Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A novel hierarchical framework for plant leaf disease detection using residual vision transformer

Sasikala Vallabhajosyula [a], Venkatramaphanikumar Sistla [b], Venkata Krishna Kishore Kolli [b,*]

[a] Department of CSE, Vignan's Nirula Institute of Technology and Science for Women, Guntur, Andhra Pradesh, India
[b] Department of CSE, Vignan's Foundation for Science, Technology, and Research, Guntur, Andhra Pradesh, India

ARTICLE INFO

Keywords:
Inception V3
MobileNetV2
Vision transformer
Deep leaning
Plant leaf disease detection

ABSTRACT

Early detection of plant leaf diseases accurately and promptly is very crucial for safeguarding agricultural crop productivity and ensuring food security. During their life cycle, plant leaves get diseased because of multiple factors like bacteria, fungi, weather conditions, etc. In this work, the authors propose a model that aids in the early detection of leaf diseases using a novel hierarchical residual vision transformer using improved Vision Transformer and ResNet9 models. The proposed model can extract more meaningful and discriminating details by reducing the number of trainable parameters with a smaller number of computations. The proposed method is evaluated on the Local Crop dataset, Plant Village dataset, and Extended Plant Village Dataset with 13, 38, and 51 different leaf disease classes. The proposed model is trained using the best trail parameters of Improved Vision Transformer and classified the features using ResNet 9. Performance evaluation is carried out on a wide aspects over the aforementioned datasets and results revealed that the proposed model outperforms other models such as InceptionV3, MobileNetV2, and ResNet50.

1. Introduction

In the recent era, automated plant disease detection has evolved as one of the key challenges in precision agriculture. However, recent advancements in Deep Learning (DL) have significantly affected the early detection and accurate plant disease prediction [1]. Automated plant disease detection tools enable the farmers to continuously monitor the plant's growth and its health in the ever-changing environmental conditions. The presence of high inter-class similarities and intra-class dissimilarities made plant disease detection a challenging task. The early detection and prediction of plant diseases are highly critical for effective management and prevention of these diseases [2]. In general, plants are being affected by factors like bacteria, insects, weeds, and many climatic conditions and the most common plant diseases include the mosaic virus, green leaf spots, and scabs. All these diseases are expected to be recognized in the early stage to prevent damage to plants and to ensure high productivity of crop yield [3].

Manual diagnosis of plant diseases by plant pathologists is an unreliable and time-consuming process. Technological advancements in the areas of Artificial Intelligence (AI) and High-Performance Computing have opened new gateways to more precise and efficient disease detection [4]. Conventional Machine Learning (ML) techniques include feature extraction and classification models that extract features from images. Further, models are trained with those features to train a classifier that can discriminate the healthy and

[*] Corresponding author.
E-mail address: kvkkishore@vignan.ac.in (V.K.K. Kolli).

https://doi.org/10.1016/j.heliyon.2024.e29912
Received 19 October 2023; Received in revised form 11 April 2024; Accepted 17 April 2024
Available online 22 April 2024

diseased leaf plants.

A CNN method based on DL has been given by the various authors to automatically classify and differentiate plant leaf diseases. However, diseases caused by abiotic stresses such as drought and nutrient deficiency are considered limitations in the accurate detection by machine learning algorithms [5]. Convolutional Neural Network(CNN) is a more popular and efficient method for learning from low-level complex features and training of deep CNN layers is highly computationally expensive. Transfer learning-based models include Visual Geometry Group (VGG-16), Residual Networks(ResNet), Densely Connected Convolutional Networks (DenseNet), and Inception have been proposed by researchers and trained with the ImageNet dataset.

The complex structure of textures and color variations in high-resolution images is a challenging task [6]. Because of this, it is more crucial to accurately detect plant diseases and their severity. As a result, a new technique for doing so is required to increase the task's adaptability and generalizability [7]. Deep Learning models can handle complex and high-resolution images and they can learn from large amounts of labeled training data and evolved as most promising in the detection of plant lesions [8].

The key contributions of the proposed work are as follows:

- Design and develop a novel architecture to extract global contextual features along with the local features using the attention module with minimum computational complexity.
- Design and development of a novel framework for capturing long-term interdependence, adaptability in handling inputs of different sizes, and generalization ability
- Performance evaluation and comparison of the proposed method with the existing deep learning models on the Plant Village Dataset.

The rest of this paper is organized as follows: Section 2 presents the thorough literature survey of various DL models used for plant leaf disease detection; Section 3 introduces the proposed methodology Hierarchical Residual Vision Transformer in detail. Then, Section 4 presents dataset details and comprehensive results using the proposed method and the comparison with widely used state-of-the-art models. Section 5 concludes this paper.

2. Related work

The major objective of this study is to provide awareness of various methods explored by various researchers in the field of Precision Agriculture. In this paper, various methods are carried out for image enhancement, segmentation, feature extraction, and classification methods adopted for diagnosing leaf diseases. The authors have gone through the research works published from July 25, 2012, to August 31, 2023, which has received major attention in this study.

Researchers have designed and developed various pre-trained models to identify plant diseases using potent feature extraction capabilities of deep learning. The recent advancements in Deep neural networks(DNN) improve efficiency but still have high computational costs. Zhang et al. [9] introduced a Novel Infrared Shape Network (ISNet) aimed at precisely detecting the shape details of infrared targets. They also presented a new benchmark, IRSTD-1k, which comprises 1000 realistic images depicting various target shapes, sizes, and diverse cluttered backgrounds, each meticulously annotated at the pixel level. Zhang et al. [10] proposed a pioneering network designed for detecting infrared small targets (IRSTD) comprised of a U-Net as backbone encoder, a context mixer decoder (CMD) integrating spatial and frequency attention (SFA), and an eyeball-shaped enhancement module (EEM).

Zhang, M et al. [11] proposed the Runge-Kutta Transformer (RKformer) model aimed to extract features with discriminative semantics while retaining fine details. Initially, a parallel encoder block (PEB) integrates transformer and convolutional techniques to capitalize on both long-range dependency modeling and locality modeling for semantic extraction and detail preservation. Mingjin [12] introduced a Feature Compensation and Cross-level Correlation Network (FC3-Net), a novel approach that delves into feature compensation and cross-level correlation to enhance Single Frame Infrared Small Target (SIRST) detection. Mingjin [13] presented the Infra-Red Prune Detection (IRPruneDet) network using wavelet structure and regularized soft channel pruning to enhance the detection of infrared small targets. Mingjin presented an Efficient Spectral correlation coefficient of the Spectrum kernel-based Self Attention transformer (ESSAformer) [14], to enhance the super-resolution of hyperspectral images. Single hyperspectral image super-resolution (single–HSI–SR) endeavors to enhance the low-resolution of a hyperspectral image into a high-resolution.

K. Ashok Kumar et al. [15] proposed a DenseNet-121 architecture, that yielded an accuracy of 98.7 % for detecting infectious diseases in plant leaves. B Sai Reddy et al. [16] proposed the Densenet model to determine the disease and damage of the leaf. In the successor stage, the Semantic segmentation algorithm was used to detect the damage to the leaf. It has produced an accuracy of 97 %. Remedial measures have been suggested depending on the amount of the plant's damage. Vinay Gautam et al. [17] have proposed the Modified Inception ResNet-V2 (MIR-V2) model for classifying diseases in tomato leaves. The prior-trained CNN model (IR-V2) and the proposed CNN model (MIR-V2) got an accuracy of 94.69 and 98.98, respectively.

Chitranjan Kumar Rai et al. [18] proposed a Deep Convolutional Neural Networks (DCNN) model to identify diseased cotton leaves and plants. The proposed model was trained and evaluated using 2293 cotton plant leaves and yielded an accuracy of 97.98 %. Vandana B. Malode et al. [19] proposed a CNN-based VGG model to detect the disease-affected leaves of grapes and tomatoes. The VGG achieved an accuracy of 98.40 % and 95.71 % for grapes and tomatoes, respectively. Ishita Bhakta et al. [20] designed a Deep convolutional neural network model to identify a prevalent illness called Leaf blight caused by bacteria in rice plants. They have gathered four sets of images as "Normal", "Stage 1 infection", "Stage 2 infection", and "Stage 3 infection". Each set contains 261 images, i.e., 261 $\times$ 4 = 1044. The accuracy achieved at these three stages for the proposed model is 95 %, 97 %, and 98 %, respectively.

To identify the diseases in all apple leaf varieties, Krishan Kumar et al. [21] developed a CNN model that produced an accuracy of

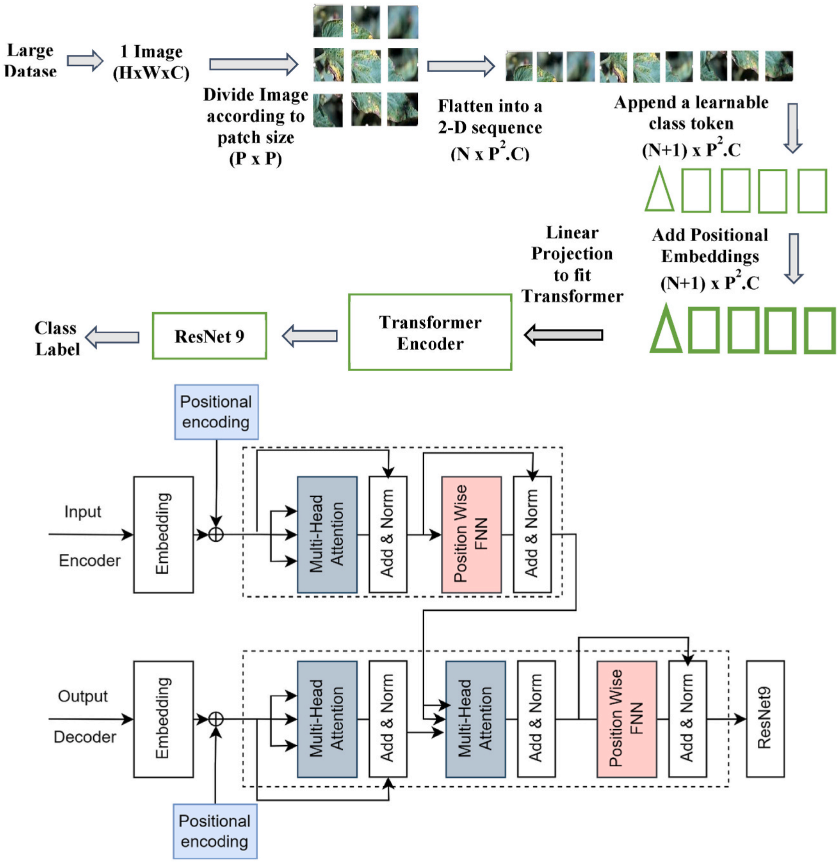


Fig. 1. Architecture of a hierarchical residual vision transformer.

98 % after training it with 1000 epochs. A lightweight deep neural network was proposed by Sabbir Ahmed et al. [22] by fusing a refined pre-trained model and a classifier network. To mitigate data leakage, runtime data augmentation methods have been applied. The proposed architecture attained 99.30 % accuracy. To recognize the bacterial leaf blight disease in rice plants, S. Vigneshwari et al. [23] suggested the Moore-Penrose pseudo-inverse weight-related deep convolutional neural network (MPW-DCNN). The accuracy of the suggested MPW-DCNN classifier is 97.5 %.

Vinay Gautam et al. [24] proposed the Mask Region Convolutional Neural Network (Mask R–CNN) to detect the disease of tomato plant leaves. The proposed design incorporates a light-headed "R–CNN" to reduce memory consumption and computational expenses. The model gained an accuracy of 98 %. A CA_DenseNet_BC_40 model was suggested by Hongbo Qiao et al. [25] to categorize the harm brought on by cotton pests. The division of cotton aphid damages was found to be classified with an accuracy of 97.3 % using the CA_DenseNet_BC_40 model, which performed better than other networks like ResNet50, DenseNet, etc. Rui Zhou et al. [26] proposed a model to identify the diseases of apple and coffee leaves by performing feature fusion and transfer learning. They first developed a deep feature description based on transfer learning to acquire a high-level latent feature representation. To capture the local texture information in photos of plant leaves, they had previously fused deep features with standard handcrafted features. This technique performed well, with 99.79 %, 92.59 %, and 97.12 % accuracy on the three datasets.

Huibin Long et al. [27] proposed an rE-GoogLeNet convolutional neural network model to identify rice leaf diseases. The dataset contains 1122 rice leaf images. This suggested model accuracy was 99.58 %, 1.72 % higher than the original GoogLeNet.Sheli Sinha Chaudhuri et al. [28] laid out Principal Component Analysis (PCA) DeepNet to identify tomato leaf diseases for agriculture. The innovative framework integrates a traditional machine learning model's PCA with a specifically designed deep neural network called PCA DeepNet. With the suggested model, classification accuracy increased to 99.60 %. Aibin Chen et al. [29] proposed LMBRNet to identify tomato leaf disease. The overall identification accuracy of the suggested model is 99.7 %. It exceeds previous models like GoogleNet (98.96 %) and ResNet (97.48 %). The suggested model's derived parameters are fewer than those of ResNet (23 million) and GoogleNet (5.7 million).

Squeeze-Net training architecture is proposed by Kantha Raju Kanaparthi et al. [30] to train chilli leaves for identifying Gemini and Mosaic viruses. Considering several training features like CNN optimizers stochastic gradient descent with momentum (SGDM), Adaptive Moment (ADAM), and Root Mean Squared Propagation (RMSPROP), the resulting training accuracy can range from 50 % to 100 %. SGDM optimizer gained an accuracy of 50 % with Max_Epochs of 40. ADAM optimizer gained an accuracy of 100 % with Max_Epochs of 40. RMSPROP optimizer gained an accuracy of 100 % with Max_Epochs of 35. The experimental results concluded that

RMSPROP is best suited for the SqueezNet architecture. Oliva Debnath et al. [31] proposed Smart farming integrating ML with the IoT network for early detection of Brown spot disease using CNN. The data is manually collected from rice fields, Tensorflow, and the Keras framework. The accuracy of the suggested model was 97.70 %. Sihan Zhou et al. [32] recommended residual-distilled transformer architecture to identify rice leaf diseases. By Refs. [33,34] residual concatenation, this approach integrates both the vision and distilled transformers, producing more effective outcomes than existing techniques. The accuracy of the advised model jumped to 92 %.

Furthermore, similarity in plant leaf diseases and complex leaf texture structure leads to errors in the detection of plant diseases [35]. [36]. All these limitations are the major constraints behind the design of an automated disease detection system for precision agriculture. Based on all these requirements, a novel deep learning model is designed by extracting discriminating features with low computational complexity. Recently [37], the attention module has produced outstanding results in terms of increasing the model's recognition accuracy in complex classification tasks. The usage of a Channel-based attention module based on Dense Net to recognize wheat stripe rust disease has increased accuracy by 5.47 % in comparison to a native model. Even though the addition of an attention module has increased the detection accuracy of plant leaf diseases, it was unable to reduce the model's parameter redundancy when features were engineered [38]. [39].

3. Methodology

This work proposes a novel customized deep learning architecture using Improved Vision Transfomer and ResNet9 for classifying diseases of various plants. The following architecture elaborates on various stages, such as image preprocessing, augmentation, model building, and testing phases on a given image dataset.

The images of plant leaves have been preprocessed to promote uniformity and learning. Further, images are typically resized to 224*224, the pixel values are normalized using methods like rotation, flipping, or zooming. Fig. 1 presents the architecture of the proposed model.

3.1. Feature extraction using Improved Vision Transformer

In recent years, the usage of transformers has rapidly increased with their parallelization and computational efficiency and is realized as the best alternative for recurrent neural networks (RNNs) in NLP tasks such as language modeling and machine translation. In this paper, the Improved Vision Transformer [40]. [41] (IVT) is used for image classification tasks. An attention mechanism is the central component of a transformer network. Irrespective of the distance between an input and an output, an attention mechanism makes it possible to describe dependencies in a series. Self-attention is an attention mechanism that links many sequence places to calculate the information of a single sequence. It has become an essential component of convincing sequence modeling and transformation models for a range of activities. Abstract generalization, textual entailment, and reading comprehension are just a few of the many tasks where self-attention has been effectively used. When a query, key, and value are vectors and are mapped from an input to an output, this mapping is called the attention function. An ordered sum of values is the result of the attention process, with each value's weight determined by comparing its value with the relevant key using the query's compatibility function. There are two types of self-attention mechanisms [42–45]: multi-head attention and scaled dot-product attention. The query matrix Q of size dk, the key matrix K, and the value matrix V of dimension d_v are the inputs to the scaled dot-product attention. The roots of Q and V are dotted, divided by d_k, and then multiplied by *V*. The query, key, and value are linearly projected into several subspaces via the multi-head attention mechanism [46] are computed using Eqs. (1)–(3).

$$\text{Attention }(\mathrm{Q},\mathrm{K},\mathrm{V}) = \text{softmax}\left(\frac{QK^t}{\sqrt{d_k}}\right)\mathrm{V} \tag{1}$$

Each subspace is a h*ead*, and each one computes its scaled dot product attention value independently. These values are then connected and projected again to acquire the final result.:

$$\text{MultiHead}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \text{Concathead}(\text{head}_1 \ldots, \text{head}_h)W^O \tag{2}$$

$$Where\ Head_i = \text{Attention}\big(\mathrm{QW}_i^Q, \mathrm{KW}_i^K, \mathrm{VW}_i^V \tag{3}$$

where the projections are parameter matrices

$$W_i^Q \epsilon\ R^{d_{model}xd_k}, W_i^k \epsilon R^{d_{model}xd_k}, W_i^k \epsilon R^{d_{model}xd_v}\ and\ W^O \epsilon R^{hd_vXd_{model}}$$

The first transaction model to completely generate its input and output representations using a self-attentive process without the use of RNNs or convolution is a transformer [47]. Fig. 1 also holds the encoder-decoder structure used by a proposed Improved Vision Transformer:

3.2. Feature classification using proposed ResNet-9 model

Consider input (W, P, K) where W represents input dimensions (n*m), K is the kernel size, P represents padding, and S is stride. Image size is reduced when transported from layer to layer [48]. To calculate the size of output dimension N_{od} obtained at each stage, using Eq. (4).

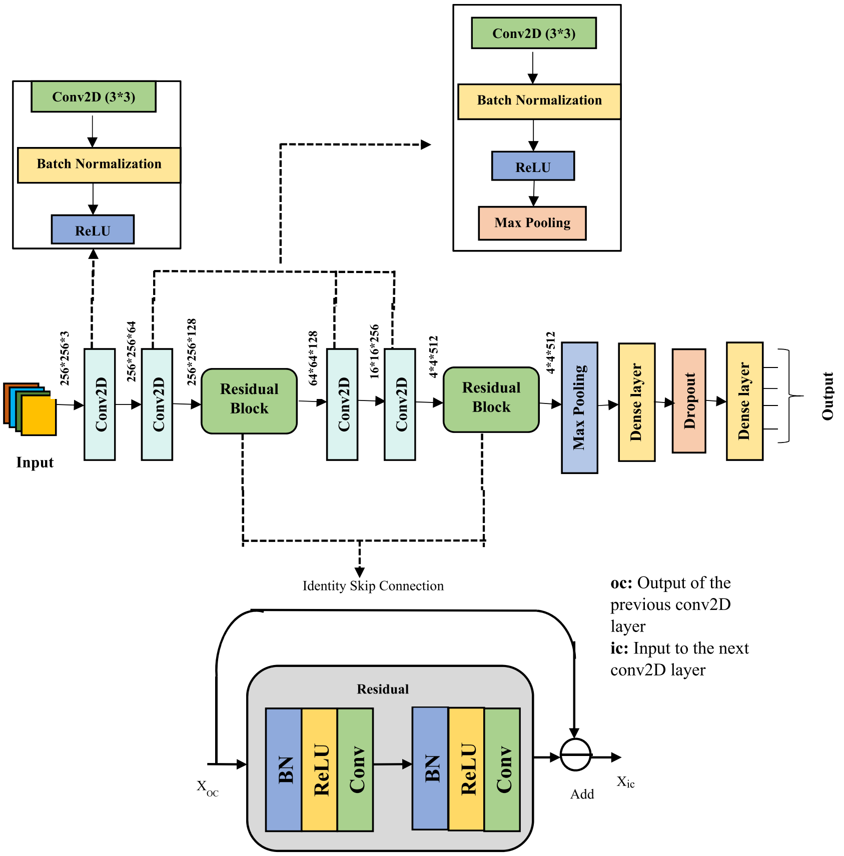


Fig. 2. Architecture of the customized ResNet9 model.

$$N_{od} = \left(\frac{(W - K + 2 * P)}{S}\right) + 1 \tag{4}$$

Moreover, The number of parameters obtained at convolutional layer N_{pc}, batch normalization N_{pbn}, and dense layer N_{pd} are computed using Eqs. (5)–(7), respectively.

$$N_{pc} = \left(\left(w_f * h_f * n_{pf}\right) + 1\right) * n_f \tag{5}$$

$$N_{pbn} = (n_{oc} * 2) \tag{6}$$

$$N_{pd} = \left(\left(n_{nc} * n_{np}\right) + 1 * n_{nc}\right) \tag{7}$$

In the above Eqs., w_f represents the filter width, h_f is the filter height, n_{pf} is the total filters in the previous layer, n_f denotes the total filters, n_{oc} is the total output channels, n_{nc} is total neurons in the current layer, and n_{np} represents the total neurons in the previous layer [49–51]. The total number of parameters obtained for the model is defined in Eq. (8).

$$N_p = \sum N_{pc} + \sum N_{pbn} + \sum N_{pd} \tag{8}$$

Each convolution layer produces the output of images with a smaller size. There is a residual layer or block after two convolution layers. In our proposal, we have also introduced a shortcut connection. There are 64,128,256 and 512 convolution kernels from layers 1 to 4. The last one features a 512-sized kernel. The kernel size of all convolutional layers is 3*3. ReLU was applied after every convolution of layers to activate neurons in the following layer. For the classifier layer, max-pooling and dropout layers with a value of 0.2 are added to avoid overfitting and improve model accuracy.

Fig. 2 shows three layers within the initial convolutional block: convolutional, batch normalization, and activation function Rectified Linear Unit (ReLU). After the initial convolutional block, each subsequent convolutional block has a max-pooling layer that produces a sharpened image compared to the original image. Shortcut networks correspond to a simple convolutional layer in the residual representation. After some weight layers, all the input values are added to the output with a skipped connection. Nine parameterized layers shall be used to implement the ResNet9 model with recurrent connections as the residual block of transmission

Table 1
List of parameters used for Hyper Parameter Tuning.

Learning Rate	Optimizer	Batch Size	Image Size	Loss Function	Weight decay	Gradient Clip
0.001	SGD	32	256 × 256	Categorical Cross Entropy	0.00018	0.13

learning. These residual [22] blocks must be applied to increase the surface depth and decrease the output size. The classifier layer, two dense layers, and activation functions ReLU and Softmax are used. The dropout layer shall be used with a value of 0.2 to avoid overfitting.

The algorithm of the ResNet 9 model for plant leaf image classification is given below:

```
#pre-processing images
begin
    for the number of input images ($X_{in}$) in the dataset:
        $X_{re}$ ←Resize ($X_{in}$)
        $X_{nr}$ ←Normalize ($X_{re}$)
        $X_{hf}$ ←RandomHorizontalFlip ($X_{nr}$)
        $X_{rr}$ ←RandomRotation ($X_{hf}$)
        $X_{CJ}$ ←ColorJitter ($X_{rr}$)
        $X_{hp}$ ← hyper_parameter_tuning($X_{CJ}$)
end for
end
#Training the network model
begin
    for the number of epochs do
begin
    for the number of $X_{CJ}$ images in a dataset:
        d1 = Conv1 ($X_{CJ}$)
        d2 = Conv2 (d1)
        d3 = residual_block (d1,d2)
        d4 = Conv3 (d3)
        d5 = Conv4 (d4)
        d6 = residual_block (d4,d5)
        d7 = classifier_block (d6)
end for
end for
end
#Testing the network model
begin
    for all input images ($X_{in}$) in the testing dataset:
        $X_t$ ← pre-processing ($X_{in}$)
end for
    for each test image ($X_t$):
        predicted_label (Y) ← test_image ($X_t$)
        target_label (T) ← input_image ($X_{in}$)
        if T = = Y($X_{in}$) then:
    The image is correctly classified
else:
    The image is not correctly classified
end for
end
```

The proposed Hierarchical Residual Vision Transformer is trained with ImageNet weights and fine-tuned with the hyperparameters mentioned in Table 1 on the Plant Village Dataset. The number of Units in the final dense layer of ResNet9 has been changed according to the number of classes.

3.2.1. Hyperparameter tuning

The collection of parameters influencing the model's learning process is termed hyperparameters. These parameters encompass aspects such as the number of layers, epochs, activation functions, learning rate, and more. The subsequent Table 1 provides an overview of the hyperparameter configurations employed in this proposed model.

Table 2
Extended plant village dataset.

Type of the Class	# Train Images	# Test Images
Local Crop Dataset		
Beans angular spot	184	25
Beans healthy	150	26
Beans rust	257	26
Cotton healthy	159	25
Cotton powdery mildew	105	25
Guava healthy	140	25
Guava red rust	99	26
Rice bacterial leaf blight	80	25
Rice brown spot	80	25
Rice leaf smut	80	25
Sugarcane healthy	150	25
Sugarcane red rot	148	26
Sugarcane red rust	150	26
Sub Total	1782	330
Plant Village Dataset		
Apple black rot	211	25
Apple black rust	217	25
Apple healthy	200	25
Apple scab	203	25
Blueberry healthy	231	25
Cherry healthy	222	25
Cherry powdery mildew	211	25
Corn(maize) common rust	218	25
Corn(maize) Gray leaf spot	158	25
Corn(maize) healthy	123	25
Corn(maize) northern leaf blight	127	25
Grape black rot	175	25
Grape Esca (black measles)	151	25
Grape healthy	183	25
Grape leaf blight	221	25
Orange Huanglongbing	151	25
Peach bacterial spot	185	25
Peach healthy	155	25
Pepper bell healthy	160	25
Pepperell bacterial spot	168	25
Potato early blight	163	25
Potato healthy	156	25
Potato late blight	163	25
Raspberry healthy	158	25
Soyabean healthy	183	25
Squash powdery mildew	169	25
Strawberry healthy	136	25
Strawberry leaf scorch	162	25
Tomato bacterial spot	135	25
Tomato yellow leaf curl virus	172	25
Tomato early blight	139	25
Tomato healthy	45	25
Tomato late blight	186	25
Tomato Leaf Mold	188	25
Tomato mosaic virus	168	25
Tomato Septoria leaf spot	195	25
Tomato spider mite	173	25
Tomato target spot	124	25
Sub Total	6485	950
Extended Plant Village Dataset		
Total Images	8267	1278

4. Experimental results and discussion

This section aims to provide a concise and informative overview of the findings derived from the conducted experiments of the Improved Vision Transformer and ResNet-9 model on the Plant Village Dataset.

4.1. About Plant Village Dataset

Images of healthy and diseased plant leaves are included in the Plant Village Dataset, which was compiled for developing various

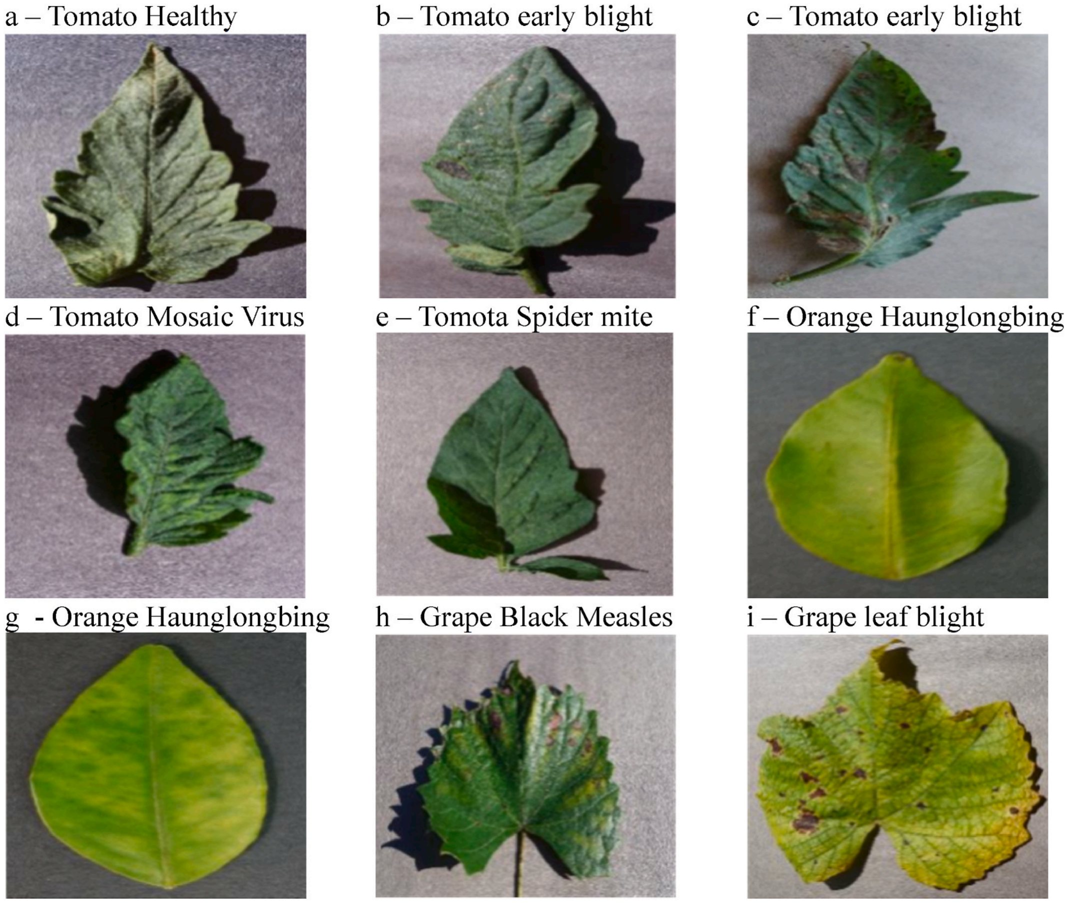


Fig. 3. Sample Plant leaf images of Plant Village Dataset - Tomotao Crop Images of Healthy and different diseases such as early blight, mosaic virus, and spotted spider mite are given in (A–E); Haunglongbing disease of Orange crop are given in (F–G) and Black measles and leaf blight disease samples of Grape crop are given in (H–I). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

learning models to detect and diagnose plant diseases. Performance evaluation of the proposed model is evaluated on three datasets i. e., Local Crop Dataset, customized Plant Village dataset, and Extended Plant Village Dataset. The local Crop dataset comprises local crops like beans, cotton, rice, sugarcane, and guava with 2112 images. We have divided this Local Crop dataset into train and test sets with 1782 and 330 images respectively. The customized Plant Village dataset includes over 7435 images of plant leaves of 38 classes from 14 different crop species. We have divided this dataset into train and test splits with 6485 and 950 images respectively. The extended Plant Village dataset consists of 9545 samples from 51 different classes collected from 19 different crop species. Out of 9545 samples, 8267 samples are used for model building and the remaining 1278 samples are used for performance evaluation. The crop-wise image details of the Local Crop dataset are given in Table 2.

To construct reliable plant disease prediction models, it is essential to gather high-quality data that is representative of the particular region, crop, and disease under investigation. Sample leaf images of Plant Village Dataset are given in Fig. 3.

4.2. Experimental setup

The experimental environment used for performing is a Windows 10 operating system with 16 GB RAM and 3 GB GPU. The Pytorch framework supports this training model. Google Colab is used to run Python code with GPU as runtime.

4.3. Performance metrics

To assess the performance of the proposed model, the following performance evaluation metrics are used.

Accuracy: The percentage of labeled samples correctly categorized in our model as a percentage of all samples is called accuracy as per Eq. (9).

Table 3
No.of Parameters of different models.

Name of the Model	No. of Parameters
InceptionV3	228,71,366
MobileNetV2	23,06,662
ResNet 50	236,08,202
Proposed model	65,96,403

Table 4
Comparative Analysis of Performance Measures for different models.

Name of the Model	Accuracy	Precision	Recall	F1-score
Local Dataset (13 Classes)				
Inception V3	91.7	0.927	0.917	0.922
MobileNet V2	92.8	0.935	0.935	0.935
Vision Transformer	78.5	0.804	0.825	0.814
ResNet 50	98.4	0.974	0.984	0.979
Proposed Model	97.5	0.989	0.984	0.986
Plant Village Dataset (38 Classes)				
Inception V3	94.3	0.9587	0.9212	0.9396
MobileNet V2	95.8	0.9647	0.9435	0.9540
Vision Transformer	80.7	0.824	0.815	0.8195
ResNet 50	99.1	0.9817	0.991	0.9863
Proposed Model	99.7	0.997	0.994	0.9955
Extended Plant Village Dataset (51 Classes)				
Inception V3	95.1	0.9687	0.95	0.9593
MobileNet V2	96.5	0.973	0.965	0.9690
Vision Transformer	81.9	0.835	0.837	0.8360
ResNet 50	99.4	0.994	0.996	0.9950
Proposed Model	99.7	0.996	0.997	0.9965

$$\text{Accuracy} = \frac{(\text{number of correctly classified samples})}{(\text{total number of samples})} \tag{9}$$

Precision: The percentage of the positive samples correctly categorized over total number of true positive and false positive samples as per Eq. (10).

$$\text{Precision} = \frac{(\text{number of correctly classified positive samples})}{(\text{total number of positive samples})} \tag{10}$$

Recall: It is a percentage of the positive samples correctly classified as the total number shall be defined as per Eq. (11).

$$\text{Recall} = \frac{(\text{number of correctly classified positive samples })}{(\text{total number of actual positive samples})} \tag{11}$$

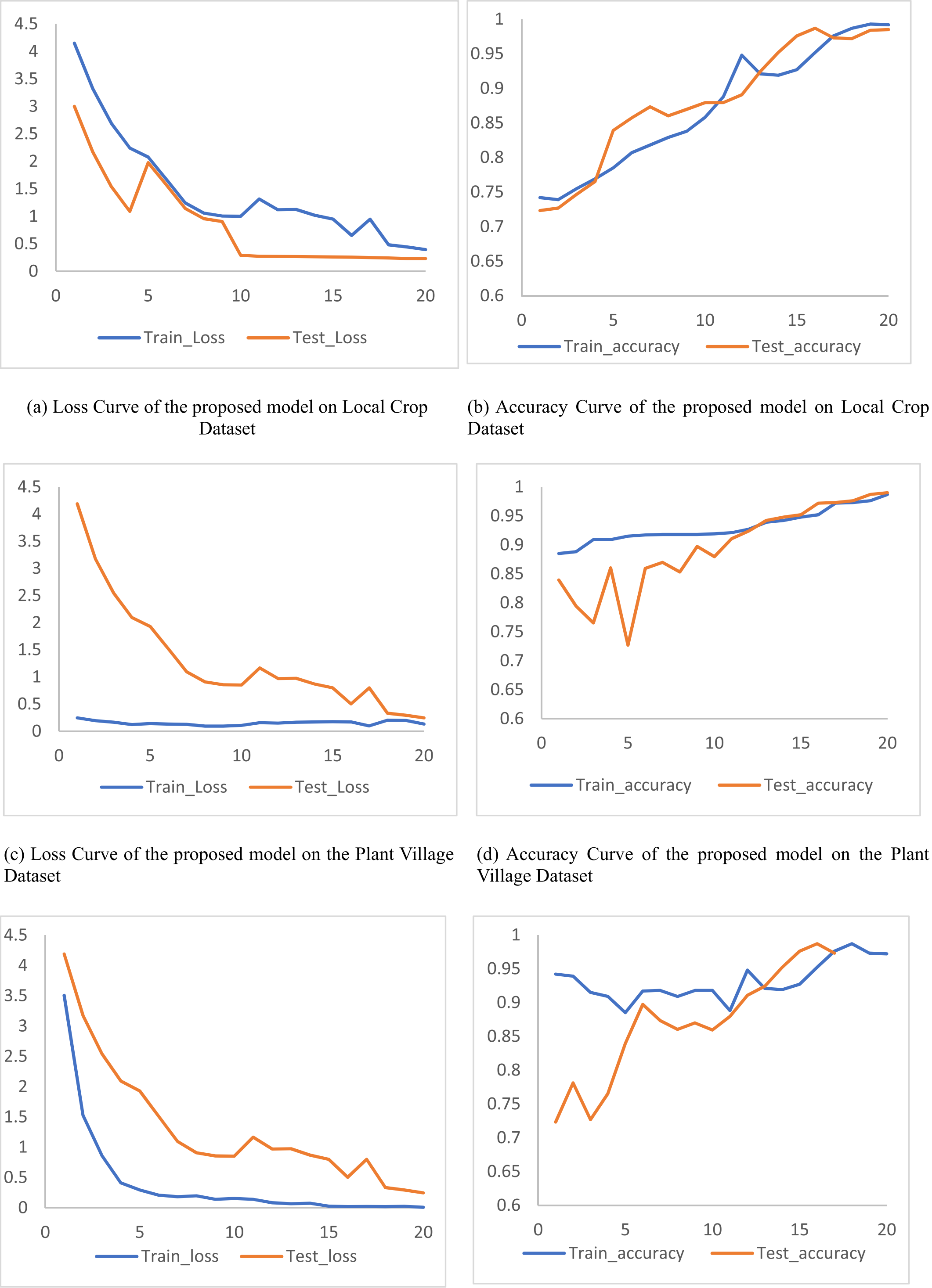
F1-score: This is a combination of both recall and precision as per Eq. (12). This can check how well your new sample could classify correctly.

$$\text{F1 score} = \frac{(\text{true positive})}{(\text{true positive}) + \left[\frac{\text{false positive}+\text{false negative}}{2}\right]} \tag{12}$$

ResNet50 is a deeper architecture with 50 layers, InceptionV3 contains multiple convolutional filters of different sizes, In MobileNetV2 computational resources are limited, whereas ResNet9 has only 9 layers. The number of parameters in a neural network generally increases with the number of layers. More layers mean more parameters because each layer typically consists of multiple neurons, and each neuron has its set of parameters (weights and biases). In the proposed work, transfer learning is adopted and a few of the layers are frozen from training. Inception V3, MobileNet V2, and ResNet 50 models have trainable parameters as 228,71,366; 23,06,662; and 236,08,202 respectively. Trainable parameters of all the above models are tabulated in Table 3.

The performance of the proposed method is compared with various deep learning models on Plant Village, Local datasets, and Extended Datasets. Accuracy, Precision, Recall, and F1-Score values with all these models are tabulated in Table 4. The proposed model outperformed all other classification models on all different datasets.

The accuracy and loss curves of the proposed Hierarchical Residual Vision Transformer on the Local Crop dataset are given in Fig. 4 (a) and (b) respectively. Further, Fig. 4(c) and (d) presents the accuracy and loss curves of proposed model on Plant Village Dataset are given. Finally, Fig. 4(e) and (f) describes the accuracy and loss curves of proposed Hierarchical Residual Vision Transformer on Extended Plant Village dataset are given.



(a) Loss Curve of the proposed model on Local Crop Dataset

(b) Accuracy Curve of the proposed model on Local Crop Dataset

(c) Loss Curve of the proposed model on the Plant Village Dataset

(d) Accuracy Curve of the proposed model on the Plant Village Dataset

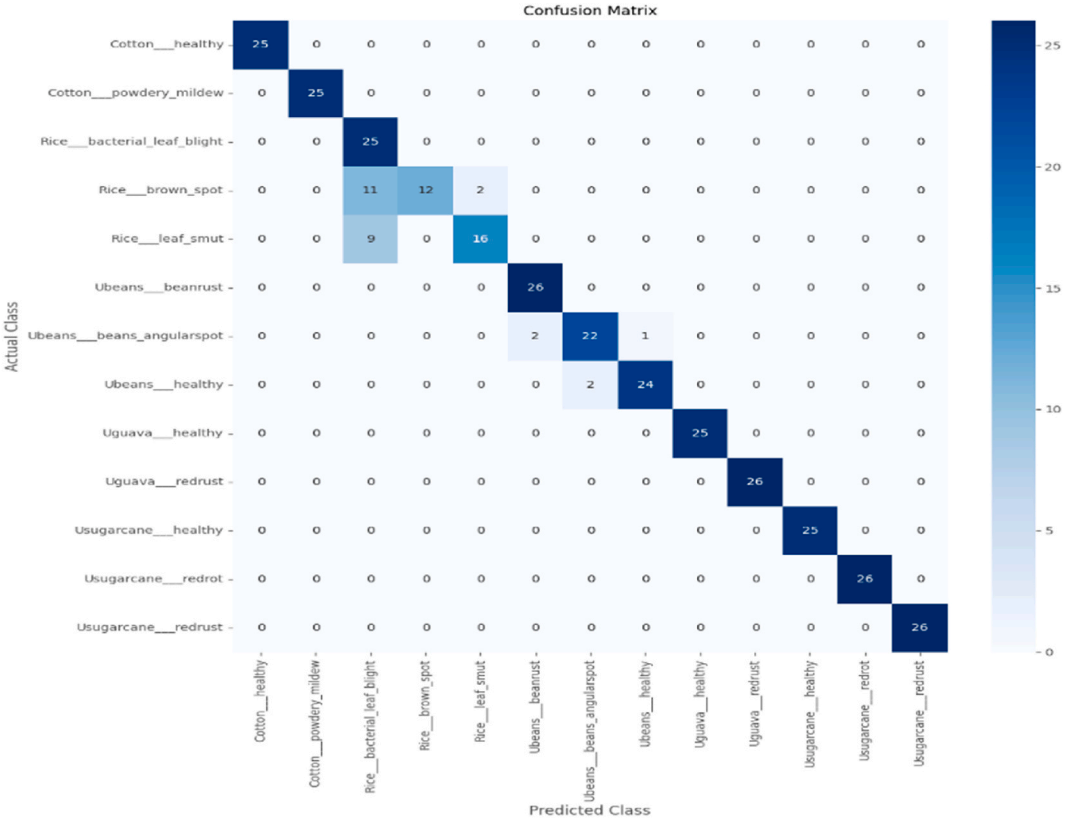
(e) Loss Curve of the proposed model on the Extended Plant Village Dataset

(f) Accuracy Curve of the proposed model on the Extended Plant Village Dataset

Fig. 4. Performance evaluation of the Proposed model on various datasets.

Table 5
Leaf disease prediction of different plants.

Crop	Leaf Image	Actual Label	Predicted Label
Cotton		Cotton Healthy	Cotton Healthy
Cotton		Cotton powdery mildew	Cotton powdery mildew
Guava		Guava red rust	Guava red rust
Rice		Rice bacterial leaf blight	Rice bacterial leaf blight
Beans		Beans angular spot	Beans angular spot
Sugarcane		Sugarcane red rust	Sugarcane red rot



	Cotton___healthy	Cotton___powdery_mildew	Rice___bacterial_leaf_blight	Rice___brown_spot	Rice___leaf_smut	Ubeans___beanrust	Ubeans___beans_angularspot	Ubeans___healthy	Uguava___healthy	Uguava___redrust	Usugarcane___healthy	Usugarcane___redrot	Usugarcane___redrust
Cotton___healthy	25	0	0	0	0	0	0	0	0	0	0	0	0
Cotton___powdery_mildew	0	25	0	0	0	0	0	0	0	0	0	0	0
Rice___bacterial_leaf_blight	0	0	25	0	0	0	0	0	0	0	0	0	0
Rice___brown_spot	0	0	11	12	2	0	0	0	0	0	0	0	0
Rice___leaf_smut	0	0	9	0	16	0	0	0	0	0	0	0	0
Ubeans___beanrust	0	0	0	0	0	26	0	0	0	0	0	0	0
Ubeans___beans_angularspot	0	0	0	0	0	2	22	1	0	0	0	0	0
Ubeans___healthy	0	0	0	0	0	0	2	24	0	0	0	0	0
Uguava___healthy	0	0	0	0	0	0	0	0	25	0	0	0	0
Uguava___redrust	0	0	0	0	0	0	0	0	0	26	0	0	0
Usugarcane___healthy	0	0	0	0	0	0	0	0	0	0	25	0	0
Usugarcane___redrot	0	0	0	0	0	0	0	0	0	0	0	26	0
Usugarcane___redrust	0	0	0	0	0	0	0	0	0	0	0	0	26

Fig. 5. Confusion matrix for the ResNet-9 model on the local crop dataset.

Actual and Predicted class labels of various test samples of the Plant Village Dataset are presented in Table 5.

In the comparison with MobileNet V2 and Inception V3, the proposed ResNet 9 has outperformed in the classification of class-wise diseases also. The number of correctly classified test images for each disease is plotted in the Confusion Matrix for locally grown crops of 13 classes of 5 crops is shown in Fig. 5. The proposed model is evaluated on the Plant Village Dataset (38 Classes) with a batch size of 32 and accuracies are presented in Fig. 6. The performance of the proposed model is very close to ResNet50 and ResNet101, but ResNet50 and 101 have high computational costs with tons of trainable parameters. The performance of the proposed model is compared with other state-of-the-art classification models such as ResNet 50, ResNet 101, MobileNet V2, GoogleNet etc.

The proposed model has outperformed GoogleNet, Inception V3, SqueezeNet, and AlexNet. But, the performance of the proposed model is very close to ResNet 50 and ResNet 101. However, the proposed VIT + ResNet9 is very shallower compared with the above ResNet 50 and ResNet 101.

In the next phase of experimentation, the performance of the proposed method is evaluated with Adam and SGD optimizers, and the results are presented in Fig. 7. The performance of the proposed model is evaluated with various initial learning rates including 10^{-3}, $10^{-4,}$ etc. During this part of experimentation, the Adam optimizer has given consistent results but SGD has given the highest accuracy at the initial learning rate of 10^{-3}.

The proposed model is evaluated for 10 epochs in 34.3 min on the customized Plant Village Dataset and acquired an accuracy of 93.5 %. Further, the model was evaluated for 20 epochs and yielded an accuracy of 99.7 % with model building time of 42.2 min. The computational complexity of the proposed model in terms of training times on various datasets is tabulated in Table 6.

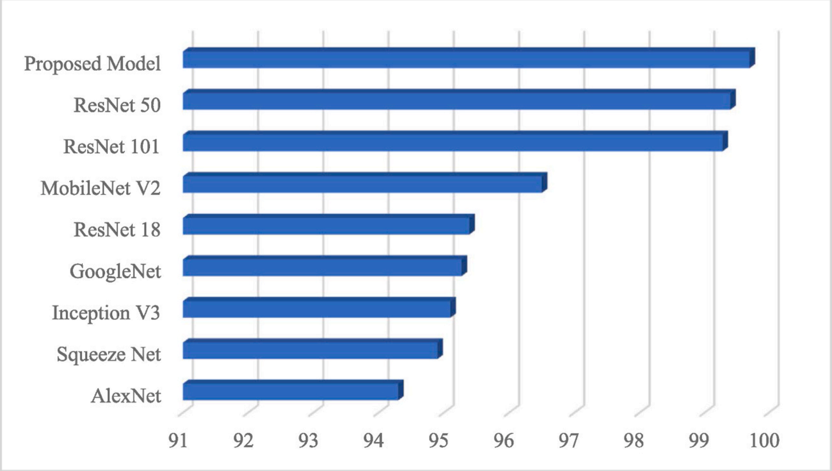


Fig. 6. Performance evaluation and comparison of the accuracy of various models on the Plant Village Dataset.

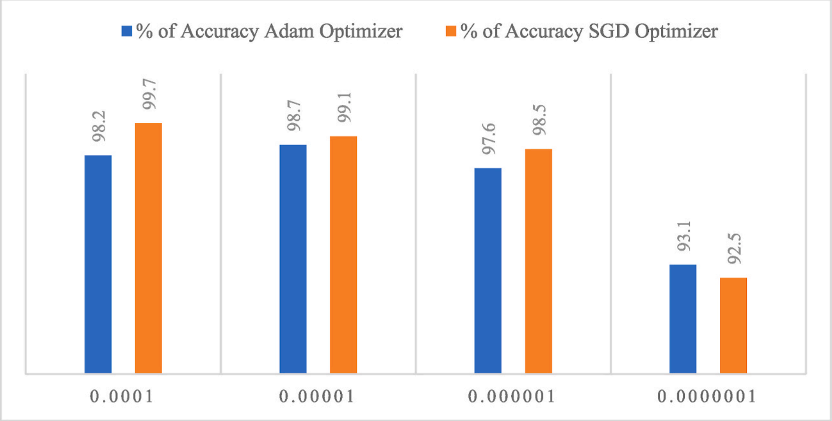


Fig. 7. Performance comparison of various optimizers on the Plant Village Dataset.

Table 6
Computational Complexity of proposed model training time.

Dataset	Proposed Model	ResNet 50	MobileNet V2	Inception V3
Local Crop Dataset	0:12:23	0:23:23	0:14:23	0:25:53
Customized Plant Village Dataset	0:34:34	0:45:29	0:26:34	0:37:22
Extended Plant Village Dataset	0:42:23	0:59:40	0:49:47	0:42:38

5. Conclusion

This research work is intended to detect different types of crop leaf diseases using the proposed Hierarchical Residual Vision Transformer model. To prevent overfitting and to reduce model complexity, a novel residual vision transformer is proposed using the Improved Vision Transformer and ResNet9. To increase the model's performance, hyperparameter tuning using Optuna was performed with 25 trials. After tuning the model, the best trail parameters were considered to train the model using Vision Transformer, and further features were classified using the lightweight ResNet9 classification model. The performance of the proposed model is evaluated on the Plant Village dataset in addition to some of the native crops including beans, sugarcane, cotton, rice, and guava. The inclusion of pre-processing tasks such as resizing, random flips, rotations, and image enhancement is supported in the extraction of discriminating features using Vision Transformer. In future work, authors aim to design and develop a lightweight deep neural network to overcome the limitations of real-time data. In addition to that, a multi-tasking classification model is needed to be designed to estimate the severity of the disease.

Data availability

The dataset are from PlantVillage (https://www.kaggle.com/datasets/emmarex/plantdisease).

CRediT authorship contribution statement

Sasikala Vallabhajosyula: Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Venkatramaphanikumar Sistla:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Venkata Krishna Kishore Kolli:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Lin, L. Gong, Y. Huang, C. Liu, J. Pan, Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network, Front. Plant Sci. 10 (2019), https://doi.org/10.3389/fpls.2019.00155.

[2] Y. Peng, Y. Wang, Leaf disease image retrieval with object detection and deep metric learning, Front. Plant Sci. 13 (2022), https://doi.org/10.3389/fpls.2022.963302.

[3] J. Liu, X. Wang, Plant diseases and pests detection based on deep learning: a review, Plant Methods 17 (2021) 22, https://doi.org/10.1186/s13007-021-00722-9.

[4] V. Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, Inf. Process. Agric. 4 (1) (2017) 41–49, https://doi.org/10.1016/j.inpa.2016.10.005.

[5] E.C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification, Comput. Electron. Agric. 161 (2019) 272–279, https://doi.org/10.1016/j.compag.2018.03.032.

[6] G. Wang, Y. Sun, J. Wang, Automatic image-based plant disease severity estimation using deep learning, Comput. Intell. Neurosci. 2017 (2017) 1–9, https://doi.org/10.1155/2017/2917536.

[7] G. Zhou, W. Zhang, A. Chen, M. He, X. Ma, Rapid detection of rice disease based on FCM-KM and faster r-CNN fusion, IEEE Access 7 (2019) 143190–143206, https://doi.org/10.1109/ACCESS.2019.2943454.

[8] F. Martinelli, R. Scalenghe, S. Davino, S. Panno, G. Scuderi, P. Ruisi, et al., Advanced methods of plant disease detection, A review. Agron. Sustain. Dev. 35 (2015) 1–25, https://doi.org/10.1007/s13593-014-0246-1.

[9] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, J. Guo, ISNet: shape matters for infrared small target detection. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 867–876, https://doi.org/10.1109/CVPR52688.2022.00095. New Orleans, LA, USA.

[10] M. Zhang, R. Zhang, J. Zhang, J. Guo, Y. Li, X. Gao, Dim2Clear network for infrared small target detection, IEEE Trans. Geosci. Rem. Sens. 61 (2023) 1–14, https://doi.org/10.1109/TGRS.2023.3263848. Art no. 5001714.

[11] Mingjin Zhang, Haichen Bai, Jing Zhang, Rui Zhang, Chaoyue Wang, Jie Guo, Xinbo Gao, RKformer: Runge-Kutta Transformer with Random-Connection Attention for Infrared Small Target Detection, 2022, pp. 1730–1738, https://doi.org/10.1145/3503161.3547817.

[12] Mingjin Zhang, Ke Yue, Jing Zhang, Yunsong Li, Xinbo Gao, Exploring Feature Compensation and Cross-Level Correlation for Infrared Small Target Detection, 2022, pp. 1857–1865, https://doi.org/10.1145/3503161.3548264.

[13] Mingjin Zhang, Handi Yang, Jie Guo, Yunsong Li, Xinbo Gao, Jing Zhang, "IRPruneDet: Efficient Infrared Small Target Detection via Wavelet Structure-Regularized Soft Channel Pruning" AAAI, 2024.

[14] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, J. Zhang, ESSAformer: efficient transformer for hyperspectral image super-resolution. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 23016–23027, https://doi.org/10.1109/ICCV51070.2023.02109. Paris, France.

[15] S. Nandhini, K. Ashokkumar, An automatic plant leaf disease identification using DenseNet-121 architecture with a mutation-based Henry gas solubility optimization algorithm, Neural Comput. Appl. 34 (2022) 5513–5534, https://doi.org/10.1007/s00521-021-06714-z.

[16] B. Sai Reddy, S. Neeraja, Plant leaf disease classification and damage detection system using deep learning models, Multimed. Tool. Appl. 81 (2022) 24021–24040, https://doi.org/10.1007/s11042-022-12147-0.

[17] P. Kaur, S. Harnal, V. Gautam, et al., A novel transfer deep learning method for detection and classification of plant leaf disease, J. Ambient Intell. Hum. Comput. (2022), https://doi.org/10.1007/s12652-022-04331-9.

[18] C.K. Rai, R. Pahuja, Classification of diseased cotton leaves and plants using improved deep convolutional neural network, Multimed. Tool. Appl. (2023), https://doi.org/10.1007/s11042-023-14933-w.

[19] Ananda S. Paymode, Vandana B. Malode, Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG, Artificial Intelligence in Agriculture 6 (2022) 23–33, https://doi.org/10.1016/j.aiia.2021.12.002.

[20] I. Bhakta, S. Phadikar, K. Majumder, et al., A novel plant disease prediction model based on thermal images using modified deep convolutional neural network, Precis. Agric. 24 (2023) 23–39, https://doi.org/10.1007/s11119-022-09927-x.

[21] V.K. Vishnoi, K. Kumar, B. Kumar, S. Mohan, A.A. Khan, Detection of apple plant diseases using leaf images through convolutional neural network, IEEE Access 11 (2023) 6594–6609, https://doi.org/10.1109/ACCESS.2022.3232917.

[22] S. Ahmed, M.B. Hasan, T. Ahmed, M.R.K. Sony, M.H. Kabir, Less is more: lighter and faster deep neural architecture for tomato leaf disease classification, IEEE Access 10 (2022) 68868–68884, https://doi.org/10.1109/ACCESS.2022.3187203.

[23] T. Daniya, S. Vigneshwari, A novel Moore-Penrose pseudo-inverse weight-based Deep Convolution Neural Network for bacterial leaf blight disease detection system in rice plant, Adv. Eng. Software 174 (2022) 103336, https://doi.org/10.1016/j.advengsoft.2022.103336.

[24] Prabhjot Kaur, Shilpi Harnal, Vinay Gautam, Mukund Pratap Singh, Santar Pal Singh, An approach for characterization of infected area in tomato leaf disease based on deep learning and object detection technique, Eng. Appl. Artif. Intell. 115 (2022) 105210, https://doi.org/10.1016/j.engappai.2022.105210.

[25] Wenxia Bao, Tao Cheng, Xin-Gen Zhou, Wei Guo, Yuanyuan Wang, Xuan Zhang, Hongbo Qiao, Dongyan Zhang, An improved DenseNet model to classify the damage caused by cotton aphid, Comput. Electron. Agric. 203 (2022) 107485, https://doi.org/10.1016/j.compag.2022.107485.

[26] Xijian Fan, Luo Peng, Yuen Mu, Rui Zhou, Tardi Tjahjadi, Yi Ren, Leaf image-based plant di image-base edification using transfer learning and feature fusion, Comput. Electron. Agric. 196 (2022) 106892, https://doi.org/10.1016/j.compag.2022.106892.

[27] Le Yang, Xiaoyun Yu, Shaoping Zhang, Huibin Long, Huanhuan Zhang, Shuang Xu, Yuanjun Liao, GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases, Comput. Electron. Agric. 204 (2023) 107543, https://doi.org/10.1016/j.compag.2022.107543.

[28] K. Roy, et al., Detection of tomato leaf diseases for agro-based industries using novel PCA DeepNet, IEEE Access 11 (2023) 14983–15001, https://doi.org/10.1109/ACCESS.2023.3244499.

[29] Mingxuan Li, Guoxiong Zhou, Aibin Chen, Liujun Li, Yahui Hu, Identification of tomato leaf diseases based on LMBRNet, Eng. Appl. Artif. Intell. 123A (2023) 106195, https://doi.org/10.1016/j.engappai.2023.106195.

[30] Kantha Raju Kanaparthi, S. Sudhakar Ilango, A survey on training issues in chili leaf diseases identification using deep learning techniques, Proc. Comput. Sci. 218 (2023) 2123–2132, https://doi.org/10.1016/j.procs.2023.01.188.

[31] Oliva Debnath, Himadri Nath Saha, An IoT-based intelligent farming using CNN for early disease detection in rice paddy, Microprocess. Microsyst. 94 (2022) 104631, https://doi.org/10.1016/j.micpro.2022.104631.

[32] Changjian Zhou, Yujie Zhong, Sihan Zhou, Jia Song, Wensheng Xiang, Rice leaf disease identification by residual-distilled transformer, Eng. Appl. Artif. Intell. 121 (2023) 106020, https://doi.org/10.1016/j.engappai.2023.106020.

[33] A. Yadav, U. Thakur, R. Saxena, et al., AFD-Net: apple Foliar Disease multi-classification using deep learning on plant pathology dataset, Plant Soil 477 (2022) 595–611, https://doi.org/10.1007/s11104-022-05407-3.

[34] M. Moussafir, H. Chaibi, R. Saadane, et al., Design of efficient techniques for tomato leaf disease detection using genetic algorithm-based and deep neural networks, Plant Soil 479 (2022) 251–266, https://doi.org/10.1007/s11104-022-05513-2.

[35] Shunping Ji, et al., 3D convolutional neural networks for crop classification with multi-temporal remote sensing images, Rem. Sens. 10 (1) (2018) 75.

[36] C. Zhou, S. Zhou, J. Xing, J. Song, Tomato leaf disease identification by restructured deep residual dense network, IEEE Access 9 (2021) 28822–28831, https://doi.org/10.1109/ACCESS.2021.3058947.

[37] Yun Zhao, Cheng Sun, Xing Xu, Jiagui Chen, RIC-Net: a plant disease classification model based on the fusion of Inception and residual structure and embedded attention mechanism, Comput. Electron. Agric. 193 (2022) 106644, https://doi.org/10.1016/j.compag.2021.106644.

[38] Xuechen Li, Xiuhua Li, Shimin Zhang, Guiying Zhang, Muqing Zhang, Heyang Shang, SLViT: Shuffle-Convolution-Based Lightweight Vision Transformer for Effective Diagnosis of Sugarcane Leaf Diseases, vols. 1319–1578, Journal of King Saud University - Computer and Information Sciences, 2022, https://doi.org/10.1016/j.jksuci.2022.09.013.

[39] K. Liu, X. Zhang, PiTLiD: identification of plant disease from leaf images based on convolutional neural network, IEEE ACM Trans. Comput. Biol. Bioinf 20 (2023) 1278–1288, https://doi.org/10.1109/TCBB.2022.3195291.

[40] Sathian Dananjayan, et al., Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images, Comput. Electron. Agric. 193 (2022) 106658.

[41] H. Alaeddine, M. Jihene, Plant leaf disease classification using Wide Residual Networks, Multimed. Tool. Appl. (2023), https://doi.org/10.1007/s11042-023-15226-y.

[42] Mehdhar SAM. Algaashani, et al., Tomato leaf disease classification by exploiting transfer learning and feature concatenation, IET Image Process. 16 (3) (2022) 913–925.

[43] Rhea Patel, Bappa Mitra, Madhuri Vinchurkar, et al., A review of recent advances in plant-pathogen detection systems, Heliyon 8 (12) (2022) e11855, https://doi.org/10.1016/j.heliyon.2022.e11855.

[44] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data 8 (2021) 1–74, https://doi.org/10.1186/s40537-021-00444-8.

[45] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495, https://doi.org/10.1109/TPAMI.2016.2644615.

[46] J. Chen, J. Chen, D. Zhang, Y. Sun, Y.A. Nanehkaran, Using deep transfer learning for image-based plant disease identification, Comput. Electron. Agric. 173 (April) (2020) 105393, https://doi.org/10.1016/j.compag.2020.105393.

[47] V.S. Dhaka, S.V. Meena, G. Rani, D. Sinwar, M.F. Ijaz, M. Woźniak, A survey of deep convolutional neural networks applied for prediction of plant leaf diseases, Sensors 21 (14) (2021) 4749, https://doi.org/10.3390/s21144749.

[48] K. Kc, Z. Yin, M. Wu, Z. Wu, Depthwise separable convolution architectures for plant disease classification, Comput. Electron. Agric. 165 (December 2018) (2019) 104948, https://doi.org/10.1016/j.compag.2019.104948.

[49] G. Kassa, T. Bekele, S. Demissew, T. Abebe, Plant species diversity, plant use, and classification of agroforestry home gardens in southern and southwestern Ethiopia, Heliyon 9 (6) (2023) e16341.

[50] R.I. Hasan, S.M. Yusuf, L. Alzubaidi, Review of the state of the art of deep learning for plant Diseases: a broad analysis and discussion, Plants 9 (2020) 1–25, https://doi.org/10.3390/plants9101302.

[51] Kabiraz, Meera Probha, Priyanka Rani Majumdar, MM Chayan Mahmud, Shuva Bhowmik, Azam Ali, Conventional and advanced detection techniques of foodborne pathogens: a comprehensive review, Heliyon 9 (4) (2023) e15482, https://doi.org/10.1016/j.heliyon.2023.e15482.