Attention-Based Deep Convolutional Neural Networks for Plant Disease Classification

Sachin B. Jadhav
sachinbjadhav84@gmail.com

Jawaharlal Nehru University https://orcid.org/0000-0002-6849-9974

Pratik Pal
Jawaharlal Nehru University

Research Article

Keywords: Deep learning, CNN, plant disease classification, Squeeze-and-Excitation block, Attention Gate, Grad-CAM, precision agriculture

Posted Date: September 12th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7463210/v1

Title Page

Attention-Based Deep Convolutional Neural Networks for Plant Disease Classification

Sachin B. Jadhav[1] Ph.D. ORCID ID: https://orcid.org/0000-0002-6849-9974

[1]School of Computer & Systems Sciences (SC&SS),

Jawaharlal Nehru University (JNU), New Delhi, India.

[1] sachinjadhav@jnu.ac.in

Prati Pal[2] (M.Tech. Scholar)

[2]School of Computer & Systems Sciences (SC&SS),

Jawaharlal Nehru University (JNU), New Delhi, India.

[2] pratik25_scs@jnu.ac.in

Corresponding Author

Sachin B. Jadhav[1] Ph.D. ORCID ID: https://orcid.org/0000-0002-6849-9974

[1]School of Computer & Systems Sciences (SC&SS),

Jawaharlal Nehru University (JNU), New Delhi, India.

[1] sachinjadhav@jnu.ac.in

Author Contributions

Conceptualization: [Sachin B. Jadhav and Pratik Pal]; **Methodology: [** Sachin B. Jadhav and Pratik Pal**]**; **Formal analysis and investigation:** [Sachin B. Jadhav and Pratik Pal]; **Writing - original draft preparation:** [Pratik Pal], **Writing - review and editing:** [Sachin B. Jadhav]; **Resources:** [Sachin B. Jadhav]; **Supervision**: [Sachin B. Jadhav]

Abstract

Plant diseases pose a significant threat to global food security and agricultural productivity. In this work, we propose a novel deep convolutional neural network (CNN) model enhanced with Squeeze-and-Excitation (SE) blocks and Attention Gates (AGs) for multi-class plant disease classification across five crops: apple, maize, grape, potato, and tomato. Leveraging a large image dataset and a comprehensive training regime, the proposed model achieves high performance across all metrics, including 99% accuracy, 0.99 F1-score, and strong specificity. Evaluation includes feature visualization and Grad-CAM interpretability. The model's robustness and interpretability make it a compelling solution for practical agricultural applications.

Keywords

Deep learning, CNN, plant disease classification, Squeeze-and-Excitation block, Attention Gate, Grad-CAM, precision agriculture.

1. Introduction

Agriculture remains a cornerstone of national economies. According to the FAO, 828 million people faced hunger last year, underscoring the urgent need to boost food production for a growing global population [1]. Agricultural sustainability is increasingly threatened by the prevalence of plant diseases, which can severely diminish crop yields and food quality. Traditional disease identification methods rely heavily on human expertise and visual inspection, making them time-consuming, error-prone, and impractical for large-scale monitoring [2]. As global food demand intensifies, there is an urgent need for scalable, accurate, and automated disease detection systems.

Deep learning, particularly convolutional neural networks (CNNs), has shown promise in plant pathology by learning complex visual patterns directly from leaf imagery [3]. CNN architectures such as AlexNet, VGG19, ResNet, GoogleNet, and EfficientNet have shown high accuracy in plant disease identification [4], [5], [6]. However, standard CNNs often struggle to generalize across diverse disease classes and crop types, especially when the diseased regions are subtle or spatially varied. Attention mechanisms have emerged as a powerful enhancement to CNNs, enabling models to focus on the most informative features and suppress irrelevant background noise.

In this study, we propose an attention-enhanced deep CNN framework that integrates **Squeeze-and-Excitation (SE) blocks** and **Attention Gates (AGs)** for fine-grained plant disease classification. Our model is trained on a multi-species dataset comprising five economically important crops—apple, maize, grape, potato, and tomato—spanning 21 disease categories. The proposed architecture not only improves classification accuracy but also enhances interpretability through feature visualization and Gradient-weighted Class Activation Mapping (Grad-CAM).

The suggested methodology presents several significant contributions that are mentioned below:

- To design a deep CNN architecture augmented with SE blocks and AGs to boost channel-wise and spatial feature discrimination.
- To evaluate the model across a comprehensive multi-class, multi-crop plant disease dataset, achieving high performance across key metrics.
- To apply Grad-CAM and intermediate feature map visualization to provide insights into the model's decision-making process, enhancing transparency and trust.

The rest of the paper is structured as follows: Section 2 reviews relevant literature, highlighting different methods that employ machine learning and deep learning techniques for the classification of various plant diseases. Section 3 describes the proposed model in detail. Section 4 provides an overview of the dataset and implementation details, examines the performance metrics used, and presents the results alongside a comparative analysis of the proposed model concerning various existing classifiers. Lastly, Section 5 concludes with the final insights derived from this paper.

2. Related work

The application of deep learning in plant disease classification has garnered significant attention due to its accuracy, scalability, and automation potential. Traditional methods relying on expert inspection are increasingly being replaced by Convolutional Neural Networks (CNNs), which can detect complex patterns in leaf imagery.

Several studies have focused on improving classification through architectural enhancements and preprocessing techniques. Ji et al. [7] achieved 95% accuracy in classifying tomato leaf diseases using a transfer learning-based model, later deployed as a mobile and web app. Ngugi et al. [8] explored preprocessing strategies (e.g., RGB to CMYK conversion with Gaussian filters), obtaining over 99% accuracy on ResNet-50. Similarly, Umamageswari et al. [9] used wavelet transforms and WNNs to identify mango leaf diseases with 98.93% accuracy.

To handle multi-species and multi-label datasets, Paul et al. [10] employed DenseNet and Xception, demonstrating the advantages of skip connections and deep feature hierarchies. Hossain et al. [11] systematically evaluated CNNs, YOLO, and Vision Transformers across datasets, noting trade-offs between accuracy and computational demand. Kabir et al. [12] designed a lightweight 12-layer CNN achieving 97% F1-score on 38 disease classes, suitable for mobile deployment.

Hybrid and ensemble models have also emerged. Krisnandi et al. [13] combined VGG-16 and MobileNet via stacking ensemble learning to classify sunflower leaf diseases. Mustofa et al. [14] and Kiran et al. [15] explored traditional machine learning techniques like Random Forest and SVM, aided by segmentation and handcrafted features, but noted limitations in generalization.

Recent innovations such as EfficientNet [16] and attention-based data augmentation [17] have further improved performance, particularly on imbalanced or synthetic datasets. These studies confirm the growing preference for deep models over traditional techniques due to their robustness and interpretability.
Between 2023 and 2024, several advanced studies [18], [19], [20] explored transfer learning-based few-shot learning (FSL) methods to markedly improve prediction accuracy in plant disease detection.

Despite extensive work across crops like tomato, mango, and wheat, few studies focus on fig trees or integrate channel- and spatial-level attention in one unified model. Our work addresses this by proposing a CNN architecture that combines Squeeze-and-Excitation (SE) blocks and Attention Gates (AGs), trained on a diverse, multi-crop dataset, and evaluated using interpretability tools like Grad-CAM.

3. Materials and methods

3.1 Proposed Architecture

We propose a deep convolutional neural network (CNN) for image classification, integrating Squeeze-and-Excitation (SE) blocks and Attention Gates (AGs) to enhance feature representation and spatial selectivity. The architecture consists of a hierarchical feature extraction backbone, followed by an attention-enhanced feature refinement module, and a fully connected classification head.

3.1.1. Convolutional Blocks

The model begins with a series of convolutional blocks to extract hierarchical spatial features. Each block consists of:

- A 3×3 convolutional layer to capture local spatial patterns,
- Batch Normalization to stabilize training,
- ReLU activation for non-linearity,
- MaxPooling (where applicable) to downsample the feature maps.

The network follows a progressive depth-wise expansion, where the number of convolutional filters increases as: 32 → 64 → 128 → 256 → 512, ensuring that low-level features (e.g., edges and textures) transition into high-level semantic representations.

3.1.2. Squeeze-and-Excitation (SE) Blocks

To refine feature representations, SE blocks are inserted at multiple stages in the network. Each SE block performs channel-wise attention by dynamically recalibrating feature map activations. The process includes:

- Global Average Pooling (GAP) to condense spatial features into a single vector per channel,
- A fully connected bottleneck layer that reduces channel dimensions (filters // ratio),
- A second fully connected expansion layer, restoring the original channel dimensions,
- A sigmoid activation to compute per-channel importance scores,
- A scaling operation, where the original feature maps are weighted element-wise by the learned importance scores.

The SE mechanism enhances discriminative channels while suppressing redundant information, improving the network's ability to focus on critical patterns.

3.1.3. Attention Gates (AGs)

In addition to SE blocks, Attention Gates (AGs) are employed to suppress irrelevant activations and enhance region-specific feature learning. Each AG consists of:

- A 1×1 convolutional gating signal, which guides attention based on high-level contextual information
- A parallel 1×1 convolution applied to the input feature map,
- An element-wise sum of the input feature map and the gating signal,
- A ReLU activation, followed by a 1×1 convolution with sigmoid activation to generate spatial attention weights,
- An element-wise multiplication between the input feature map and the attention weights.

The AGs allow the network to selectively enhance important spatial regions, improving feature selectivity and reducing background noise.

3.1.4. Classification Head

After feature extraction and attention refinement, the final feature maps are flattened and passed through fully connected layers:

- 256 neurons with ReLU activation,
- 128 neurons with ReLU activation,
- Softmax activation with num_classes=21, producing probability distributions over target classes.

3.1.5. Overall Pipeline

The overall network architecture follows:

- Initial feature extraction using convolutional blocks,
- Feature refinement via SE blocks,
- Spatial attention enhancement via AGs,
- Final classification through fully connected layers.

This approach enables the model to efficiently learn complex hierarchical features, while leveraging attention mechanisms for improved feature selectivity and interpretability

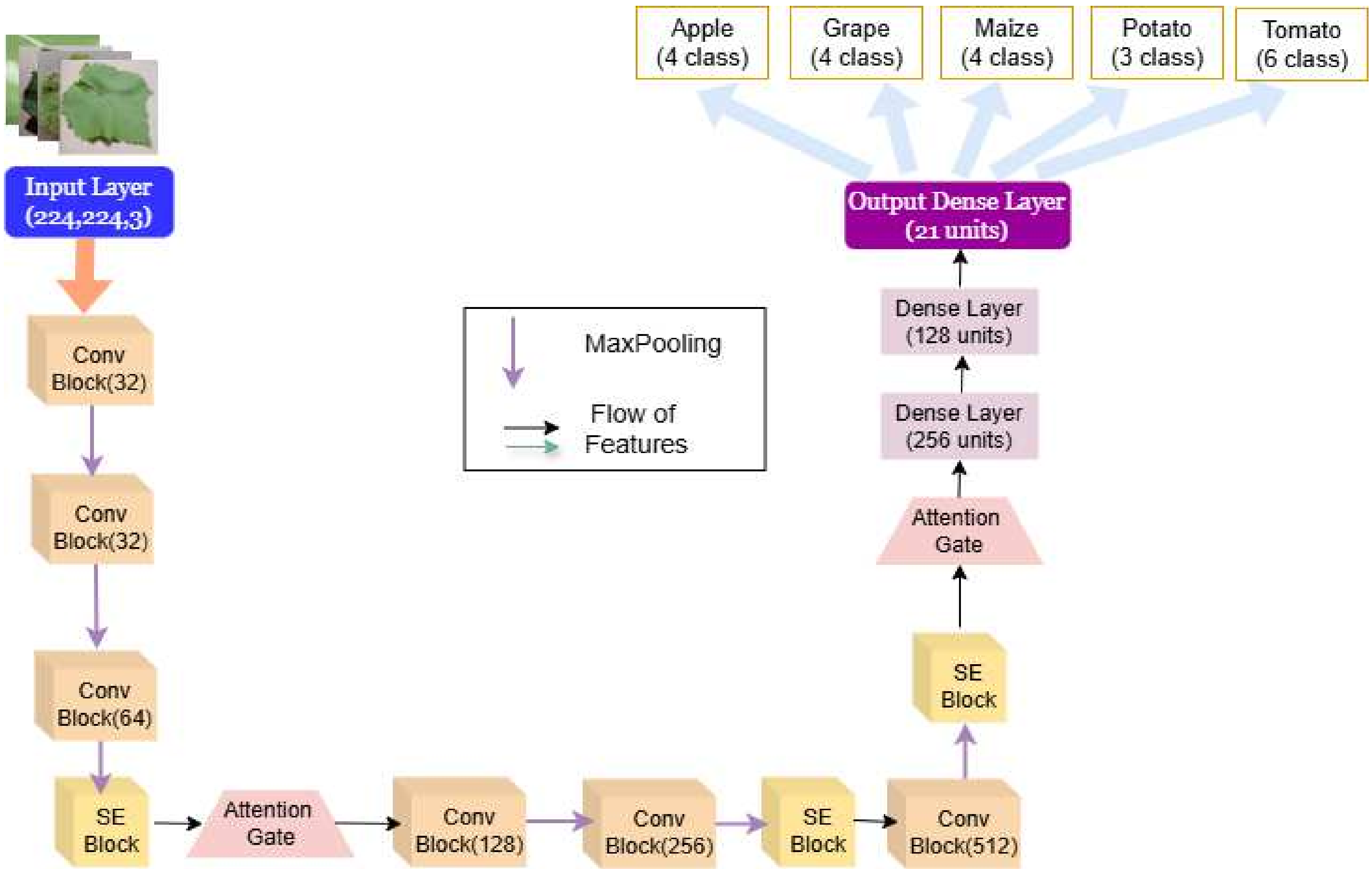


Fig. 1 Block diagram of the overall architecture of the proposed model (the values in brackets indicate number of filters)

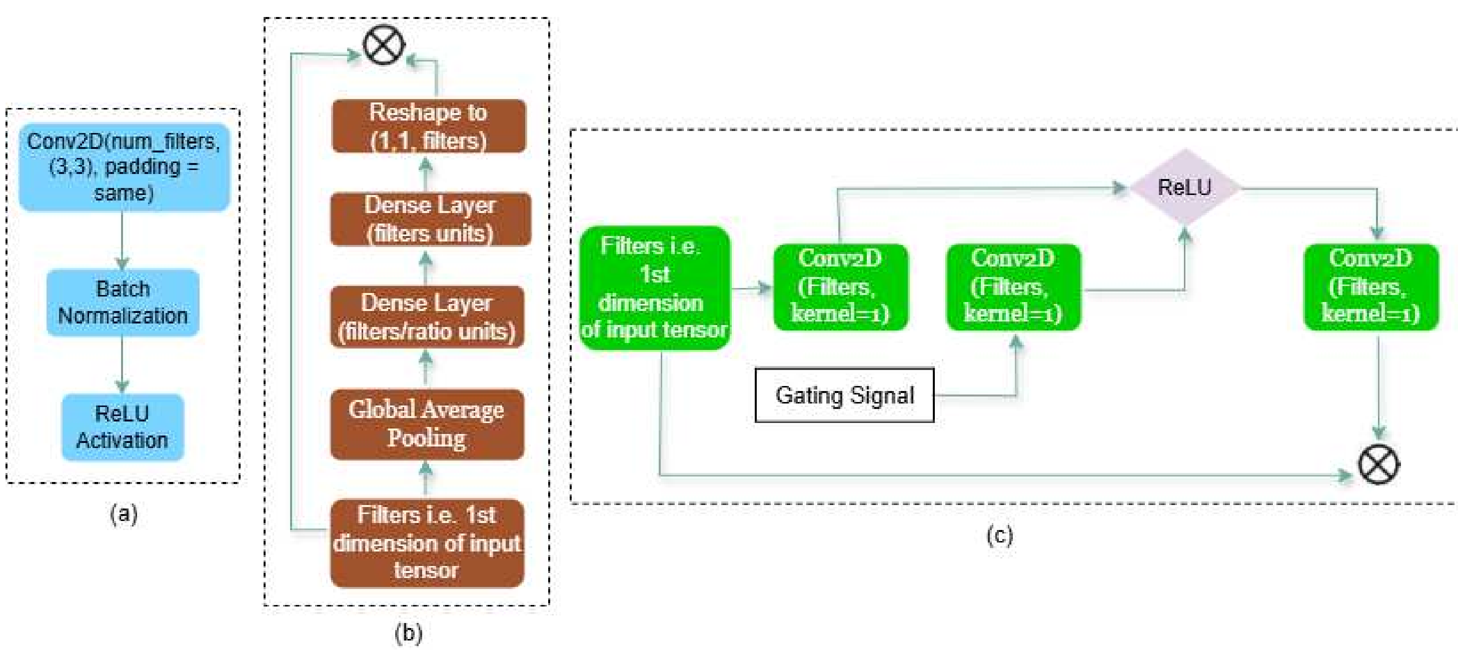


Fig. 2 Block diagram of other blocks of the proposed framework. (a) Conv block (b) SE block (c) Attention gate

Algorithm 1 outlines the sequential process for implementing the CNN model

Algorithm 1 Proposed Model Approach for Implementation

Require: Input Image $X \in R^{H \times W \times C}$

Ensure: Segmentation Mask $Y \in R^{H \times W}$

Initialize: Encoder weights, Positional Embedding, Attention Weights

Step 1: Encoder Path

for $i = 1$ to L **do**

  $X_i \leftarrow$ Apply two convolution layers with Batch Normalization and ReLU

  $X_i \leftarrow$ Apply Squeeze-and-Excitation Block

  if $i < L$ **then**

    $X_i \leftarrow$ MaxPooling operation

  end if

end for

Step 2: Position Embedding

$X_{pos} \leftarrow$ Apply Positional Embedding to last encoder feature map

Step 3: Spatial Attention

Compute Query (Q), Key (K), and Value (V) matrices using learned projections

Compute attention weights using Scaled Dot-Product Attention

Apply attention to update feature representations

$X_{att} \leftarrow$ Weighted sum of values based on attention scores

Step 4: Bottleneck

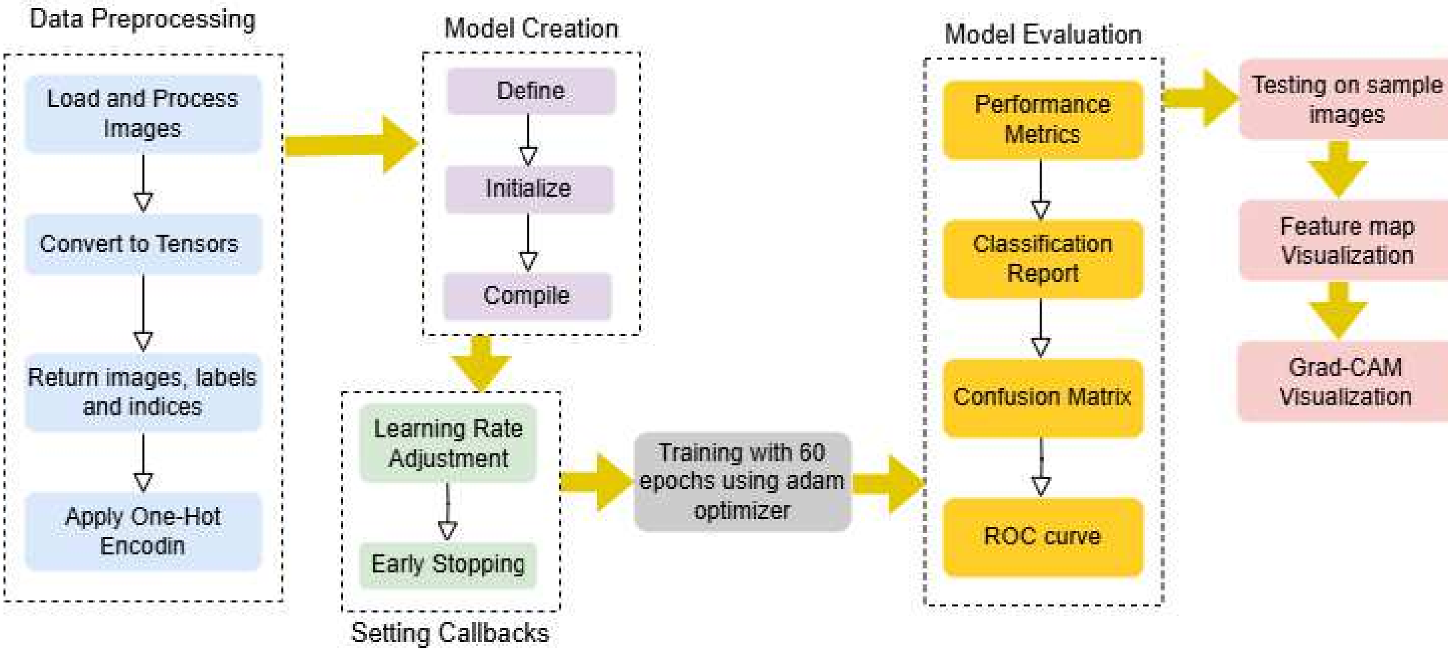
$Xbottleneck \leftarrow Xpos + Xatt$ (Residual Connection)

Step 5: Decoder Path

Perform upsampling and concatenation with corresponding encoder features

Apply convolutional layers and activation functions

Generate final segmentation mask Y



'ig. 3 Workflow of the proposed deep learning framework for plant disease classification. The pipeline consists of four nain stages: data preprocessing, model creation with attention mechanisms (SE blocks and AGs), model training using ›ptimized callbacks, and comprehensive model evaluation, including Grad-CAM-based interpretability

3.2 Dataset

The dataset used in this study is a curated subset of the publicly available PlantVillage dataset [21], originally hosted on Kaggle. The complete dataset contains over 54,000 images spanning multiple crop species and disease categories. For this research, we extracted and utilized images from five major crops—potato, tomato, grape, maize, and apple—due to their agricultural importance and rich class diversity.

Each crop includes images of both healthy and diseased leaves, with diseases such as early and late blight in potatoes; early and late blight as well as bacterial and septoria leaf spots in tomatoes; leaf blight, black rot and esca black measles in grapes; common rust, northern leaf blight and cercospora leaf spot in maize; and cedar rust and scab in apples. The selected subset covers 21 distinct classes, ensuring a balanced representation of disease types across species.

All images are RGB, high-resolution, and captured under controlled lighting conditions with plain backgrounds, which simplifies preprocessing and enhances the model's focus on leaf-specific features. The dataset was split into 80% for training, 18% for validation, and 2% for testing, maintaining class balance throughout.

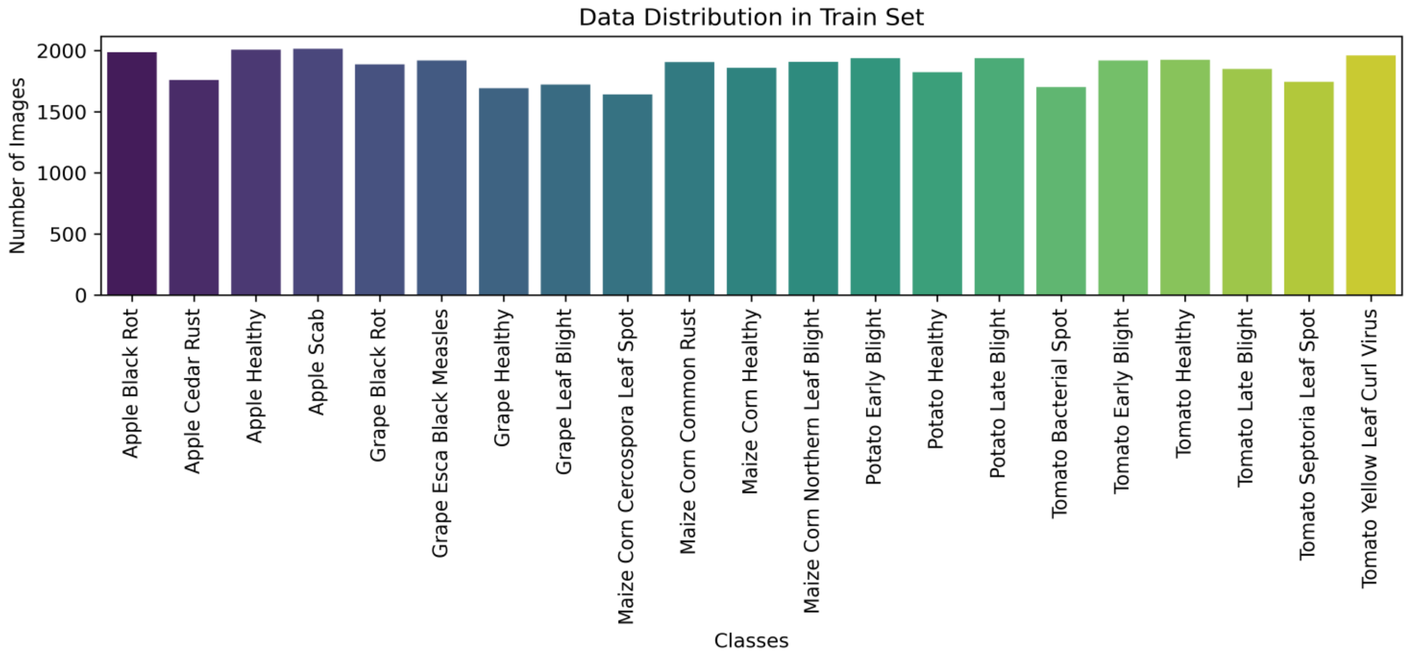


Fig. 4 Class-wise distribution of training images for the selected 21 plant diseases

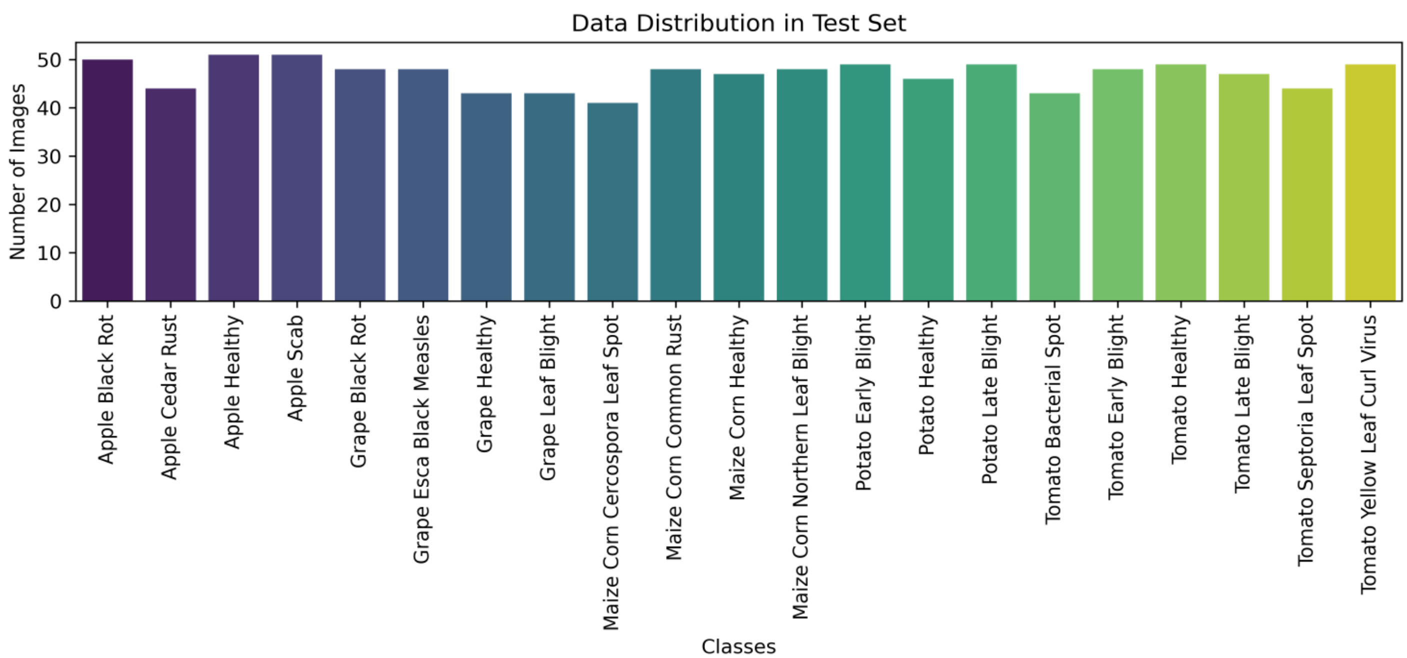


Fig. 5 Class-wise distribution of test images for the selected 21 plant diseases

Figures 4,5 and 6 show the distribution of images across the 21 selected plant disease and healthy classes. The dataset exhibits relatively balanced representation, with each class containing between approximately 1600 and 2100 images

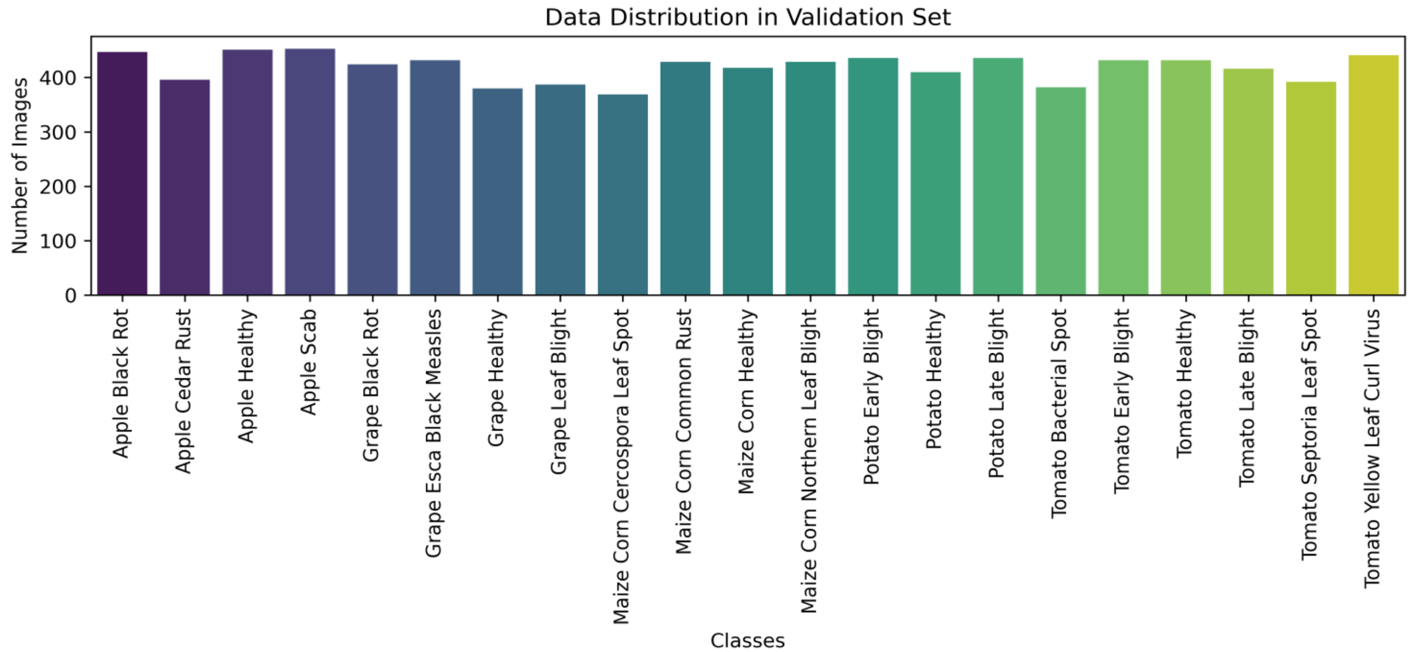


Fig. 6 Class-wise distribution of validation images for the selected 21 plant diseases

3.3 Implementation details

Table 1 Summary of proposed model's hyperparameters settings

Hyperparameters	**Values**
Optimizer	Adam
Learning Rate (lr)	1e-3 (0.001)
Batch Size	16
Learning Rate Reduction	factor=0.2, patience=6, min_lr=1e-6
Early Stopping	Patience = 8, min_delta = 0.001, restore_best_weights=True
Epochs	60

The model was implemented using the Keras API and trained on two Tesla T4 GPUs (15 GB RAM each). Input images were resized to 256 × 256 pixels to maintain consistency with prior studies. The Adam optimizer was selected for efficient weight updates, with an initial learning rate of 0.001. The training was conducted over 60 epochs, utilizing early stopping to prevent overfitting. The weight update rule for Adam optimizer is given below in equations 1 and 2. The key hyperparameter configurations employed during training, including optimizer type, learning rate, batch size, and epoch count, are summarized in Table 1.

$$m_t = \beta_1 m_t - 1 + (1-\beta_1) g_t \qquad v_t = \beta_2 v_{(t-1)} + (1-\beta_2) g_t 2 \tag{1}$$

$$\widehat{m_t} = \frac{m_t}{1-\beta_1^t}, \widehat{v_t} = \frac{v_t}{1-\beta_2^t}, \; w_t = w_{t-1} - \frac{\eta}{\sqrt{\widehat{v_t}}+\epsilon}\widehat{m_t} \tag{2}$$

where: g_t is the gradient; $\beta_1$and β_2 are momentum parameters; η is the learning rate.

A batch size of 16 was employed, adapting to distributed training. Learning rate reduction (ReduceLROnPlateau, factor = 0.2, patience = 6, min_lr = 1e-6) was applied to dynamically adjust learning rates when validation loss plateaued. For validation, 15% of the training data was randomly selected, ensuring a robust assessment of model generalization.

3.4 Performance metrics

The assessment of the model's performance primarily relied on accuracy, a widely recognized metric in classification tasks. In addition, Precision, Recall, F1 score, and Specificity were calculated, all derived from true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values to provide a comprehensive evaluation of the classification outcomes. A confusion matrix and ROC curve were also employed in this analysis002E

Accuracy: The measurement of accuracy is utilized to assess the effectiveness of the proposed method. Classification accuracy is defined as in equation 6:

$$Accuracy = (TN + TP)/(TN + FP + TP + FN) \tag{3}$$

In this equation, True Positive (TP) refers to data that is accurately labeled and classified, True Negative (TN) pertains to data that is correctly predicted, False Positive (FP) indicates MR images that have been incorrectly classified, and False Negative (FN) denotes data that has been wrongly classified as negative.

Precision: As defined in equation 7, precision quantifies the ratio of correctly predicted positive instances to the total number of predicted positive instances. It reflects the accuracy of positive predictions made by the model.

$$Precision = TP / (TP + FP) \tag{4}$$

Recall: Recall measures the proportion of correctly predicted positive instances against the total number of actual positive instances. It indicates the model's effectiveness in identifying all positive instances and is depicted in equation 8 :

$$Recall = TP / (TP + FN) \tag{5}$$

F1 Score: The F1 score, which is the harmonic mean of precision and recall, is calculated as shown in equation 9 and provides a comprehensive assessment of a model's performance.

$$F1\,Score = 2 * (Precision * Recall) / (Precision + Recall) \tag{6}$$

Specificity: This metric represents the proportion of actual negatives that are accurately identified, serving as an indicator of a model's capability to minimize false positives. It is calculated as follows:

$$Specificity = TN / (TN + FP) \tag{7}$$

3.5 Loss function

The proposed model is trained using the Categorical Cross-Entropy (CCE) loss function, which is widely used for multi-class classification tasks. It measures the dissimilarity between the true class labels and predicted probabilities. The CCE loss can be defined as in equation 11:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{i,j}\log\widehat{y_{\iota,J}} \qquad (8)$$

where $y_{i,j}$ is the one-hot encoded ground truth, and $\widehat{y_{\iota,J}}$ is the predicted probability for the class j of sample i. N is the total number of samples, and C is the number of classes. This loss penalizes incorrect predictions, ensuring that the model assigns higher confidence to the correct class.

4. Results and discussion

4.1 Experimental results

Figure 7 presents the training and validation performance of the proposed attention-enhanced CNN model across 28 epochs. The model demonstrates rapid convergence within the first 5 epochs across all key metrics. Training and validation loss decrease sharply, stabilizing near zero, indicating effective learning and minimal overfitting.

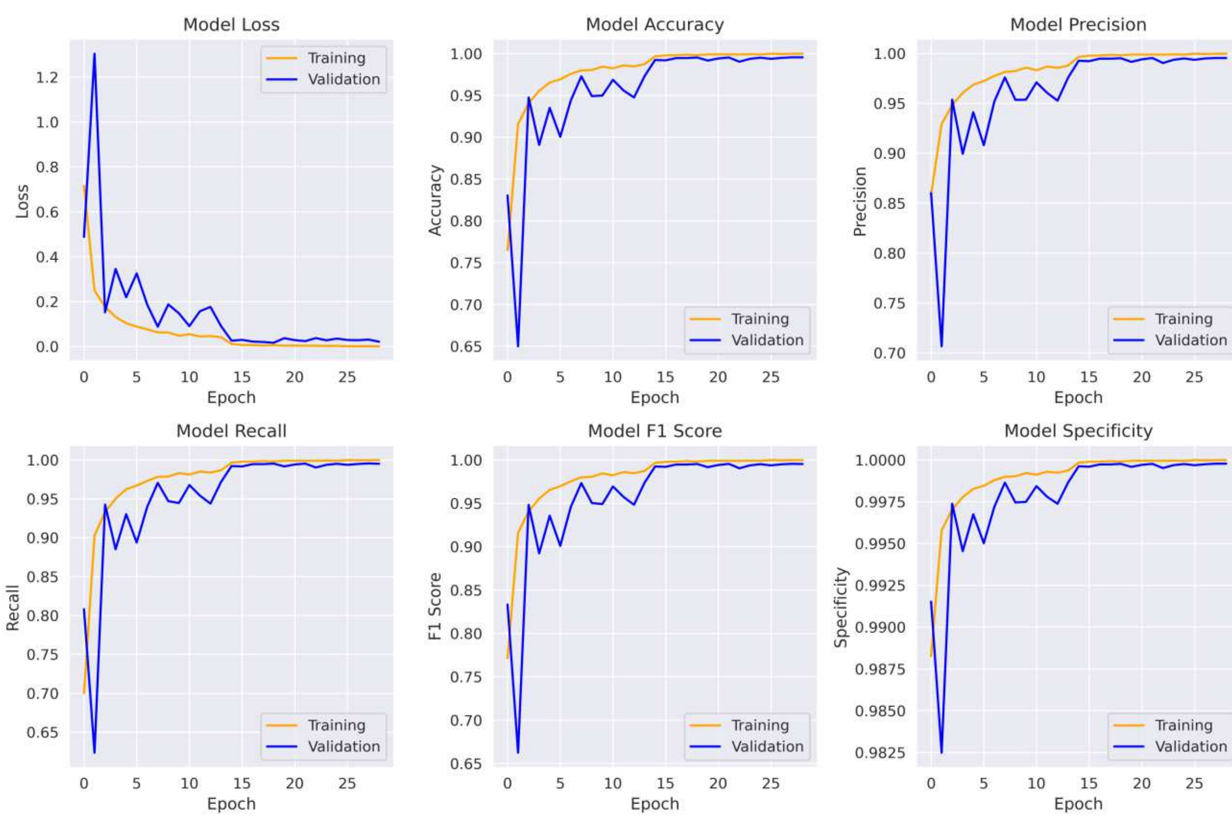


Fig. 7 Training and validation curves of the proposed CNN model across the epochs

The accuracy consistently exceeds 99% for both training and validation sets, confirming strong generalization across all classes. Similar trends are observed in precision, recall, and F1-score, all converging above 98%, reflecting the model's balanced performance in handling false positives and false negatives.

The specificity curve remains high throughout, exceeding 99.8%, suggesting the model effectively avoids misclassifying healthy leaves as diseased.

These results affirm the effectiveness of the integrated SE blocks and Attention Gates (AGs), which enhance the model's focus on critical spatial and channel-level features, ultimately improving classification robustness and interpretability.

Table 2 Classification report of apple diseases

Class	**precision**	**recall**	**f1-score**	**support**
Black Rot	1	1	1	50
Cedar Rust	1	0.977273	0.988506	44
Healthy	1	1	1	51
Scab	1	0.980392	0.990099	51
micro avg	1	0.989796	0.994872	196
macro avg	1	0.989416	0.994651	196
weighted avg	1	0.989796	0.994843	196

Table 3 Classification report of grape diseases

Class	**precision**	**recall**	**f1-score**	**support**
Black Rot	1	1	1	48
Esca Black Measles	1	1	1	48
Healthy	1	1	1	43
Leaf Blight	1	1	1	43
accuracy	1	1	1	1
macro avg	1	1	1	182
weighted avg	1	1	1	182

4.1.1 Apple Disease Classification Report

The model demonstrates exceptional performance in classifying apple leaf diseases. As shown in the classification report of Table 2, the model achieved perfect precision, recall, and F1-score (1.00) for Black Rot, Healthy, and Scab, indicating flawless identification with zero false positives or false negatives in these categories.

For Cedar Rust, the model maintained 100% precision, but recall slightly dropped to 0.977, leading to an F1-score of 0.9885. This suggests that a small fraction of Cedar Rust instances was misclassified as another class (likely similar in appearance), though no other classes were falsely predicted as Cedar Rust.

The macro average F1-score for apple classes was 0.9946, and the micro and weighted averages were similarly high at 0.9948, indicating overall consistency and robustness in classification across all apple leaf conditions.

4.1.2 Grape Disease Classification Report

The model demonstrates outstanding performance in classifying grape leaf diseases. As shown in the classification report of Table 3, the model achieved perfect precision, recall, and F1-score (1.00) for all the diseases, indicating flawless identification with zero false positives or false negatives in these categories.

The macro and weighted average F1-score for grape classes is also perfectly 1.00, which shows overall consistency and robustness in classification across all leaf conditions.

4.1.3 Maize Corn Disease Classification Report

The classification performance on maize corn leaf diseases reflects high accuracy and reliable class-level predictions. The model correctly identified Common Rust and Healthy leaves with perfect scores (precision, recall, F1 = 1.00), demonstrating no misclassifications for these two categories, as shown in Table 4.

For Cercospora Leaf Spot, the model maintained perfect precision (1.00) but showed a slightly reduced recall of 0.9268, resulting in an F1-score of 0.9620. This indicates that a small number of Cercospora cases were not detected (i.e., false negatives), but none were incorrectly predicted as such (i.e., no false positives).

Interestingly, for Northern Leaf Blight, the model achieved perfect recall (1.00), meaning all cases were identified, yet the precision dropped to 0.9412, leading to an F1-score of 0.9697. This suggests that while all Northern Leaf Blight cases were detected, a few instances of other classes may have been wrongly labelled as this class.

At the aggregate level, the model yielded an overall accuracy of 98.37%, with macro and weighted average F1-scores of 0.9823 and 0.9836, respectively. These results underscore the model's strong generalization and ability to maintain balance across multiple maize disease classes.

Table 4 Classification report of maize corn diseases

Class	precision	recall	f1-score	support
Cercospora Leaf Spot	1	0.926829	0.962025	41
Common Rust	1	1	1	48
Healthy	1	1	1	47
Northern Leaf Blight	0.941176	1	0.969697	48
accuracy	0.983696	0.983696	0.983696	0.983696
macro avg	0.985294	0.981707	0.982931	184
weighted avg	0.984655	0.983696	0.983633	184

Table 5 Classification report of potato diseases

Class	precision	recall	f1-score	support
Early Blight	1	1	1	49
Healthy	1	1	1	46
Late Blight	1	0.979592	0.989691	49
micro avg	1	0.993056	0.996516	144
macro avg	1	0.993197	0.996564	144
weighted avg	1	0.993056	0.996492	144

Table 6 Classification report of tomato diseases

Class	precision	recall	f1-score	support
Bacterial Spot	1	1	1	43
Early Blight	1	0.979167	0.989474	48
Healthy	1	1	1	49
Late Blight	0.979167	1	0.989474	47
Septoria Leaf Spot	1	1	1	44
Yellow Leaf Curl Virus	1	1	1	49
accuracy	0.996429	0.996429	0.996429	0.996429
macro avg	0.996528	0.996528	0.996491	280
weighted avg	0.996503	0.996429	0.996429	280

4.1.4 Potato Disease Classification Report

The classification performance on potato leaf diseases reflects high accuracy and reliable class-level predictions. The model correctly identified Early Blight and Healthy leaves with perfect scores (precision, recall, F1 = 1.00), demonstrating no misclassifications for these two categories, as shown in Table 5.

For Late Blight, the model maintained perfect precision (1.00) but showed a slightly reduced recall of 0.9795, resulting in an F1-score of 0.9896. This indicates that a small number of cases were not detected (i.e., false negatives), but none were incorrectly predicted as such (i.e., no false positives).

At the aggregate level, the model yielded the same micro, macro and weighted average F1-scores of 0.9965. These results underscore the model's strong generalization and ability to maintain balance across multiple potato disease classes.

4.1.5 Tomato Disease Classification Report

The classification performance on tomato leaf diseases reflects high accuracy and reliable class-level predictions. The model correctly identified Bacterial and Septoria Leaf Spot, Yellow Leaf Curl Virus, and Healthy leaves with perfect scores (precision, recall, F1 = 1.00), demonstrating no misclassifications for these two categories, as shown in Table 6.

For Early and Late Blight, the model maintained precisions of 1.00 and 0.9792 but showed the reverse recall of 0.9792 and 1.00, resulting in an F1-score of 0.9895 for both diseases. This indicates that a small number of cases were not detected (i.e., false negatives).

At the aggregate level, the model yielded an impressive overall accuracy of 99.64%, with macro and weighted average F1-scores of 0.9965 and 0.9964, respectively. These results underscore the model's strong generalization and ability to maintain balance across multiple tomato disease classes.

Figure 8 presents the confusion matrices for apple, maize, grape, potato, and tomato diseases, reflecting the model's performance across 21 distinct classes. The matrices confirm the high classification accuracy reported in the evaluation metrics, with most predictions aligning perfectly with true labels.

- For apple diseases, the model shows perfect predictions for Black Rot, Healthy, and Scab, while Cedar Rust has minimal misclassification.
- In maize corn, Cercospora Leaf Spot exhibits a few misclassifications into Northern Leaf Blight, suggesting some inter-class similarity. However, Common Rust, Healthy, and Northern Leaf Blight are classified with perfect accuracy.

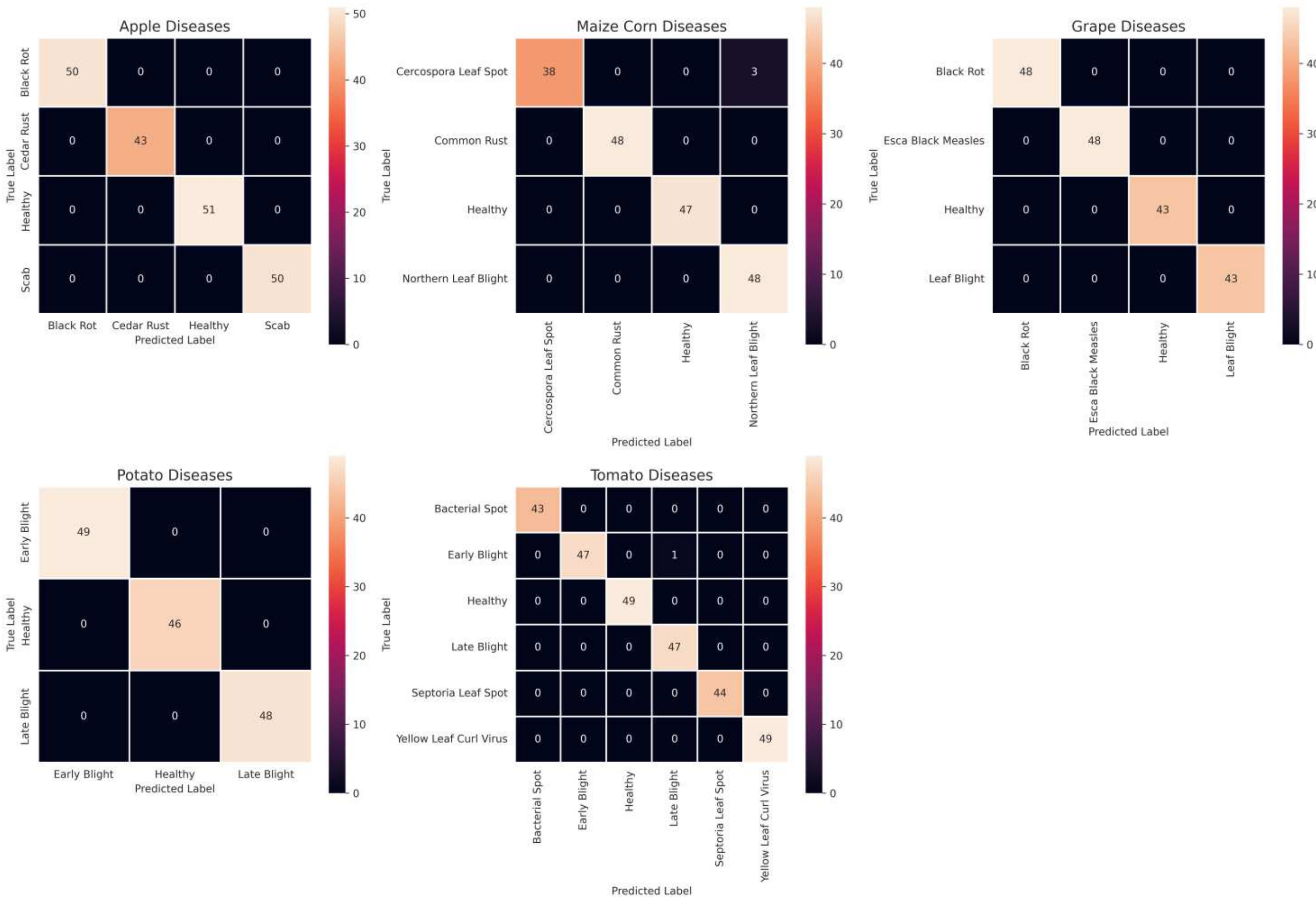


Fig. 8 Confusion matrix heatmaps of the proposed model for various plant leaf diseases.

- Grape disease classification shows 100% accuracy across all four categories, with no observed false positives or false negatives.
- For potato, all three classes—Early Blight, Healthy, and Late Blight—are predicted correctly, except for one instance of Early Blight misclassified as Healthy.
- The tomato confusion matrix shows near-perfect results across six classes, with a single misclassification of Early Blight as Healthy. All other classes, including Yellow Leaf Curl Virus and Bacterial Spot, are perfectly identified.

Overall, these matrices indicate that the proposed attention-enhanced CNN model maintains excellent class discrimination with minimal confusion, even among visually similar diseases. This confirms the model's robustness and highlights its potential for real-world application in crop disease diagnostics. Figure 9 illustrates a selection of correctly predicted samples, showcasing the model's accuracy and confidence across various plant disease classes.

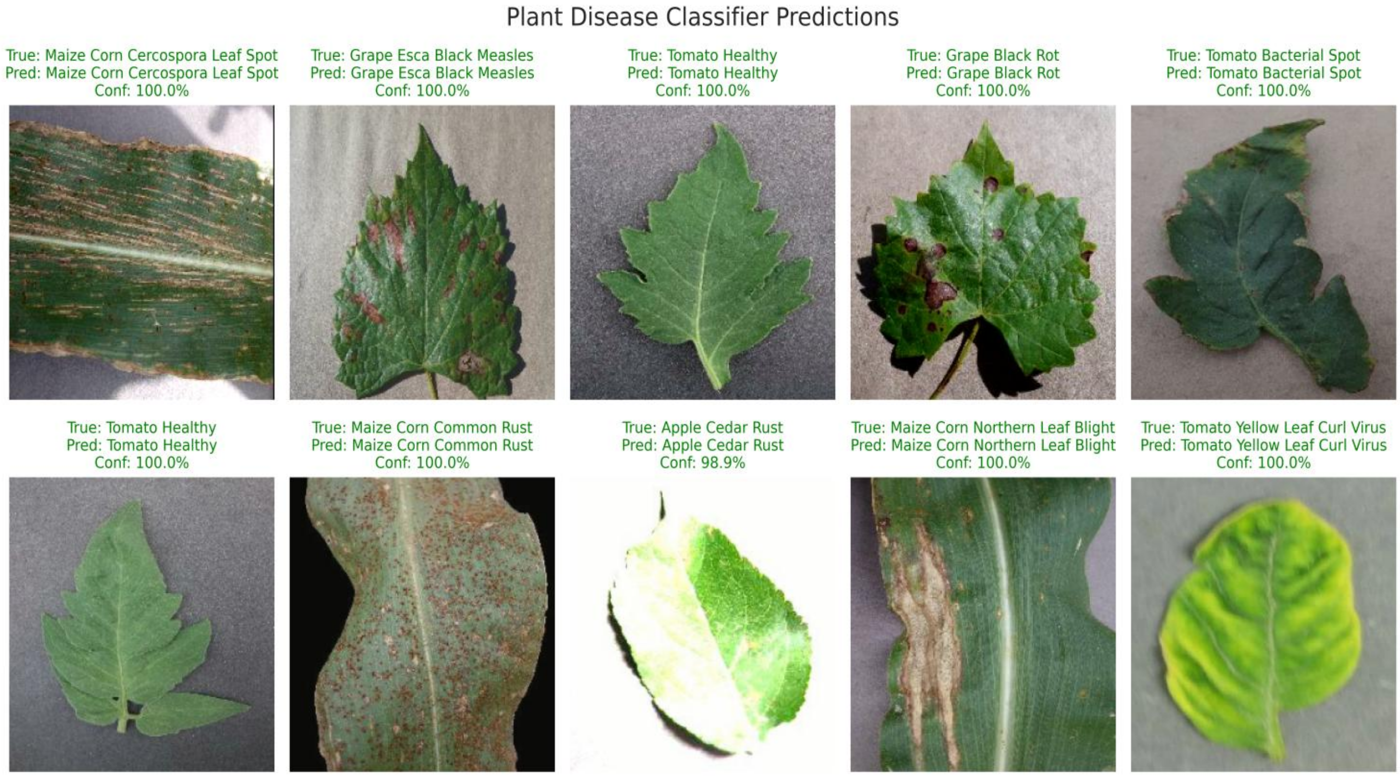


Fig. 9 Sample visualization of correctly classified leaf images with predicted labels and model confidence scores across multiple crop disease categories.

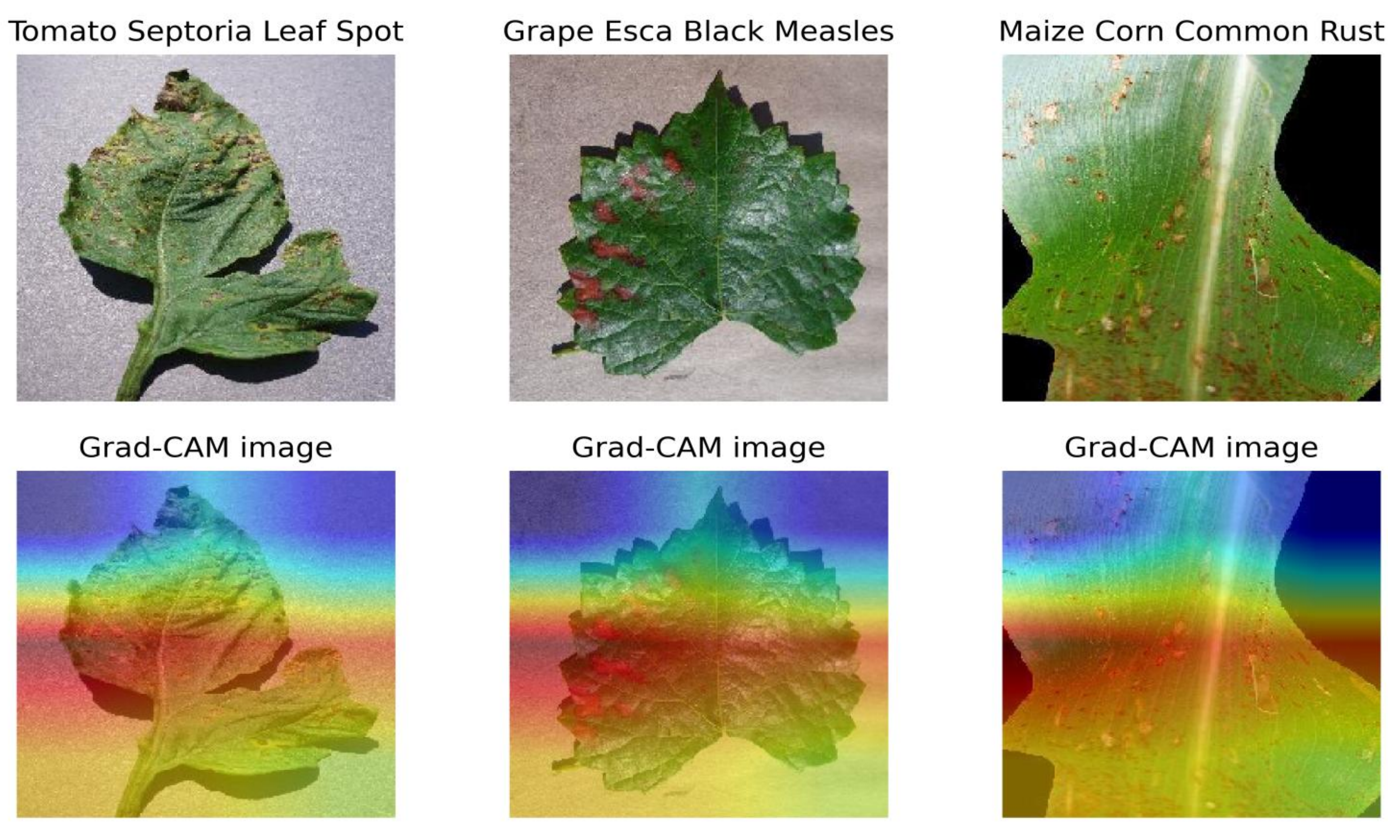


Fig. 10 Grad-CAM visualization of some sample leaf images

4.1.6 Grad-CAM Visualization Analysis

Figure 10 presents Grad-CAM visualizations for three sample predictions—Tomato Septoria Leaf Spot, Grape Esca Black Measles, and Maize Corn Common Rust. Each row displays the original input image followed by the corresponding Grad-CAM overlay, highlighting the regions that most influenced the model's classification decision.

In all three cases, the Grad-CAM heatmaps reveal that the model successfully attends to the disease-affected regions of the leaf surfaces. The red and yellow areas indicate higher activation, corresponding precisely to visible symptoms such as lesions, discoloration, or texture changes. For instance:

- In the Tomato Septoria Leaf Spot image, the model focuses on the dense cluster of dark spots across the leaf's midsection.
- In the Grape Esca Black Measles case, attention is drawn to the reddish patches, which are key indicators of the disease.
- In Maize Corn Common Rust, the model accurately highlights the linear brown pustules running along the leaf veins.

These visualizations validate the model's interpretability and provide confidence that predictions are not arbitrary but grounded in meaningful visual cues. Grad-CAM thus enhances trust in the model's decision-making and supports its potential deployment in real-world agricultural diagnostics.

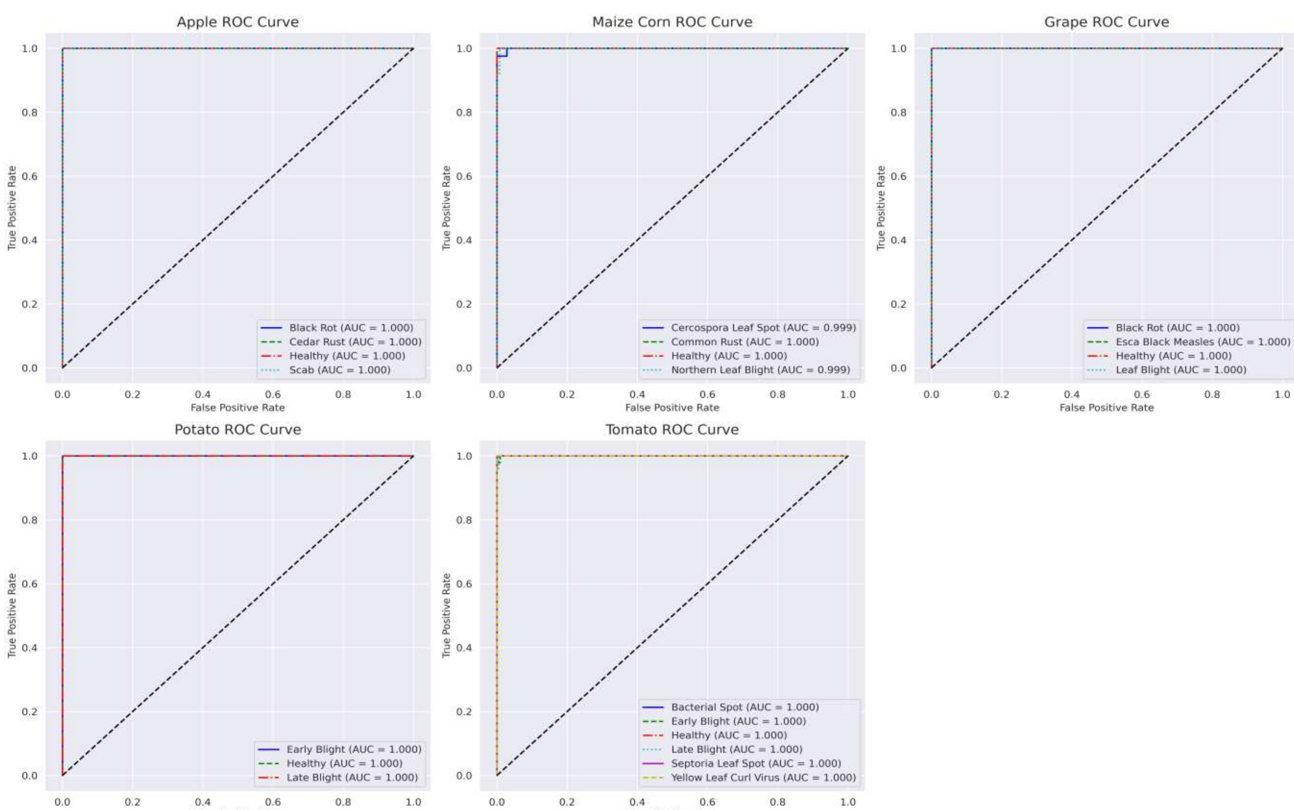


Fig. 11 ROC curves for disease classification across five plant species: apple, maize corn, grape, potato, and tomato

4.1.7 ROC Curve Analysis

Figure 11 illustrates the Receiver Operating Characteristic (ROC) curves for each plant category—apple, maize, grape, potato, and tomato—evaluating the performance of the proposed CNN classifier across all disease and healthy classes.

Across all plant types, the ROC curves exhibit near-ideal behaviour, with curves sharply rising toward the top-left corner, indicative of high sensitivity and low false positive rates. The Area Under the Curve (AUC) values for most classes reach the maximum of 1.000, signifying perfect classification capability.

- For apple, all four classes (Black Rot, Cedar Rust, Healthy, and Scab) achieve AUC = 1.000, confirming complete separability between classes.
- In maize, the model shows almost perfect classification, with Cercospora Leaf Spot and Northern Leaf Blight yielding AUC scores of 0.999, while the remaining classes attain a full 1.000.
- The grape disease classifier achieves perfect AUC scores for all classes, highlighting the model's strong generalization across fungal and healthy conditions.
- Similarly, the model delivers AUC = 1.000 across all potato and tomato disease categories, demonstrating robust discrimination for multiple disease types, including complex classes like Yellow Leaf Curl Virus and Septoria Leaf Spot.

These ROC results further validate the model's effectiveness beyond standard metrics like accuracy and F1-score, reinforcing its capability to perform consistently across diverse crop and disease categories with minimal overlap or ambiguity.

4.2. Performance comparison with other techniques

4.3 Misclassified cases in the plant leaf disease detection

Figure 12 highlights a selection of misclassified test samples, offering valuable insight into the limitations and edge cases of the proposed model. Despite high overall accuracy, a few instances were incorrectly labelled, often between visually similar disease classes or across crops with overlapping symptom features.

For example, Potato Late Blight was misclassified as Tomato Late Blight with a confidence of 60.1%, likely due to similar lesion patterns. Similarly, Cercospora Leaf Spot in maize was misclassified as Northern Leaf Blight and vice versa, both with confidence exceeding 84%, indicating inter-class visual similarity.

A more challenging case involved Apple Cedar Rust being predicted as Potato Healthy, potentially due to high brightness and lack of distinct lesion colour contrast. The model also confused Tomato Bacterial Spot with Yellow Leaf Curl Virus, demonstrating the difficulty in distinguishing disease-specific textural features on leaves with variable damage patterns.

Fig. 12 Examples of misclassified test images along with true and predicted labels and the model's confidence scores.

These errors emphasize the importance of incorporating more diverse samples, improving illumination normalization, and potentially expanding model interpretability tools. Nonetheless, the confidence scores in these misclassifications are generally lower than correct predictions, aligning with the model's uncertainty in borderline cases.

5. Conclusion

In this study, we presented a deep learning framework integrating Squeeze-and-Excitation (SE) blocks and Attention Gates (AGs) for multi-class classification of plant leaf diseases across five major crops. The model demonstrated excellent performance across evaluation metrics, achieving high accuracy, precision, recall, and interpretability through Grad-CAM visualizations. The use of attention mechanisms significantly enhanced the model's ability to focus on disease-specific features, leading to robust generalization across diverse disease classes.

Future work will focus on expanding the dataset to include real-field images under varying environmental conditions, improving generalization in uncontrolled settings. Additionally, deploying the model in mobile or edge devices for real-time disease diagnosis, integrating multi-modal data such as weather or soil parameters, and exploring self-supervised or few-shot learning techniques could further strengthen its applicability in practical agricultural scenarios.

Declaration

Funding:

The authors did not receive support from any organization for the submitted work

Competing Interest:

The authors have no relevant financial or non-financial interests to disclose.

References

[1] Secretariat, I. P. P. C. (2023). International Plant Health Conference-Report.

[2] Rani, K. A., & Gowrishankar, S. (2023). Pathogen-based classification of plant diseases: A deep transfer learning approach for intelligent support systems. *IEEE Access*, *11*, 64476-64493.

[3] Salman, Z., Muhammad, A., Piran, M. J., & Han, D. (2023). Crop-saving with AI: latest trends in deep learning techniques for plant pathology. *Frontiers in Plant Science*, *14*, 1224709.

[4] Jung, M., Song, J. S., Shin, A. Y., Choi, B., Go, S., Kwon, S. Y., ... & Kim, Y. M. (2023). Construction of deep learning-based disease detection model in plants. *Scientific reports*, *13*(1), 7331.

[5] Gupta, D., Sinha, A., & Kumar, H. (2023, June). Detection and classification of pathogens causing various plant diseases using supervised machine learning approaches based on hybrid feature sets. In *AIP Conference Proceedings* (Vol. 2782, No. 1). AIP Publishing.

[6] Berwo, M. A., Khan, A., Fang, Y., Fahim, H., Javaid, S., Mahmood, J., ... & MS, S. (2023). Deep learning techniques for vehicle detection and classification from images/videos: A survey. *Sensors*, *23*(10), 4832.

[7] Ji, M., Zhang, L., & Wu, Q. (2020). Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks. *Information Processing in Agriculture*, *7*(3), 418-426.

[8] Ngugi, L. C., Abelwahab, M., & Abo-Zahhad, M. (2021). Recent advances in image processing techniques for automated leaf pest and disease recognition–A review. *Information processing in agriculture*, *8*(1), 27-51.

[9] Umamageswari, A., Bharathiraja, N., & Irene, D. S. (2023). A novel fuzzy C-means based chameleon swarm algorithm for segmentation and progressive neural architecture search for plant disease classification. *ICT Express*, *9*(2), 160-167.

[10] Paul, S. G., Biswas, A. A., Saha, A., Zulfiker, M. S., Ritu, N. A., Zahan, I., ... & Islam, M. A. (2023). A real-time application-based convolutional neural network approach for tomato leaf disease classification. *Array*, *19*, 100313.

[11] Hossain, M. I., Jahan, S., Al Asif, M. R., Samsuddoha, M., & Ahmed, K. (2023). Detecting tomato leaf diseases by image processing through deep convolutional neural networks. *Smart Agricultural Technology*, *5*, 100301.

[12] Uddin, M. S., & Bansal, J. C. (Eds.). (2022). *Computer Vision and Machine Learning in Agriculture, Volume 2*. Springer Singapore.

[13] Krisnandi, D., Pardede, H. F., Yuwana, R. S., Zilvan, V., Heryana, A., Fauziah, F., & Rahadi, V. P. (2019). Diseases classification for tea plant using concatenated convolution neural network. *CommIT (Communication and Information Technology) Journal*, *13*(2), 67-77.

[14] Mustofa, S., Munna, M. M. H., Emon, Y. R., Rabbany, G., & Ahad, M. T. (2023). A comprehensive review on Plant Leaf Disease detection using Deep learning. *arXiv preprint arXiv:2308.14087*.

[15] Kiran, S. M., & Chandrappa, D. N. (2023). Plant leaf disease detection using efficient image processing and machine learning algorithms. *Journal of Robotics and Control (JRC)*, *4*(6), 840-848.

[16] Imanulloh, S. B., Muslikh, A. R., & Setiadi, D. R. I. M. (2023). Plant diseases classification based leaves image using convolutional neural network. *Journal of Computing Theories and Applications (JCTA)*, *1*(1), 1-10.

[17] Malik, A., Vaidya, G., Jagota, V., Eswaran, S., Sirohi, A., Batra, I., ... & Asenso, E. (2022). Design and evaluation of a hybrid technique for detecting sunflower leaf disease using deep learning approach. *Journal of Food Quality*, *2022*(1), 9211700.

[18] Alharbi, A., Khan, M. U. G., & Tayyaba, B. (2023). Wheat disease classification using continual learning. *IEEE Access*, *11*, 90016-90026.

[19] Wu, W. (2023). Plant leaf diseases classification based on few-shot learning. In *Academic Journal of Science and Technology* (Vol. 5, No. 2).

[20] Rezaei, M., Diepeveen, D., Laga, H., Jones, M. G., & Sohel, F. (2024). Plant disease recognition in a low data scenario using few-shot learning. *Computers and electronics in agriculture*, *219*, 108812.

[21] Plant Village Dataset – A dataset of diseased plant leaf images and corresponding labels. https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset/data