

# Interpretable Plant Leaf Disease Detection Using Attention-Enhanced CNN

Balram Singh<sup>1</sup>, Ram Prakash Sharma<sup>1\*</sup>, Somnath Dey<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, National Institute of Technology Hamirpur, Anu, Hamirpur, 177005, Himachal Pradesh, India.

<sup>2</sup>Department of Computer Science and Engineering, Indian Institute of Technology Indore, Simrol, Indore, 453552, Madhya Pradesh, India.

\*Corresponding author(s). E-mail(s): [ram.sharma@nith.ac.in](mailto:ram.sharma@nith.ac.in); Contributing authors: [23mcs104@nith.ac.in](mailto:23mcs104@nith.ac.in); [somnath@iiti.ac.in](mailto:somnath@iiti.ac.in);

## Abstract

Plant diseases pose a significant threat to global food security, necessitating accurate and interpretable disease detection methods. This study introduces an interpretable attention-guided Convolutional Neural Network (CNN), CBAM-VGG16, for plant leaf disease detection. By integrating Convolution Block Attention Module (CBAM) at each convolutional stage, the model enhances feature extraction and disease localization. Trained on five diverse plant disease datasets, our approach outperforms recent techniques, achieving high accuracy (up to 98.87%) and demonstrating robust generalization. Here, we show the effectiveness of our method through comprehensive evaluation and interpretability analysis using CBAM attention maps, Grad-CAM, Grad-CAM++, and Layer-wise Relevance Propagation (LRP). This study advances the application of explainable AI in agricultural diagnostics, offering a transparent and reliable system for smart farming. The code of our proposed work is available at <https://github.com/BS0111/PlantAttentionCBAM>

**Keywords:** Agriculture, Plant leaf disease, Artificial intelligence, Deep learning, Attention, Interpretability

## 1 Introduction

The agriculture industry is essential in maintaining food security worldwide, but different crops suffer from a variety of diseases due to varying weather conditions, posing a major threat to crop yield and quality. These diseases are often caused by factors such as extreme temperatures, microbial infections, and changes in humidity or soil conditions. Farmers have always relied on human examination to identify diseases, which is labour-intensive, prone to mistakes, and ineffective on large farms. As a result, among the most crucial areas of research in smart agriculture is the automation of plant disease identification and classification utilizing various Artificial Intelligence (AI) approaches [1]. Machine learning methods in agriculture has drawn a lot of interest with applications ranging from yield prediction and crop recommendation to disease detection. In recent years, Convolutional Neural Networks (CNNs) and transformers have shown impressive performance in a variety of computer vision applications such as segmentation, object recognition, and image denoising [2]. CNN-based models in agriculture have performed well on challenges such as classifying plant diseases. However, despite their effectiveness, CNNs often operate as black boxes, offering limited transparency into how predictions are made, hampering trust and broader adoption.

In order to tackle this, Explainable AI (XAI) methods have been developed enabling visual interpretation of model decisions through attention maps [3], Grad-CAM [4], and LRP [5]. While GradCAM and GradCAM++ [6] generate class-discriminative localization maps using gradients from the final convolutional layers, they may lack resolution and be susceptible to noisy activations. LRP, in contrast, provides pixel-level attributions by backpropagating the model output through a set of layer-specific relevance propagation rules. This offers a more granular explanation of predictions, which is crucial in medical and agricultural diagnostics.

In this work, an explainable deep learning approach is proposed for the detection of plant leaf disease. Our architecture is based on the VGG16 [7] backbone enhanced with the CBAM [8, 9] which introduces attention layers for inherent interpretability of the model with an emphasis on the most relevant features at both the channel and spatial levels. Following each of the five convolutional layers, CBAM modules are added to improve classification accuracy and localization of relevant features by capturing both spatial and channel-wise attention. Five distinct datasets are used to train the model, namely Apple, Plant Village, Embrapa, Maize and Rice to ensure the generalizability and applicability of our proposed method across diverse set of crops. Apart from the inherent interpretability of the proposed method's decision-making process through CBAM layers, we also demonstrate interpretability using advanced explainability methods like LRP, Grad-CAM, and Grad-CAM++. We have also employed high-dimensional feature visualization techniques such as t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) to visualize feature in lower dimension for visualization of extracted features. Overall, this study advances the application of XAI for agricultural use by presenting an interpretable and performance-robust framework for plant disease classification. The following is a summary of the main contributions of the proposed work:

- A novel inherently interpretable technique, CBAM-VGG16, is proposed to identify plant leaf disease in various types of crops.
- The proposed method enables explainability through feature-level attention across all five convolutional stages, which can be visualized using CBAM attention maps.
- Post-hoc explainability of the proposed method is demonstrated using multiple techniques, including Grad-CAM, Grad-CAM++, and a suite of LRP variants.
- Feature space analysis using UMAP and t-SNE is conducted to visualize class-wise clustering and inter-model differences.
- The proposed method demonstrates better performance compared to other recent cutting-edge techniques on all five datasets considered in this work ensuring generalization and adaptive capability of proposed method for any crop leaf disease identification.

In rest of the paper, Section 2 reviews existing literature, highlighting key limitations and motivations. Section 3 details the architecture of the proposed CBAM-VGG16 model, including the integration of attention mechanisms, and interpretability approaches. Section 4 describes the complete results including the experimental setup, dataset description, comparative results, and the explainability analysis. Finally, Section 5 brings the study to a conclusion and suggests possible directions for further research in this domain.

## 2 Related Works

Deep learning combined with image processing has been frequently utilized for the automatic detection of diseases in plant leaves. Numerous Research has explored various models, datasets, and image processing pipelines for this task. This section offers a thorough analysis of the methods employed for multiclass leaf disease classification, emphasizing traditional techniques, deep learning methods, and recent advances in vision transformers [10, 11].

[12] put together a large dataset known as plant village that included 38 distinct pictures of crop leaf diseases. They assessed the dataset using several popular deep learning frameworks like AlexNet and GoogleNet, employing a transfer learning approach. [13] applied conventional Convolutional Neural Network (CNN) architectures to analyze how variations in samples and dataset size influence The efficiency of identifying plant diseases via transfer learning. For this purpose, a dataset comprising 1,383 background-free images across 56 distinct disease types was compiled. [14] developed a Multilayer Convolutional Neural Network (MCNN) to classify mango leaves infected by Anthracnose disease. They utilised a dataset of 1,070 images, comprising both healthy and infected leaves, captured at India's Shri Mata Vaishno Devi University which is located in Katra, Jammu and Kashmir. The suggested MCNN model demonstrated higher classification accuracy of 97.13% compared to other state-of-the-art approaches.

[15] have put out the Grape Leaf Disease Detection Network (GLDDN) approach. Their proposed network uses Residual Convolution Neural Network (R-CNN) and

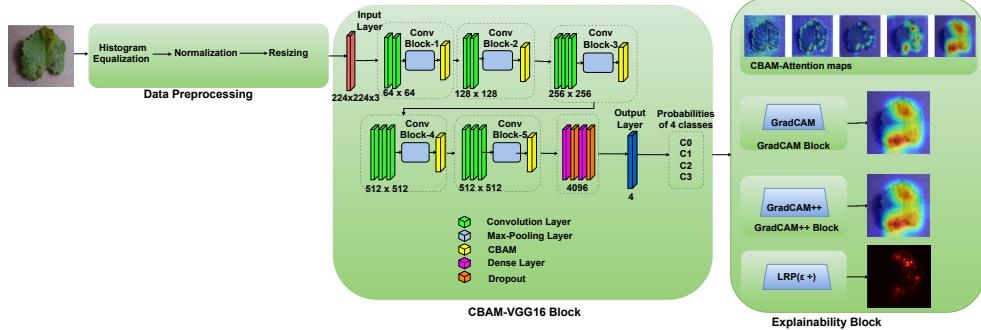
multitask learning to identify esca, isariopsis, and black-rot in grape photos that are thoroughly explained in the proposed grape leaf disease detection network which is relies on multitasking, spatial and channel-wise attention processes, and Faster R-CNN for multi-level feature extraction. [16] utilized a residual CNN enhanced with an attention mechanism to identify diseases in tomato leaves. An evaluation conducted on a dataset with 95,999 categorized tomato leaf images revealed a 98% accuracy rate for their approach. [17] created a residual CNN with self-attention architecture for disease classification, achieving 98% accuracy on the MK-D2 dataset and an accuracy of 95.33% on the AES-CD9214 dataset. Researchers in [18][19] applied different attention strategies within well-known architectures such as MobileNet, Inception, and residual CNNs to identify crop diseases. While these approaches delivered strong results, their evaluation focused on only a limited selection of plant diseases, unlike broader models addressing multiple crops. [20] has proposed LeafGAN architecture for the augmentation of diseased leaf images via transformation for improving the plant disease diagnosis system using large scale dataset. Their architecture has own attention mechanism to transform only relevant areas from plant leaf images with different backgrounds, which ensures the diversity of training data. [21] proposed CROPCARE, which is a technology for detecting and preventing crop diseases in real time. This system integrates Computer Vision (CV), Google Cloud services, and the Internet of Things (IoT) for the detection of crop diseases in the plant village dataset. It uses the pre-trained model MobileNet-V2 and Super resolution Convolution Neural Network (SRCNN) for generating decisions of the model trained on various diseases. To solve the data imbalance problem of health vs unhealthy images [22] have used a DoubleGAN architecture. Their architecture uses Wasserstein Generative Adversarial Networks (WGAN) and Superresolution Generative Adversarial Network (SRGAN) to balance the dataset. [23] has proposed a technique to categorize the severity of damaged plants, called Agriculture Detection (AgriDet). Their method combines traditional Inception-Visual Geometry Group Networks (INC-VGGN) with Kohonen-based deep learning networks. A solution for the smaller dataset is provided by [24] where they have proposed a network called PiTLiD. Their proposed network that uses a small sample size and pretrained Inception-V3 convolutional neural network and transfer learning to identify plant leaf diseases from plant leaf phenotypic data. Gehlot et al. [25] has presented a new dataset for multiclass tomato specific disease which was evaluated using difference CNN architecture models. Zeng et al. [26] has presented a Squeeze-and-Excitation Vision Transformer (SEViT) which used ResNet with channel attention module for the classification of disease in different crop dataset crawled from different internet sources. Their results establishes the supremacy for many similar types of diseases.

The above-discussed methods rely on deep learning algorithms that operate as black boxes, making it challenging to comprehend how they make their decisions. While they achieve high accuracy, they lack interpretability, limiting trust and dependability in practical uses. These methods do not provide any explainable solution for deep learning models' opaque nature that makes it difficult to understand their judgements. Therefore, explainability is necessary for trust, transparency, and validation, ensuring that model predictions are reliable and interpretable, especially in critical

applications like plant disease detection. The introduction of XAI techniques, such as GradCAM and LRP, has addressed this issue by visualizing model decisions. [27] developed a Computer-Aided Diagnostic System (CADS) incorporating an XAI framework to build trust in the system's decisions. EfficientNet B5 outperformed VGG19, VGG16, Inception v3, ResNet, MobileNet, and other EfficientNet variants among the models used for transfer learning. The predictions made by EfficientNet-B5 were explained using XAI methods like LIME and Grad-CAM, with validation performed through YOLOv4. [28] introduced a separate XAI-driven approach for using deep learning for disease detection. They found the illness and indicated the crucial leaf areas for the GradCAM++ explainer model's classification. They began by utilising images from the Plant Village dataset of both healthy and damaged plant leaves. A pre-trained ImageNet system with a VGG16 deep learning model as a base was used by the authors. They have used GradCAM++'s visualisation technique for model interpretation, which effectively identified the afflicted areas and used heatmap overlays to highlight them. Although methods like GradCAM [29], [30] and LRP are commonly utilized to interpret CNN outputs, their application in detecting plant leaf diseases is still relatively new. GradCAM has shown success in areas like medical imaging, image categorisation, and object detection and has recently been explored for plant disease identification. Another widely used technique for analyzing deep neural networks is LRP, which assigns relevance scores backward from the output to the input features, providing detailed explanations of CNN decisions [31]. In the context of plant disease identification, Layer-wise Relevance Propagation (LRP) has been utilized to highlight significant regions of a leaf, aiding in the detection of disease symptoms and affected zones [32]. Nevertheless, the resulting visual outputs from LRP can be intricate and challenging to understand, which may hinder their usability in real-world applications. To improve interpretability, a number of studies have looked into combining GradCAM using additional methods of explanation, including Lime [33], DeepLIFT [34], and Shapley Value Sampling [35], with the goal of addressing GradCAM's shortcomings, including its tendency to produce noisy or vague visual outputs. Despite progress, issues like unclear GradCAM explanations and the complexity of LRP visualizations remain unresolved. Integrating various explainable AI techniques offers a promising direction for improving model transparency and fostering trust and dependability in sensitive domains such as plant disease diagnosis. [36] has presented an comprehensive review of rice disease detection techniques using AI based solutions. A detailed survey of recent advances in plant disease detection is provided by [37]. They have highlighted the issue and challenges in plant leaf disease detection using machine learning and deep learning based solutions.

### 3 Proposed Methodology

The overview of interpretable architecture used for identifying plant leaf disease is provided in Fig. 1. The input plant images undergoes through data preprocessing to improve quality and guarantee alignment with the input structure of the proposed CBAM-VGG16 model. This involves histogram equalization to improve contrast, normalization of pixel values to the  $[0, 1]$  range and resizing images to  $224 \times 224$  pixels.



**Fig. 1:** System overview of plant disease classification using CBAM-VGG16.

Following preprocessing, the enhanced VGG16 [7] model integrated with CBAM is employed for disease classification. In this architecture, CBAM modules are added following every convolutional stage of the VGG16 model. The CBAM mechanism adaptively refines the feature maps by applying both channel and spatial attention, thereby enhancing the model's ability to focus on disease-affected regions. Once the input image's class label is predicted by the model, multiple explainability modules are used to interpret its decision-making process. CBAM attention maps, generated by the integrated CBAM modules in the VGG16 architecture, provide insights into the spatial and channel-wise focus of the network across layers. For class-discriminative localization, Grad-CAM [29] and Grad-CAM++ [6] are used to generate class-discriminative heatmaps that draw attention to the input image's most significant areas. Additionally LRP is employed using multiple propagation rules i.e. including  $\epsilon$ -rule,  $\epsilon\gamma$ -box,  $\alpha_2\beta_1$ , and excitation backpropagation to assign pixel-level relevance scores, further enriching the interpretability of the model's predictions through diverse attribution perspectives. This entire pipeline from data preprocessing to classification and visual explanation forms an end-to-end plant leaf disease detection system designed for improved interpretability and trust in AI-based agricultural diagnostics. The internal modules of the architecture are explained in further subsections.

### 3.1 Data Preprocessing

To improve the quality, consistency, and compatibility of the input images to the CBAM-enhanced VGG16 model, we apply a series of preprocessing steps, i.e., histogram equalization, normalisation, and resizing, visually presented in Fig 2.

#### 3.1.1 Histogram Equalization

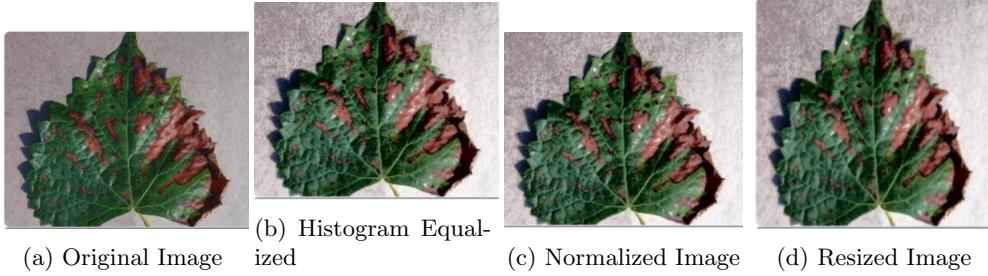
Histogram equalization using the Contrast Limited Adaptive Histogram Equalization (CLAHE) [38] is used to improve the grape leaf picture's contrast by redistributing pixel intensity values, as shown in Fig. 2. Unlike traditional histogram equalization, CLAHE operates on small image regions (tiles) and applies contrast enhancement while limiting noise amplification in homogeneous areas, making it more effective for preserving fine details in medical and agricultural images.

### 3.1.2 Normalization

Normalization scales pixel intensity values in the range  $[0, 1]$ , ensuring uniformity across the dataset as illustrated in Fig. 2. This prevents biased learning caused by varying pixel ranges and improves model convergence and training stability.

### 3.1.3 Resizing

All images are resized to  $224 \times 224$  pixels, which matches the expected input size of the CBAM-VGG16 architecture. This resizing step maintains structural integrity while ensuring consistent dimensions for batch processing.



**Fig. 2:** Pre-processing of the input images

## 3.2 CBAM VGG16 Architecture

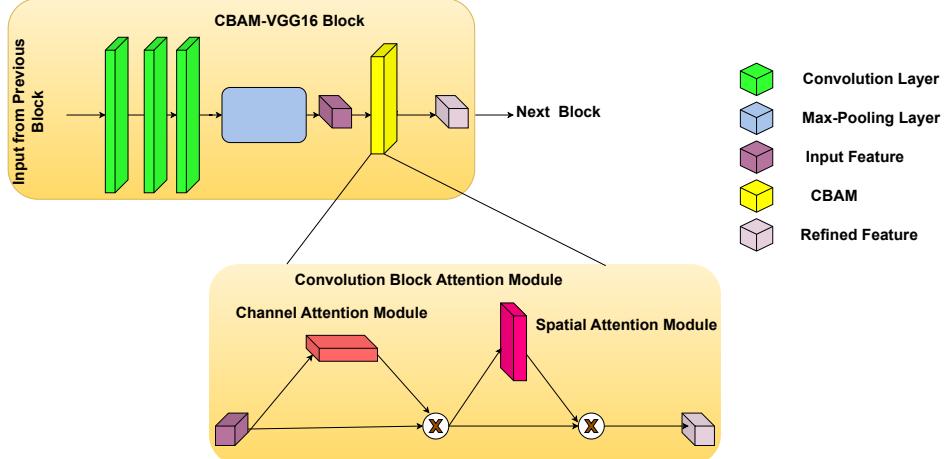
As illustrated in Fig. 1, the proposed CBAM-VGG16 architecture begins by utilising input layer of size  $224 \times 224 \times 3$ . In the architecture, CBAM layers are added after each MaxPooling2D layer. Even though CBAM is a lightweight architecture, overuse could lead to overfitting, which is why the CBAM layer is added after each MaxPooling layer. The channel module and the spatial module are the two primary parts of the CBAM developed by [8]. Given the input  $I_f \in \mathbb{R}^{c \times W \times H}$ , a 1D channel attention map is generated by CBAM  $C_m \in \mathbb{R}^{c \times 1 \times 1}$  as well as a 2D spatial attention map  $S_m \in \mathbb{R}^{c \times W \times H}$ , where  $c$  in our scenario stands for three channel input,  $W$  and  $H$  stand for the input feature map's width and height, respectively. Our suggested CBAM VGG16's attention process is given in Eq. 1 and 2.

$$I_c f' = C_m(I_f) \otimes I_f \quad (1)$$

$$I_s f'' = S_m(I_c f') \otimes I_c f' \quad (2)$$

where  $\otimes$  denotes element-wise multiplication,  $I_c f'$  is the channel multiplied output, and  $I_s f''$  denote the final refined output.

The placement of the CBAM layer after each MaxPooling layer is illustrated in Fig. 3. The internal workings of CBAM's channel attention and spatial attention mechanisms are detailed in the following subsections.



**Fig. 3:** Internal structure of CBAM layers

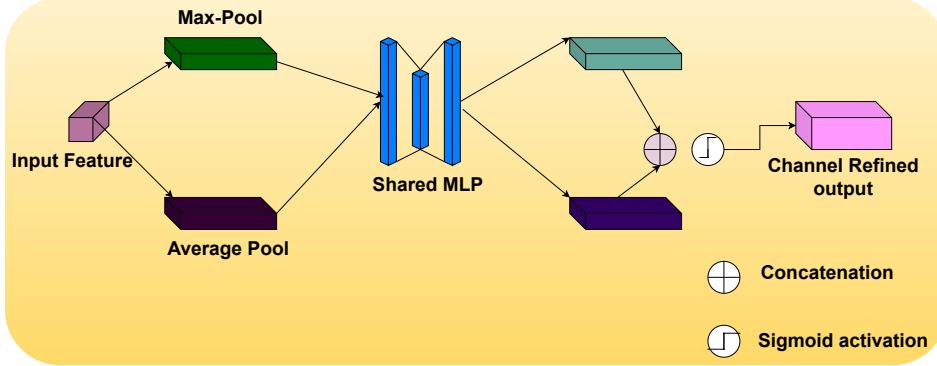
### 3.2.1 Channel Attention Module

The composition of the channel attention module is given in Fig. 4. A map of channel attention  $C_m \in \mathbb{R}^{c \times 1 \times 1}$  is applied to the input features by the channel attention module. To generate the average pooled feature,  $I_f^{\text{avg}}$ , and the max pooled feature,  $I_c^{\text{max}}$ , the average and max pooling procedures are applied independently. Thereafter, a shared Multilayer Perceptron (MLP) network with one hidden layer receives  $I_c^{\text{avg}}$  and  $I_c^{\text{max}}$ . In our CBAM VGG16 design, the activation size of the hidden layer is set to  $\mathbb{R}^{c/r \times 1 \times 1}$ , where the reduction ratio denoted by  $r$ , which is set to eight. This value provides offers an equitable trade-off between model complexity and representational capacity, as empirically validated in [8], ensuring sufficient channel inter dependencies are captured without incurring significant computational overhead. The channel attention method used in our proposed architecture is provided as given in Eq. 3 and 4.

$$C_m(I_f) = \sigma(\text{MLP}(\text{AvgPool}(I_f)) + \text{MLP}(\text{MaxPool}(I_f))) \quad (3)$$

$$C_m(I_f) = \sigma(W_1(W_0(I_c^{\text{avg}})) + W_1(W_0(I_c^{\text{max}}))) \quad (4)$$

In the above equations,  $\sigma$  denotes the sigmoid activation function. The weights  $W_0 \in \mathbb{R}^{c \times c/r}$  and  $W_1 \in \mathbb{R}^{c/r \times c}$  are shared parameters of the multi-layer perceptron (MLP). To generate the channel attention map  $C_m(I_f)$ , the input feature map  $I_f$  undergoes both average pooling and max pooling operations across its spatial dimensions, resulting in  $I_c^{\text{avg}}$  and  $I_c^{\text{max}}$ , respectively. These are then passed through the shared MLP, and their outputs are summed and activated using the sigmoid function. The resulting channel attention map is element-wise multiplied with the original input  $I_f$  to obtain the refined feature map  $I_c f'$ , as defined in Eq. 1. The structure of the channel attention block in CBAM is depicted in Fig. 4.



**Fig. 4:** Channel attention module used in CBAM layers.

### 3.2.2 Spatial Attention Module

Fig. 5 illustrates the spatial attention module integrated into the proposed CBAM-VGG16 architecture. This module generates a 2D spatial attention map  $S_m \in \mathbb{R}^{c \times W \times H}$ , which is applied to the channel-refined feature map  $I_c f'$ . To compute spatial attention, the module first applies average pooling and max pooling operations along the channel axis, resulting in two 2D feature maps:  $I_s^{\text{avg}} \in \mathbb{R}^{1 \times W \times H}$  and  $I_s^{\text{max}} \in \mathbb{R}^{1 \times W \times H}$ . These two maps are then concatenated along the channel dimension and passed through a convolutional layer to produce the spatial attention map. The resulting map is used to refine the features by performing an element-wise multiplication with  $I_c f'$ , yielding the final spatially-refined output  $I_s f''$ . The spatial attention computation in CBAM is formally described in Eq. 5 and 6.

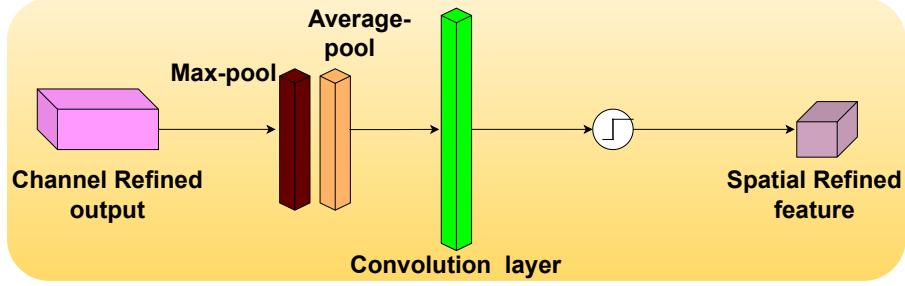
$$S_m(I_c f') = \sigma(\nu_{7 \times 7}([\text{AvgPool}(I_c f'); \text{MaxPool}(I_c f')])) \quad (5)$$

$$S_m(I_c f') = \sigma(\nu_{7 \times 7}([I_s^{\text{avg}}; I_s^{\text{max}}])) \quad (6)$$

In the above equations,  $\sigma$  represents the sigmoid activation function, and  $\nu_{7 \times 7}$  denotes a convolution operation with a kernel size of  $7 \times 7$ . The spatial attention module computes a 2D attention map by first applying average pooling and max pooling operations along the channel dimension of the input  $I_c f'$ , resulting in two spatial descriptors  $I_s^{\text{avg}}$  and  $I_s^{\text{max}}$ . These are concatenated and passed through the convolution layer to obtain the spatial attention map  $S_m(I_c f')$ , which is then element-wise multiplied with the input  $I_c f'$  to produce the final spatially-refined feature map  $I_s f''$ .

The intermediate layers of the model utilize the Rectified Linear Unit (ReLU) activation function [39], defined as:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$



**Fig. 5:** Spatial attention module used in CBAM layers

Here,  $x$  is the input to the activation function. If  $x > 0$ , the output remains  $x$ ; otherwise, the output is zero. ReLU is preferred over sigmoid and tanh due to its ability to mitigate the vanishing gradient problem. It also aids in learning complex, non-linear patterns and accelerates convergence during training by maintaining sparse activation.

The final output layer employs the Softmax activation function [40], a normalized exponential function commonly used for multi-class classification. Softmax transforms the input vector into a probability distribution where the sum of all output probabilities is one, as defined in Eq. 8.

$$\sigma(\vec{z}_i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (8)$$

In our work, as indicated by Eq. 9, we have utilised the Cross Entropy loss function [41], for model optimization.

$$Loss = - \sum_{i=1}^n y_i \log(\hat{y}_i) \quad (9)$$

Our proposed architecture have five CBAM layers which makes it a complex model and can cause overall classification of plant leaf disease detection to become less generic. To reduce the model's complexity, L2 regularization [42] is employed where extreme changes in weights during the training phase are penalized by regularization reducing the probability of over-fitting. During training, the L2 regularizer additionally simplifies the input features and stabilises performance. The definition of the L2 regularizer utilised in our CBAM VGG16 is given in Eq. 10.

$$R_2 = \frac{\lambda}{2N} \sum_{i=1}^n \|\omega_i\|^2 \quad (10)$$

where  $\omega_i$  denotes the two-dimensional weight matrix of the  $i$ th layer,  $N$  is the number of input samples, and  $\lambda$  is the regularization hyperparameter.

### 3.3 Explainability Techniques

#### 3.3.1 Attention Based Technique

The inherent CBAM [8] layers of the models leverages with attention mechanisms to focus on the most informative spatial and channel-wise features. Unlike gradient-based approaches, these methods generate attention maps inherently from the model's structure, offering insight into where the model looks without relying on backpropagation. CBAM attention maps embeds attention mechanisms directly within the network architecture unlike the post-hoc interpretability methods like GradCAM and Grad-CAM++. CBAM sequentially applies two types of attention, channel attention and spatial attention to refine the intermediate feature maps during the forward pass. The attention maps from CBAM can also be visualized to interpret the model's focus at various stages. After each MaxPooling layer, the CBAM module generates:

- Channel attention map ( $C_m \in \mathbb{R}^{C \times 1 \times 1}$ ): These maps highlights the most informative feature channels.
- Spatial attention map ( $S_m \in \mathbb{R}^{1 \times H \times W}$ ): These maps highlight important spatial locations within the feature maps.

These attention maps offer real-time interpretability during training and inference by highlighting where the network focuses at each layer.

#### 3.3.2 Post-hoc Attribution Based Explainability Methods

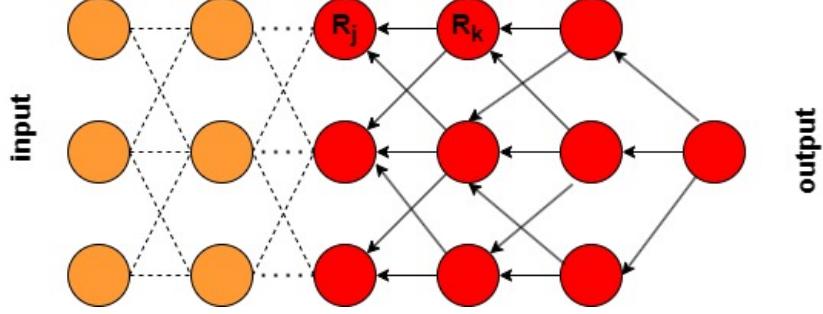
In this section, we explore different post-hoc explainability techniques used to interpret the model's decision-making process.

##### *Class-Discriminative Attribution Technique*

Class-discriminative methods, such as Grad-CAM [29] and Grad-CAM++ [6], utilize the gradients of class scores with respect to intermediate feature maps to produce localization maps. Grad-CAM is a widely-used technique to show the areas of an input image that are distinctive to a class and have the biggest influence on a model's judgment. Grad-CAM++ is a generalized version of Grad-CAM that enables more accurate localization, especially in scenarios with several instances of the same item. These techniques highlight areas of the input that have the most influence for predicting a specific class, offering visually interpretable, class-specific explanations.

##### *Layer-wise Relevance Propagation (LRP)*

LRP [5], [43] belongs to the class of additive explanation techniques, which also includes Shapley values [35], Gradient  $\times$  Input [7], and DeepLIFT [34]. These methods operate under the assumption that a function  $f_j$  with  $N$  input features  $x = \{x_i\}_{i=1}^N$  can be expressed as a sum of contributions from each input variable, represented as  $R_{i \leftarrow j}$ , referred to as *relevance scores*. Here,  $R_{i \leftarrow j}$  quantifies how much the  $i$ th input contributes to the  $j$ -th output. The total function value can be approximated (or exactly recovered) as shown in Eq. 11.



**Fig. 6:** Visualization of the LRP mechanism. Relevance score  $R_j$  or  $R_k$ , denotes how much an individual neuron contributes to the final output prediction

$$f_j(x) \propto R_j = \sum_i R_{i \leftarrow j} \quad (11)$$

When an input  $x_i$  influences multiple outputs  $j$ , as is common in multidimensional functions, the total relevance attributed to  $x_i$  is the sum over its contributions from each output, as defined in Eq. 12.

$$R_i = \sum_j R_{i \leftarrow j} \quad (12)$$

These relevance scores are propagated from the output layer back through the network to the input layer by the LRP algorithm. This backward redistribution is proportional to the contribution of each neuron in the forward computation, effectively assigning an importance score to each input feature. Unlike other interpretability approaches, LRP explicitly treats the neural network as a hierarchical, layer-wise acyclic graph where each unit  $j$  in layer  $l$  is associated with a local function  $f_j^l$ . Relevance values from the output layer  $L$ , where  $R_j^L \propto f_j^L$ , are passed backward, layer by layer, through the network until it reaches the input. This reverse mapping follows the same activation path used during the forward inference process, moving from the output node  $f^L$  down to the initial input  $f^1$ , as illustrated in Fig. 6.

A key characteristic of LRP is its relevance conservation property, shown in Eq. 13, which ensures that the total relevance remains consistent across all layers of the model. This makes the interpretation more robust, as it maintains a one-to-one correspondence with the original model output  $f^L$ .

$$R^{l-1} = \sum_i R_i^{l-1} = \sum_{i,j} R_{i \leftarrow j}^{(l-1,l)} = \sum_j R_j^l = R^l \quad (13)$$

guaranteeing that the total relevance across each layer stays the same enables interpretable attribution, as each relevance score can be directly associated with the model's original output  $f^L$ .

To address limitations of standard LRP, several rule-based variants have been proposed that tailor the relevance redistribution rules to specific layer types and activation behaviors. These methods include the below variants of LRP.

i. Epsilon Plus Rule ( $\varepsilon^+$ )<sup>[5]</sup>: This variant stabilizes relevance propagation by adding a small term  $\varepsilon$  in the denominator, preventing division by small values and suppressing noise in low-activation neurons as given in Eq. 14.

$$R_i = \sum_j \frac{z_{ij}}{z_j + \varepsilon \cdot \text{sign}(z_j)} R_j \quad (14)$$

where  $z_{ij} = x_i w_{ij}$  is the contribution of input neuron  $i$  to output neuron  $j$ ,  $z_j = \sum_i z_{ij}$  is the total input to neuron  $j$ , and  $\text{sign}(z_j)$  denotes the sign of  $z_j$ .

ii. Epsilon Plus Gamma Box Rule ( $\varepsilon^+ \gamma \square$ )<sup>[44]</sup>: This composite rule introduces both the  $\varepsilon$ -stabilizer and a  $\gamma$ -scaling factor that increases the relevance assigned to positively activated neurons, thereby enhancing the interpretability of positively contributing features as given in Eq. 15.

$$R_i = \sum_j \left( \frac{z_{ij}^+}{z_j^+ + \varepsilon} + \gamma \cdot \frac{z_{ij}^-}{z_j^- + \varepsilon} \right) R_j \quad (15)$$

where  $z_{ij}^+ = \max(0, z_{ij})$  and  $z_{ij}^- = \min(0, z_{ij})$  represent the positive and negative contributions respectively,  $z_j^+ = \sum_i z_{ij}^+$  and  $z_j^- = \sum_i z_{ij}^-$ , and  $\gamma$  is a scaling parameter that amplifies the relevance of positive contributions.

iii. Epsilon Plus Flat Rule ( $\varepsilon_{\text{flat}}^+$ )<sup>[43]</sup>: This flat variant evenly distributes relevance across all connected inputs after applying  $\varepsilon$ -stabilization. It is particularly useful in fully connected or linear layers where uniform contribution is desired as given in Eq. 16.

$$R_i = \sum_j \frac{1}{N_j} R_j \quad (16)$$

where  $N_j$  is the number of input neurons connected to neuron  $j$ .

iv. Epsilon Alpha2 Beta 1 Flat Rule ( $\varepsilon_{\text{flat}}^{\alpha=2, \beta=1}$ )<sup>[44]</sup>: A more aggressive redistribution scheme combining  $\varepsilon$ -stabilization with a relevance split governed by  $\alpha = 2$  and  $\beta = 1$ , prioritizing positive contributions and suppressing negative evidence. It is effective for visualizing fine-grained relevance in deeper layers and attention-enhanced models like CBAM-VGG16 as given in Eq. 17.

$$R_i = \sum_j \left( \alpha \cdot \frac{z_{ij}^+}{z_j^+} - \beta \cdot \frac{z_{ij}^-}{z_j^-} \right) R_j \quad (17)$$

where  $\alpha$  and  $\beta$  are relevance weighting parameters (with  $\alpha = 2$ ,  $\beta = 1$ ), and  $z_{ij}^+$ ,  $z_{ij}^-$ ,  $z_j^+$ , and  $z_j^-$  are defined as in Eq. 23.

These rules enable a more flexible and stable relevance propagation pipeline, making them well-suited for diverse architectural components in convolutional neural networks.

**Table 1:** Description of datasets used for plant disease classification

Dataset	Total Images	Training	Testing	Total Disease	Name of Disease/Crops
PlantVillage	54,303	43,442	10,861	38	Apple Scab, Grape Esca, Tomato Early Blight, etc.
Embrapa	46,376	37,100	9,276	93	Coffee, Cotton, Soybean, Citrus, Grapevine etc.
Maize	3,852	3,082	770	4	Gray Leaf Spot, Northern Leaf Blight, and Common Rust.
Apple	3,644	2921	723	4	Cedar Apple Rust, Healthy, Apple Scab etc.
Rice	9,000	7,200	1,800	10	Brown Spot, Blast, Tungro etc.

## 4 Experimental Results

The outcomes of our proposed methodology are shown in this section, together with information on the dataset used, experimental setup, evaluation metrics, results, and explainability analysis.

### 4.1 Datasets

We have selected five different dataset for to evaluate our proposed method. These five datasets, namely, PlantVillage, Embrapa, Maize, Apple, and Rice ensures the generalizability of our approach in plant leaf disease detection for other crops. All the datasets are divided in almost 80:20 ration in the training and testing set. The overall composition of the datasets is provided in Table 1.

### 4.2 Evaluation Metrics

The F1 score (F1), Area Under Curve (AUC), Recall (REC), Accuracy (ACC), Precision (PREC), and Cohen’s kappa score (KAPPA) are common classification metrics that are used to assess and compare the performance of the proposed CBAM-VGG16 model with state-of-the-art (SOTA) models. The evaluation is conducted across five publicly available datasets to ensure robustness and generalizability. The model’s feature representation capability is further assessed through high-dimensional visualization techniques, t-SNE [45] and UMAP [46].

### 4.3 Results and Discussion

The evaluation findings of the suggested CBAM-VGG16 model on five publicly accessible plant disease datasets—Apple, Embrapa, Maize, PlantVillage, and Rice—are shown within this section. According to Table 2, the model demonstrates strong generalization capabilities, achieving high accuracy across all datasets, with the best

<b>Dataset</b>	<b>Loss</b>	<b>Acc</b>	<b>Prec</b>	<b>Rec</b>	<b>F1</b>	<b>AUC</b>	<b>Kappa</b>
Apple	0.41	95.42	95.27	95.42	95.27	99.05	0.90
Embrapa	0.32	94.20	94.70	93.90	94.30	99.75	0.95
Maize	0.12	95.00	95.05	95.00	95.01	99.62	0.93
PlantVillage	0.03	98.72	98.72	98.72	98.72	99.95	0.99
Rice	0.01	98.87	98.87	98.87	98.87	99.94	0.99

**Table 2:** Performance of CBAM-VGG16 model on different datasets

performance observed on the Rice (98.87%) and PlantVillage (98.72%) datasets. These accuracy scores are complemented by consistently high values of precision, recall, F1-score, and AUC, indicating balanced and reliable classification performance across classes. In addition, the model’s effectiveness is further supported by Cohen’s Kappa statistics, which range from 0.90 to 0.99. Notably, the highest Kappa values (0.99) on the Rice and PlantVillage datasets show almost perfect consistency between the actual and predicted labels, outperforming chance and confirming the validity of the suggested approach. Together, these findings show how reliable and successful the CBAM-VGG16 model is at identifying plant diseases in a variety of datasets.

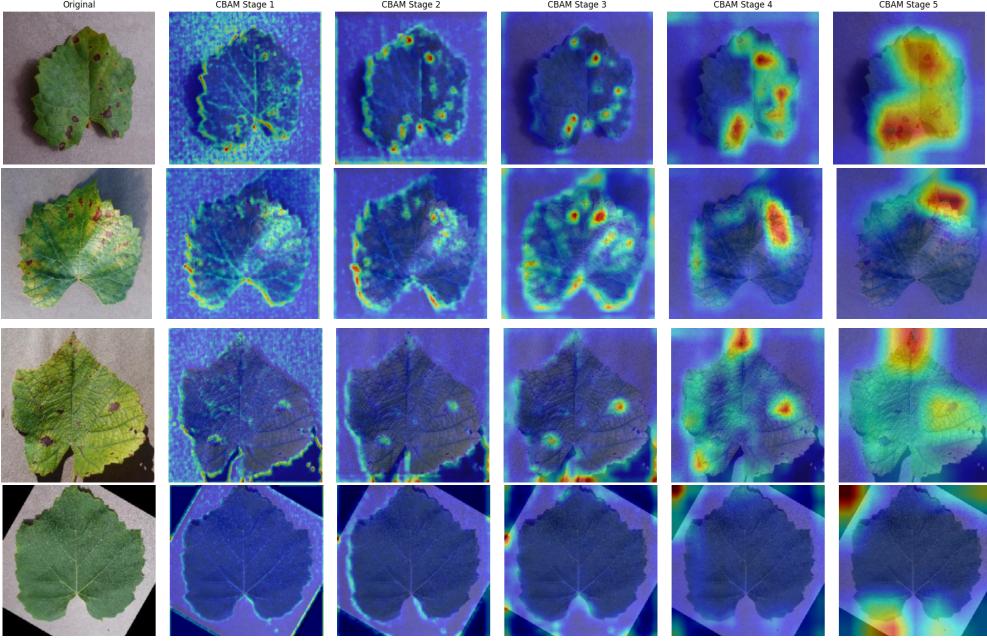
Additionally, we assess our model’s performance by comparing it with recent techniques. The comparative performance is summarized in Table 3. It can be observed from the results that across all five datasets, our CBAM-VGG16 consistently outperforms existing models [16], [19], [47], [51], [48], [49], [50], [52], [53], [55], [56], [54]. On the Apple dataset, the our method performs better with the highest accuracy of 95.42%. In other evaluation metrics also our proposed method demonstrates superior precision, recall, and F1-score, along with an AUC of 99.05% and a Cohen’s kappa score of 0.90. On the Embrapa dataset, CBAM-VGG16 again achieves best performance, with an accuracy of 94.20%. The similar superiority performance is observed in other evaluation metrics as well, indicating the robust generalization capability of our proposed method across different datasets. For the Maize dataset, our model obtains a notable accuracy of 95.00%, outperforming all the methods while also achieving the best F1-score, AUC (99.62%), and Kappa score (0.93). In the PlantVillage dataset, the proposed model maintains a competitive edge with nearly perfect scores across all metrics. Notably, it again exceeds in performance, reaffirming its superior feature learning and attention capabilities as compared with other methods. Lastly, on the Rice dataset, our approach attains the maximum accuracy of 98.87%, alongside the best precision, recall, and AUC (99.94%), demonstrating its effectiveness in capturing fine-grained disease patterns even in challenging samples. The obtained results on all the datasets exhibits that the suggested approach has good generalizability and can be adapted to identify leaf diseases in any other crop.

#### 4.4 Explainability Analysis

We have selected grape leaf dataset from plant village dataset for detailed explainability analysis of our achieved results to validate the model’s interpretability.

Apple Dataset								
Author	LOSS	ACC	PREC	REC	F1	AUC	KAPPA	
Karthik et al. [16]	1.56	62.37	62.84	61.83	62.35	83.53	0.45	
Chen et al. [19]	0.79	83.33	83.70	82.80	83.25	94.10	0.76	
Chen et al. [47]	0.63	83.33	85.47	82.26	83.83	95.59	0.76	
Li et al. [48]	1.34	41.94	44.59	35.48	39.52	71.68	0.16	
Chen et al. [49]	0.73	87.10	87.10	87.10	87.10	95.02	0.81	
Yang et al. [50]	0.49	85.48	85.71	83.87	84.78	96.44	0.79	
Thakur et al. [51]	<b>0.30</b>	93.55	93.55	93.55	93.55	97.01	0.91	
Thakur et al. [52]	0.43	94.62	94.62	94.62	94.62	98.08	<b>0.92</b>	
Tong Li et al. [53]	—	93.89	93.84	—	93.80	—	—	
<b>Proposed work</b>	0.41	<b>95.42</b>	<b>95.27</b>	<b>95.42</b>	<b>95.27</b>	<b>99.05</b>	0.90	
Embrapa Dataset								
Author	LOSS	ACC	PREC	REC	F1	AUC	KAPPA	
Karthik et al. [16]	0.77	80.29	83.07	78.38	80.60	97.80	0.82	
Chen et al. [19]	1.11	73.63	80.12	67.86	73.48	98.02	0.73	
Chen et al. [47]	1.12	74.88	82.93	66.06	73.18	98.20	0.75	
Li et al. [48]	0.90	73.17	79.23	67.59	72.95	98.38	0.72	
Chen et al. [49]	0.28	93.19	94.33	93.47	93.90	99.08	0.94	
Yang et al. [50]	0.62	88.48	90.91	86.53	88.67	99.51	0.88	
Thakur et al. [51]	0.46	89.24	91.17	88.27	89.70	98.73	0.89	
Thakur et al. [52]	0.34	93.83	94.78	93.21	93.99	99.68	0.94	
<b>Proposed work</b>	<b>0.32</b>	<b>94.20</b>	<b>94.70</b>	<b>93.90</b>	<b>94.30</b>	<b>99.75</b>	<b>0.95</b>	
Maize Dataset								
Author	LOSS	ACC	PREC	REC	F1	AUC	KAPPA	
Karthik et al. [16]	1.17	54.32	59.72	53.09	56.21	83.19	0.39	
Chen et al. [19]	0.60	87.65	87.65	87.65	87.65	96.03	0.84	
Chen et al. [47]	0.50	88.89	91.03	87.65	87.65	96.78	0.85	
Li et al. [48]	1.38	50.62	51.47	43.21	46.98	72.47	0.34	
Chen et al. [49]	0.54	91.36	91.36	91.36	91.36	97.29	0.89	
Yang et al. [50]	1.52	76.54	78.48	76.54	77.50	94.53	0.69	
Thakur et al. [51]	0.34	92.59	<b>93.67</b>	91.36	92.50	97.21	0.90	
Thakur et al. [52]	<b>0.04</b>	92.59	92.59	92.59	92.59	96.27	0.90	
Alpsalaz et al. [54]	—	94.97	—	—	—	—	—	
<b>Proposed work</b>	0.12	<b>95.00</b>	95.05	<b>95.00</b>	<b>95.01</b>	<b>99.62</b>	<b>0.93</b>	
PlantVillage Dataset								
Author	LOSS	ACC	PREC	REC	F1	AUC	KAPPA	
Karthik et al. [16]	0.16	95.83	96.20	95.60	95.89	99.70	0.96	
Chen et al. [19]	1.07	74.63	80.92	69.64	74.03	98.18	0.74	
Chen et al. [47]	0.17	96.68	97.49	95.83	96.64	99.26	0.97	
Li et al. [48]	0.46	86.51	88.85	85.08	86.92	99.40	0.85	
Chen et al. [49]	0.12	97.27	97.29	97.27	97.27	99.63	0.97	
Yang et al. [50]	0.09	97.69	97.83	97.69	97.71	99.82	0.98	
Thakur et al. [51]	0.04	98.46	98.52	98.46	98.46	99.91	0.98	
Thakur et al. [52]	0.06	98.29	98.30	98.29	98.29	<b>99.95</b>	0.98	
Shafik et al. [55]	—	97.8	97.7	96.8	94.4	—	—	
<b>Proposed work</b>	<b>0.03</b>	<b>98.72</b>	<b>98.72</b>	<b>98.72</b>	<b>98.72</b>	<b>99.95</b>	<b>0.99</b>	
Rice Dataset								
Author	LOSS	ACC	PREC	REC	F1	AUC	KAPPA	
Karthik et al. [16]	0.13	96.30	96.41	96.30	96.30	99.76	0.96	
Chen et al. [19]	0.45	85.92	89.25	85.92	85.96	98.90	0.86	
Chen et al. [47]	0.14	95.06	95.19	95.06	95.06	99.41	0.95	
Li et al. [48]	0.54	91.01	92.69	91.01	91.01	99.61	0.91	
Chen et al. [49]	0.08	97.20	97.31	97.20	97.20	99.81	0.97	
Yang et al. [50]	0.14	95.67	95.79	95.67	95.67	99.84	0.95	
Thakur et al. [51]	0.04	97.59	97.59	97.59	97.59	99.88	0.97	
Thakur et al. [52]	0.03	98.10	98.11	98.10	98.10	99.93	0.98	
Bhuyan et al. [56]	—	98.41	98.58	98.39	98.19	—	—	
<b>Proposed work</b>	<b>0.01</b>	<b>98.87</b>	<b>98.87</b>	<b>98.87</b>	<b>98.87</b>	<b>99.94</b>	<b>0.99</b>	

**Table 3:** Comparison of the work with other SOTA methods on five publicly available datasets



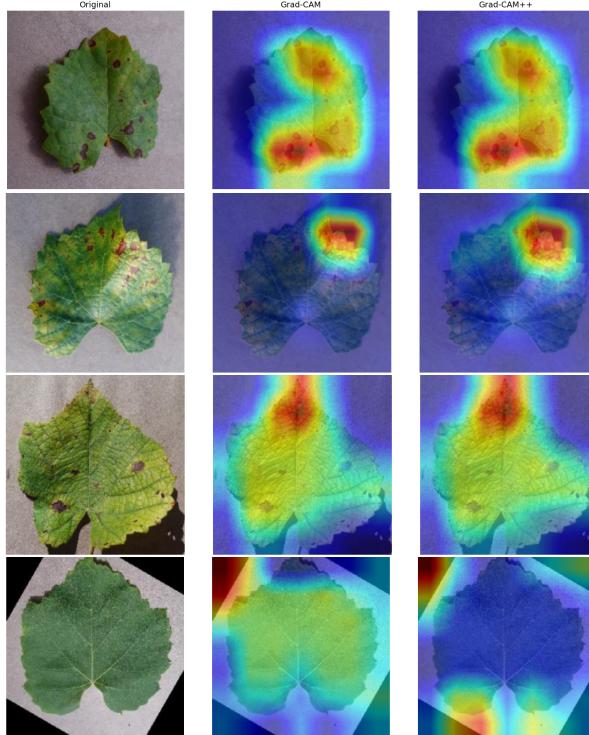
**Fig. 7:** Illustrations of CBAM attention maps generated for inputs of different classes on a CBAM-VGG16. Rows from top to bottom corresponds to images of class black rot, esca, leaf blight and healthy, respectively.

#### 4.4.1 CBAM Attention Maps Visualization

In the Fig. 7, CBAM attention maps generated across all the five convolutional stages inserted in the proposed CBAM-VGG16 model for grape leaf classification are shown. These maps provide insight into how the model attends to informative regions at various depths of the network. It can be seen that the first CBAM block (CBAM1) provides an attention maps highlighting low-level features such as leaf textures, edges, and background suppression. These early filters focus on enhancing basic visual cues while discarding irrelevant background. The subsequent CBAM blocks, further refines these maps where focus is diverted to distinguishable disease prone areas in the leaf. Overall, these visualizations demonstrate how CBAM refines feature learning progressively across the network, resulting in improved focus, interpretability, and diagnostic accuracy.

#### 4.4.2 Attribution Based Explainability Methods

To better understand the internal reasoning of the CBAM-VGG16 model predictions, we examine and compare visual explanations generated using, Grad-CAM, and Grad-CAM++ (post-hoc explainability techniques). Fig. 8 displays a visual comparison of these heatmaps for various leaf samples across different disease classes. Grad-CAM computes gradients reaching the last convolutional layers to determine which

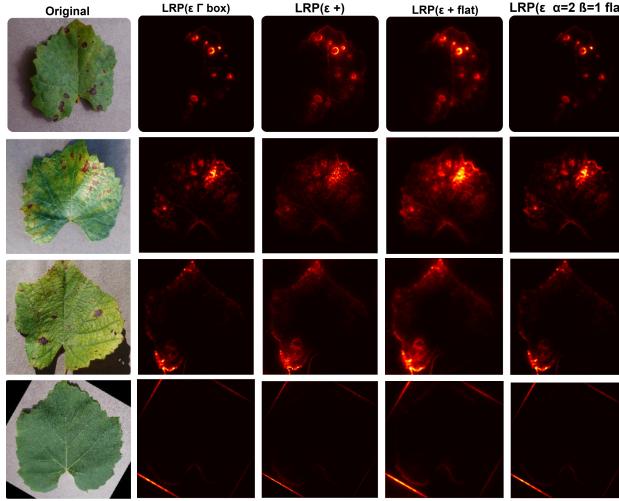


**Fig. 8:** Visualization of GradCAM, GradCAM++ outputs. Rows from top to bottom corresponds to images of class black rot, esca, leaf blight and healthy, respectively.

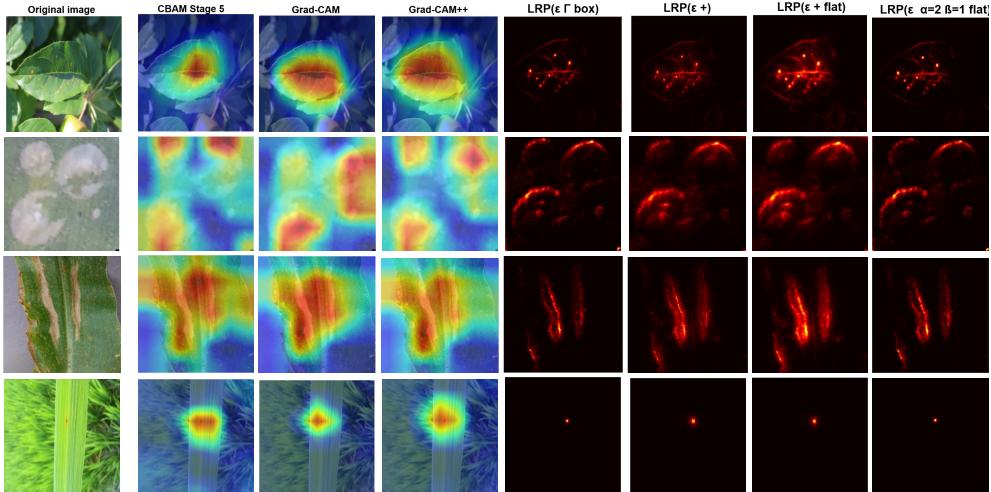
input locations most affected the model’s decision-making. However, it typically produces relatively coarse, blob-like heatmaps. In contrast, Grad-CAM++ refines this by incorporating second-order gradients, resulting in more spatially accurate and class-discriminative attention maps. Despite their usefulness, both are post-hoc techniques, which do not influence or enhance the training process directly.

#### 4.4.3 Layerwise Relevance Propagation (LRP)

Fig. 9 illustrates the class-wise attribution heatmaps generated for the CBAM-VGG16 model using a comprehensive suite of LRP techniques. The columns represent different attribution methods of LRP ,namely *Epsilon Gamma Box*, *Epsilon Plus*, *Epsilon Plus Flat*, and *Epsilon Alpha2 Beta1 Flat*. The LRP family consistently produces sparse yet discriminative saliency maps, aligning well with regions exhibiting disease-related symptoms. Among them, *Epsilon Plus Flat* and *Epsilon Alpha2 Beta1 Flat* stand out for their capacity to sharply localize high-relevance areas, typically corresponding to lesions, necrotic margins, or discolored patches symptomatic of disease. These variants effectively suppress irrelevant background and vein structures, enhancing focus on pathologically significant textures.

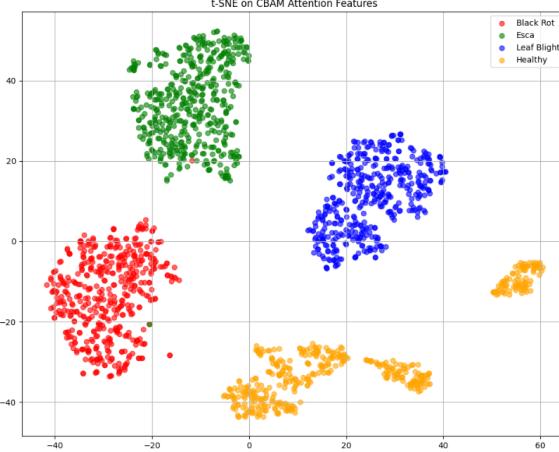


**Fig. 9:** LRP visualization across multiple rules and leaf samples. Each row shows a different test image of class black rot, esca, leaf blight and healthy, respectively; each column corresponds to a different LRP technique.



**Fig. 10:** Explainability analysis for other crop leafs used in the experiments. Top to bottom row represents leafs of Apple, Embrapa (Coconut), Maize, and Rice.

We have shown in Fig. 10 the explainable visualization of proposed method on some other disease of other dataset used in this work. These visualization also exhibits the same properties as observed in the detailed explainability analysis for grape leaf dataset.



**Fig. 11:** Visualization of extracted features using t-SNE plots.

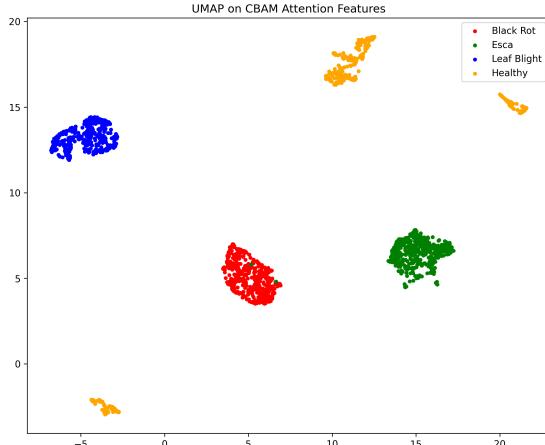
## 4.5 Feature Analysis

### 4.5.1 t-SNE Visualization

To evaluate the discriminative capability of the CBAM-enhanced VGG16 model, we employed t-SNE to project high-dimensional feature representations into a 2-D space. Fig. 11 presents the t-SNE plots on the four grape leaf classes: *Esca*, *Healthy*, *Leaf Blight*, and *Black Rot*. Each point represents a sample image, colored according to its ground truth class label. It can be interpreted from the t-SNE figure that our proposed model is providing a clear distinguishable cluster separation with minimal class overlap, indicating better feature discrimination. These findings demonstrate how well CBAM works to improve the VGG16 backbone’s representational quality, resulting in improved class-wise separability and model interpretability.

### 4.5.2 Uniform Manifold Approximation and Projection (UMAP) Feature Visualization

We have also employed UMAP to display features with great dimensions in a two-dimensional area. Fig. 12 presents the UMAP projections of features extracted from the CBAM-VGG16 model. In the CBAM-based UMAP plot, class boundaries are clearly defined, with each category forming dense and distinct clusters. For instance, Healthy samples are sharply isolated from all disease classes, while Black Rot and Esca show minimal overlap, indicating improved intra-class consistency. Leaf Blight forms a distinct cluster far from the rest, highlighting CBAM’s ability to capture unique disease-specific features. These visual differences affirm the effectiveness of CBAM in refining deep feature representations by leveraging spatial and channel-wise attention.



**Fig. 12:** Visualization of extracted features using UMAP plots.

## 5 Conclusion and Future Work

The study in this work provides an interpretable CBAM-integrated VGG16 approach for detection of plant leaf diseases. By incorporating attention modules at each convolutional stage, the model not only enhances disease-specific feature extraction but also offers improved performance across five public datasets. The performance on five different datasets exhibits the generalizability of our proposed solution for any other crop. The proposed architecture achieves high classification accuracy while maintaining compatibility with multiple attribution methods for visual explanation. In order to evaluate the model's interpretability, we carried this in-depth evaluation using different attribution techniques. The qualitative study highlighted that LRP variants produced the most visually clear, localized, and class-discriminative heatmaps with minimal noise. Future directions include refining the attention mechanism to enhance class-awareness and reduce interpretability gaps, exploring global or transformer-inspired attention integration for better contextual representation, and conducting human-in-the-loop evaluations to assess the trust and reliability of visualizations. Furthermore, combining attention maps with attribution heatmaps can lead to a more unified and semantically rich interpretability framework suitable for real-world agricultural issues.

### Author Contribution:

Balram singh has implemented the work and written the first draft of the work. Ram Prakash Sharma has helped to formulate the problem, develop the solution, and proofread the manuscript. Somnath Dey has improved the problem solution with proofreading of the manuscript.

### Funding:

No funding has been received to carry out this work.

## **Data Availability:**

This study has not generated any datasets during the current study.

## **Declarations**

### **Conflict of interest:**

The authors declare that they have no conflict of interest.

## **References**

- [1] Xiang, S., Liang, Q., Sun, W., Zhang, D., Wang, Y.: L-csms: Novel lightweight network for plant disease severity recognition. *Journal Of Plant Diseases And Protection* **128**(2), 557–569 (2021)
- [2] Lin, C., Zou, C., Xu, H.: Scnet: A dual-branch network for strong noisy image denoising based on swin transformer and convnext. *Computer Animation and Virtual Worlds* **36**(3), 70030 (2025)
- [3] Lin, X., Sun, S., Huang, W., Sheng, B., Li, P., Feng, D.D.: Eapt: Efficient attention pyramid transformer for image processing. *IEEE Transactions on Multimedia* **25**, 50–61 (2023)
- [4] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626 (2017)
- [5] Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W.: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE* **10**(7), 1–46 (2015)
- [6] Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 839–847 (2018)
- [7] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR), 1–14 (2015)
- [8] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. *Computer Vision – ECCV*, 3–19 (2018)
- [9] Labiad, I., Boubchir, L., Seddik, H.: Optimization of 2d and 3d facial recognition through the fusion of cbam alexnet and resnext models. *The Visual Computer* **41**, 5235–5250 (2025)

- [10] Zhang, H., Ren, G.: Intelligent leaf disease diagnosis: image algorithms using swin transformer and federated learning. *The Visual Computer* **41**, 4815–4838 (2025)
- [11] Zhang, M., Tian, X.: Transformer architecture based on mutual attention for image-anomaly detection. *Virtual Reality & Intelligent Hardware* **5**(1), 57–67 (2023)
- [12] Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. *Frontiers in Plant Science* **7**, 1–10 (2016)
- [13] Barbedo, J.G.A.: Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers And Electronics In Agriculture* **153**, 46–53 (2018)
- [14] Singh, U.P., Chouhan, S.S., Jain, S., Jain, S.: Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease. *IEEE Access* **7**, 43721–43729 (2019)
- [15] Dwivedi, R., Dey, S., Chakraborty, C., Tiwari, S.: Grape disease detection network based on multi-task learning and attention features. *IEEE Sensors Journal* **21**(16), 17573–17580 (2021)
- [16] Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., Menaka, R.: Attention embedded residual cnn for disease detection in tomato leaves. *Applied Soft Computing* **86**, 1–12 (2020)
- [17] Zeng, W., Li, M.: Crop leaf disease recognition based on self-attention convolutional neural network. *Computers and Electronics in Agriculture* **172**, 1–7 (2020)
- [18] Chen, J., Zhang, D., Suzauddola, M., Nanehkaran, Y.A., Sun, Y.: Identification of plant disease images via a squeeze-and-excitation mobilenet model and twice transfer learning. *IET Image Processing* **15**(5), 1115–1127 (2021)
- [19] Chen, J., Zhang, D., Suzauddola, M.I., Zeb, A.: Identifying crop diseases using attention embedded mobilenet-v2 model. *Applied Soft Computing* **113**, 1–12 (2021)
- [20] Cap, Q.H., Uga, H., Kagiwada, S., Iyatomi, H.: Leafgan: An effective data augmentation method for practical plant disease diagnosis. *IEEE Transactions on Automation Science and Engineering* **19**(2), 1258–1267 (2022)
- [21] Garg, G., Gupta, S., Mishra, P., Vidyarthi, A., Singh, A., Ali, A.: Cropcare: An intelligent real-time sustainable iot system for crop disease detection using mobile vision. *IEEE Internet of Things Journal* **10**(4), 2840–2851 (2023)
- [22] Zhao, Y., Chen, Z., Gao, X., Song, W., Xiong, Q., Hu, J., Zhang, Z.: Plant disease

- detection using generated leaves based on doublegan. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **19**(3), 1817–1826 (2022)
- [23] Pal, A., Kumar, V.: AgriNet: Plant leaf disease severity classification using agriculture detection framework. *Engineering Applications of Artificial Intelligence* **119**, 105754 (2023)
  - [24] Liu, K., Zhang, X.: Pitlid: Identification of plant disease from leaf images based on convolutional neural network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **20**(2), 1278–1288 (2023)
  - [25] Gehlot, M., Saxena, R.K., Gandhi, G.C.: “tomato-village”: a dataset for end-to-end tomato disease detection in a real-world environment. *Multimedia Systems* **29**, 3305–3328 (2023)
  - [26] Zeng, Q., Niu, L., Wang, S., Ni, W.: Sevit: a large-scale and fine-grained plant disease classification model based on transformer and attention convolution. *Multimedia Systems* **29**, 1001–1010 (2023)
  - [27] Arvind, C.S., Totla, A., Jain, T., Sinha, N., Jyothi, R., Aditya, K., Keerthan, Farhan, M., Sumukh, G., Ak, G.: Deep learning based plant disease classification with explainable ai and mitigation recommendation. *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8 (2021)
  - [28] Kinger, S., Kulkarni, V.: Explainable ai for deep learning based disease detection. *Thirteenth International Conference on Contemporary Computing (IC3)*, 209–216 (2021)
  - [29] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision* **128**(2), 336–359 (2019)
  - [30] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929 (2016)
  - [31] Binder, A., Montavon, G., Lapuschkin, S., Müller, K.-R., Samek, W.: Layer-wise relevance propagation for neural networks with local renormalization layers. *Artificial Neural Networks and Machine Learning – ICANN*, 63–71 (2016)
  - [32] Tariq, M., Ali, U., Abbas, S., Hassan, S., Naqvi, R.A., Khan, M.A., Jeong, D.: Corn leaf disease: Insightful diagnosis using vgg16 empowered by explainable ai. *Frontiers in Plant Science* **15**, 1–12 (2024)
  - [33] Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you? explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144

(2016)

- [34] Shrikumar, A., Greenside, P., Kundaje, A.: Learning important features through propagating activation differences. In: Proceedings of the 34th International Conference on Machine Learning (ICML), pp. 3145–3153 (2017)
- [35] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS), pp. 4768–4777 (2017)
- [36] Mukherjee, R., Ghosh, A., Chakraborty, C., De, J.N., Mishra, D.P.: Rice leaf disease identification and classification using machine learning techniques: A comprehensive review. *Engineering Applications of Artificial Intelligence* **139**, 109639 (2025)
- [37] Qadri, S.A.A., Huang, N.-F., Wani, T.M., Bhat, S.A.: Advances and challenges in computer vision for image-based plant disease detection: A comprehensive survey of machine and deep learning approaches. *IEEE Transactions on Automation Science and Engineering* **22**, 2639–2670 (2025)
- [38] Pizer, S.M., Johnston, R.E., Erickson, J.P., Yankaskas, B.C., Muller, K.E.: Contrast-limited adaptive histogram equalization: Speed and effectiveness. In: Proceedings of the First Conference on Visualization in Biomedical Computing, pp. 337–345 (1990)
- [39] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML), pp. 807–814 (2010)
- [40] Xiao, Z., Xu, P., Wang, X., Chen, L., An, F.: A multi-class objects detection coprocessor with dual feature space and weighted softmax. *IEEE Transactions on Circuits and Systems II: Express Briefs* **67**(9), 1629–1633 (2020)
- [41] Mao, A., Mohri, M., Zhong, Y.: Cross-entropy loss functions: Theoretical analysis and applications. In: Proceedings of the 40th International Conference on Machine Learning, pp. 23803–23828 (2023)
- [42] Yang, M., Lim, M.K., Qu, Y., Li, X., Ni, D.: Deep neural networks with l1 and l2 regularization for high dimensional corporate credit risk prediction. *Expert Systems with Applications* **213**, 1–20 (2023)
- [43] Montavon, G., Lapuschkin, S., Binder, A., Samek, W., Müller, K.-R.: Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition* **65**, 211–222 (2017)
- [44] Anders, C., Neumann, D., Samek, W., Müller, K.-R., Lapuschkin, S.: Software for dataset-wide xai: From local explanations to global insights with zennit, corelay,

- and virelay. arXiv e-prints, 2106–13200 (2021)
- [45] Maaten, L., Hinton, G.: Visualizing data using t-sne. *Journal of Machine Learning Research* **9**(86), 2579–2605 (2008)
- [46] McInnes, L., Healy, J., Saul, N., Großberger, L.: Umap: Uniform manifold approximation and projection. *Journal of Open Source Software* **3**(29), 1–2 (2018)
- [47] Chen, J., Zhang, D., Suzauddola, M., Zeb, A.: Identification of rice plant diseases using lightweight attention networks. *Expert Systems with Applications* **169**, 1–13 (2021)
- [48] Li, H., Li, S., Yu, J., Han, Y., Dong, A.: Plant disease and insect pest identification based on vision transformer. In: *Proceedings of the International Conference on Internet of Things and Machine Learning*, vol. 12174, pp. 1–10 (2022)
- [49] Chen, J., Chen, W., Zeb, A., Yang, S., Zhang, D.: Lightweight inception networks for the recognition and detection of rice plant diseases. *IEEE Sensors Journal* **22**(14), 14628–14638 (2022)
- [50] Yang, Y., Liu, Y., Wang, H., Zhang, Y., Zhang, Y.: Googlenet based on residual network and attention mechanism identification of rice leaf diseases. *Computers and Electronics in Agriculture* **204**, 1–11 (2023)
- [51] Thakur, P.S., Chaturvedi, S., Khanna, P., Sheorey, T., Ojha, A.: Vision transformer meets convolutional neural network for plant disease classification. *Ecological Informatics* **77**, 1–21 (2023)
- [52] Thakur, P. Shubhangi Khanna, Sheorey, T., Ojha, A.: Real-time plant disease identification: Fusion of vision transformer and conditional convolutional network with c3gan-based data augmentation. *IEEE Transactions on AgriFood Electronics* **2**(2), 576–586 (2024)
- [53] Li, T., Zhang, L., Lin, J.: Precision agriculture with yolo-leaf: advanced methods for detecting apple leaf diseases. *Frontiers in Plant Science* **15**, 1–15 (2024)
- [54] Alpsalaz, F., Özüpak, Y., Aslan, E., Uzel, H.: Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence. *Chemometrics and Intelligent Laboratory Systems* **262**, 1–14 (2025)
- [55] Shafik, W., Tufail, A., De Silva Liyanage, C., Apong, R.A.A.H.M.: Using transfer learning-based plant disease classification and detection for sustainable agriculture. *BMC Plant Biology* **24**(1), 1–19 (2024)
- [56] Bhuyan, P., Singh, P.K.: Evaluating deep cnns and vision transformers for plant leaf disease classification. In: *Distributed Computing and Intelligent Technology*,

pp. 293–306 (2024)