Journal of Agricultural Sciences (JAS) 32(3):759-775
DOI: 10.15832/ankutbd.1738222

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

Deep Transformer-Based Visual Framework for Early Detection of Plant Leaf Pathologies

Abhishek Mathur[a*], Shailendra Kumar Shrivastava[a]

[a]*Department of Information Technology, Samrat Ashok Technological Institute, Vidisha, M.P, INDIA*

ARTICLE INFO
Research Article
Corresponding Author: Abhishek Mathur, E-mail: abhishekmathur1185@gmail.com
Received: 09 July 2025 / Revised: 13 January 2026 / Accepted: 03 February 2026 / Online: 28 July 2026

Cite this article
Mathur A, Shrivastava S K (2026). Deep Transformer-Based Visual Framework for Early Detection of Plant Leaf Pathologies. *Journal of Agricultural Sciences (Tarim Bilimleri Dergisi)*, 32(3):759-775. DOI: 10.15832/ankutbd.1738222

ABSTRACT
Accurate and quick detection of plant leaf diseases is essential for precision agriculture to intervene promptly and boost crop yields. A new deep learning model called ResVNet has been introduced in this study. It combines the powerful local feature detection of ResNet152 with the global attention capabilities of Vision Transformer (ViT) and utilises Low-Rank Adaptation (LoRA) to accelerate fine-tuning. The PlantVillage dataset, which contains both healthy and diseased tomato samples, was used to train and test ResVNet. Experimental evaluation on the PlantVillage tomato dataset using stratified 5-fold cross-validation demonstrates that the proposed ResVNet model achieves a mean classification accuracy of 97.45%, along with superior macro-precision, macro-recall, and macro-F1 scores compared to existing deep learning architectures. The results of the confusion matrix and the ROC analysis validate its discriminatory power. The results highlight the potential of architectures strengthened with transformers in agricultural diagnostics. For real-time disease detection in the field, ResVNet is perfect for edge device deployment on drones and smartphones thanks to its high accuracy and adaptability. The application of Explainable AI (XAI) technologies for interpretability, integration with the Internet of Things (IoT), and multi-crop classification will all be explored in future studies. We will also look into model compression approaches so we can deploy efficiently in low-resource settings without sacrificing performance.

Keywords: Tomato Leaf Disease, Deep Learning, ResVNet, Vision Transformer (ViT), ResNet152, Low-Rank Adaptation (LoRA), PlantVillage Dataset, Image Classification, Smart Agriculture, Convolutional Neural Networks (CNN), Plant Disease Detection

1. Introduction

Agricultural biodiversity is essential to human civilization because it provides food and raw resources. The disease may be caused by pathogenic organisms such as fungi, bacteria, and nematodes, as well as soil pH, temperature extremes, moisture changes, air humidity, and other conditions that harm plants (Demilie 2024). Diseases represent a disruption to the normal condition of the plant, altering or interfering with essential functions such as photosynthesis, transpiration, pollination, fertilization, and germination. These diseases arise from pathogens such as fungi, bacteria, and viruses, as well as from unfavourable environmental conditions. Thus, early diagnosis of plant diseases by leaf analysis is an absolutely vital task. Farmers depend on constant supervision by experts, but it may be quite expensive and time-consuming. Therefore, the search for a quick, affordable, and exact method to automatically detect illnesses depending on the symptoms shown on plant leaves has great pragmatic value (Gavhale & Gawande 2014). The existence of plant diseases in the leaves reduces agricultural output. Ignorance of plant diseases early on will cause food insecurity to grow. Successful prevention and treatment of plant diseases depend on early identification of problems in leaves, which greatly affects agricultural output plans and decision-making procedures (L. Li et al. 2021).. It is almost impossible for people to evaluate every single plant, as the procedure of spotting and protecting crops against diseases calls for much effort and time. As such, research on the integration and application of creative technologies for the effective detection of illnesses in plant leaves has been conducted; currently, studies focusing on the use of artificial intelligence (AI) for diagnosing plant diseases in leaves are ongoing. The continuous development of improved classification models, particularly those for plant health monitoring and disease detection, can support AI-assisted decision-making systems in the field of smart agriculture. A range of investigations has been conducted to enhance the application of deep learning algorithms in disease detection, including the implementation of newly developed architectures for the automatic detection and classification of lesions in plant images (Jung et al. 2023).

Transformer-based detectors are gaining interest for their versatility in detection. DEtection TRansformer (DETR) and its variation algorithms excel in object detection. Enhanced DETR by combining multi-scale features to handle issues with overlapping leaves, different diseases, and complicated backgrounds in disease detection (Zhang et al. 2023; Li et al. 2024). In order to guarantee agricultural output and food security, it is essential to detect plant diseases. Traditional monitoring methods

are time-consuming and costly. Advances in deep learning, particularly transformer-based models like DETR, offer powerful capabilities for accurately identifying complex leaf diseases, supporting automation in smart agricultural systems.

1.1. Aim and Contributions of the Study

The primary aim of this study is to develop an accurate and efficient deep learning-based system for detecting and classifying tomato leaf diseases using image analysis. Leveraging advanced Convolutional Neural Network (CNN) architectures and a ResVNet model that integrates ResNet152 with Vision Transformer (ViT) via LoRA, the proposed system enhances diagnostic precision, aiding early disease detection and contributing to agricultural sustainability and improved crop management.

This study introduces a novel deep learning architecture, ResVNet, which synergistically combines the spatial feature extraction capabilities of ResNet152 with the global contextual understanding of the Vision Transformer (ViT). To enhance training efficiency and performance, Low-Rank Adaptation (LoRA) is integrated for optimized fine-tuning. The model was evaluated extensively against well-established architectures, including ResNet152, DenseNet121, and EfficientNetB0, using the PlantVillage dataset comprising high-resolution images across ten tomato disease categories and healthy leaves. The proposed ResVNet consistently outperformed its counterparts, achieving the highest classification accuracy along with superior precision, recall, and F1-score.

A comprehensive performance evaluation was conducted using robust metrics and visual tools such as accuracy/loss plots, confusion matrices, classification reports, and ROC-AUC curves, ensuring the model's reliability and generalizability. To support qualitative analysis, predicted outputs were showcased, highlighting both correct and misclassified disease instances. These visualizations underscored the model’s strong interpretability and classification strength. Finally, the real-world utility of ResVNet was demonstrated by its ability to facilitate automated and early detection of plant diseases, provide an efficient and extensible method for improving crop health monitoring systems and push the field of precision agriculture forward.

1.2. Organization of the paper

The structure of this paper is organized into six main sections. Section I introduces the background, motivation, and significance of utilizing transformer-driven deep learning architectures for accurate plant disease detection through leaf image analysis. Section II provides a comprehensive review of related literature, highlighting advancements in convolutional and transformer-based models for plant disease classification. Section III describes the proposed methodology in detail, including dataset overview, preprocessing techniques, the architectural design of the ResVNet model—which integrates ResNet152 with the Vision Transformer (ViT) via Low-Rank Adaptation (LoRA)—and the model training process. Section IV presents the experimental setup details, outlines the evaluation metrics, statistical significance analysis, and Ablation study of ResVNet and provides a comparative performance analysis of ResVNet against other deep learning models. Section V discusses the results and concludes by emphasizing the superior performance and diagnostic precision achieved by the proposed architecture. Finally, Section VI explores future research opportunities and the real-world applicability of this work in precision agriculture and smart farming systems.

2. Literature Review

The existing work on AI-based Plant Disease Detection through Machine Learning (ML) and Deep Learning (DL) methodologies is discussed here. This work suggests a new deep learning methodology that can efficiently classify and identify different apple leaf diseases with a small loss. Using the 'Plant Pathology 2020 - FGVC7' dataset with 1821 images belonging to Scab, Rust, Healthy, and Multiple diseases, our research initiates a thorough exploration of apple leaf health. In addition, the problem of class imbalance is avoided by using the SMOTE algorithm. To select the best model, various techniques are used, such as altering learning rate values, using various batch sizes, and using varied optimizer functions. The research covers the use of a basic CNN model and transfer models like DenseNet, InceptionV3, VGG16, ResNet50, and ResNet101. Upon very extensive testing, it is indicated that the best performing model is the DenseNet model with an impressive validation accuracy of 97.18%. The research can indicate the capacity of deep learning algorithms in revolutionizing apple disease identification, presenting farmers and agriculturists with an effective means to fight plant disease (Hana & Yassmina 2024). This research offers a new way to classify four big cotton leaf diseases: Utilizing a unique VGG16 fine-tuned deep learning model for diagnosing and assessing cotton plant diseases like Cotton Bacterial Blight, Verticillium Wilt, Fusarium Wilt, and Cotton Leaf Curl Disease. It utilized a leaf image dataset to cover an extremely wide range of signs, symptoms, and conditions to pre-train and fine-tune the model. By conducting thorough pre-training consisting of image enhancement and model adaptation for the specific task of detecting cotton leaf disease, it is realized that the VGG16 architecture can be adapted sufficiently to our scenario. The performance of the model was put through a rigorous examination with the aid of an alternate testing dataset, establishing the validity of a complete classification of 93.8% (Kaur et al. 2024). This work compares ResNet with Inception-v3, SVM in image classification based on the CIFAR-10 dataset. Our experiment results prove that ResNet performs much better than Inception-v3. ResNet scored 99.81% classification accuracy, while Inception-v3 scored 87%, and SVM scored 99.88%. We also compare the performance of the two models when trained under different hyperparameters and see that ResNet is less sensitive to adjusting these parameters compared to Inception-v3 (Hanif et al. 2024). In the current proposed study, the development of four DL models: Visual

Geometry Group VGG16, VGG19, EfficientNet-B0, and MobileNet is implemented on a subset of the Plant Village database, which consists of 7,233 images of apple and grape leaves belonging to 8 distinct classes. The evaluation is carried out using different metrics, i.e., Precision, Recall, F1-score and Specificity. The deep learning models performed exceptionally, attaining an accuracy exceeding 95%. Among all the deep learning models, MobileNet achieved the best accuracy at 99.38%. Accuracy rates of 96.75%, 98.69%, and 98.76% were achieved by EfficientNet-B0, VGG19, and VGG16, respectively (Joshi et al. 2024). In order to efficiently diagnose potato leaf diseases from photos, an advanced deep learning (DL) technology is implemented in the proposed methodology. The AutoEncoder was used for feature learning and dimensionality reduction, while the task of classification was carried out using the Xception model. The AutoEncoder-Xception model recorded 97.2% testing accuracy, better than conventional machine learning (ML) models like Random Forest (RF) (91.5%), SVM (89.3%), and KNN (87.1%). The model showed better precision, recall, and F1-score with the values ranging above 95% for all the classes (Murthy et al. 2024). Researchers are able to generate and implement effective methods for recognizing and classifying plant disease signs by means of CNN models and visualization approaches. Mostly based on leaf characteristics—that is, colour and damage—the technique. Models were trained using a 54,306-picture plant-village dataset comprising 26 infection kinds and 14 crop classifications. On this dataset, the CNNs and other machine learning models were trained and obtained an amazing accuracy rate of 95% in recognizing disease symptoms on plant leaves (Thotad et al. 2023). This paper analyses the results using the Kaggle rice leaf dataset, which includes several leaf diseases as well as healthy leaves. The CNN model that is used here sets its parameters using the Adam optimizer. Furthermore, the results show that with a learning rate of 0.0002, the proposed approach performs at its best. The model's performance was assessed using the accuracy parameter in conjunction with loss metrics. When combined with the ADAM optimizer, the CNN model achieved an accuracy of 95.72% and a loss value of 0.1630, according to the data (Saini et al. 2023).

Early diagnosis of tomato leaf diseases is critical since tomatoes are a global vegetable. The suggested approach uses tomato leaf photos for disease categorization. Bacterial Leaf Spot, Tomato Mosaic, Late Blight, Yellow Curved, and Septoria Leaf Spot affect tomatoes. The online Plant Village initiative provided the dataset. This article uses a dataset of diseased tomato leaf photos. The objective is to train a CNN model to achieve optimal performance, subsequently applying the learned weights to evaluate new tomato leaf images. The hybrid method, the fusion of VGG16 with an attention model, optimized model testing and validation weights, and had an accuracy of 95.90%. The top model with high accuracy and minimal overfitting was chosen based on performance analysis (Meeradevi et al. 2020). Table 1 shows the summary of a comprehensive comparative analysis of AI-driven plant disease detection techniques, which is categorized in the methodology and performance metrics with respect to their data set.

Table 1- Summary of Comprehensive Comparative Analysis of AI-Driven Plant Disease Detection Techniques

References	***Methodology***	***Dataset***	***Performance***	***Limitations & Future Work***
Hana et al. (2024)	CNN, Transfer Learning (DenseNet, InceptionV3, VGG16, ResNet50, ResNet101), SMOTE for dataset balancing	Plant Pathology 2020 - FGVC7 (1821 apple leaf images: Scab, Rust, Healthy, Multiple diseases)	DenseNet achieved **97.18% validation accuracy**	Limited dataset; future work involves testing on larger, real-world datasets and improving generalization.
Kaur et al. (2024)	Fine-tuned VGG16 deep learning model	Cotton leaf dataset (Bacterial Blight, Verticillium Wilt, Fusarium Wilt, Cotton Leaf Curl Disease)	93.8% accuracy	Focuses only on cotton; future research should expand to different crop diseases and hybrid models
Joshi et al. (2024)	VGG16, VGG19, EfficientNet-B0, MobileNet on leaf images	Plant Village (7233 apple and grape leaf images, 8 classes)	MobileNet **99.38%,** EfficientNet-B0 **98.76%,** VGG19 **98.69%,** VGG16 **96.75%**	Requires a more diverse dataset; future improvements include real-time disease detection models
Murthy et al. (2024)	AutoEncoder for feature extraction, Xception for classification	Plant Village (3000 potato leaf images: Healthy, Early Blight, Late Blight)	AutoEncoder-Xception achieved **97.2% accuracy,** RF **91.5%,** SVM **89.3%,** KNN **87.1%**	Needs real-world validation; future enhancements should integrate edge AI for field applications.
Thotad et al. (2023)	CNN-based model for plant disease detection	Plant Village (54,306 images of 14 crop classes, 26 diseases)	CNN achieved **95% accuracy**	More real-time testing is required; future work includes improving computational efficiency for large-scale deployment.
Saini et al. (2023)	CNN with Adam optimizer	Kaggle Rice Leaf Dataset (Healthy & diseased rice leaves)	CNN achieved **95.72% accuracy, a loss value of 0.1630**	Lacks comparison with other DL models; future work should include multi-crop classification
Meeradevi et al. (2020)	VGG16 with an attention mechanism for tomato disease detection	Plant Village (Tomato leaf images: Bacterial Leaf Spot, Yellow Curved, Late Blight, Tomato Mosaic, Septoria Leaf Spot)	The hybrid VGG16-attention model achieved **95.90% accuracy**	Limited to tomato; future research should extend to more plant species and field images

3. Methodology

The proposed methodology for plant disease classification via leaf images leverages advanced deep learning architectures, including ResNet152, DenseNet121, EfficientNetB0, and a hybrid ResVNet model. The study utilizes the Plant Village dataset, which contains labelled high-resolution images of healthy and diseased tomato leaves, to train and evaluate model performance. Image preprocessing involves resizing to 224×224 pixels, normalization using ImageNet statistics, and augmentation techniques such as Random Rotation, Random Horizontal Flip, Colour Jitter, and Random Resized Crop to enhance model generalization. The centrepiece of the architecture, ResVNet, fuses the powerful spatial feature extraction of ResNet152 with the global attention capabilities of the Vision Transformer (ViT). This hybrid model is further optimized using Low-Rank Adaptation (LoRA), enabling efficient fine-tuning through selective parameter updates. All models are trained using categorical cross-entropy loss and evaluated using key performance metrics, including accuracy, precision, recall, and F1-score. Figure 1 illustrates the comprehensive architecture of the proposed transformer-driven tomato leaf disease detection framework.

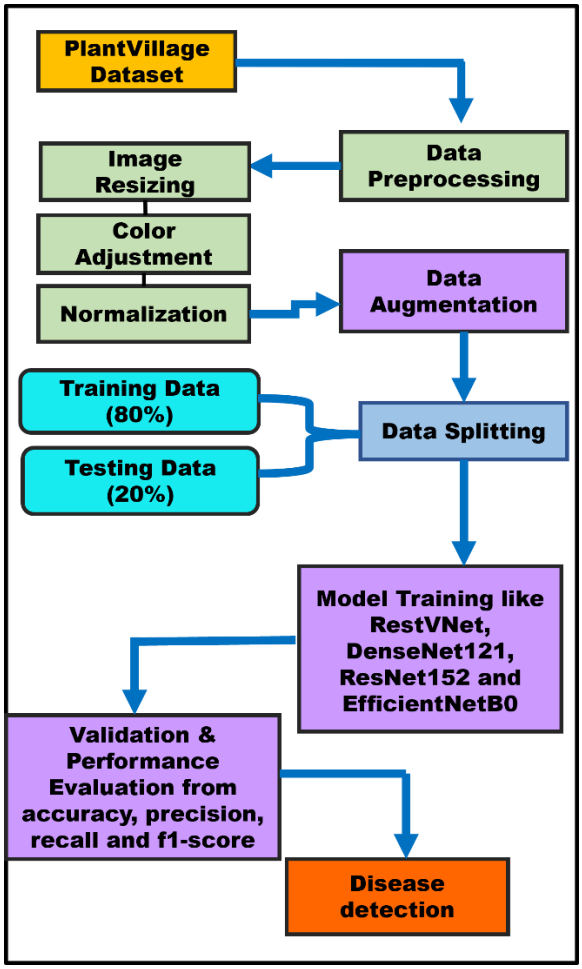


Figure 1- Data Flow Diagram of Plantvillage Disease Detection

3.1. Data collection

The PlantVillage dataset, a standard benchmark in plant pathology, was used in this research to provide thorough coverage of tomato leaf diseases. It contains high-resolution images of ten different categories, such as Tomato Mosaic Virus, Target Spot, Septoria Leaf Spot, Early Blight, Bacterial Spot, Tomato Yellow Leaf Curl Virus, Spider Mites (Two-Spotted Spider Mite), Late Blight, Leaf Mould, and Healthy Leaves. Seized under experimental settings, the photos have low noise and great feature definition. The dataset was split 80:20 into testing and training sets as a kind of experiment. Figure 2 shows the sample images of the plant's leaf.

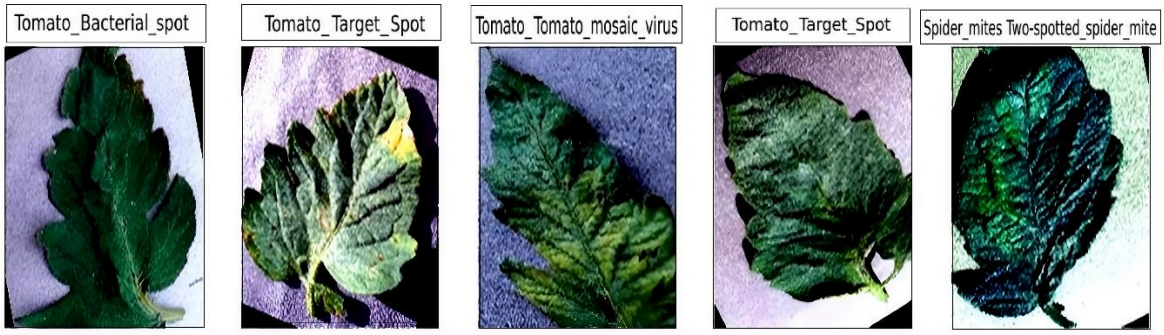


Figure 2- Sample Images of the Plant's Leaf

3.2. Data preprocessing

Realizing that optimal model performance depends on the process of turning raw data into an organized and polished form is crucial for deep learning (DL) techniques being crucial. Several preprocessing techniques were used for the ResV-Net model learnt on the PlantVillage dataset in order to raise data quality and consistency. All photos were specifically scaled to 224×224 pixels, therefore retaining consistent input sizes necessary for convolutional neural networks. To dynamically change brightness, contrast, saturation, and hue using Colour Jittering, thereby strengthening the model's resistance against various lighting situations. For aligning the input distribution with the pretrained model assumptions and accelerating convergence during training, pixel values were also standardized using the ImageNet mean and variance. These preprocessing methods significantly improve the generalizing capacity and resilience of the ResV-Net architecture overall.

3.3. Data augmentation

PyTorch's strong transformation capabilities are used with an augmentation generator to enrich the training set and so improve the generalizing capacity of the model on fresh data. By employing a variety of augmentation strategies—random horizontal flipping, rotation, cropping, and colour jittering—the training data is synthetically diversified, enabling the model to learn more invariant and discriminative features. Whereas Random Rotation (30) adds orientation variation by rotating images between ±30 degrees, Random Horizontal Flip () adds left-right symmetry with a 50% probability, hence increasing variability. Random Resized Crop (224) not only generates spatial diversity but also imposes a standard output size of 224×224 pixels, which is necessary for models as ResNet and Vision Transformer (ViT). By means of modifications to brightness, contrast, saturation, and hue under control, colour jitter improves the model's resistance to real lighting variations.

The images also undergo some additional preparation procedures to preserve consistency and fit with deep learning processes. In convolutional and transformer-based networks, the Resize ({224, 224} operation normalizes image sizes to ready them for batch processing. The pre-processed images are then converted by the To Tensor () transformation into PyTorch tensors, which enable effective GPU or CPU processing. Normalize (mean=\ [0.485, 0.456, 0.406], std=\ [0.229, 0.224, 0.225]) adjusts pixel intensities to ImageNet standards, therefore aligning the distribution of inputs with pre-trained models and accelerating convergence during training. Taken collectively, these changes are fundamental in improving the learning dynamics of the model, lowering overfitting, and ensuring superior performance on various and demanding image datasets.

3.4. Data splitting

The PlantVillage collection is split between training and testing sets using an 80/20 ratio. This guarantees that the ResNet-ViT Fusion model is tested on unseen data from the training set and learns from it rather effectively. This helps to prevent overfitting and raise the generalizing ability of the model towards fresh photos of plant diseases.

3.5. Classification models

The classification is performed using advanced DL models, Such as ResNet152, DenseNet121, EfficientNetB0 and ResVnet Fusion Model. Which are briefly described below:

1)ResVNet

The ResV-Net model is a deep learning architecture that combines the powerful convolutional capabilities of ResNet152 with the global contextual modelling strength of the Vision Transformer (ViT), as shown in Figure 3. To enhance efficiency during training, Low-Rank Adaptation (LoRA) is applied to the ViT component, enabling lightweight fine-tuning by updating only a small subset of parameters. This architecture aims to fuse the local spatial features learned by CNNs with the long-range dependencies captured by Transformers, improving overall performance in image classification tasks.

ResNet152 Feature Extraction

$F_{Res} = ResNet152_{trunc}(I)$ …….(1)

Where: I is the input image and F_{Res} is the extracted spatial feature vector.

ViT Feature Extraction with LoRA Fine-Tuning: ViT computes attention using query (Q), key (K), and value (V) matrices. With LoRA, it is possible to adapt only the query and value projections shown in equation (2):

$Q = X(W_Q + A_Q B_Q),\ V = X(W_V + A_V B_V)$ …….(2)

Here, X is the input token matrix, and W_Q, W_V These are the pretrained query and value weight matrices. LoRA introduces low-rank matrices A_Q, B_Q, A_V, and B_V of size $\mathbb{R}^{d\times r}$, where $r \ll d$, enabling efficient fine-tuning by updating only these lightweight components while keeping the main weights frozen.

Feature Fusion and Classification: After extracting both CNN and transformer features, they are concatenated and passed through a fully connected layer with SoftMax, as shown in equations (3) and (4):

$F_{fused} = Concat(F_{Res}, F_{ViT})$ ……(3)

$\hat{y} = Softmax(W_f F_{used} + b_f)$ ……. (4)

Where: W_f and b_f are the parameters of the final classification layer, and $\hat{y}$ is the output probability distribution.

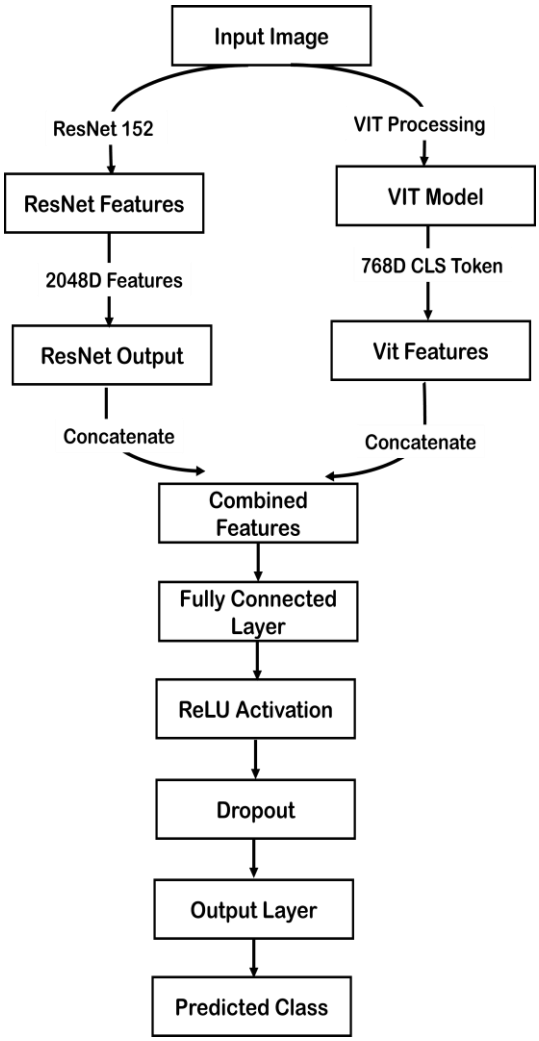


Figure 3- ResVNet Model Architecture

2)DenseNet121

DenseNet-121 is a deep CNN that is optimized for image classification using densely connected layers, such that each layer gets feature maps from all the previous layers using concatenation (J. et al. 2022). The architecture promotes feature reuse efficiently and also overcomes the vanishing gradient problem. The network does not become bulky due to the decrease in the number of channels through a controlled growth rate referred to as k. Each composite layer performs batch normalization, ReLU activation, and convolution on the k output channels. Operations of pooling and transformation between layers are controlled by Equation (5):

$$Y = W_3\{x, h_1(x), \quad h_2(x), \quad h_3(x)\} \ldots\ldots\ldots (5)$$

3)ResNet152

ResNet-152 is a deep convolutional neural network model to classify images and identify objects. The model has 152 layers and is much deeper than the original ResNet-50 model (Samundiswary, 2023). It uses skip connections and residual blocks to enable the training of deeper networks without experiencing the vanishing gradient problem. The residual blocks enable the model to learn residual mappings, i.e., the distinction between a layer's input and output, which can be easily optimized during training. Instead of passing the output of each layer directly to the next layer, the output of the current layer is added to the actual output of the previous layers, allowing gradients to propagate more quickly through the network. The architecture of ResNet152 consists of a convolutional layer followed by several blocks of convolutional layers, each containing residual connections. The network also includes global average pooling and a fully connected layer for classification.

4)EfficientNetB0

EfficientNet-B0 is utilized as the baseline model for the suggested approach owing to its efficiency in terms of computation and suitability for use in environments with resource constraints (Krishna et al. 2025). EfficientNet-B0 preserves key characteristics of the Efficient Net family, including MBConv blocks and squeeze-and-excitation (SE) modules, while implementing a more streamlined network architecture. Though architecturally equivalent to EfficientNet-B3, EfficientNet-B0 has fewer layers and fewer parameters, hence appropriate for use in low-latency inference requirements and hardware use minimization without losing considerable classification accuracy.

3.6. Performance matrix

Classification model performance is commonly based on confusion matrix-derived metrics, which are a powerful tool for evaluating prediction accuracy in both binary and multiclass classification problems. The confusion matrix provides a tabular representation of the model-generated predicted class labels against actual ground truth labels, from which key classification performance metrics can be derived, as shown in Table 2. These comprise true negatives (TN), correctly classified negative occurrences, and true positives (TP), those in which the model gets the positive class right. Type I errors, sometimes known as false positives, occur when the model misclassifies a negative occurrence as positive. Type II errors—false negatives—FN—occur when a positive event is misclassified as negative. Together, these components form the foundation for generating several assessment measures like accuracy, precision, recall, and F1-score, therefore offering a whole picture of the predictive ability of the model.

Table 2- Confusion Metrix

Value	***Actual positive***	***Actual Negative***
Predictive positive	TP	FP
Predictive negative	FN	TN

From these fundamental counts, a variety of performance metrics can be derived, such as:

1) Accuracy

The assessment evaluates the model's overall accuracy. The calculation of the proportion of correct predictions relative to the total predictions is presented in equation (6):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

2) Precision

This metric evaluates the accuracy of the model's positive predictions by determining the proportion that were indeed correct. The calculation of the proportion of true positives relative to all positive predictions is demonstrated in equation (7):

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots\dots\dots\dots\dots\dots (7)$$

3) Recall

This quantifies the number of true positive instances accurately identified by the model. The calculation of the proportion of true positives among all actual positives is presented in equation (8):

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots\dots\dots\dots\dots\dots\dots (8)$$

4) **F1-Score**

The balance between precision and recall is very important when the two are imbalanced. Equation (9) calculates the precision and recall harmonic mean:

$$F1 - Score = \frac{2 * Precision * Recall}{(Precision + Recall)} \dots\dots\dots (9)$$

3.7. Experimental implementation details

All models were trained for 50 epochs using the AdamW optimizer with an initial learning rate of 3×10^{-5} and a weight decay of 1×10^{-4}. The batch size was set to 32. A cosine annealing learning rate scheduler with a warm-up period of five epochs was employed. Early stopping with a patience of five epochs based on validation loss was applied to prevent overfitting. Experiments were conducted on a system equipped with an Intel Core i7 processor, 32 GB RAM, and NVIDIA GPU acceleration.
For the Vision Transformer branch, Low-Rank Adaptation (LoRA) was applied to the query and value projection matrices. The LoRA rank was set to r = 16, with a scaling factor α = 32 and a dropout rate of 0.1. During fine-tuning, only the LoRA parameters and the final classification head were updated, significantly reducing the number of trainable parameters while maintaining high classification performance.

3.8. Algorithm: ResVNet Transformer Model for Plant Leaf Diseases Detection

Input: plant leaf diseases detection prediction
Output: Trained model capable of effectively classifying plant leaf disease detection

1. **Import Libraries**:
   - Import necessary libraries (NumPy, matplotlib. Pytorch, transform etc.
2. **Load Dataset**:
   - Load the tomato leaf disease detection from Kaggle.
   - Extract image file paths and labels.
   - Use the Data Loader from PyTorch to load and batch the data
3. **Exploratory Data Analysis (EDA)**:
   - Display some sample images from the dataset.
4. **Data Pre-Processing**:
   - **Image Resizing**: Resize all images to a uniform size of 224x224.
   - **Tensor Conversion:** convert the image to a tensor
   - **Normalize:** Normalize the image for better training
5. **Data Augmentation**:
   - Use Pytorch Transform method to perform various transformations
   - RandomHorizontalFlip,
   - Random Rotation (30),
   - RandomResizedCrop (224),
   - ColorJitter (brightness=0.2, contrast=0.2, saturation=0.2, hue=0.2),
6. **Split Dataset**:
   - Split the dataset into a training set (80%) and a testing set (20%).
7. **Feature Extraction Using Pre-Trained Models**:
   - Load pre-trained models: **ResNet-50 and Vision Transformer**
   - Fine-tune ResNet by unfreezing all layers
   - Use LORA for fine-tuning Vision Transformer

8. **Fusion of Features (ResVnet Model)**:
    - Concatenate **Features**:
        - These two models are combined by concatenating their output features (ResNet's and ViT's), which are passed to a custom fully connected layer.
        - Pass them from a custom fully connected layer
    - **Train ResVnet Model:**
        - Train a model using the concatenated features.
        - Use optimizer 'adam' and loss function 'categorical_crossentropy'.
        - Set batch size = 32 and epochs = 20.
        - Evaluate the model
9. **Model Evaluation**:
    - Calculate evaluation metrics: **Accuracy**, **Precision**, **Recall**, **F1-Score**, **Sensitivity**, **Specificity**.
    - Plot **ROC Curve** for each model and compute **AUC** (Area Under the Curve).
10. **Results and Analysis**:
    - Display training and validation learning curves (loss and accuracy) for all models.

4. Results and Discussion

This section presents the experimental setup and outcomes of the proposed models for tomato leaf disease classification using image-based deep learning techniques. The implementation was carried out using Python programming on a system equipped with an Intel Core i7 processor (2.8 GHz) and 32 GB RAM. On interactive platforms like Google Colab and Jupyter Notebook, which offer a flexible and effective setting for running machine learning processes, model construction and testing were done. Essential Python libraries and frameworks were used in the experiments: TensorFlow and Keras for deep learning models; Scikit-learn for evaluation metrics and utilities; NumPy for effective numerical computation; and Matplotlib for result presentation.

The study used the PlantVillage collection, which included high-resolution pictures of tomato leaves afflicted by several illnesses. Under consistent experimental settings, several transformer-based deep learning architectures—ResNet152, DenseNet121, EfficientNet B0, and the proposed ResVNet—were trained and evaluated. Together with a comparative study with other current state-of-the-art methods, the following sections offer comprehensive categorization results of every model.

4.1. Experiment results

The experimental results of the proposed models (ResNet152, DenseNet121, EfficientNetB9 and ResVnet model) are discussed below according to accuracy, precision, recall, f1score, confusion and matrix classification report for Plant leaf diseases detection. Table 3 provides the performance of each model according to the performance measure with 5-Fold CV. A proposed ResVnet model gets the highest accuracy of 97.45% in comparison to other models for plant disease classification. The PlantVillage tomato dataset exhibits a degree of class imbalance, with certain disease categories (e.g., Tomato Yellow Leaf Curl Virus) containing substantially more samples than others (e.g., Tomato Mosaic Virus). To mitigate potential evaluation bias arising from this imbalance, macro-averaged performance metrics were adopted, ensuring that all disease classes contribute equally to the final evaluation. As shown in Table 3, the proposed ResVNet model achieves not only the highest mean accuracy but also the lowest standard deviation (±0.25), indicating more stable and consistent performance across different data splits.

Table 3- Averaged Performance Metrics (5-Fold CV)

Model	***Accuracy (%)***	***Macro-Precision (%)***	***Macro-Recall (%)***	***Macro-F1 (%)***
DenseNet121	93.12 ± 0.51	93.05 ± 0.62	92.89 ± 0.55	92.95 ± 0.58
EfficientNetB0	94.05 ± 0.38	94.12 ± 0.45	93.95 ± 0.41	94.01 ± 0.43
ResNet152	95.23 ± 0.41	95.18 ± 0.39	95.10 ± 0.42	95.14 ± 0.40
ResVNet (Proposed)	**97.45 ± 0.25**	**97.38 ± 0.31**	**97.41 ± 0.28**	**97.39 ± 0.29**

All performance metrics reported in this study are computed using macro-averaging. In this approach, precision, recall, and F1-score are calculated independently for each class and then averaged across all classes. This ensures that minority classes contribute equally to the final evaluation and mitigates bias caused by class imbalance in the dataset.

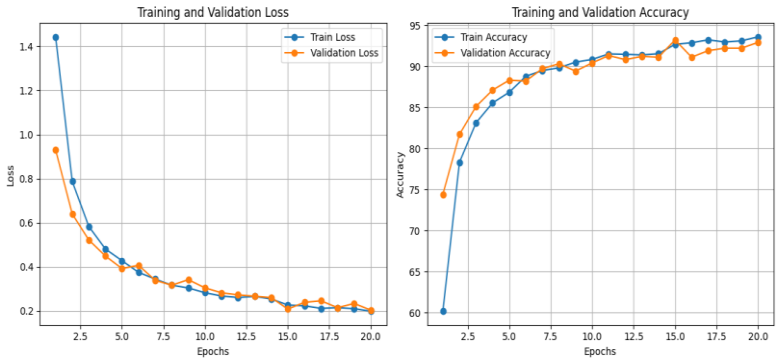


Figure 4- Plotting accuracy and loss curve of Densenet121

The training and validation curves demonstrate that DenseNet121 achieved stable convergence, with both loss functions declining steadily to below 0.2 by the 20th epoch, as shown in Figure 4. Concurrently, the model's accuracy improved consistently, attaining approximately 94%, indicating strong learning capability and generalization for plant disease classification.

The confusion matrix highlights DenseNet121's classification performance across ten tomato leaf disease categories and healthy samples, as shown in Figure 5.

High true positive counts, notably 204 for Leaf_Mold and 202 for Septoria_leaf_spot, reflect the model's strong diagnostic precision. Minimal misclassifications indicate excellent discriminative capability and robust disease-specific generalization.

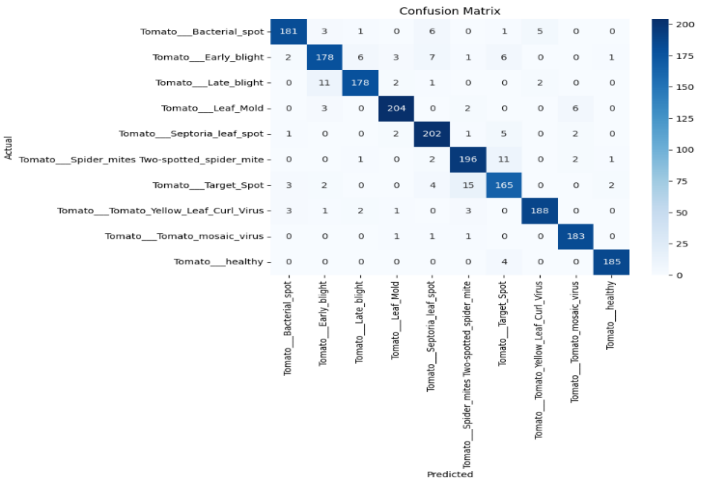


Figure 5- Classification report of Densenet121

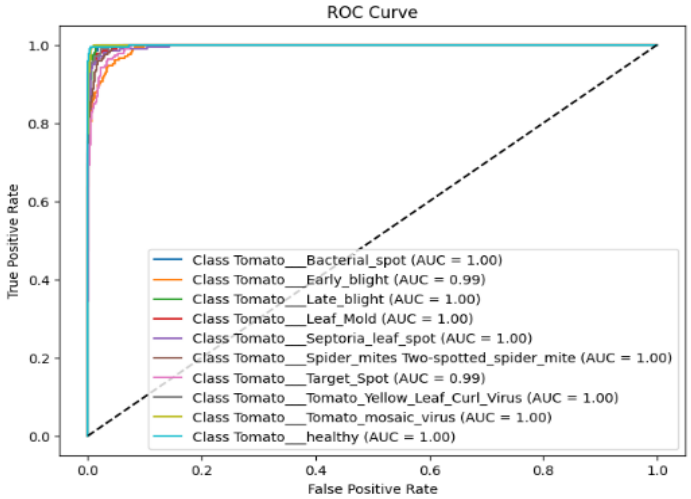


Figure 6- ROC curve of Densenet121

The ROC curve illustrates DenseNet121's exceptional classification performance, achieving AUC scores of 1.00 for the majority of tomato disease classes, indicating near-perfect separability in Figure 6. The slightly lower AUC values of 0.99 for *Early blight* and *Target Spot* still affirm the model's high diagnostic accuracy and reliability.

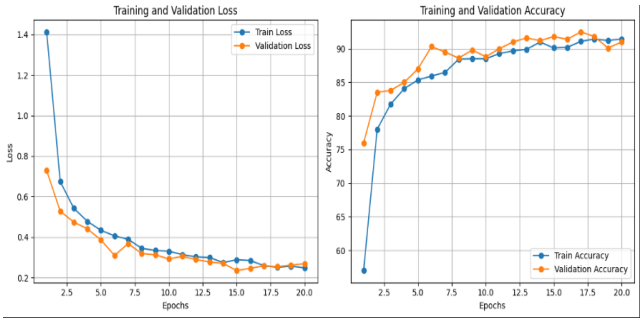


Figure 7- plotting accuracy and loss curve of EfficientNeB0

The convergence behaviour of EfficientNetB0 demonstrates a well-optimized learning process, with both training and validation loss declining progressively and accuracy increasing significantly over 20 epochs, suggesting effective feature extraction and generalization capacity with negligible overfitting, thereby confirming the architectural efficiency and stability of the model's learning trajectory in Figure 7.

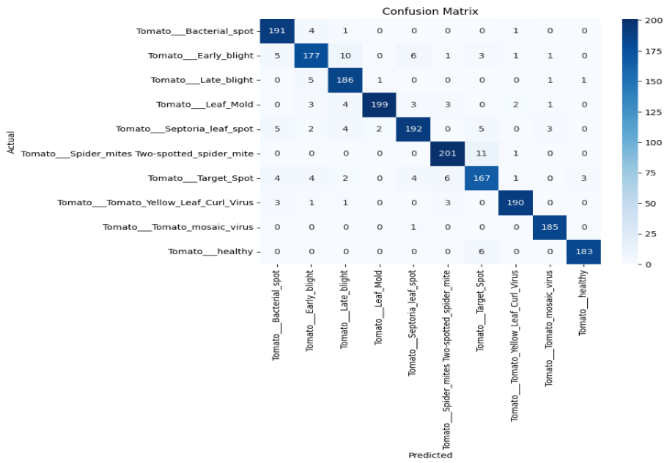


Figure 8- Confusion Matrix of efficientnetB0

The EfficientNetB0 model demonstrates high classification precision, with class-wise accuracies such as 191/200 for Bacterial Spot, 186/200 for Late Blight, and 183/200 for the Healthy class, indicating strong predictive consistency as depicted in the confusion matrix in Figure 8. Misclassifications remain minimal, showcasing robust model generalization.

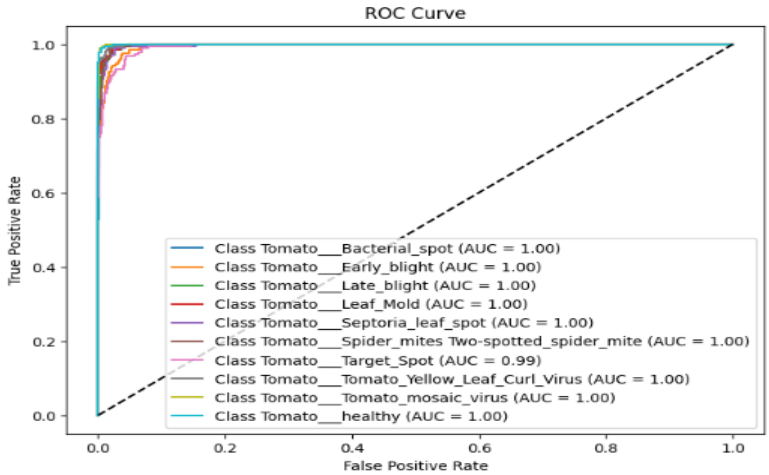


Figure 9- ROC curve of efficientnetB0

The ROC curve of EfficientNetB0 illustrates superior classification capability, with most classes achieving an AUC of 1.00, including Bacterial Spot, Leaf Mold, and Healthy categories, indicating near-perfect discrimination (Figure 9). Only two classes, Early Blight and Target Spot, recorded slightly lower AUC values of 0.99.

Figure 10 presents the training and validation performance metrics for the ResNet152 model. The initial plot demonstrates a steady decline in both training and validation loss, indicating effective learning and convergence. The plot on the right illustrates an enhancement in accuracy, which stabilizes at approximately 95%. The close alignment of training and validation curves indicates minimal overfitting, reinforcing the model's robust generalization capability.

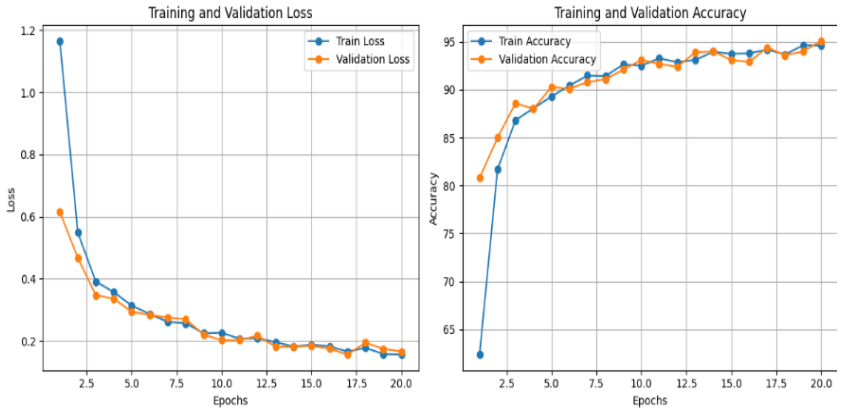


Figure 10- Loss and accuracy curve of Resnet152

The confusion matrix for the ResNet152 model, shown in Figure 11, illustrates its classification performance across different tomato leaf disease categories. The strong diagonal dominance indicates high accuracy, with minimal misclassifications. Some confusion exists among similar diseases, but overall, the model demonstrates robust predictive capability.

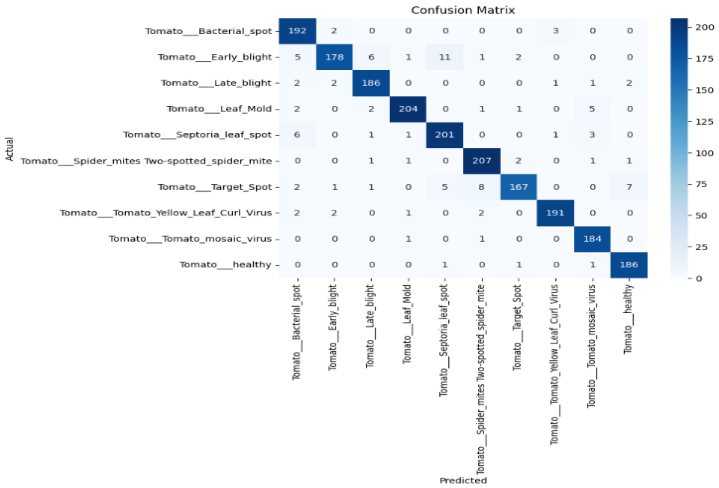


Figure 11- Confusion matrix of Resnet152

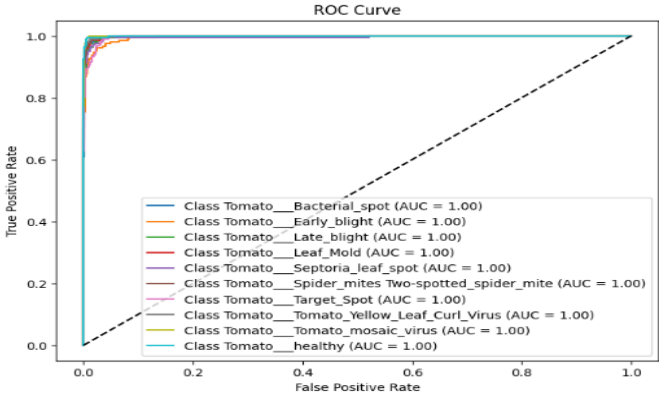


Figure 12- ROC curve of Resnet152

The ROC curve for the ResNet152 model, shown in Figure 12, illustrates its ability to distinguish between different tomato leaf disease classes. The model achieves high AUC values across all classes, indicating strong classification performance. While most classes approach an AUC of 1.00, minor variations suggest slight differences in predictive capability. The overall strong true positive rates highlight the model's effectiveness and reliability in disease detection.

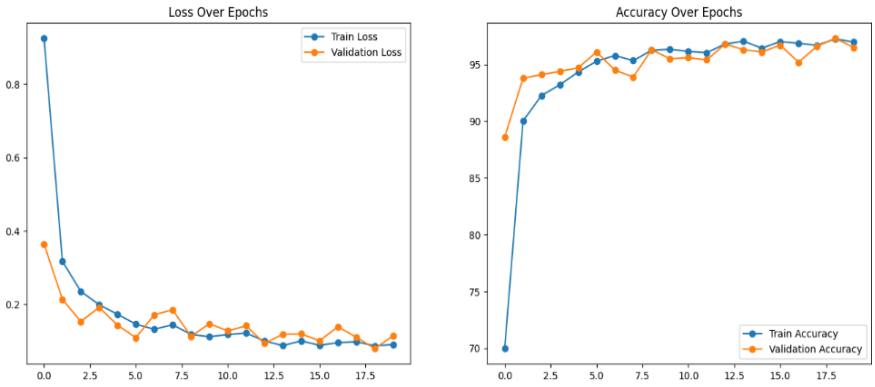


Figure 13- Loss and accuracy curve of ResVNet

The training and validation performance of the ResVNet model, shown in Figure 13, demonstrates effective learning. The left plot indicates a steady decline in loss, confirming convergence. The right plot shows accuracy rapidly improving before stabilizing above 95%, suggesting strong generalization with minimal overfitting.

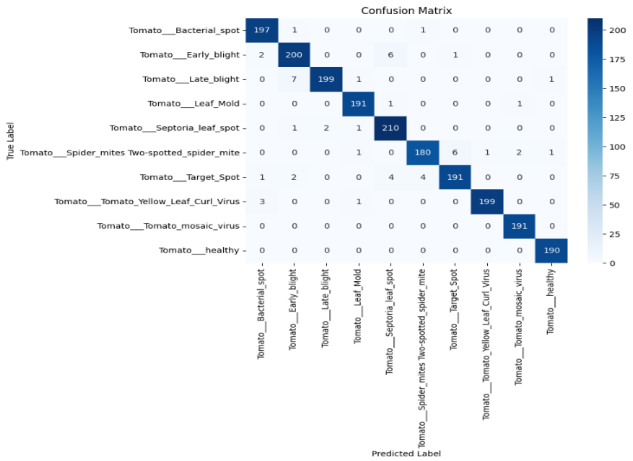


Figure 14- Confusion Matrix of ResVNet

Figure 14 illustrates the confusion matrix of the proposed ResV-NeT model, emphasizing its effectiveness in categorizing different tomato leaf diseases. The model shows high accuracy across all classes, with minimal misclassifications. Notably, classes like "Tomato Early Blight" and "Tomato Yellow Leaf Curl Virus" achieved nearly perfect predictions, demonstrating the model's effectiveness and robustness in plant disease detection through leaf image analysis.

The ROC curve of the ResV-NeT model, demonstrating outstanding classification performance across all tomato disease classes in Figure 15. Every class attains an AUC of 1.00, signifying flawless differentiation between true positive and false positive rates. This demonstrates the model's remarkable capacity to differentiate among various categories of plant diseases with impressive precision.

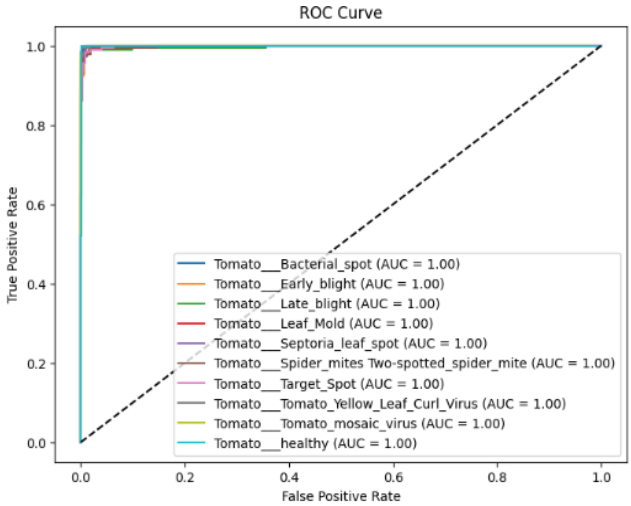


Figure 15- ROC curve of ResVNet

Figure 16 showcases predicted leaf disease classifications, highlighting accurate and misclassified cases for bacterial spot, early blight, and late blight.



Figure 16- Predicted Images of Leaf Disease Classes (Tomato Mosaic Virus, Target Spot, Bacterial Spot, Late Blight, and Healthy Leaves)

4.2. Statistical significance analysis

To assess whether the observed performance improvements are statistically significant, McNemar's test was conducted between the proposed ResVNet model and the strongest baseline (ResNet152). The test resulted in a p-value less than 0.001, indicating that the performance improvement achieved by ResVNet is statistically significant and not attributable to random variation.

4.3. Ablation study of ResVNet components

The ablation study's detailed explanation in Table 4 confirms that the hybrid integration of convolutional and transformer-based feature extraction significantly enhances classification performance. While full fine-tuning of the hybrid architecture yields marginally higher accuracy, the LoRA-based configuration achieves comparable performance with substantially reduced training time and computational cost, demonstrating the effectiveness of parameter-efficient fine-tuning.

4.4. Comparative analysis

The comparative analysis underscores the exceptional performance of the proposed ResVNet model when evaluated against various leading models across multiple datasets, showcasing its reliability and efficiency in accurately identifying plant diseases from leaf images.

The proposed deep learning model, ResVNet, demonstrates superior performance in plant disease detection from leaf images. As shown in Table 5, ResVNet achieved the highest scores across all evaluation metrics with 97.45% accuracy, precision, recall, and F1-score using the PlantVillage dataset. It outperforms baseline models like ResNet152 (95%), EfficientNetB0 (94%), and DenseNet121 (93%). Other existing approaches, such as CNN + SVM (93%) and VGG-16 (92.3%), o performed well but fell short of the proposed model. Meanwhile, traditional models like XGBClassifier and VGG19 reported lower accuracies of 88.99% and 84%, respectively. These results highlight the effectiveness of ResVNet in accurately identifying plant diseases from leaf images.

Table 4- Ablation Study models for Plant disease detection

Configuration	*Architecture*	*Fine-Tuning Strategy*	*Accuracy (%)*	*Training Time (hrs)*
A	ResNet152 only	Full fine-tuning	95.23	4.5
B	ViT-B/16 only	Full fine-tuning	96.10	6.2
C	ResNet + ViT	Full fine-tuning	97.62	8.5
D	**ResVNet (Proposed)**	**LoRA (r=16)**	**97.45**	**5.1**

Table 5- Comparative analysis of the proposed model with other models for Plant disease detection

Author	*Model*	*Dataset*	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Proposed model	**ResVNet**	**PlantVillage dataset**	**97.45**	**97.38**	**97.41**	**97.39**
Baseline model	Resnet152	PlantVillage dataset	95.00	95.00	95.00	95.00
Baseline model	EfficientNetB0	PlantVillage dataset	94.00	94.00	94.00	94.00
Baseline model	Densenet121	PlantVillage dataset	93.00	93.00	93.00	93.00
(Nagpal et al. 2024)	CNN + SVM	Leaf Image Dataset	93.00	91.00	94.00	92.00
(Shinde et al. 2024)	VGG-16	PlantVillage dataset	92.30	91.80	91.10	91.40
(Poornima et al. 2023)	CNN	Village plant dataset	90.45	96.45	95.12	94.67
(Salman & Ali 2025)	XGBClassifier	PlantVillage dataset	88.99	88.79	88.71	88.75
(Khade & Patil 2024)	VGG19	PlantVillage dataset	84.00	-	84.00	84.00

5. Conclusions

This work introduced ResVNet, a hybrid CNN-Transformer framework for early tomato leaf disease detection that combines the local feature learning strength of ResNet152 with the global contextual modelling capability of Vision Transformers, enhanced by Low-Rank Adaptation (LoRA) for parameter-efficient fine-tuning. This design enables the effective capture of fine-grained texture cues and long-range disease dependencies, which are critical for accurate visual diagnosis.

Extensive experiments on the PlantVillage tomato dataset using stratified 5-fold cross-validation demonstrated that ResVNet consistently outperforms state-of-the-art CNN architectures, achieving a mean accuracy of 97.45% with superior macro-precision, macro-recall, and macro-F1 scores. McNemar's statistical test confirmed that these gains over the strongest baseline are statistically significant ($p < 0.001$). Ablation studies further validated the superiority of the hybrid CNN-Transformer architecture and showed that LoRA substantially reduces trainable parameters and training time while maintaining near-optimal performance.

While the results are promising, the evaluation was limited to the PlantVillage dataset, which consists of laboratory-acquired leaf images with uniform backgrounds and controlled illumination; thus, performance under real agricultural conditions involving complex backgrounds, variable lighting, partial occlusions, and co-occurring disease symptoms remains to be validated. Future work will focus on evaluating the proposed framework on real-field datasets such as PlantDoc and self-acquired in-field images, along with edge-device deployment and explainable AI integration. Overall, this study establishes a statistically robust and computationally efficient framework for intelligent plant disease diagnosis in precision agriculture.

Author Declarations

Authors' contributions: A. M. contributed to the conceptualization, model development, experimentation, and manuscript writing. S. K. supervised the research, provided technical guidance, and contributed to the review and refinement of the manuscript.

Conflict of interest/Competing interests: The author(s) declare no conflict of interest.

Ethics approval/declarations: No need/ Not applicable.

Funding: This research received no external funding.

Consent to participate: No need/ Not applicable.

Consent for publication: No need/ Not applicable.

Data availability: Data are available on request due to privacy or other restrictions.

References

Demilie W B (2024). Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*, 11(1). https://doi.org/10.1186/s40537-023-00863-9

Gavhale M K R & Gawande P U (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. *IOSR Journal of Computer Engineering*, 16(1): 10–16. https://doi.org/10.9790/0661-16151016

Hana B & Yassmina S (2024). Deep Learning-Based Detection of Apple Leaf Diseases using Image Analysis. *2024 1st International Conference on Innovative and Intelligent Information Technologies (IC3IT)*, 1–6. https://doi.org/10.1109/IC3IT63743.2024.10869435

Hanif M A, Khadimul Islam Zim M & Kaur H (2024). ResNet vs Inception-v3 vs SVM: A Comparative Study of Deep Learning Models for Image Classification of Plant Disease Detection. *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, 2: 1–6. https://doi.org/10.1109/IATMSI60426.2024.10502832

J A, Eunice J, Popescu D E, Chowdary M K & Hemanth J (2022). Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. *Agronomy* 12(10). https://doi.org/10.3390/agronomy12102395

Joshi A, Badola A & Saini R (2024). Deep Learning Based Fruit Crop Disease Classification Using Plants Leaf Imagery. *2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*, 1–6. https://doi.org/10.1109/ICECSP61809.2024.10698492

Jung M, Song J S, Shin A Y, Choi B, Go S, Kwon S Y, Park J, Park S G & Kim Y M (2023). Construction of deep learning-based disease detection model in plants. *Scientific Reports*, 13(1): 1–13. https://doi.org/10.1038/s41598-023-34549-2

Kaur A, Sharma R, Chattopadhyay S & Joshi K (2024). Cotton Leaf Disease Classification Using Fine-Tuned VGG16 Deep Learning Model. *2024 2nd World Conference on Communication & Computing (WCONF)*, 1–4. https://doi.org/10.1109/WCONF61366.2024.10692185

Khade V C & Patil S B (2024). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A Hybrid Deep Learning Approach for Crop Disease Severity Level Prediction. *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, 2024(9s): 215–224

Krishna M S, Machado P, Otuka R I, Yahaya S W, dos Santos F & Ihianle I K (2025). Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach. *J*, 8(1). https://doi.org/10.3390/j8010004

Li L, Zhang S & Wang B (2021). Plant Disease Detection and Classification by Deep Learning—A Review. *IEEE Access*, *9*, 56683–56698. https://doi.org/10.1109/ACCESS.2021.3069646

Li W, Zhu L & Liu J (2024). PL-DINO: An Improved Transformer-Based Method for Plant Leaf Disease Detection. *Agriculture*, 14(5). https://doi.org/10.3390/agriculture14050691

Meeradevi V R, Mundada M R, Sawkar S P, Bellad R S & Keerthi P S (2020). Design and Development of Efficient Techniques for Leaf Disease Detection using Deep Convolutional Neural Networks. *2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 153–158. https://doi.org/10.1109/DISCOVER50404.2020.9278067

Murthy T S, Krishnamoorthi K, Selvarani P, Srinivasulu R, Saranya G & Kausalya A K (2024). MDISL: Experimental Evaluation of Learning Based Potato Plant Disease Detection Using Modified Deep Image Segmentation Logic. *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 1–6. https://doi.org/10.1109/ICSES63760.2024.10910396

Nagpal J, Goel L & Shekhawat P S (2024). A hybrid deep learning neural network for early plant disease diagnosis using a real-world Wheat–Barley vision dataset: challenges and solutions. *International Journal of Data Science and Analytics*, *March*. https://doi.org/10.1007/s41060-024-00578-x

Poornima S, Sripriya N, Alrasheedi A F, Askar S S & Abouhawwash M (2023). Hybrid Convolutional Neural Network for Plant Diseases Prediction. *Intelligent Automation and Soft Computing*, 36(2): 2393–2409. https://doi.org/10.32604/iasc.2023.024820

Saini A, Guleria K & Sharma S (2023). Multiclass Classification of Rice Leaf Disease Using Deep Learning Based Model. *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 1–6. https://doi.org/10.1109/ASIANCON58793.2023.10270229

Salman R D & Ali Y H (2025). *Hybrid Model for Plant Disease Diagnosis: Transfer Learning with Dynamic Feature Selection*. *10*.

Samundiswary P (2023). Wheat Disease Detection and Classification Using Resnet152. *International Journal of Creative Research Thoughts*, 11(5): 2320–2882

Shinde, R. P., Shah, S. B., Belani, R., & Nune, N. C. (2024). *Hybrid Deep Learning Approach for Automated Plant Disease Detection in Precision Agriculture* 7(12): 114–121

Thotad P N, Kallur S & Nandeppanavar A (2023). An Efficient Model for Plant Disease Detection in Agriculture Using Deep Learning Approaches. 2023 4th *IEEE Global Conference for Advancement in Technology (GCAT)*, 1–6. https://doi.org/10.1109/GCAT59970.2023.10353343

Zhang X, Dong H, Gong L, Cheng X, Ge Z & Guo L (2023). Multiple paddy disease recognition methods based on deformable transformer attention mechanism in complex scenarios. *International Journal of Computers and Applications* 45(10): 660–672



Copyright © 2026 The Author(s). This is an open-access article published by Faculty of Agriculture, Ankara University under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is properly cited.