Ingénierie des Systèmes d'Information
Vol. 31, No. 3, March, 2026, pp. 781-790
Journal homepage: http://iieta.org/journals/isi

A Hybrid Attention and Transformer-Based Deep Network with Multi-Scale Feature Fusion for Black Pepper Leaf Disease Detection

Saritha Suvarna*, Demian Antony Dmello

Department of Computer Science and Engineering, Canara Engineering College (Affiliated to Visvesvaraya Technological University, Belagavi), Benjanapadavu 574219, India

Corresponding Author Email: sarithasuvarna25@gmail.com

https://doi.org/10.18280/isi.310312

Received: 11 December 2025
Revised: 15 February 2026
Accepted: 20 March 2026
Available online: 31 March 2026

Keywords:
black pepper, deep attention-based transformer network, deep learning, vision transformer, convolutional block attention module, atrous spatial pyramid poolin, feature fusion, plant disease

ABSTRACT

Plant diseases cause substantial yield losses in global agriculture, and black pepper cultivation in India is severely affected by leaf diseases such as slow wilt, anthracnose, quick wilt, and nutrient deficiencies. Traditional manual inspection methods are often inaccurate and delayed, resulting in improper pesticide usage and reduced crop productivity. The objective of this work is to develop an automated and accurate deep learning model for black pepper leaf disease identification by jointly learning structural and contextual disease features. To achieve this, a Deep Attention-Based Transformer Network (DATNet) is proposed, which integrates convolutional and transformer-based learning in a unified framework. The model employs image preprocessing and Sobel-based edge detection to enhance structural lesion boundaries, while Atrous Spatial Pyramid Pooling (ASPP) extracts multi-scale spatial features. Convolutional Block Attention Module (CBAM) is used to emphasize disease-relevant regions, and a Data-Efficient Image Transformer (DeiT) captures global contextual relationships. These complementary features are fused to form a discriminative representation for disease classification. The proposed DATNet model was trained and evaluated on a custom dataset collected from black pepper farms in India and achieved an accuracy of 98.04%, outperforming conventional deep learning models including CNN, ResNet, DenseNet, Inception, and VGG16. The results demonstrate that the synergistic integration of edge-aware spatial learning, attention refinement, and transformer-based global modeling significantly enhances black pepper leaf disease recognition, making DATNet suitable for practical and early-stage disease diagnosis in precision agriculture.

1. INTRODUCTION

Agriculture is backbone of many economies across the world, and crop health plays a critical role in sustaining food security and farmer's livelihoods. However, plant diseases are responsible for significant global crop losses every year, leading to a substantial decrease in yield, reduced food supply, and economic strain on farming communities [1, 2]. According to the Food and Agriculture Organization (FAO), nearly 20% to 40% of global crop production is lost annually due to pests and diseases, highlighting critical need for early and accurate disease detection mechanisms [3]. These plant diseases are primarily caused by a combination of biotic factors such as fungi, bacteria, viruses, pests, and abiotic stressors like nutrient deficiency, poor irrigation, and unsuitable soil conditions [4, 5]. Among these, fungal and bacterial pathogens remain the most common and destructive causes of plant health deterioration [6, 7]. Moreover, traditionally, farmers have relied on manual and experience-based methods to identify and predict plant diseases. This involves visual inspection of leaves, stems, and fruits, often guided by past farming knowledge or informal consultations with local experts. While this method is time-tested, it is highly subjective and prone to human error. In many cases, the disease symptoms appear similar, leading to misidentification. As a result, farmers may apply incorrect chemicals or pesticides, which not only fail to resolve the issue but also degrade soil health and contribute to environmental pollution. The lack of timely and accurate diagnosis leads to progressive crop damage, reduced productivity, and significant economic losses.

Among the various high-value crops, black pepper holds prominent position, particularly in India, which is one of the largest producers and exporters of this spice globally. Often referred to as the "King of Spices", black pepper is widely used in culinary, medicinal, and industrial applications due to its pungent flavor and therapeutic properties [8, 9]. The spice has been cultivated in India for centuries, with Kerala, Karnataka, Tamil Nadu, and parts of the Northeastern states serving as the major cultivation zones. The crop thrives in tropical climates with high humidity and well-drained soil, but its growth and yield are heavily influenced by environmental and management conditions [10]. Despite its importance, black pepper cultivation faces severe threat from various diseases affecting its leaves, which are first visible indicators of underlying plant health issues. The diversity in soil types, use

of different fertilizers, pesticides, and varying irrigation practices across regions contributes to a range of leaf diseases. Some of the most common ones include algal leaf spot, anthracnose, bacterial blight, chlorosis (due to iron or magnesium deficiency), nitrogen deficiency, slow wilt, quick wilt, and pest-induced damage [11, 12]. These diseases not only affect the quality of the leaves but also hinder photosynthesis and plant development, leading to reduced yield, poor spice quality, and sometimes total crop failure.

In recent years, the advancement of Deep Learning (DL) techniques has opened new avenues for automated and precise plant disease identification. Convolutional Neural Networks (CNNs) have become a popular tool for image-based disease classification due to their powerful feature extraction capabilities [13, 14]. Several works have utilized CNNs for identifying diseases in crops like rice, tomato, grapevine, and apple [15-17]. However, very limited research has focused specifically on black pepper leaf disease classification. The few existing studies primarily aim at disease classification, rather than early prediction or detailed feature analysis, which is critical for timely intervention and disease management [18-24].

Despite these advances, existing CNN- and transformer-based approaches still face important technical limitations in black pepper leaf disease detection. Most current models rely on a single-stream learning strategy that extracts features either from raw RGB images or from deep convolutional layers, without explicitly modeling disease boundary structures and fine-grained lesion morphology. Moreover, transformer-based models, while effective in capturing global context, often lack mechanisms to emphasize local structural cues such as vein distortions, edge irregularities, and lesion contours that are crucial for differentiating visually similar black pepper diseases. In addition, multi-scale symptom variations and background clutter remain insufficiently addressed by conventional architectures, leading to reduced robustness under real-field conditions. These unresolved challenges highlight the need for a unified framework that can jointly exploit edge-based structural information, attention-guided contextual features, and global dependency modeling for more reliable black pepper leaf disease identification.

To address these limitations, the present work introduces a novel model, Deep Attention-Based Transformer Network (DATNet) for accurate identification of black pepper leaf diseases. Unlike traditional CNN-based models, DATNet incorporates a dual-path architecture that not only captures low-level features like edges and textures through Sobel edge detection, but also learns high-level contextual information using Data-efficient Image Transformer (DeiT). Furthermore, it integrates Atrous Spatial Pyramid Pooling (ASPP) for multi-scale feature learning and Convolutional Block Attention Module (CBAM) to enhance the attention on disease-relevant features. By fusing features from both paths, DATNet provides a more comprehensive understanding of disease patterns, leading to improved performance in identifying diverse and complex disease types affecting black pepper leaves. This approach thus offers a robust, scalable, and efficient solution for early detection and effective management of black pepper diseases, ultimately supporting farmers with smarter and more sustainable agricultural practices.

The main novelty of this work lies in the proposed DATNet framework, which introduces a dual-stream learning strategy that explicitly separates structural and contextual feature representations for black pepper disease detection. Unlike conventional single-stream CNN or transformer-based models, DATNet integrates Sobel-based edge enhancement with transformer-based global modeling in one stream, while simultaneously employing ASPP and CBAM to extract multi-scale and attention-refined contextual features in a parallel stream. These complementary representations are fused to form a disease-aware embedding that captures both lesion boundaries and texture-level symptoms. This pipeline-level orchestration establishes a task-driven inductive bias for agricultural disease imaging and demonstrates how edge priors, attention refinement, and transformer self-attention can be synergistically combined under limited data conditions. The proposed model achieves superior performance over several state-of-the-art deep learning baselines, confirming the effectiveness of this integrated design.

The contributions of DATNet are presented below.

•A novel model DATNet is proposed for accurate identification of black pepper leaf diseases.

•A dataset has been collected, which consists of over 800 high-resolution images of both healthy and diseased black pepper leaves, captured from real-world farms in India.

•The model integrates both CNN and Transformer-based architectures using a dual-path approach to enhance disease feature learning.

•Techniques like Sobel edge detection, ASPP, CBAM and DeiT are utilized to capture both low-level and high-level features.

•The DATNet model significantly outperforms traditional DL models such as CNN, ResNet, DenseNet, Inception, and VGG16 in terms of accuracy, precision, recall, and F1-score.

•The work emphasizes not just classification but also early prediction of diseases, addressing a key limitation in previous research.

The manuscript is organized in the following manner. In Section II, the literature survey is discussed, where different DL-based approaches presented for black pepper are discussed. In Section III, the DATNet model is discussed in detail. Section IV discusses results of DATNet model compared with existing DL approaches. Finally, Section V presents conclusion and future work.

2. LITERATURE SURVEY

This section discusses the existing approaches presented for black pepper leaf disease. Dai et al. [18], presented an Improved-Lightweight CNN approach for pepper-leaf disease prediction. The approach was based on an improved GoogLeNetwork (GoogLeNet), where InceptionNetwork (InceptionNet) was compressed for minimizing parameters, providing better prediction speed and decreasing memory-consumption time, called as GoogLeNet Enhanced Lightweight (GoogLeNet-EL). Also, Spatial-Pyramid Pooling layer was incorporated for capturing local-global features. For testing their IL-CNN approach, a pepper leaf disease dataset which had 9,183 images having 6 classes were used for training and testing. Findings showed that for pepper leaf disease classification, GoogLeNet with InceptionNet-V1 achieved 91.87% accuracy, GoogLeNet with InceptionNet-V3 achieved 91.23% accuracy and GoogLeNet-EL achieved 97.87% accuracy. When compared to different CNN approaches like MobileNetwork-V2, ResidualNetwork-50 and AlexNetwork, the GoogLeNet-EL achieved less memory. Bezabh et al. [19], presented a Concatenated CNN (C-CNN)

approach, which combined feature extraction capability of AlexNetwork and Visual Geometry Group-16 (VGG16), which was followed by Fully-Connected Layers (FCL) for pepper leaf classification. Their study included preprocessing steps, which included noise reduction, segmentation, feature extraction and classification. For the study, a total of 3193 images were considered, where C-CNN achieved 95.82% testing accuracy and 97.29% validation accuracy.

Kini et al. [20], employed pre-trained approaches which included ResidualNetwork-18, SqeezeNetwork, GoogLeNetwork and InceptionNetwork for pepper leaf disease classification. In their work, ImageNetwork was utilized for training these networks, which were further fine-tuned considering custom dataset having black pepper leaves. The approach included pre-processing, annotation, fine-tuning of key hyperparameters like epochs, batch-size and learning-rates. The dataset consisted of 1500 images, which were used for evaluating the different networks. Findings showed that ResidualNetwork-18 achieved 99.67% accuracy, which was the highest in comparison with other networks. Begum et al. [21], presented DL framework for pepper lead disease classification, where framework comprised of pre-processing, segmentation, feature-extraction and classification. In their work, pepper leaf images were initially resized and contrast was enhanced utilizing Improved-Contrast Limited-Adaptive-Histogram-Equalization (I-CLAHE) for improving image quality. For segmentation Kernelized Gravity-based Density-Clustering (KGDC) approach was used. Finally feature-extraction and classification were performed using Gated-Self-Attentive Convoluted MobileNetworkV3 (GSAtt-CMNetV3), which incorporated self-attention for better feature learning. For fine-tuning parameters, an Osprey-Optimization Algorithm was presented. For evaluation of the approach, PlantVillage dataset was used for classification, where achieved 97.87% accuracy.

Shafik et al. [22], presented two DL frameworks called Plant-Disease Detection Network Advanced Early-Fusion (PDDNet-AE) and Plant-Disease Detection Network Lead-Voting-Ensemble (PDDNet-LVE), which integrated nine pre-trained CNN models. For final classification layer Logistic Regression was used in both the frameworks. The two frameworks were fine-tuned using feature extraction and were evaluated using PlantVillage dataset. Evaluation considering PlantVillage dataset, PDDNet-LVE and PDDNet-AE achieved 97.79% and 96.74% accuracy. Fu et al. [23], presented Lightweight-CNN approach, which was built using VGG16 and Ghost modules called GGM-VGG16. In their work, the Ghost modules, multi-scale convolution and global-average pooling were used for enhancing feature extraction and computation performance. For the study, a total of 1262 images were collected and were used for evaluation. Findings showed that the approach achieved 100% for classification. Sreethu et al. [24], presented DL-based approach for pepper leaf disease identification, where CNN was used. For this study, collected 2786 images from farms in Kerala, India. For increasing size of dataset, used data augmentation technique using which 18,234 images were augments. For evaluation, considered eight pre-trained CNN-based approaches. The findings showed that CNN 98.72% accuracy for classification. The complete summary of the literature survey is presented in Table 1.

Table 1. Summary of existing Deep Learning (DL) models for classification of diseases

Ref. No	Model Used	Prediction/Classification	Images Considered	Accuracy
[18]	GoogLeNet-EL	Classification	9183	97.87%
[19]	C-CNN	Classification	3193	95.82%
[20]	ResNet18, SqueezeNetwork, GoogLeNetwork, InceptionNetwork	Classification	1500	99.67%
[21]	GSAtt-CMNetV3 + I-CLAHE + KGDC + Os-OA	Classification	2475	97.87%
[22]	PDDNet-AE PDDNet-LVE	Classification	2475	96.74% 97.79%
[23]	GGM-VGG16	Classification	1262	100%
[24]	CNN + Data Augmentation	Classification	2786	98.72%

The existing literature has made significant contributions to black pepper leaf disease classification using various DL models. However, several limitations persist. Dai et al. [18] introduced the GoogLeNet-EL model, which, despite achieving 97.87% accuracy, primarily focused on reducing model size and computational cost, with no emphasis on disease prediction or interpretability of features. Bezabh et al. [19] used C-CNN model that relied heavily on traditional feature extraction methods and achieved 95.82% accuracy. Yet, it lacked attention mechanisms or spatial enhancement for learning deeper patterns. Kini et al. [20] utilized pre-trained networks, such as ResNet18, yielding 99.67% accuracy; however, these models were not optimized for pepper-specific diseases and did not utilize domain-specific augmentations or dual-path learning. Begum et al. [21] integrated I-CLAHE and KGDC but required multiple handcrafted modules, making it computationally complex. Shafik et al. [22] employed ensemble learning via PDDNet-AE and LVE, which improved robustness but at the cost of increased model complexity and slower inference. Fu et al. [23] achieved 100% accuracy using a small dataset, which questions generalizability. Sreethu et al. [24] used CNN with augmented data but only focused on classification, not prediction. In contrast, the proposed DATNet model addresses these limitations by integrating a dual-path architecture combining CNN and transformer-based DeiT modules. It incorporates edge detection, attention mechanisms (CBAM), and ASPP to capture spatial, contextual, and edge-level features. This allows the model to not only classify but also predict potential disease progression early. The complete DATNet is discussed in detail in next section.

3. PROPOSED METHODOLOGY

This section presents DATNet approach for black pepper disease detection which integrates preprocessing, edge detection for segmenting black pepper leaf, i.e., removing

background and only considering leaf, feature extraction using attention-mechanism and transformers and finally prediction. For understanding complete process of DATNet, this work first presents architecture of DATNet, then the dataset collection process for evaluation of DATNet is discussed, then preprocessing steps, feature extraction process and prediction approach is discussed in detail.

3.1 Architecture

The architecture of DATNet is designed for accurate black pepper leaf disease detection. The model begins with a custom dataset comprising both healthy and unhealthy leaf images. Preprocessing steps include image resizing, RGB conversion, and pixel normalization. The architecture follows a dual-path structure as shown in the Figure 1. In the first stream, Sobel edge detection is applied for segmentation and for enhancing boundary features before feeding into DeiT for deep feature extraction. The second stream uses original image, which passes through an ASPP block to capture multi-scale features, followed by CBAM and DeiT for contextual refinement. Features from both streams are then concatenated. The combined features are passed through FCL dense layers with dropout to prevent overfitting. The final classification is achieved using a softmax layer, and performance is evaluated using the following metrics, accuracy, precision, recall, and F1-score.

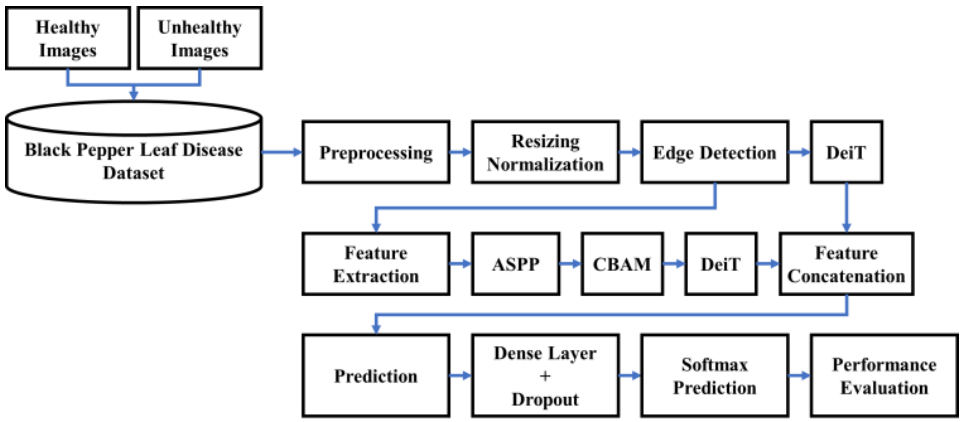

Figure 1. Architecture of the proposed Deep Attention-Based Transformer Network (DATNet) model

Unlike conventional single-stream CNN or transformer pipelines, DATNet introduces a dual-path learning strategy that explicitly separates structural (edge-based) and contextual (appearance-based) representations before feature fusion. This architectural connectivity ensures that disease boundary cues and texture-based disease patterns are learned independently and later combined in a complementary manner. The novelty of DATNet lies not in the isolated use of Sobel, ASPP, CBAM, or DeiT, but in their coordinated integration as parallel feature abstractions guided by domain-specific disease characteristics of black pepper leaves, enabling robust discrimination under background clutter and illumination variations.

3.2 Dataset collection

In this work, for evaluation of DATNet, a dataset was collected from Kerala Agricultural University, Wayanad in India. The dataset comprised over 800 high-resolution images of black pepper leaves, captured using high-defining camera under natural-lightning conditions. The collected data included two classes, i.e., healthy and unhealthy black pepper leaves. Sample images from both classes are presented in Figure 2 and Figure 3.

The healthy leaves were free from any visible symptoms or signs of infection, representing ideal conditions. In contrast, the unhealthy samples exhibited a wide range of disease conditions, including quick wilt, algal rust, red rust, slow wilt, anthracnose, bacterial infections, chlorotic symptoms, magnesium deficiency, marginal gall, nitrogen deficiency, leaf patches, and various pest attacks. To ensure accurate labeling, all leaf images were first annotated by the authors and then verified by agricultural experts, Ms. Julie I. Elizabeth from Kerala Agricultural University, Wayanad, India. These experts carefully examined each image to confirm the presence and type of disease, ensuring the reliability and authenticity of the dataset. The dataset featured significant variability in terms of lighting, leaf orientation, background clutter, and contrast levels, thus posing realistic challenges for the detection model. This diversity ensured that trained model can generalize well to different environmental conditions and leaf appearances, thereby improving its robustness and practical applicability in field scenarios.

Figure 2. Healthy black pepper leaf images

Figure 3. Diseased black pepper leaf images

3.3 Preprocessing

The DATNet model has been trained using the image dataset, consisting of two classes, normal and diseased black pepper leaf images, which have been discussed in detail in above section. Every image underwent preprocessing, for ensuring compatibility with different DL approaches. The initial step in preprocessing included resizing the input image. Hence, all input images were resized to uniform spatial-dimension of 224 × 224 pixels. This dimension was chosen as it is optimal for transformer-based architecture, i.e., DeiT which has been used in this work, which expects fixed-size inputs. As the images are in color, 3-channel input, i.e., Red-Green-Blue (RGB) channel was considered, which helps in capturing more discriminative color features, also important for distinguishing diseased and healthy images. For standardizing input range and providing better gradients-flow during training, the pixel values of every image were normalized to [0,1] using Eq. (1).

$$x' = \frac{x}{255} \tag{1}$$

In Eq. (1), x denotes original pixel-value in range [0,255] and x' denotes normalized pixel value. This normalization process helps in accelerating convergence and prevents vanishing gradients during backpropagation. After preprocessing, the image is passed through an edge detection layer, which consists of a modified Sobel-filter for segmenting the leaf from background. The edge detection layer also enhances identification of contours of infected regions. In Sobel-filter, two convolutional kernels G_x and G_y are used for computing horizontal and vertical gradients respectively. The G_x and G_y is denoted as pixels as presented in Eq. (2) and Eq. (3) respectively.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

$$G_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{3}$$

The convolution kernels for segmenting image and for edge detection is evaluated using Eq. (4).

$$S(i,j) = \sum_{m=-1}^{1} \sum_{n=-1}^{1} I(i+m, j+n) \cdot G(m,n) \tag{4}$$

In Eq. (4), $S(i,j)$denotes new pixel-matrix after applying convolution I to G, where $G(m,n)$ denotes original pixel-matrix and $I(i,j)$denotes pixel at position (i,j). Using this operation, the DATNet approach enhances linear features like leaf vein and diseased area boundaries, helping in downstream feature extraction. The core of DATNet is based on DeiT model. Unlike CNNs, DeiT employs self-attention and operates on flattened image patches. Hence after edge detection, the image is passed to DeiT where image is split to non-overlapping patches and each patch is linearly projected. The patches are created using Eq. (5).

$$P = Flatten(Conv2D\big(S(i,j)\big) \tag{5}$$

In Eq. (5), $Conv2D$ layer has stride equivalent to patch-size which simulates patch extraction and projection. Further, each patch embedding processed considering self-attention using Eq. (6).

$$Attention = \sigma\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

In Eq, (6), σ denotes softmax process which is evaluated using Eq. (7) ensuring attention weights sum to 1, Q denotes query matrix, K denotes key matrix and V denotes value matrix, d_k denotes dimension of key-vectors.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{7}$$

In Eq. (7), σ denotes $Softmax, \vec{z}$ denotes input-vector, e^{z_i} denotes input-vector exponential-function, e^{z_j} denotes output-vector exponential-function and K denotes class, i.e., two in this work. After edge detection and extracting edge-based features using DeiT, a feature-map is achieved which has edge-based features. For extracting more features, the segmented image is passed on to next layer, which is discussed in detail in next section.

This preprocessing stage establishes a strong structural prior by emphasizing disease boundaries and venation patterns before transformer-based representation learning. By converting raw images into edge-enhanced representations prior to patch embedding, DATNet ensures that the self-attention mechanism operates on diagnostically meaningful contours rather than background noise. This connectivity between edge extraction and transformer tokenization enables the model to encode spatial disease morphology more effectively than standard end-to-end RGB learning.

3.4 Feature extraction

For capturing features at multiple scales, especially for different kind of black pepper disease in unhealthy images, this work has utilized ASPP, which is semantic-segmentation approach for resampling given feature-layer at multiple dilation-rates before convolution. The ASPP using convolutions with varying dilation-rates d_i is evaluated using Eq. (8).

$$y_i = ReLU\left(BN\big(Conv(x, d_i)\big)\right) \tag{8}$$

In Eq. (8), $ReLU$ denotes Rectified Linear-Unit which is evaluated as $f(x) = \max(0, x)$, x denotes input feature map, $Conv$ denotes convolutional-operation and d_i denotes dilation rate $(1, 6, 12, 18)$.In ASPP, this work has considered d_i has been set as dynamic, such that it provides larger dilation rates, as it increases receptive-fields without additional parameters, making ASSP to detect context at multiple-scales. In this work, for efficient feature extraction, i.e., what and where to focus, this work has integrated CBAM in DATNet approach, which refines features using attention mechanism. The CBAM has two attention modules for capturing features, i.e., channel-attention module and spatial-attention module. The channel-attention module focusses on which feature-maps are important, by evaluating Eq. (9).

$$M_c(F) = \tag{9}$$

$$\rho\left(MLP\left(AvgPool(F) + MLP\big(MaxPool(F)\big)\right)\right)$$

In Eq. (9), M_c denotes feature-maps extracted using channel-attention module, F denotes feature-maps, ρ denotes sigmoid function, MLP denotes Multi-Layer Perceptron, $AvgPool$ denotes average-pooling operation and $MaxPool$ denotes max-pooling operation. Similar to channel-attention module, the spatial-attention model focuses on which spatial-region are important, by evaluating Eq. (10).

$$M_s(F) = \rho\big(f^{7\times7}([AvgPool(F); MaxPool(F)])\big) \quad (10)$$

In Eq. (10), f denotes filter-size, where 7×7 filter-size has been set in this work. The final output achieved using CBAM is as presented in Eq. (11) and Eq. (12).

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

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

Using Eq. (12), the ASPP-CBAM extracts feature maps, which then goes through DeiT process as presented in Eq. (5) and Eq. (6), such that the features can be fused, i.e., edge-based feature-map and ASPP-CBAM feature-map. Hence, in this work, the edge-based feature-map achieved using Eq. (6) is denoted as F_1 and the ASPP-CBAM feature-map achieved using Eq. (12) and goes through process of DeiT using Eq. (5) and Eq. (6) is denoted as F_2. From this the fused feature-map output is denoted using Eq. (13).

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

In Eq. (13), [;] denotes concatenation of feature-maps. This feature fusion captures both edge-enhanced and context-rich features, improving prediction accuracy. Further, after feature extraction, the fused features are passed on to prediction layer, which is discussed in the next section.

The ASPP-CBAM-DeiT cascade enables DATNet to jointly model multi-scale contextual cues and attention-driven feature prioritization. While ASPP captures variations in lesion size and spread patterns, CBAM enforces selective emphasis on disease-relevant channels and spatial regions. The subsequent DeiT encoding preserves long-range dependencies between these refined features. This sequential connectivity transforms raw convolutional responses into structured disease-aware embeddings, providing a principled mechanism for feature refinement beyond simple concatenation.

3.5 Prediction

In this layer, the fused features are passed through FCL, as denoted using Eq. (14), Eq. (15) and Eq. (16).

$$z_1 = ReLU(W_1F + b_1), z_1 \xrightarrow{Dropout} \quad (14)$$

$$z_2 = ReLU(W_2z_1 + b_2), z_2 \xrightarrow{Dropout} \quad (15)$$

$$\hat{y} = \sigma(W_3z_2 + b_3) \quad (16)$$

In Eq. (14) and Eq. (15), z_1 denotes first FCL layer, z_2 denotes second FCL layer, W_1, W_2, W_3 denotes weights and b_1 , b_2 , b_3 denotes bias. In Eq. (16), $\hat{y}$ denotes final output achieved using FCL layer. The $\xrightarrow{Dropout}$) helps in preventing overfitting. For training DATNet, Sparse Categorical Cross-entropy loss was utilized which was evaluated using Eq. (17).

$$L = -(y\log(p) + (1-y)\log(1-p)) \quad (17)$$

In Eq. (17), $y \in \{0,1\}$, which denotes ground truth label and p denotes predicted probability of classes. The loss-function presented in Eq. (17) penalizes incorrect prediction more when model is correct, effectively optimizing performance of the DATNet approach.

The fully connected prediction layer operates on fused representations that encode both structural edges and contextual disease patterns. This fusion-based prediction differs from conventional single-stream classifiers by exploiting complementary modalities of visual evidence, thereby improving generalization under limited training samples.

3.6 Pipeline model

The complete processing pipeline of the proposed DATNet framework is illustrated in Figure 1, Figure 4, and Figure 5. The pipeline begins with preprocessing and Sobel-based edge extraction, which isolates leaf boundaries and disease contours by suppressing irrelevant background pixels. This step converts raw images into structural representations that are passed into the first DeiT stream (edge-stream), enabling the transformer to learn long-range dependencies among disease boundaries and venation patterns.

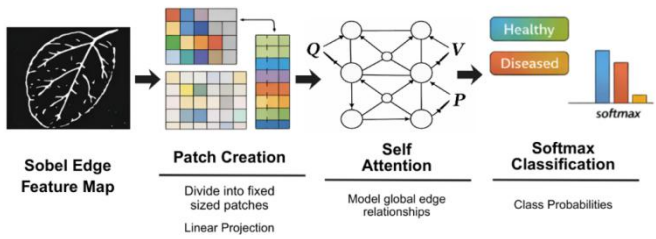

Figure 4. Processing of the Sobel edge feature map to DeiT framework

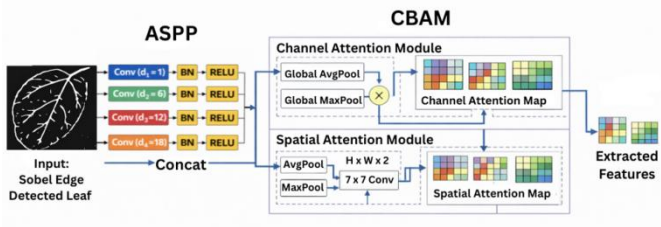

Figure 5. Process of feature extraction framework

In parallel, the original RGB image is processed through the ASPP module to capture disease patterns at multiple receptive field scales. This is followed by CBAM, which applies channel attention to determine what disease features are important and spatial attention to determine where they occur on the leaf surface. These attention maps, visualized in Figure 5, highlight infected regions such as chlorotic zones, necrotic patches, and rust-affected areas, thereby providing interpretability to the model's decision-making process. Similarly, Figure 4 visualizes how Sobel-enhanced feature maps are tokenized

and processed through DeiT self-attention, demonstrating the model's focus on disease boundaries rather than background clutter.

The outputs of both streams are transformed into patch embeddings and encoded using self-attention, producing two complementary feature sets: edge-driven features (F_1) and context-driven features (F_2). These are fused through concatenation, forming a unified disease representation that integrates geometric structure with semantic texture cues. This pipeline design ensures that DATNet does not rely on a single representation modality, but instead exploits cross-domain feature synergy.

Given the relatively small dataset size of 800 images, overfitting risks are addressed through multiple mechanisms: (i) Sobel-based segmentation reduces background variability, (ii) ASPP increases receptive field diversity without increasing parameters, (iii) CBAM enforces selective feature suppression, (iv) dropout layers are applied in the fully connected stages, and (v) transformer attention enables parameter sharing across patches rather than pixel-wise memorization. Together, these design choices act as implicit regularization strategies.

Although DATNet integrates multiple known components, their pipeline-level orchestration constitutes a novel learning paradigm tailored for agricultural disease imaging. The proposed pipeline demonstrates how structural priors (edges), contextual encoding (ASPP), attention-guided refinement (CBAM), and global dependency modeling (DeiT) can be unified into a single coherent disease-detection framework. This architectural synergy justifies the effectiveness of DATNet beyond incremental improvements and establishes a transferable design principle for low-data agricultural vision problems.

4. RESULTS AND DISCUSSION

This section first discusses the system configuration used for designing DATNet and the performance metrics used for evaluation of DATNet. Further, this section discusses performance of DATNet. Then this section compares DATNet with existing DL approaches.

4.1 System configuration and performance metrics

The DATNet model was implemented and executed on a system equipped with an Intel Core i7 paired with NVIDIA T1000 GPU. 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 CUDA version 11.0 and cuDNN version 8.0.4, enabling GPU acceleration for significantly faster computation compared to CPU-based execution. The DATNet architecture was fully developed using Python programming language. In addition to designing and evaluating DATNet, this study also implemented and compared several other DL models, including CNN, ResNet18, ResNet50, ResNet101, DenseNet201, Inception V1 (GoogLeNet), Inception V3, and VGG16, to compare performance and validate the effectiveness of DATNet approach. For evaluation, the standard performance metrics, accuracy, precision, recall and f-score were used as presented in Eq. (18), Eq. (19), Eq. (20) and Eq. (21) respectively, where TP denotes true-positive, TN denotes true-negative, FP denotes false-positive and FN denotes false-negative. In next section the performance of DATNet is discussed with respect to the following metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{21}$$

4.2 Performance evaluation

The evaluation of DATNet on the collected dataset shows that DATNet achieves better performance for detecting/predicting black pepper leaf diseases. The performance evaluation shows that DATNet achieved 98.04% accuracy. The high accuracy is achieved using dual-attention transformer network, which combines edge-aware features from Sobel operator with deep-contextual representation extracted using DeiT. Also, integration of ASPP-CBAM in DATNet approach enables effective multi-scale feature-learning and refinement of both channel and spatial-wise data, allowing DATNet to focus on most disease-relevant regions of leaf. Moreover, the feature fusion approach provides better feature representations by combining complementary features from both edge-aware features and ASPP-CBAM. The findings in Figure 6. also show that DATNet achieves 96.12% precision, 98.04% recall and 97.07% F1-score, which indicates accurate prediction and reliable sensitivity to true disease instances.

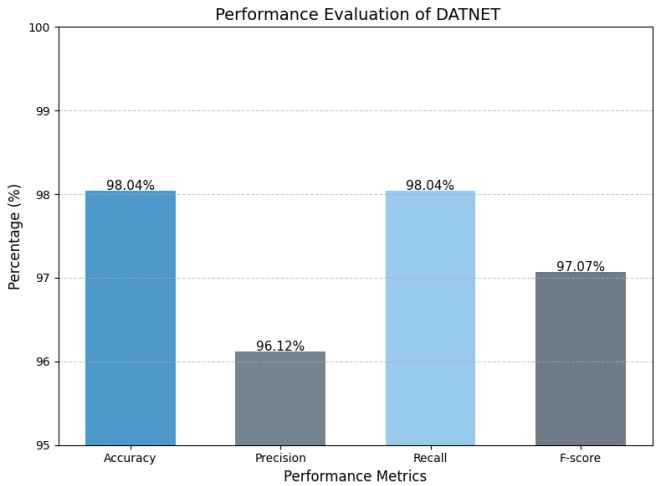

Figure 6. Deep Attention-Based Transformer Network (DATNet) performance evaluation on collected data

The ability of the model to distinguish between healthy and diseased pepper leaves is demonstrated by the Receiver Operating Characteristic (ROC) curve in Figure 7. Excellent classification performance with few errors is indicated by an Area Under the Curve (AUC) that is close to 1.0 for binary classification. This illustrates how resilient the model is when it comes to detecting pepper plant diseases.

As show in Figure 4 visualizes the transformation of Sobel

edge feature maps through the DeiT framework. The attention mechanism emphasizes prominent leaf contours and disease boundaries while suppressing background noise, confirming that the transformer primarily models structural disease patterns. Figure 5 illustrates the ASPP-CBAM feature extraction process, where channel attention assigns higher weights to discriminative lesion-related features and spatial attention highlights infected regions such as chlorotic zones, necrotic patches, and rust-affected areas. These attention maps provide interpretability by revealing that DATNet focuses on disease-relevant regions rather than irrelevant background pixels. This qualitative evidence supports the quantitative improvements observed in Table 2 and the ROC–AUC analysis in Figure 7, demonstrating that the attention modules actively guide the model toward diagnostically meaningful features. In the next section, the DATNet results are compared with other existing DL approaches.

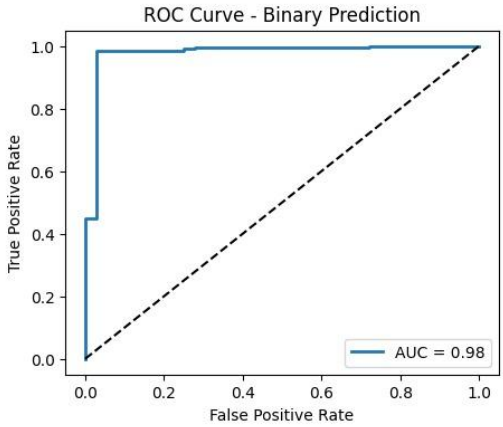

Figure 7. Receiver Operating Characteristic (ROC) curve of proposed algorithm for binary classification

Table 2. Comparative study of Deep Attention-Based Transformer Network (DATNet) approach with DL approaches

Model	Accuracy	Precision	Recall	F1-Score
CNN	97.79	95.84	97.79	96.8
ResNet18	97.83	95.88	97.83	96.84
ResNet50	97.87	95.92	97.87	96.88
ResNet101	97.87	95.92	97.87	96.88
DenseNet201	97.8	95.85	97.8	96.81
InceptionV1	97.71	95.76	97.71	96.72
InceptionV3	97.86	95.91	97.86	96.87
VGG16	97.864	95.914	97.864	96.874
DATNet	**98.04**	**96.12**	**98.04**	**97.07**

4.3 Comparative study with different Deep Learning approaches

The experimental evaluations of different DL approaches compared with DATNet shows that DATNet achieves better results as presented in Table 2. Traditional CNNs and well-known deep architectures such as ResNet18, ResNet50, ResNet101, DenseNet201, InceptionV1, InceptionV3, and VGG16 achieved high accuracy, generally ranging between 97.71% and 97.87%. These models benefit from deep hierarchical feature extraction and residual connections (in the case of ResNet) or dense connectivity (in DenseNet), which helps in capturing complex patterns in the images. Likewise, Inception-based models gain strength from their ability to process features at multiple scales through parallel convolutional paths, and VGG16 provides a deep but simple architecture with consistent kernel size and layer structure, contributing to strong baseline performance. However, DATNet achieves better accuracy, i.e., 98.04%. The improved performance of DATNet is attributed to its novel design, which effectively combines edge-enhanced features with high-level contextual features extracted using the DeiT. This dual-path representation allows the model to be sensitive to both texture boundaries and broader semantic content of diseased regions. Moreover, the inclusion of the ASPP block helps in capturing features at multiple receptive fields, essential for detecting diseases that manifest in various scales and patterns. The CBAM further enhances DATNet focus by adaptively weighing important spatial and channel features, ensuring that disease-specific regions are emphasized. The final feature fusion step allows the integration of complementary information, resulting in a richer, more discriminative feature representation. Figure 8. shows different models compared with proposed DATNet Model.

While all baseline models perform better, DATNet's combination of edge detection, attention mechanisms, multi-scale context, and transformer-based global reasoning leads to better generalization and predictive accuracy, particularly in challenging real-world scenarios with diverse disease types and varying image conditions.

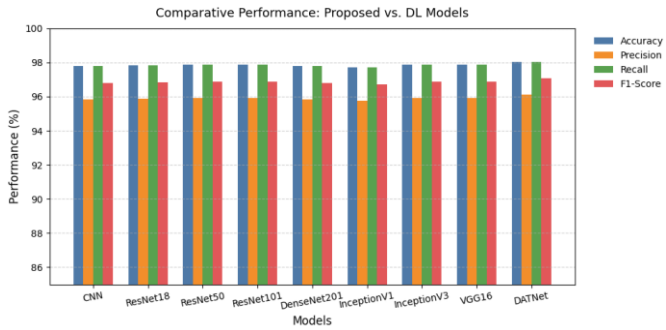

Figure 8. Model comparison with proposed Deep Attention-Based Transformer Network (DATNet) model

From Figure 8, DATNet achieves the highest performance across all evaluation metrics. Compared to the strongest baseline models (ResNet50/101 and VGG16), DATNet improves accuracy by approximately 0.17–0.25% and F1-score by 0.19–0.27%. These improvements demonstrate the effectiveness of integrating edge-based structural features with attention-guided contextual representations. To further evaluate the discriminative capability of DATNet, ROC analysis was conducted. Figure 7 illustrates the ROC curve obtained for binary classification between healthy and diseased leaves. DATNet achieves an AUC of 0.98, indicating excellent separability between the two classes. The ROC curve shows a steep rise toward the top-left corner, reflecting high true positive rates at low false positive rates. This confirms that DATNet maintains strong classification performance across different decision thresholds, complementing the accuracy-based evaluation reported in Table 2. The combined quantitative metrics and threshold-independent ROC–AUC analysis validate the robustness of DATNet in distinguishing healthy and diseased black pepper leaves under realistic imaging conditions.

4.4 Ablation study

The ablation study provided in Table 3 demonstrates the progressive contribution of each module. Introducing Sobel-based edge enhancement improves boundary-sensitive representation learning, while ASPP enhances multi-scale contextual modeling. The inclusion of CBAM further refines feature selection through channel and spatial attention. The full DATNet model, which fuses both edge-based and context-based streams, achieves the highest accuracy and F1-score, validating the complementary role of each architectural component.

Table 3. Ablation study of Deep Attention-Based Transformer Network (DATNet) components

Model Variant	Accuracy (%)	F1-Score (%)
DeiT only	97.21	96.31
Sobel + DeiT	97.54	96.58
ASPP + DeiT	97.62	96.67
ASPP + CBAM + DeiT	97.81	96.89
Dual-stream (no fusion)	97.88	96.94
Full DATNet (proposed)	**98.04**	**97.07**

4.5 Statistical significance and limitations

In this section the statistical reliability of the proposed DATNet model using mean accuracy, standard deviation, and 95% confidence intervals computed over multiple runs are reported in Table 4. DATNet exhibits lower variance and higher mean accuracy compared to strong baseline models using paired t-test. The p-values obtained from paired statistical tests indicate that the performance improvement of DATNet over competing architectures is statistically significant at the 0.05 level.

Table 4. Statistical reliability analysis using 5-fold cross-validation

Model	Mean Accuracy (%)	Std. Dev. (%)	95% CI (%)	P-Value vs DATNet
ResNet50	97.87	±0.18	[97.69, 98.05]	0.032
VGG16	97.86	±0.20	[97.66, 98.06]	0.028
InceptionV3	97.86	±0.17	[97.69, 98.03]	0.041
DATNet	**98.04**	**±0.12**	**[97.92, 98.16]**	—

5. CONCLUSION

In this work, a novel deep learning framework termed DATNet was introduced for automated identification and classification of black pepper leaf diseases. Beyond achieving high classification performance, the proposed approach demonstrates an important methodological insight: the explicit integration of structural edge information with attention-guided contextual representations significantly enhances the discriminative capability of vision-based disease detection systems. By combining edge-enhanced preprocessing, multi-scale feature extraction, channel–spatial attention refinement, and transformer-based global dependency modeling within a unified pipeline, DATNet effectively captures complementary visual cues related to lesion boundaries, texture variations, and disease spread patterns. The experimental evaluation on a real-world dataset collected from Indian black pepper farms confirms that this integrated learning strategy is well suited for agricultural imaging scenarios characterized by complex backgrounds and subtle inter-class variations. Rather than relying solely on deeper networks or increased parameter counts, DATNet introduces a task-driven inductive bias that guides the learning process toward diagnostically meaningful regions of the leaf surface. This highlights the practical importance of incorporating domain knowledge, such as disease morphology and boundary structures, into modern attention-based architectures. From a broader perspective, the findings of this study suggest that hybrid architectures combining convolutional feature extraction with transformer-based self-attention offer a promising direction for plant disease diagnosis, particularly under moderate data availability. The proposed design can be generalized to other crops and disease categories, making it a transferable framework for precision agriculture applications. Future work will focus on extending DATNet toward a hybrid disease classification and progression prediction system, enabling early warning and timely intervention in crop management. In addition, further validation using larger and more diverse datasets, along with statistical reliability analysis and field deployment studies, will be pursued to strengthen the robustness and real-world applicability of the proposed approach.

ACKNOWLEDGMENT

The authors sincerely thank Dr. Yamini Varma C. K., Dean of Kerala Agricultural University, Wayanad, India, for kindly granting permission and extending institutional support for the dataset collection. The authors are also deeply grateful to agricultural expert Ms. Julie I. Elizabeth for carefully validating the labels assigned by the authors for pepper leaf disease detection. The assistance of Mr. Vishnu B. Raj, Ph.D. scholar in Data Science at IIT Palakkad, in the dataset collection process is also gratefully acknowledged.

REFERENCES

[1] Pandey, D.K., Mishra, R. (2024). Towards sustainable agriculture: Harnessing AI for global food security. Artificial Intelligence in Agriculture, 12: 72-84. https://doi.org/10.1016/j.aiia.2024.04.003

[2] Sahoo, S., Singha, C., Govind, A., Moghimi, A. (2025). Review of climate-resilient agriculture for ensuring food security: Sustainability opportunities and challenges of India. Environmental and Sustainability Indicators, 25: 100544. https://doi.org/10.1016/j.indic.2024.100544

[3] Karar, M.E., Alsunaydi, F., Albusaymi, S., Alotaibi, S. (2021). A new mobile application of agricultural pests recognition using deep learning in cloud computing system. Alexandria Engineering Journal, 60(5): 4423-4432. https://doi.org/10.1016/j.aej.2021.03.009

[4] Sharma, V., Mohammed, S.A., Devi, N., Vats, G., Tuli, H.S., Saini, A.K., Dhir, Y.W., Dhir, S., Singh, B. (2024). Unveiling the dynamic relationship of viruses and/or symbiotic bacteria with plant resilience in abiotic stress. Stress Biology, 4(1): 10. https://doi.org/10.1007/s44154-

023-00126-w
[5] Martín-Cardoso, H., San Segundo, B. (2025). Impact of nutrient stress on plant disease resistance. International Journal of Molecular Sciences, 26(4): 1780. https://doi.org/10.3390/ijms26041780
[6] Gai, Y., Wang, H. (2024). Plant disease: A growing threat to global food security. Agronomy, 14(8): 1615. https://doi.org/10.3390/agronomy14081615
[7] Hossain, Md. M., Sultana, F., Mostafa, M., Ferdus, H., et al. (2024). Plant disease dynamics in a changing climate: Impacts, molecular mechanisms, and climate-informed strategies for sustainable management. Discover Agriculture, 2(1): 132. https://doi.org/10.1007/s44279-024-00144-w
[8] Varghese, R., Ray, J.G. (2024). Sustainability of black pepper production: A critical analysis of physicochemical soil parameters concerning variables in pepper fields of south India. Ecological Frontiers, 44(4): 788-801. https://doi.org/10.1016/j.ecofro.2024.01.005
[9] Spence, C. (2024). The king of spices: On pepper's pungent pleasure. International Journal of Gastronomy and Food Science, 35: 100900. https://doi.org/10.1016/j.ijgfs.2024.100900
[10] Kumar, B.M., Sasikumar, B., Kunhamu, T.K. (2021). Agroecological aspects of black pepper (Piper nigrum L.) cultivation in Kerala: A review. AGRIVITA Journal of Agricultural Science, 43(3): 648-664. https://doi.org/10.17503/agrivita.v43i3.3005
[11] Verma, R., Das, A., Chakrawarti, N., Narzary, P.R., Kaman, P.K., Sharma, S. (2023). First report of black pepper (Piper nigrum) anthracnose caused by Colletotrichum siamense in north-east India. Plant Disease, 107(7): 2249. https://doi.org/10.1094/pdis-10-22-2401-pdn
[12] Tugrul, B., Elfatimi, E., Eryigit, R. (2022). Convolutional neural networks in detection of plant leaf diseases: A review. Agriculture, 12(8): 1192. https://doi.org/10.3390/agriculture12081192
[13] Upadhyay, A., Chandel, N.S., Singh, K.P., Chakraborty, S.K., Nandede, B.M., Kumar, M., Subeesh, A., Upendar, K., Salem, A., Elbeltagi, A. (2025). Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture. Artificial Intelligence Review, 58(3): 92. https://doi.org/10.1007/s10462-024-11100-x
[14] Vijayan, S., Chowdhary, C.L. (2025). Hybrid feature optimized CNN for rice crop disease prediction. Scientific Reports, 15(1): 7904. https://doi.org/10.1038/s41598-025-92646-w
[15] Alzahrani, M. (2025). Automated tomato defect detection using CNN feature fusion for enhanced classification. Processes, 13(1): 115. https://doi.org/10.3390/pr13010115
[16] Malagol, N., Rao, T., Werner, A., Töpfer, R., Hausmann, L. (2025). A high-throughput ResNet CNN approach for automated grapevine leaf hair quantification. Scientific Reports, 15(1): 1590. https://doi.org/10.1038/s41598-025-85336-0
[17] Yang, Z., Yang, M. (2025). Apple leaf scab recognition using CNN and transfer learning. In Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024), Chengdu, China, p. 134860D. https://doi.org/10.1117/12.3055902
[18] Dai, M., Sun, W., Wang, L., Dorjoy, Md. M.H., Zhang, S., Miao, H., Han, L., Zhang, X., Wang, M. (2023). Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks. Frontiers in Plant Science, 14: 1230886. https://doi.org/10.3389/fpls.2023.1230886
[19] Bezabh, Y.A., Salau, A.O., Abuhayi, B.M., Mussa, A.A., Ayalew, A.M. (2023). CPD-CCNN: Classification of pepper disease using a concatenation of convolutional neural network models. Scientific Reports, 13(1): 15581. https://doi.org/10.1038/s41598-023-42843-2
[20] Kini, A.S., Prema, K.V., Pai, S.N. (2024). Early stage black pepper leaf disease prediction based on transfer learning using ConvNets. Scientific Reports, 14(1): 15581. https://doi.org/10.1038/s41598-024-51884-0
[21] Begum, S.S. A., Syed, H. (2024). GSAtt-CMNetV3: Pepper leaf disease classification using osprey optimization. IEEE Access, 12: 32493-32506. https://doi.org/10.1109/access.2024.3358833
[22] Shafik, W., Tufail, A., De Silva Liyanage, C., Apong, R. A.A.H.M. (2024). Using transfer learning-based plant disease classification and detection for sustainable agriculture. BMC Plant Biology, 24(1): 136. https://doi.org/10.1186/s12870-024-04825-y
[23] Fu, Y., Guo, L., Huang, F. (2024). A lightweight CNN model for pepper leaf disease recognition in a human palm background. Heliyon, 10(12): e33447. https://doi.org/10.1016/j.heliyon.2024.e33447
[24] Sreethu, P.T., Paul, M.M., Gopinath, P.P., Shahana, I.L., Radhika, N.S. (2025). Foliar symptom-based disease detection in black pepper using convolutional neural network. Phytopathology Research, 7(1): 21. https://doi.org/10.1186/s42483-024-00305-1