from field-grown pepper plants and were annotated under the supervision of a qualified agricultural expert, which ensured accurate labeling of each disease type. During dataset collection, special attention was given

while capturing variations in background interferences, lighting, and leaf orientation. This variability introduced real-world complexities like shadows and occlusions, making the dataset ideal for training the DAT-OWH-XGBNet model. By capturing this diversity, the dataset supported generalization, providing DAT-OWH-XGBNet to perform reliably across different environments and visual scenarios, which is typically encountered in real agricultural settings.

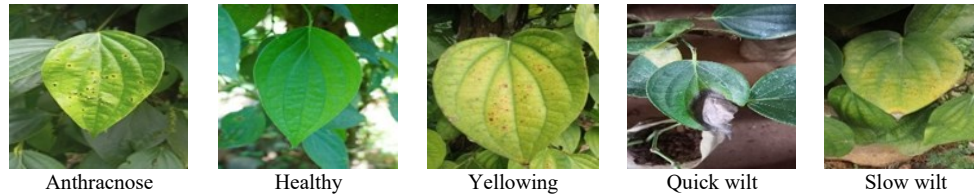


Figure 2. Samples of collected black pepper leaf disease dataset

Table 1. Black pepper leaf disease collected data samples

SL. No	Class	Samples
1	Anthracnose	170
2	Healthy	180
3	Quick wilt	600
4	Slow wilt	240
5	Yellowing	450
Total samples		1,640

3.3. Preprocessing

This section discusses the preprocessing steps considered in this study before inputting the images to the DAT-OWH-XGBNet model, as preprocessing has important role for preparing raw images for efficient and accurate model training. In this work, each image collected from dataset was resized to a standard dimension of 224×224 pixels. This specific size was chosen to ensure compatibility with DeiT, which requires fixed-size inputs for optimal processing. The resizing process not only standardized input data, but also reduced computation complexity providing DAT-OWH-XGBNet model to learn more efficiently from uniform shaped images. Further, all pixel values were normalized to range $[0,1]$, which improved convergence during model training and prevented issues related to vanishing gradients. The normalization was carried out using (1).

$$I_{norm}(i,j) = \frac{I(i,j)}{255} \quad (1)$$

In (1), $I(i,j)$ represents the original pixel intensity at (i,j) pixel coordinates, $I_{norm}(i,j)$ denotes the corresponding normalized pixel value. After the normalization process using (1), each image retained three color channels, i.e., red, green, blue (RGB). These color channels are important as it helps in preserving visual features like texture and discoloration, which are key indicators of healthy leaf and diseased leaf. In this work, data augmentation technique was applied, i.e., contrast adjustment, flipping and rotation to increase variability and robustness during training of the DAT-OWH-XGBNet model. This step ensured the DAT-OWH-XGBNet model generalized better when exposed to real-world field images, which may vary in lighting and orientation. During training, the DeiT feature extractor was optimized using the Adam optimizer with mini-batch learning to ensure stable convergence. Data augmentation was applied only to the training images, while the validation and testing datasets remained unchanged. Regularization was achieved through data augmentation, weight decay in the optimizer, and the regularization term incorporated in the OWH-XGBoost objective function, reducing overfitting and improving model generalization.

3.4. Feature extraction

This section presents the novel feature extraction approach included in DAT-OWH-XGBNet for extracting features from the black pepper leaves for classification. In ML or DL, feature extraction is an important step, where this process captures the most relevant patterns that differentiate healthy and diseased black pepper leaves. This work has employed a hybrid feature extraction approach that combines edge-based features and contextual features using Sobel filters, DeiT, ASPP, and CBAM, to ensure both local-global structure and texture details are capture efficiently. In this work, for isolating leaf contours, diseased patches

and leaf veins, the Sobel filter has been employed for detecting horizontal and vertical gradients, denoted as G_x and G_y respectively. The two filters are represented as given in (2).

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (2)$$

The filters in (2) are convolved with image for highlighting gradients using (3).

$$S(i, j) = \sum_{m=-1}^1 \sum_{n=-1}^1 I(i + m, j + n) \cdot G(m, n) \quad (3)$$

In (3), $S(i, j)$ denotes the resulting pixel after edge enhancement, $I(i, j)$ denotes input image pixel and $G(m, n)$ denotes the gradient kernel (either G_x or G_y). Following edge detection, the image is divided into patches for transformer processing using (4).

$$P = \text{Flatten}(\text{Conv2D}(S(i, j), \text{stride} = p)) \quad (4)$$

In (4), p denotes the patch size, which has been considered as 16×16 in this study and P denotes flattened patch embeddings, which is fed to first branch, i.e., DeiT. As transformers rely on self-attention for capturing relationships between different patches, the DeiT attention is evaluated using (5).

$$\text{Attention} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

In (5), Softmax is function that converts vector of numbers into a probability distribution, i.e., it takes list of numbers and transforms them into set of probabilities that add up to 1, Q denotes query, K denotes key, V denotes value matrices and d_k denotes key-vector dimensions. In this work, the Softmax function used in (5) is used for normalizing weights, which is evaluated using (6).

$$\text{Softmax} = \sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (6)$$

In (6), the z_i denotes i^{th} patch score and K denotes the total attention heads. After the process of the first branch, the image is forwarded to second branch, where ASPP is utilized for capturing features at multiple scales through parallel convolutions having varying dilation rates, using (7).

$$y_i = \text{ReLU}\left(\text{BN}(\text{Conv}(x, d_i))\right) \quad (7)$$

In (7), y_i denotes output of ASPP block, ReLU denotes activation function, i.e., rectified linear unit, BN denotes batch normalization, x denotes input feature map, d_i denotes dilation rate, i.e., $d_i \in \{1, 6, 12, 18\}$ and Conv denotes convolution operation. In (7), the activation function ReLU is applied as presented in (8).

$$\text{ReLU}(x) = \max(0, x) \quad (8)$$

Further, after ASPP block, the feature map is constructed and passed on to CBAM, where it refines feature map using channel-spatial attention approach. The channel and spatial attention process are done using (9) and (10) respectively.

$$M_c(F) = \sigma\left(\text{MLP}\left(\text{AvgPool}(F) + \text{MLP}(\text{MaxPool}(F))\right)\right) \quad (9)$$

$$M_s(F') = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (10)$$

In (9) and (10), M_c , M_s are attention masks, F denotes input feature map, σ denotes Softmax operation, which is evaluated using (6). The MLP denotes multi-layer perceptron process, AvgPool is average pooling operation and MaxPool is max-pooling operation. Further, these attentions modulate input as given in (11) and (12).

$$F' = M_c(F) \cdot F \quad (11)$$

$$F'' = M_s(F') \cdot F' \quad (12)$$

From the two branches, two types of features are extracted, i.e., F_1 , i.e., from DeiT after patch and attention-based processing and F_2 , i.e., from ASPP-CBAM after attention and multi-scale filtering. These two features are then fused into one feature map as presented in (13).

$$F = [F_1; F_2] \quad (13)$$

In (13), $[\cdot; \cdot]$ denotes concatenation operation, F denotes final hybrid feature vector. By integrating edge-focused features with context-aware features, the Sobel with DeiT captures spatial structure, boundaries and vein patterns, whereas the ASPP-CBAM highlights regions with multiscale dilation and adaptive focus. This multi-perspective representation improves black pepper leaf disease classification, mainly in condition with overlapping symptoms, background clutter and varying lighting conditions. The resulting F feature map is further passed to OWH-XGBoost for final classification, which is discussed in the next section.

3.5. Optimized weighted hyper-tuned XGBoost for classification

This study presents a novel classifier called as OWH-XGBoost for classifying black pepper leaves. In this work, XGBoost was chosen as it handles class imbalance issues and provides better outcomes for classification [26]. Similar to XGBoost approach [27], the OWH-XGBoost is designed for handling class imbalance, complex patterns and divers leaf conditions using three key components, i.e., class-weighted loss, hyperparameter tuning and leaf-optimized tree-learning. The core of OWH-XGBoost is its objective function, which provides balance between prediction loss and model complexity. The objective function is defined using (14).

$$L(\theta) = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (14)$$

In (14), $l(\hat{y}_i, y_i)$ denotes logistic loss between predicted $\hat{y}_i$ and true label y_i , $\Omega(f_k)$ denotes regularization for preventing overfitting. The regularization term is evaluated using (15).

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (15)$$

In (15), T denotes the total leaf nodes present in OWH-XGBoost, λ denotes $L2$ regularization parameter, γ denotes tree-complexity penalty and w denotes leaf-weights vector. In this work, as the black pepper disease classes as presented in Table 1 are imbalanced, with fewer samples for rare diseases and healthy, this work has applied a class-weighted loss, defined using (16).

$$l_w = -\sum_{i=1}^n w_i \cdot y_i \log(\hat{y}_i) \quad (16)$$

The class weights in (16) denoted as w_i are computed inversely to their frequency as presented in (17).

$$w_i = \frac{1}{freq(y_i)} \quad (17)$$

This ensures that minority classes receive higher gradient contributions during training. Moreover, in this work, the OWH-XGBoost utilizes second-order optimization, where the first and second derivatives of gradient and hessian of loss function are evaluated using (18).

$$g_i = G = \frac{\partial l(\hat{y}_i, y_i)}{\partial \hat{y}_i}, h_i = H = \frac{\partial^2 l(\hat{y}_i, y_i)}{\partial \hat{y}_i^2} \quad (18)$$

These derivatives help in efficiently finding optimal tree splits and improving model convergence. In OWH-XGBoost, for deciding where to split a node, the *Gain* function is utilized for evaluating benefit of split using accumulated g_i and h_i , as presented in (19).

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (19)$$

In (19), G_L , G_R denote sum of gradients on left and right child, H_L , H_R denotes sum of Hessians on left and right child, λ and γ denotes regularization parameters. This gain function provides efficient and

well-separated splits. The final output of OWT-XGBoost is aggregated result of all DTs followed by a sigmoid activation for converting probabilities, as presented in (20).

$$\hat{y} = \text{sigmoid}(\sum_{k=1}^K f_k(x)) \quad (20)$$

In (20), $f_k(x)$ denotes the output of k^{th} tree for input x and $\hat{y}$ denotes the predicted probability of disease and healthy classes. For improving performance, hyperparameter tuning was applied using grid search and random search approach. The parameters tuned for the OWH-XGBoost included learning rate η , which controlled how each tree contributed, max depth d which provided depth of each DT, subsample ratio r which provided fraction of samples used for each tree, min-child weight which provided the minimum weight of samples in child node and lambda λ which used $L2$ regularization for preventing overfitting. By searching for the optimal combination of these parameters, the OWH-XGBoost provided better classification of black pepper leaves. In the next section, the complete process of DAT-OWH-XGBNet is discussed using the algorithm.

3.6. Algorithm

This section presents the algorithm for DAT-OWH-XGBNet, which is presented in Algorithm 1. The DAT-OWH-XGBNet integrates DL and ensemble learning for accurately classifying black pepper leaves. For evaluation of the DAT-OWH-XGBNet, the performance metrics used are discussed in the next section.

Algorithm 1. DAT-OWH-XGBNet for black pepper leaf disease classification

Input: Black Pepper Leaf Images, Labels

Output: Classified disease label for each input image $\ominus$

- 1: Start
- 2: For each input image
- 3: Resize each input image to 224×224
 Normalize pixel intensities using (1).
 Apply Sobel Filter for extracting edges
- 4: End For
- 5: Extract features using first branch (DeiT), i.e., extract global features using self-attention
- 6: Extract features using second branch (ASPP + CBAM), i.e., capture multi-scale features and refine attention spatially and channel-wise
- 7: Concatenate outputs from both branches creating a feature map
- 8: Use feature map for classifying black pepper leaf disease using OWH-XGBoost
- 9: For each image during training
- 10: Compute class-weighted loss
 Calculate gradients and Hessians
 Perform tree splits based on *Gain*
- 11: End for
- 12: Predict black pepper leaf disease using (20).
- 13: Use grid and random search for tuning η , d , r , λ , min_child_weight
- 14: Return final predicted label for each image
- 15: End

3.7. Performance metrics

The DAT-OWH-XGBNet for black pepper disease classification was evaluated using four metrics, which include accuracy measured in (21).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (21)$$

The precision performance is measured using (22).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (22)$$

The recall performance is measured in (23).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (23)$$

The F-score performance is measured in (24).

$$F - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (24)$$

The results of the DAT-OWH-XGBNet are discussed in detail in the next section.

4. RESULTS AND DISCUSSION

This section outlines the implementation setup for DAT-OWH-XGBNet. Then assess the black pepper disease prediction performance using DAT-OWH-XGBNet. Further, how DAT-OWH-XGBNet can classify the diseases is studied. Finally, a comparative analysis with existing models and an in-depth discussion is provided.

4.1. System configurations

The DAT-OWH-XGBNet model was developed and executed on a system equipped with an Intel Core i7 paired with an NVIDIA T1000 GPU, and the system had 32 GB of RAM and operated on the Windows 11 platform. The development environment was based on Python 3.6, configured with compute unified device architecture (CUDA) version 11.0 and cuDNN version 8.0.4, enabling GPU acceleration for significantly faster computation compared to CPU-based execution. This setup ensured efficient model training and execution, with significantly improved performance over traditional CPU-based computation. The entire architecture was implemented in Python.

4.2. Prediction performance

The performance evaluation of the DAT-OWH-XGBNet model on binary classification (healthy vs disease) of black pepper leaf disease is presented in Figure 3. The DAT-OWH-XGBNet achieved 99.98% accuracy, 99.86% precision, 99.87% recall, and 99.87% F-score. The findings show the DAT-OWH-XGBNet capability in correctly classifying diseased and healthy leaves with minimal error. This performance is attributed to the integration of edge-enhanced patch embeddings, the attention-rich DeiT, and optimized hyper-tuned XGBoost for final classification. By incorporating class weights and tuning parameters effectively, the model handles class imbalance robustly while extracting highly discriminative features, leading to better prediction outcomes.

4.3. Classification performance

The evaluation of DAT-OWH-XGBNet on five-class classification of black pepper leaf diseases—anthracnose, yellowing, quick wilt, slow wilt, and healthy—demonstrated highly accurate performance, achieving 99.82% accuracy, 99.57% precision, 99.66% recall, and a 99.61% F-score as presented in Figure 4. This performance is attributed to the fusion of deep attention-based transformers with edge features from ASPP-CBAM, enabling extraction of both global and fine-grained features. Additionally, the OWH-XGBoost classifier effectively handled class imbalance and enhanced learning through leaf-wise optimization and hyperparameter tuning, resulting in more robust and generalizable classification across multiple disease categories.

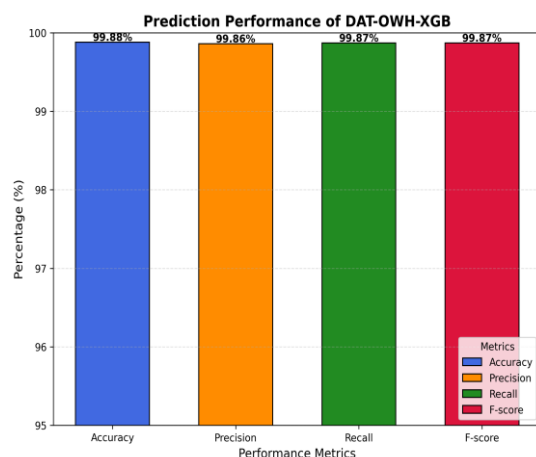


Figure 3. Performance evaluation of DAT-OWH-XGBoost on two classes

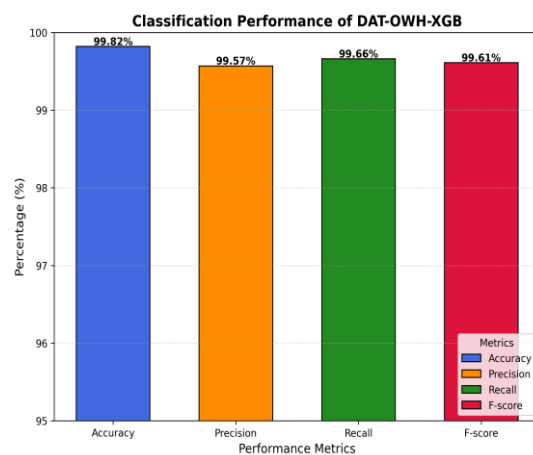


Figure 4. Performance evaluation of DAT-OWH-XGBoost on five classes

4.4. Comparative study

This section presents the comparative study, where the DAT-OWH-XGBNet is compared with existing approaches. In Table 2, class-wise performance comparison reveals that existing approaches, such as convolutional networks (ConvNets) and CNNs, demonstrate good accuracy and recall across different black pepper diseases. For instance, the CNN-based method [16] achieved recall values up to 100% for advanced anthracnose, while ConvNets [14] reached up to 99.59% accuracy for early-stage phytophthora. However, some disease classes, such as yellowing and slow wilt, were relatively less accurately classified with ConvNets, yielding F-scores of 88.89% and 88.28%, respectively. This inconsistency highlights the limitations of previous models in generalizing across varying disease manifestations. In contrast, the DAT-OWH-XGBNet significantly outperformed these approaches across all five classes. It achieved 99.71% to 99.82% accuracy, with consistent performance. These results are due to the fusion of transformer-based deep attention networks and ASPP-CBAM for dual-stream feature extraction, which captures both global and local patterns effectively. Additionally, the OWH-XGBoost classifier addresses class imbalance and overfitting through weight optimization and hyperparameter tuning.

As shown in Table 3, a comparison of models indicates that while traditional CNN architectures (e.g., DenseNet121 at 96.72%, MobileNetV2 at 96.05%) perform relatively well, their generalization is limited. Even the high-performing CNN model from [16] achieved 98.72% accuracy, but fell short in precision and F-score compared to DAT-OWH-XGBNet, which achieved 99.82% accuracy, 99.57% precision, 99.66% recall, and 99.61% F-score. The improvement stems from hybrid architecture's ability to effectively learn discriminative features while optimizing classification through tree-based gradient boosting. Hence, DAT-OWH-XGBNet provides a more robust and accurate solution for black pepper leaf disease diagnosis. The receiver operating characteristic (ROC) curve in Figure 5 illustrates the model's effectiveness in differentiating between various disease categories and healthy leaves. An area under the curve (AUC) nearing 1.0 across all classes signifies outstanding classification performance with minimal errors. This demonstrates the model's resilience in addressing multi-class plant disease detection. The confusion matrix in Figure 6 clearly shows perfect classification across all five categories: anthracnose, healthy, quick wilt, slow wilt, and yellowing. All samples fall along the diagonal with no misclassifications in the off-diagonal cells. This result indicates that the model achieves complete accuracy in distinguishing between multiple plant disease types.

Table 2. Performance comparison of class-wise black pepper leaf disease classification using existing methods and DAT-OWH-XGBNet

Reference	Class	Model	Accuracy	Precision	Recall	F-score
Kini <i>et al.</i> [14]	Anthracnose	ConvNets	98.43	94.29	95.65	94.96
	Early-stage phytophthora		99.59	100	97.53	98.75
	Healthy		98.62	100	92.77	96.25
	Phytophthora		97.34	91.78	91.78	91.78
	Slow wilt		96.2	90.14	86.49	88.28
	Yellowing		96.42	83.12	95.52	88.89
Sreethu <i>et al.</i> [16]	Initial anthracnose	CNN	99.33	98.89	99.33	99.11
	Advanced anthracnose		97.82	98.93	100	99.46
	Foot rot		97.42	99.05	97.42	98.23
	Healthy		100	97.82	97.82	97.82
	Anthracnose		99.71	99.38	99.57	99.47
Proposed	Healthy	DAT-OWH-XGBNet	99.82	99.71	99.29	99.5
	Yellowing		99.65	98.92	99.72	99.32
	Quick wilt		99.48	98.66	99.2	98.93
	Slow wilt		99.37	97.88	98.92	98.4

Table 3. Existing model performance comparison of different DL architectures and DAT-OWH-XGBNet for black pepper leaf disease classification

Reference	Models	Accuracy	Precision	Recall	F-score
Sreethu <i>et al.</i> [16]	VGG-16	92.55	92.64	92.55	92.54
	ResNet50	65.13	65.52	65.13	64.9
	VGG-19	91.16	91.38	91.16	91.1
	MobileNetV2	96.05	96.07	96.05	96.05
	DenseNet121	96.72	96.71	96.72	96.71
	InceptionV3	91.32	91.44	91.32	91.29
	Xception	91.94	92.01	91.94	91.86
	ResNet50 V2	93.16	93.31	93.16	93.19
	CNN	98.72	99.28	97.65	98.66
Proposed	DAT-OWH-XGBNet	99.82	99.57	99.66	99.61

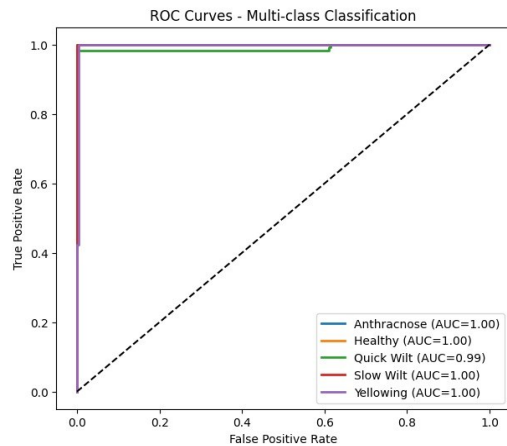


Figure 5. ROC curve of the proposed algorithm against five different classes of pepper leaves

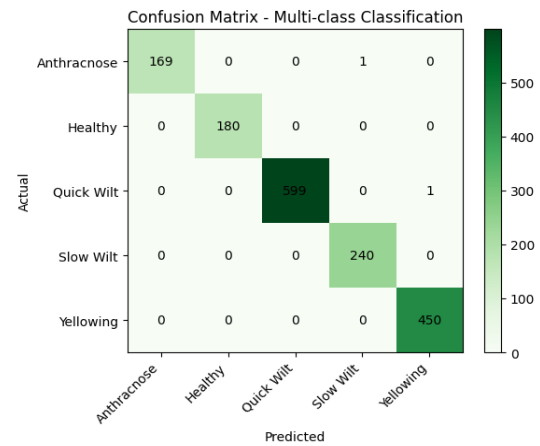


Figure 6. Confusion matrix for five different classes of pepper

4.5. Discussion

The DAT-OWH-XGBNet demonstrates better performance in black pepper leaf disease classification by effectively addressing key limitations of previous models, such as class imbalance, poor generalization, and limited feature representation. Unlike conventional CNN or single-path TL approaches, DAT-OWH-XGBNet integrates a dual-branch attention-based transformer backbone with ASPP and CBAM, enabling robust feature extraction at both global-local scales. This architectural enhancement allows DAT-OWH-XGBNet to capture fine-grained visual differences between disease types that often exhibit similar visual symptoms, such as yellowing and early-stage wilt. The hybrid design is further strengthened by OWH-XGBoost, which incorporates class weighting, gradient-level learning, and hyperparameter optimization to minimize overfitting and boost classification reliability. In addition to its high classification accuracy, the model is capable of estimating disease severity levels. This is achieved by analyzing the distribution and extent of affected regions within leaf images, allowing the classifier not only to predict the disease category but also to infer its progression stage. Such granularity is crucial for timely intervention and informed decision-making in precision agriculture. The severity estimation is implicit in the model's attention maps and gradient-based optimization, making it adaptable to varying field conditions. Overall, DAT-OWH-XGBNet not only improves accuracy metrics across multiple disease classes but also adds practical value by offering early-stage detection and severity insights that are vital for disease management and yield protection.

4.6. Model generalization on Kaggle black pepper and PlantVillage bell pepper leaf disease dataset

To evaluate the robustness and generalization ability of the proposed DAT-OWH-XGBNet model, external validation was conducted using multiple publicly available black pepper leaf disease datasets. First, the PlantVillage Bell Pepper dataset [28] was employed, comprising 2,475 images, including 1,478 healthy and 997 bacterial spot-infected samples. The dataset is publicly available via Mendeley Data at <https://data.mendeley.com/datasets/tywbtsjrjv/1> and accessible through Kaggle at <https://www.kaggle.com/datasets/zienabesam/pepper-belly-crop-plantvillage-ds>. All images were resized to 224×224 pixels and normalized before being fed into the model. The trained DAT-OWH-XGBNet was directly tested on this external dataset without retraining, ensuring a fair evaluation of its robustness and adaptability to unseen data distributions. The comparative performance of various DL architectures is summarized in Table 4. Traditional models such as MobileNetV2, DenseNet121, and ResNet50 V2 achieved accuracies between 93% and 97%, while the proposed DAT-OWH-XGBNet demonstrated a superior accuracy of 99.15%, outperforming existing approaches across all metrics, including precision (98.972%), recall (98.98%), and F-score (99.01%).

Additionally, two Kaggle black pepper datasets were merged for extended validation. Kaggle black pepper sample 1 [29] includes 500 images each of foot rot (quick wilt), pollu disease, and slow decline, while Kaggle black pepper sample 2 [30] comprises 273 images each of healthy, leaf blight, and yellow mottle virus. All images were standardized to 256×256 pixels and augmented via rotation and mirroring. The comparative performance of various DL architectures is summarized in Table 5. On the combined dataset, DAT-OWH-XGBNet achieved 98.01% accuracy, significantly outperforming baseline models (ConvNet,

MobileNetV2, DenseNet121, ResNet50V2, and standard CNN), demonstrating high generalization across different disease types, imaging conditions, and geographic origins. These results confirm that DAT-OWH-XGBNet reliably generalizes to unseen data, demonstrating its strong discriminative capability and potential for automated crop disease diagnosis across diverse agricultural settings.

Table 4. Performance comparison of DAT-OWH-XGBNet on PlantVillage bell pepper leaf datasets

Models	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
ConvNets [14]	98.01	98.23	96.91	97.1
MobileNetV2 [16]	96.02	96.03	96.02	96.02
DenseNet121 [16]	96.78	96.8	96.8	96.81
ResNet50 V2 [16]	93.19	93.48	93.15	93.29
CNN [16]	97.93	98.95	97.15	98.01
DAT-OWH-XGBNet [proposed]	99.15	98.97	98.98	99.01

Table 5. Performance comparison of DAT-OWH-XGBNet on merged black pepper leaf datasets

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ConvNet [14]	95.34	95.12	94.85	95.00
MobileNetV2 [16]	92.87	92.65	92.40	92.52
DenseNet121 [16]	93.45	93.20	93.15	93.17
ResNet50 V2 [16]	91.68	91.45	91.32	91.38
CNN [16]	94.80	94.55	94.30	94.42
DAT-OWH-XGBNet [proposed]	98.01	97.81	97.77	97.79

5. CONCLUSION

Black pepper, often termed the “king of spices,” plays a vital role in the global spice economy, with significant cultivation in countries like India, Vietnam, and Indonesia. However, its production is severely impacted by various leaf diseases such as anthracnose, yellowing, quick wilt, and slow wilt. Traditional identification methods are labor-intensive, time-consuming, and prone to human error. While ML and DL have recently shown promise in agricultural diagnostics, a major gap remains in the form of limited and imbalanced datasets specific to black pepper leaf diseases. Most prior works relied on small, non-standardized datasets and lacked robust classification accuracy. To address these challenges, this study proposed a novel architecture, DAT-OWH-XGBNet. The objectives of this work included collecting a comprehensive dataset of over 1,640 annotated images across five disease classes, implementing advanced feature extraction using DeiT and ASPP-CBAM branches, and optimizing classification through a customized XGBoost with class-weighted learning and hyperparameter tuning. The findings show that the proposed DAT-OWH-XGBNet achieved performance of 99.88% accuracy for prediction and 99.82% accuracy for classification, outperforming traditional CNN models. Its architecture efficiently captures both global and local features, enhancing generalization and robustness under diverse field conditions. From an application perspective, the proposed DAT-OWH-XGBNet can support AI-enabled mobile advisory systems by providing rapid and accurate black pepper disease diagnosis. Such systems can assist farmers in timely disease identification, enabling early intervention, reducing crop losses, and promoting precision agriculture practices. The proposed DAT-OWH-XGBNet also exhibits potential for deployment on mobile and edge computing platforms after appropriate model optimization and compression. Such deployment would enable low-latency, on-field disease diagnosis without continuous cloud connectivity, making the framework suitable for practical precision agriculture applications. Future work will focus on three directions. In the short term, the proposed model will be validated on larger black pepper datasets collected from diverse geographical regions. In the medium term, disease severity estimation and cross-dataset evaluation will be investigated to further improve model robustness. In the long term, DAT-OWH-XGBNet will be integrated into lightweight mobile and edge-based advisory systems for real-time disease diagnosis and precision agriculture.

FUNDING INFORMATION

No funding has been received for current research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Saritha Suvarna	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Demian Antony Dmello	✓	✓			✓	✓				✓		✓	✓	

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

No financial institute has any financial competing interest from the research work. On behalf of all authors, the corresponding author states that there is no conflict of interest.

ETHICAL APPROVAL

No humans or animals have been used in this research study.

DATA AVAILABILITY

All tools and data used for experimenting will be made available to the author on genuine request through mail which must be cited properly.

REFERENCES

[1] C. Spence, "The king of spices: on pepper's pungent pleasure," *International Journal of Gastronomy and Food Science*, vol. 35, Mar. 2024, doi: 10.1016/j.ijgfs.2024.100900.

[2] J. N.-Guz and M. Śmiechowska, "Health benefits and risks of consuming spices on the example of black pepper and cinnamon," *Foods*, vol. 11, no. 18, Sep. 2022, doi: 10.3390/foods11182746.

[3] G. Velarde *et al.*, "Tree boosting methods for balanced and imbalanced classification and their robustness over time in risk assessment," *Intelligent Systems with Applications*, vol. 22, Jun. 2024, doi: 10.1016/j.iswa.2024.200354.

[4] R. Varghese and J. G. Ray, "Sustainability of black pepper production: a critical analysis of physicochemical soil parameters concerning variables in pepper fields of South India," *Ecological Frontiers*, vol. 44, no. 4, pp. 788–801, Aug. 2024, doi: 10.1016/j.ecofro.2024.01.005.

[5] R. S. Bhai, A. I. Bhat, C. N. Biju, A. K. Vijayan, and K. A. Saju, "Diseases of black pepper and cardamom," in *Handbook of Spices in India: 75 Years of Research and Development*. Singapore: Springer Nature Singapore, 2024, pp. 3623–3673, doi: 10.1007/978-981-19-3728-6_55.

[6] X. Zheng, Z. Shao, Y. Chen, H. Zeng, and J. Chen, "MSPB-YOLO: high-precision detection algorithm of multi-site pepper blight disease based on improved YOLOv8," *Agronomy*, vol. 15, no. 4, Mar. 2025, doi: 10.3390/agronomy15040839.

[7] P. Sajitha, A. D. Andrushia, N. Anand, and M. Z. Naser, "A review on machine learning and deep learning image-based plant disease classification for industrial farming systems," *Journal of Industrial Information Integration*, vol. 38, Mar. 2024, doi: 10.1016/j.jii.2024.100572.

[8] A. Singla *et al.*, "Exploration of machine learning approaches for automated crop disease detection," *Current Plant Biology*, vol. 40, Dec. 2024, doi: 10.1016/j.cpb.2024.100382.

[9] S. U. Khan, A. Alsuhaibani, A. Alabduljabbar, F. Almarshad, Y. N. Altherwy, and T. Akram, "A review on automated plant disease detection: motivation, limitations, challenges, and recent advancements for future research," *Journal of King Saud University Computer and Information Sciences*, vol. 37, no. 3, May 2025, doi: 10.1007/s44443-025-00040-3.

[10] A. Adam *et al.*, "Dr. LADA: diagnosing black pepper pests and diseases with decision tree," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 4–2, pp. 1584–1590, Oct. 2018, doi: 10.18517/ijaseit.8.4-2.6818.

[11] C. Y. Khew, Y. Q. Teow, E. T. Lau, S. S. Hwang, C. H. Bong, and N. K. Lee, "Evaluation of deep learning for image-based black pepper disease and nutrient deficiency classification," in *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, Sep. 2021, pp. 1–6, doi: 10.1109/AiDAS53897.2021.9574346.

[12] O. C. H. Chen, C. H. Bong, and N. K. Lee, "Benchmarking CNN models for black pepper diseases and malnutrition prediction," in *2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAJET)*, Sep. 2023, pp. 146–151, doi: 10.1109/IICAJET59451.2023.10291604.

[13] A. S. Kini, P. K. V., and S. N. Pai, "State of the art deep learning implementation for multiclass classification of black pepper leaf diseases," *Preprint Research Square*, Aug. 25, 2023, doi: 10.21203/rs.3.rs-3272019/v1.

[14] A. S. Kini, K. V. Prema, and S. N. Pai, "Early-stage black pepper leaf disease prediction based on transfer learning using ConvNets," *Scientific Reports*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-024-51884-0.

[15] D. Sasi, A. P. J. Jinso, J. Paul, and D. V., "Black pepper leaf disease classification using deep learning models," *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2024, pp. 1-6, doi: 10.1109/RAICS61201.2024.10690146.

[16] P. T. Sreethu, M. M. Paul, P. P. Gopinath, I. L. Shahana, and N. S. Radhika, "Foliar symptom-based disease detection in black pepper using convolutional neural network," *Phytopathology Research*, vol. 7, no. 1, Mar. 2025, doi: 10.1186/s42483-024-00305-1.

[17] J. B. G., R. Hegde, and A. Padyana, "Black pepper leaf disease detection using deep learning," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 2, pp. 897–907, Mar. 2025, doi: 10.53894/ijirss.v8i2.5389.

- [18] A. S. Kini and R. Kamath, "Integrating deep learning and mSVM for category of class prediction of spice plant leaf diseases," *Preprint Research Square*, Jun. 3, 2025, doi: 10.21203/rs.3.rs.
- [19] S. S. A. Begum and H. Syed, "GSAtt-CMNetV3: pepper leaf disease classification using osprey optimization," *IEEE Access*, vol. 12, pp. 32493–32506, 2024, doi: 10.1109/ACCESS.2024.3358833.
- [20] S. Çetinkaya and A. T. Gursel, "Pep-VGGNet: a novel transfer learning method for pepper leaf disease diagnosis," *Applied Sciences*, vol. 15, no. 15, Aug. 2025, doi: 10.3390/app15158690.
- [21] Z. Cai, N. Farhana, A. M. Karim, F. Zhai, W. Huang, and M. Guo, "IMNM: integrated multi-network model for identifying pepper leaf diseases," *Frontiers in Plant Science*, vol. 16, Sep. 2025, doi: 10.3389/fpls.2025.1558349.
- [22] A. Shrotriya, A. K. Sharma, S. Prabhu, and A. K. Bairwa, "An approach toward classifying plant-leaf diseases and comparisons with the conventional classification," *IEEE Access*, vol. 12, pp. 117379–117398, 2024, doi: 10.1109/ACCESS.2024.3411013.
- [23] S. Huddar, K. Prabhushetty, J. Jakati, R. Havaladar, and N. Sirdeshpande, "Deep autoencoder-based image enhancement approach with hybrid feature extraction for plant disease detection using supervised classification," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 4, Aug. 2024, doi: 10.11591/ijece.v14i4.pp3971-3985.
- [24] K. Taji, A. Sohail, T. Shahzad, B. S. Khan, M. A. Khan, and K. Ouahada, "Corrections to 'an ensemble hybrid framework: a comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plant disease classification,'" *IEEE Access*, vol. 13, pp. 31659–31659, 2025, doi: 10.1109/ACCESS.2025.3542202.
- [25] M. Khalid and M. A. Talukder, "A hybrid deep multistacking integrated model for plant disease detection," *IEEE Access*, vol. 13, pp. 116037–116053, 2025, doi: 10.1109/ACCESS.2025.3583793.
- [26] S. Kienzle, G. Velarde, and B. Gannon, "AI and prediction of criminal justice risk: revisiting the COMPAS dataset, experiments, and the problem of algorithmic bias and fairness," *AI and Ethics*, vol. 6, no. 3, Jun. 2026, doi: 10.1007/s43681-026-01116-0.
- [27] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
- [28] J. A. Pandian and G. Geetharamani, "Data for: identification of plant leaf diseases using a 9-layer deep convolutional neural network," *Mendeley data*, vol. 1, 2019, doi: 10.17632/tywbtsrjrv.1.
- [29] T. G. Adithyan, "Black pepper leaf disease mini-dataset," Kaggle, 2025. [Online]. Available: <https://www.kaggle.com/datasets/adithyantg/black-pepper-leaf-disease-mini-dataset>
- [30] U. Melan, "Black pepper leaf blight and yellow mottle virus," Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/udi17live/black-pepper-leaf-blight-and-yellow-mottle-virus>

BIOGRAPHIES OF AUTHORS



Saritha Suvarna    is an assistant professor in the Department of Computer Science and Engineering, Canara Engineering College, Mangalore, Karnataka, India. She has 10 years of teaching experience. She is pursuing her Ph.D. in Computer Science and Engineering at Visvesvaraya Technological University, Belgaum, Karnataka, India. Her teaching and research interest areas are machine learning, data science, visualization, and computer vision. She has presented papers in various conferences and published papers in International Journals. She can be contacted at email: sarithasuvarna25@gmail.com.



Demian Antony Dmello    is a professor in Department of Computer Science and Engineering, Canara Engineering College, Mangalore, Karnataka, India. Education: Ph.D. from National Institute of Technology Karnataka, Mangalore. M.Tech. from National Institute of Technology Karnataka, Mangalore. Current position: professor in the Department of Computer Science and Engineering and Dean Academics at Canara Engineering College, Mangalore, Karnataka, India. Teaching experience: over 25 years of teaching various subjects in the Department of Computer Science and Engineering. Research: more than 95 papers published in journals and conferences. Areas of interest: web services, artificial intelligence, machine learning, computer vision, and big data. He can be contacted at email: demian.antony@gmail.com.