agronomy

MDPI

Article

ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture

Utpal Barman [1], Parisimita Sarma [2], Mirzanur Rahman [2], Vaskar Deka [2], Swati Lahkar [2], Vaishali Sharma [2] and Manob Jyoti Saikia [3,*]

[1] Faculty of Computer Technology, Assam down town University, Guwahati 781026, India
[2] Department of Information Technology, Gauhati University, Guwahati 781014, India
[3] Department of Electrical Engineering, University of North Florida, Jacksonville, FL 32224, USA
* Correspondence: manob.saikia@unf.edu

Abstract: Invading pests and diseases always degrade the quality and quantity of plants. Early and accurate identification of plant diseases is critical for plant health and growth. This work proposes a smartphone-based solution using a Vision Transformer (ViT) model for identifying healthy plants and unhealthy plants with diseases. The collected dataset of tomato leaves was used to collectively train Vision Transformer and Inception V3-based deep learning (DL) models to differentiate healthy and diseased plants. These models detected 10 different tomato disease classes from the dataset containing 10,010 images. The performance of the two DL models was compared. This work also presents a smartphone-based application (Android App) using a ViT-based model, which works on the basis of the self-attention mechanism and yielded a better performance (90.99% testing) than Inception V3 in our experimentation. The proposed *ViT-SmartAgri* is promising and can be implemented on a colossal scale for smart agriculture, thus inspiring future work in this area.

Keywords: vision transformer; tomato diseases; ViT; Inception V3; android app; smart agriculture; deep learning; plant village

check for updates

Citation: Barman, U.; Sarma, P.; Rahman, M.; Deka, V.; Lahkar, S.; Sharma, V.; Saikia, M.J. *ViT-SmartAgri*: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture. *Agronomy* **2024**, *14*, 327. https://doi.org/10.3390/agronomy14020327

Academic Editor: Andrea Sciarretta

Received: 4 January 2024
Revised: 30 January 2024
Accepted: 2 February 2024
Published: 4 February 2024

1. Introduction

Tomato farming is one of the most profitable agribusinesses. After potatoes, it is the second most important crop in the world. Tomatoes are high in vitamins like A and C as well as antioxidants. Various diseases and pests can affect tomatoes during their growing stage. Lower yields or even crop loss could follow if control is not carried out promptly. The most important stage is to precisely detect disease types and insect pests in order to effectively apply mitigation to diseases and pests to help vegetable farmers increase tomato yields [1].

Conventional methods of tomato disease identification rely on visual inspection by skilled agronomists, which can be time-consuming, labor-intensive, and subject to human error. As a result, there is a growing interest in leveraging the advancement in computer technologies [2] to automate and enhance the process of disease identification in plants. The goal of this work is to innovate technologies for preventing diseases in plant leaves and to increase farmers' knowledge of these technologies. Since tomato is a commonly available vegetable, this work focused on machine learning and smartphone-based technology for developing an effective solution for tomato plant leaf disease detection. As opposed to manual plant leaf disease recognition, computer vision-based technology has the potential for directly detecting plant leaf diseases using raw input from camera images captured using a smartphone.

Convolutional Neural Network (CNN) has emerged as a predominant architecture for visual recognition in recent years, demonstrating notable success. The incorporation of convolution operations, imparting crucial inductive biases such as translation equivalence, is widely acknowledged as a key factor contributing to this success in object identification. Recent studies by Barman et al. [3] and Agarwal et al. [4] have reported successful identification of plant diseases in citrus and tomato leaves, respectively through the utilization of MobileNet and CNN. Again, MobileNet is reported as demonstrating 99.30% accuracy by Ahmed et al. [5]. While CNNs and pretrained models have been extensively employed for tomato plant disease detection [5–7], noteworthy research by Zhao et al. [7] proposed a recognition model for tomato disease without relying on convolutions. In a similar vein, Karthik et al. [8] introduced an attention-based mechanism for detecting diseases in tomato leaves. These advancements underscore the diverse approaches within the realm of neural networks for addressing the challenges of plant disease recognition.

Recent developments in deep learning have shown astounding success in a variety of computer vision applications, including object detection and image classification. One such groundbreaking development is the Vision Transformer (ViT) architecture [9], which has shown exceptional performance on large-scale image classification tasks [10–12]. Unlike CNNs, which have been widely used in computer vision, ViT operates solely on self-attention mechanisms, enabling it to capture global relationships between image patches and learn representations effectively [13].

Like ViT, Inception was used by Sakkarvarthi et al., 2022 [14], to detect tomato illness with a validation accuracy of 80%, which is lower than the accuracy reported in this study for Inception V3. They also reported a very low accuracy using VGG and ResNet. ResNet is again reported as a good technique for tomato leaf disease detection by Zhang et al. [15]. A total of 92% accuracy was reported by Liu and Wang [16] in their study using the improved Yolo V3. An accuracy of up to 95.65% for DensNet161, 94.93% for DensNet121, and 90.58% for VGG16 was reported by Ouhami et al. [17].

In the context of tomato leaf disease detection through deep learning, despite various techniques explored by researchers [18,19], several questions persist:

(i) The challenge of assigning equal weights to all deep features in deep learning models prompts the question of whether an attention mechanism is essential [20].
(ii) Researchers question whether integrating attention mechanisms in deep learning or transfer learning-based models could address the risk of overlooking significant features during classification.

The Vision Transformer (ViT) model holds significance in addressing the unanswered questions in tomato leaf disease detection using deep learning:

(i) ViT introduces a transformative approach by leveraging self-attention mechanisms, allowing the model to focus on relevant features without manual extraction. Its architecture inherently addresses the need for efficient automated feature extraction, potentially mitigating concerns about excluding vital information in the detection process [21].
(ii) ViT inherently incorporates attention mechanisms through its self-attention mechanisms. Unlike traditional models, ViT dynamically assigns weights to different regions of the input image during processing. This addresses the challenge of equal weight distribution in deep features, potentially ensuring that significant features receive adequate attention during the classification of tomato leaf diseases.

In essence, the ViT model, with its attention mechanisms and unique architecture, provides a promising avenue for advancing automated feature extraction and addressing the challenges posed by traditional approaches. It aligns with the evolving needs of deep learning in the context of tomato leaf disease detection, offering potential solutions to the identified questions in the field.

Considering the above points, ViT was used in this study, which applies the pure self-attention-based architecture to sequences of picture patches and achieves competitive performance on the tough ImageNet classification task when compared to CNNs [5]. The outcomes of this research have the potential to revolutionize the field of tomato disease management by providing an efficient, automated, and accurate solution for disease detection.

By enabling early detection and intervention, farmers can minimize crop losses, optimize resource allocation, and contribute to sustainable and smart agricultural practices [22].

2. Materials and Methods

Identifying diseases in tomato leaves is a complex task due to the diverse range of diseases that affect tomatoes. Recently, machine and deep learning have become powerful tools for computer-assisted disease detection in tomatoes. These techniques use deep neural networks, which can understand intricate patterns from extensive image data. The recommended approach of this study for detecting tomato diseases through deep learning involves several essential steps. Firstly, a dataset of tomato leaf images is compiled, featuring examples of both diseased and healthy leaves. In this study, the dataset is a plant village dataset. Following basic preprocessing, these images were used to train the inception V3 and ViT models to recognize detailed patterns within the images. Once trained, the model can classify newly acquired images into either the healthy class or one of the disease classes. Following the procedure, the best model was deployed in the android application for real-time testing.

2.1. Research Steps

The research outlines two steps, both of which are illustrated in the diagrams in Figure 1.

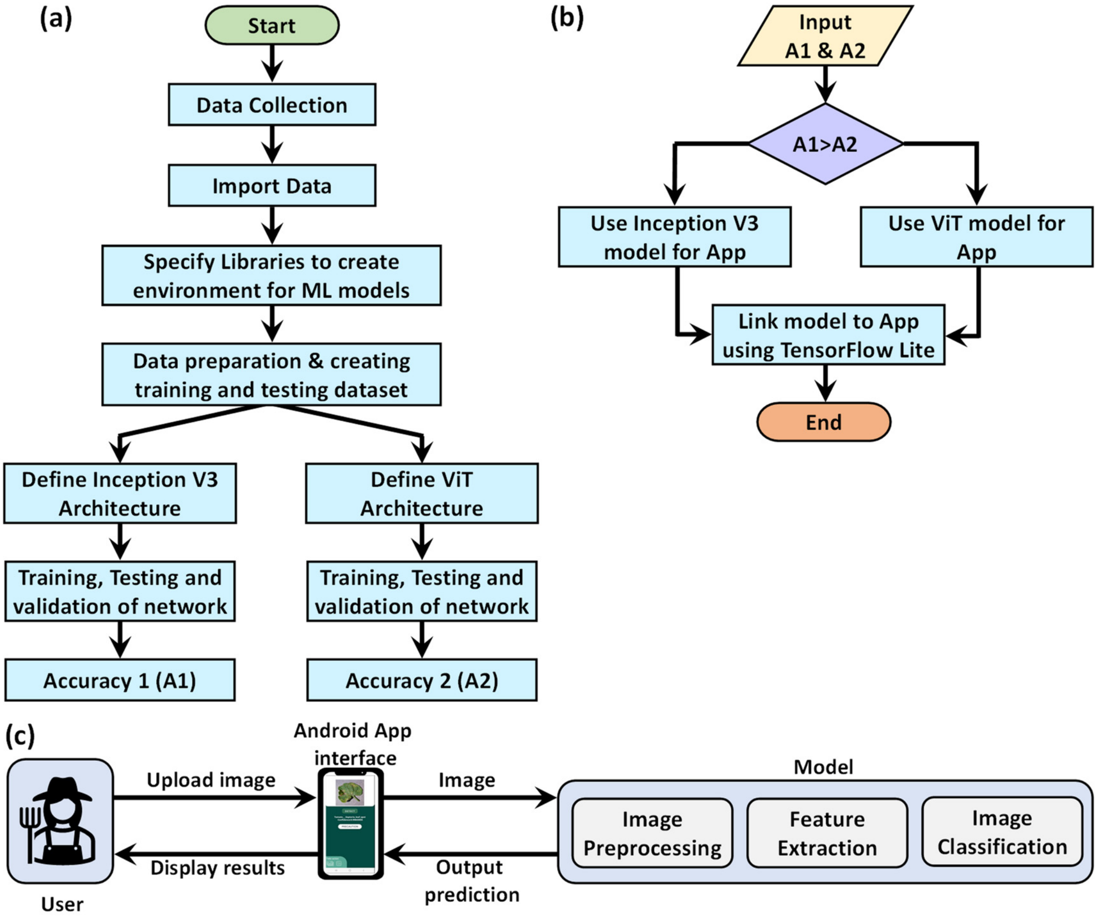

(a)
Start
Data Collection
Import Data
Specify Libraries to create environment for ML models
Data preparation & creating training and testing dataset
Define Inception V3 Architecture
Training, Testing and validation of network
Accuracy 1 (A1)
Define ViT Architecture
Training, Testing and validation of network
Accuracy 2 (A2)
(b)
Input A1 & A2
A1>A2
Use Inception V3 model for App
Use ViT model for App
Link model to App using TensorFlow Lite
End
(c)
User
Upload image
Display results
Android App interface
Image
Output prediction
Model
Image Preprocessing
Feature Extraction
Image Classification

Figure 1. Flowchart and schematic diagram of the entire system for tomato disease detection. (**a**) CNN- and ViT-based model development; (**b**) model evaluation and implementation; and (**c**) Smartphone Android App-based system for model deployment.

I. After building the CNN- and ViT-based models, this work evaluated their accuracy so that the highest accuracy model could be considered as the best model.

II. Adopting the best model allows for development of the smartphone application (Android App).

Step I involved the tasks shown in Figure 1a,b, and Figure 1c shows the schematic diagram for the entire system developed in Step II.

2.2. Dataset Information

The dataset was obtained from the plant village dataset [23]. The plant village dataset comprises 54,303 images of both healthy and unhealthy leaves, categorized into 38 groups based on species and disease, including tomato leaves. In this study, tomato leaves were considered as the plant of study. Our dataset consists of 10 different classes of 10,010 images. The tomato diseases that our developed system detected are shown in Table 1. In addition to this, we had 1 remaining class, consisting of healthy leaves. 10 different images of 10 different classes are shown in Figure 2.

Table 1. List of tomato diseases in the dataset.

Sl No.	Disease Name	Total Number of Images
1	Bacterial spot disease	1010
2	Early Blight	1000
3	Late Blight	1000
4	Leaf Mold	1000
5	Septoria Leaf Spot	1000
6	Spider Mites	1000
7	Target Spot	1000
8	Mosaic Virus	1000
9	Yellow Leaf Curl Virus	1000
10	Healthy Leaf	1000

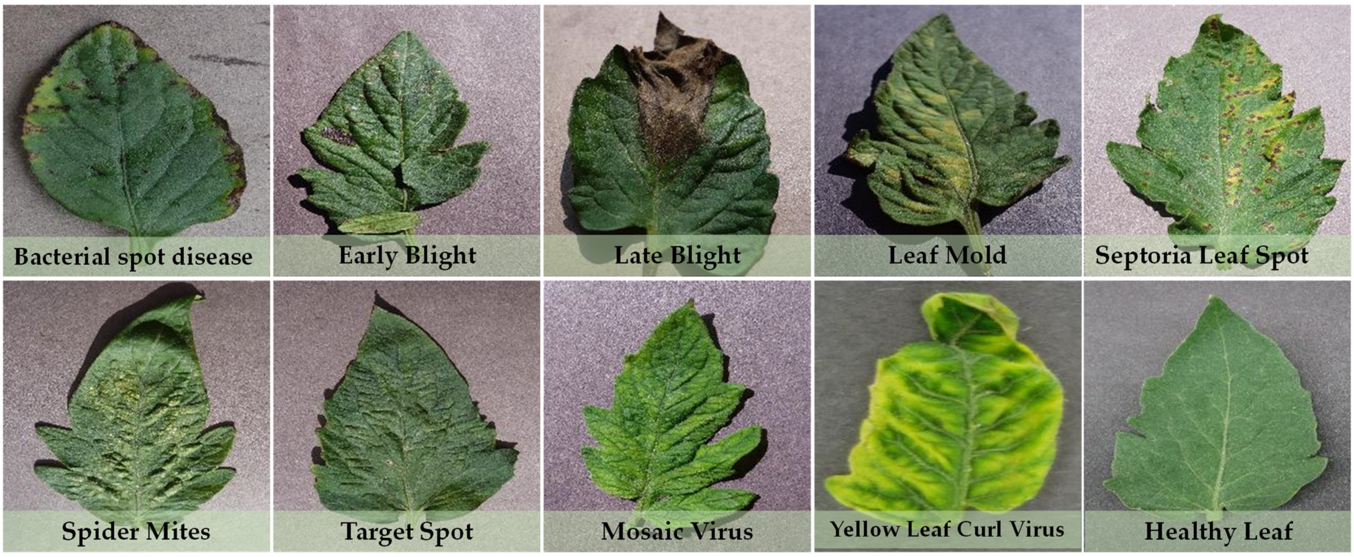


Figure 2. Samples of tomato leaf diseases.

2.3. Enhancing Tomato Disease Classification

Inception V3 is a deep learning architecture that employs amalgamation of convolution, polling, and fully connected layers [3]. It employs a unique module called the "Inception module" that enables the network to capture features at different scales by using multiple filter sizes. Inception V3 is pre-trained on a large-scale dataset, such as ImageNet, which enhances its ability to extract meaningful features from images. To employ the Inception V3 model for tomato disease classification, a dataset containing both infected and healthy leaves was built. The dataset was resized and split by assigning the training and validation path. The Inception V3 model was imported through the transfer learning technique. Here, we used ImageNet weights and excluded the first layer so that the parameters depended on our image size. Moreover, it was not required to train existing weights; therefore, the layer.trainable value was kept FALSE during the implementation. The layers were flattened. The softmax activation function was used in the output layer of the model. It guesses a multinomial probability distribution. In general, softmax is applied to multiclass classification where each sample may be a member of only one class. Each resultant value signifies predictable probability of an input piece to the corresponding class. The Adam optimizer with categorical_crossentropy was used in the compilation step. Lastly came the fitting step, where the model was compiled through a maximum of 30 epochs to avoid overfitting. The model summary of the Inception V3 is presented in Table 2.

Table 2. Model summary of Inception V3 for tomato disease classification.

Layers	Output Shape	Parameters
inception_v3 (Functional)	(none, 2048)	21,802,748
Dense Layer	(None, 10)	20,490

Like Inception V3, we implemented the ViT model and started importing the dataset. Then, the data was pre-processed, which included resizing the images into 64 × 64 and splitting the dataset into a training and testing set. The total number of images in the dataset was 10,010, comprising the training, validation, and testing sets. There was no duplication of images between these sets. The training set, which included 9000 images, was distinct from the testing set, which contained 1010 images. Out of 9000 images of training, 10% of the images were used for validation. This division ensures that the model is trained on a specific set of data and then tested on a separate set to evaluate its generalization performance. The allocation of images into training and testing sets was random, maintaining a clear distinction between the datasets to ensure the validity of the model's assessment. Once the data to be used were prepared, we moved on to creating a conducive environment for implementing ViT. The ViT model was implemented according to the Figure 3 and using the hyperparameters shown in Table 3. The preliminary learning rate was fixed at 0.0001, initial batch size was 256, and the model was trained for 30 epochs.

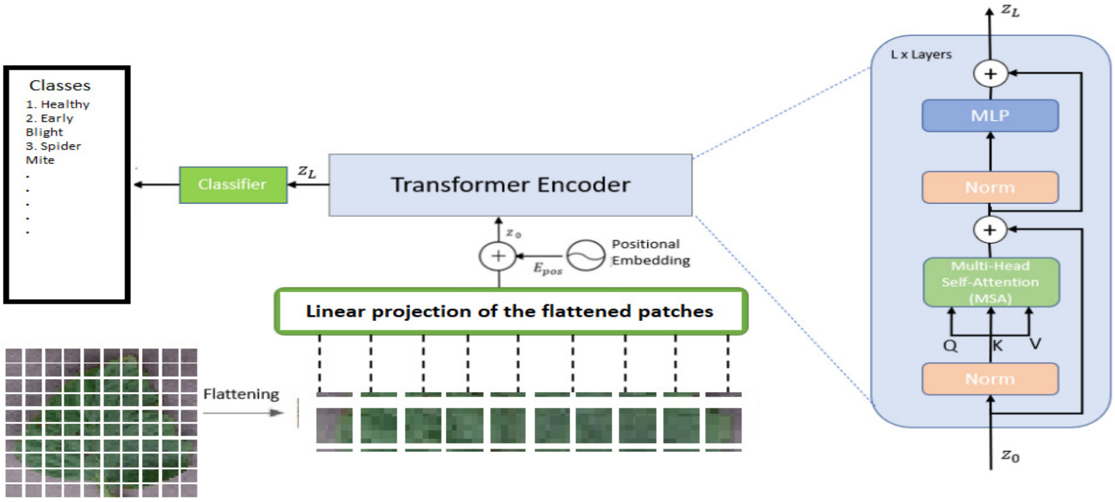


Figure 3. Vison transformer (ViT) architecture for tomato leaf disease detection.

Table 3. Hyperparameters of ViT model used in this study.

Serial No.	Hyperparameter Name	Value
1	Input size	64
2	Learning rate	0.001
3	Weight_decay	0.0001
4	Batch_Size	256
5	Transformer Dim	2
6	Projection Layers	8

In a number of image identification tasks, the ViT model displayed impressive performance, occasionally even exceeding conventional convolutional neural networks (CNNs). It is scalable to greater image sizes, since it can handle images with a fixed-size set of patches. The general ViT model works using the following steps:

- Input image standardization: One dimensional vector is created by linearly flattening the patches of fixed-size blocks prepared from the input image, and it serves as the model's input tokens.
- Embedding: Every patch is converted into an embedding, a fixed-dimensional vector representation.
- Positional information: To determine the relative location of the patches, positional encodings are used.
- Transformer Encoder: The nucleus of the ViT model is the Transformer Encoder. It consists of a number of layers, each of which contains neural network building blocks for feed-forward and self-attention.
- Classification: After going by the transformer, a single feature vector is created and sent for classification to a fully connected CNN layer.
- Training: The weights used in embedding and transformer layers are revised using optimization techniques like stochastic gradient descent or the Adam optimizer.
- Inference: An input is sent to the already trained model, and the result is achieved from the forward pass and output probabilities of the model.

The model summary of the ViT is presented in Table 4.

Table 4. Model summary of ViT for tomato disease classification.

Layers	Output Shape	Parameters
Add_14 (Add)	(none, 100, 64)	0
Layer Normalization	(none, 100, 64)	128
Dense Layer	(None, 100, 128)	8320
Dropout Layer	(None, 100, 128)	0
Dense Layer	(None. 100, 64)	8256
Dropout Layer	(None. 100, 64)	0
Add	(None. 100, 64)	0
Layer Normalization	(None. 100, 64)	128
Flatten	(None, 6400)	0
Dropout Layer	(None, 100, 64)	0
Dense Layer	(None, 2048)	13,109,248
Dropout Layer	(None, 2048)	2,098,176
Dense Layer	(None, 1024)	0
Dropout Layer	(None, 1024)	10,250
Dense Layer	(None, 10)	

3. Results and Discussion

As discussed above, the Inception V3 and ViT models were used for tomato disease detection. At first, the Inception V3 model was implemented with 30 epochs. The model yielded an accuracy of 94.80% and a 1.00% loss. The sparse categorical cross-entropy loss-based Adam optimizer was used in each layer for the optimization of the error and calculation of the loss in each layer. Initially, Inception V3 reported 65% accuracy in training and 70% accuracy in validation. With the increase of the epochs, the overall accuracy of the model was increased, and the losses were also reduced accordingly. The overall accuracy of the model from epoch 1 to 30 is shown in Table 5 and Figure 4.

Table 5. Training and validation accuracy of the Inception V3 for tomato disease classification.

Number of Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
01/30	3.0451	0.6467	2.1386	0.7072
02/30	1.9651	0.7683	1.4701	0.8226
03/30	1.8504	0.8018	1.5862	0.8269
04/30	1.6156	0.8311	2.8977	0.7591
05/30	1.6753	0.8430	1.8865	0.8792
06/30	1.7831	0.8443	1.0311	0.8912
07/30	1.6320	0.8590	1.7337	0.8567
08/30	1.5297	0.8720	1.4035	0.8873
09/30	1.5834	0.8715	1.0898	0.8992
10/30	1.3658	0.8924	1.2282	0.8898
11/30	1.3528	0.8951	1.1552	0.8896
12/30	1.2853	0.9012	1.1625	0.8925
13/30	1.2614	0.9025	1.1836	0.9053
14/30	1.2685	0.9058	1.5232	0.8903
15/30	1.2456	0.9073	1.1034	0.9021
16/30	1.2123	0.9112	1.0567	0.9057
17/30	1.1987	0.9135	1.0243	0.9083
18/30	1.1752	0.9174	1.0056	0.9105
19/30	1.1598	0.9201	0.9867	0.9132
20/30	1.1405	0.9232	0.9745	0.9156
21/30	1.1287	0.9254	0.9632	0.9181
22/30	1.1124	0.9283	0.9456	0.9203
23/30	1.0982	0.9310	0.9324	0.9230
24/30	1.0856	0.9332	0.9211	0.9254
25/30	1.0723	0.9359	0.9102	0.9278
26/30	1.0589	0.9381	0.9023	0.9302
27/30	1.0456	0.9409	0.8921	0.9325
28/30	1.0321	0.9430	0.8809	0.9351
29/30	1.0189	0.9458	0.8712	0.9376
30/30	1.0056	0.9480	0.8605	0.9402

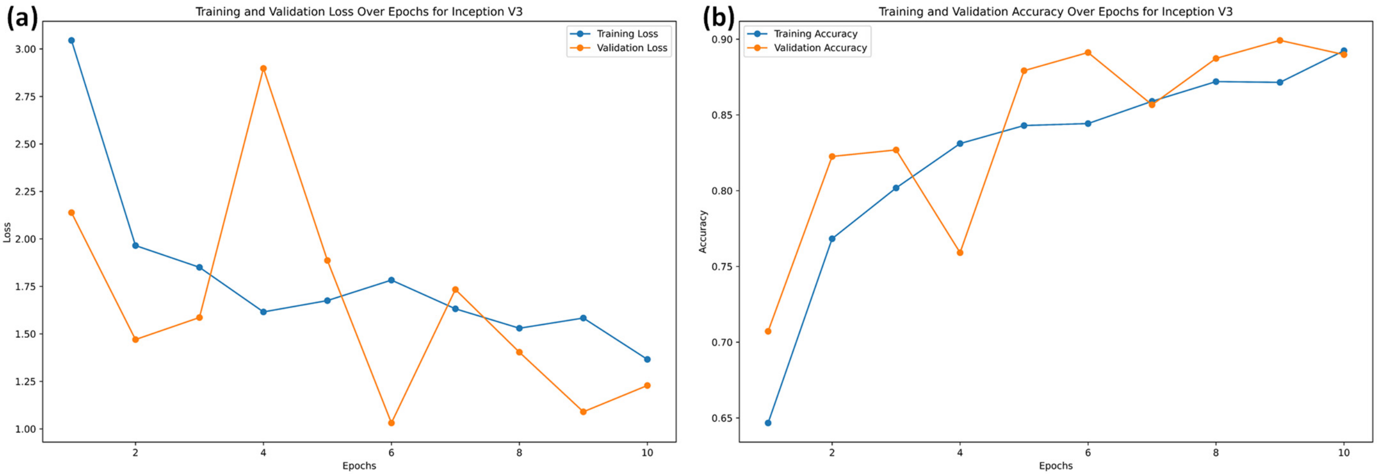


Figure 4. Training and validation (**a**) loss and (**b**) accuracy of Inception V3 for tomato disease detection.

Like Inception V3, the keras preprocessing layers were used for data augmentation and then embed in the ViT model. A multilayer perceptron (MLP) network with a dense layer and a dropout layer was introduced. A network that can change the images into patches was introduced after the MLP layer. For this, we used of the extract_patches module offered by TensorFlow. After the image was divided into patches, a patch encoder that executes the image patches' linear transformation and adds a learnable position embedding to the projected vector was reported.

Precedingly, in the compilation step, a sparse categorical cross-entropy loss-based Adam optimizer was used for optimization and calculating the losses. Lastly was the training step: the model ran for 30 epochs and achieved an accuracy of 95% with 0.15% loss, as shown in Table 6 and Figure 5.

Table 6. Training and testing accuracy of the ViT model for tomato disease classification.

Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
01/30	1.4200	0.5894	0.6276	0.6446
02/30	1.1243	0.5994	0.8276	0.7246
03/30	0.8851	0.6853	0.7326	0.7283
04/30	0.7577	0.7198	0.5725	0.8170
05/30	0.6027	0.7022	0.4366	0.8436
06/30	0.4928	0.8155	0.3871	0.8702
07/30	0.4448	0.8420	0.3977	0.8453
08/30	0.4350	0.8442	0.3599	0.8810
09/30	0.4124	0.8502	0.3343	0.8927
10/30	0.3347	0.8806	0.3299	0.8852
11/30	0.2942	0.8948	0.2026	0.9010
12/30	0.2632	0.9045	0.2383	0.9185
13/30	0.2274	0.9197	0.2422	0.9140
14/30	0.2230	0.9201	0.2502	0.9210
15/30	0.2016	0.9245	0.2382	0.9225
16/30	0.1548	0.9449	0.2682	0.9260
17/30	0.1789	0.9264	0.2510	0.9201
18/30	0.2237	0.9245	0.2414	0.9118
19/30	0.1533	0.9480	0.1968	0.9309
20/30	0.1386	0.9528	0.2100	0.9368
21/30	0.1221	0.9597	0.2232	0.9268
22/30	0.1055	0.9659	0.1460	0.9559
23/30	0.0891	0.9700	0.1567	0.9542
24/30	0.1059	0.9617	0.2567	0.9126
25/30	0.1157	0.9599	0.2173	0.9393
26/30	0.1172	0.9612	0.2121	0.9443
27/30	0.0987	0.9673	0.1426	0.9584
28/30	0.0953	0.9698	0.2189	0.9376
29/30	0.0425	0.9712	0.1439	0.9468
30/30	0.0767	0.9734	0.1585	0.9576

From Table 5, it is observed that the ViT reported very good results from epoch number 10 onwards, with a minimum error in the model. The comparison data for accuracy and loss of Inception V3 and ViT are tabulated below, in Table 7.

Table 7. Differences between Inception V3 and ViT in terms of accuracy and loss.

Model Name	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
Inception V3	94.80%	1.00%	94.02%	0.86%
ViT	97.37%	0.07%	95.76%	0.15%

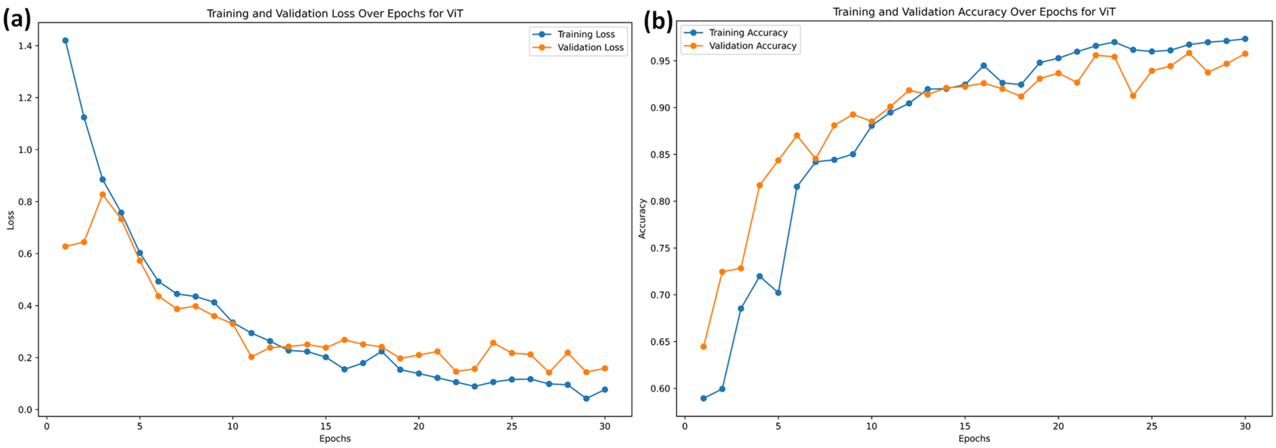


Figure 5. Training and validation (**a**) loss and (**b**) accuracy of ViT for tomato disease detection.

Table 6 compares the performance metrics between Inception V3 and Vision Transformer (ViT) models in terms of accuracy and loss. The ViT model outperforms Inception V3, achieving a significantly higher accuracy of 97.37% compared to Inception V3's accuracy of 89.24%. This indicates that ViT has a superior ability to make correct predictions during the training phase. In addition to accuracy, the loss metrics are compared. The loss is a measure of how well the model is performing, with lower values indicating better performance. ViT demonstrates a remarkably lower loss of 0.07% compared to Inception V3's loss of 1.36%. This signifies that the ViT model exhibits better optimization and is able to minimize errors more effectively during training.

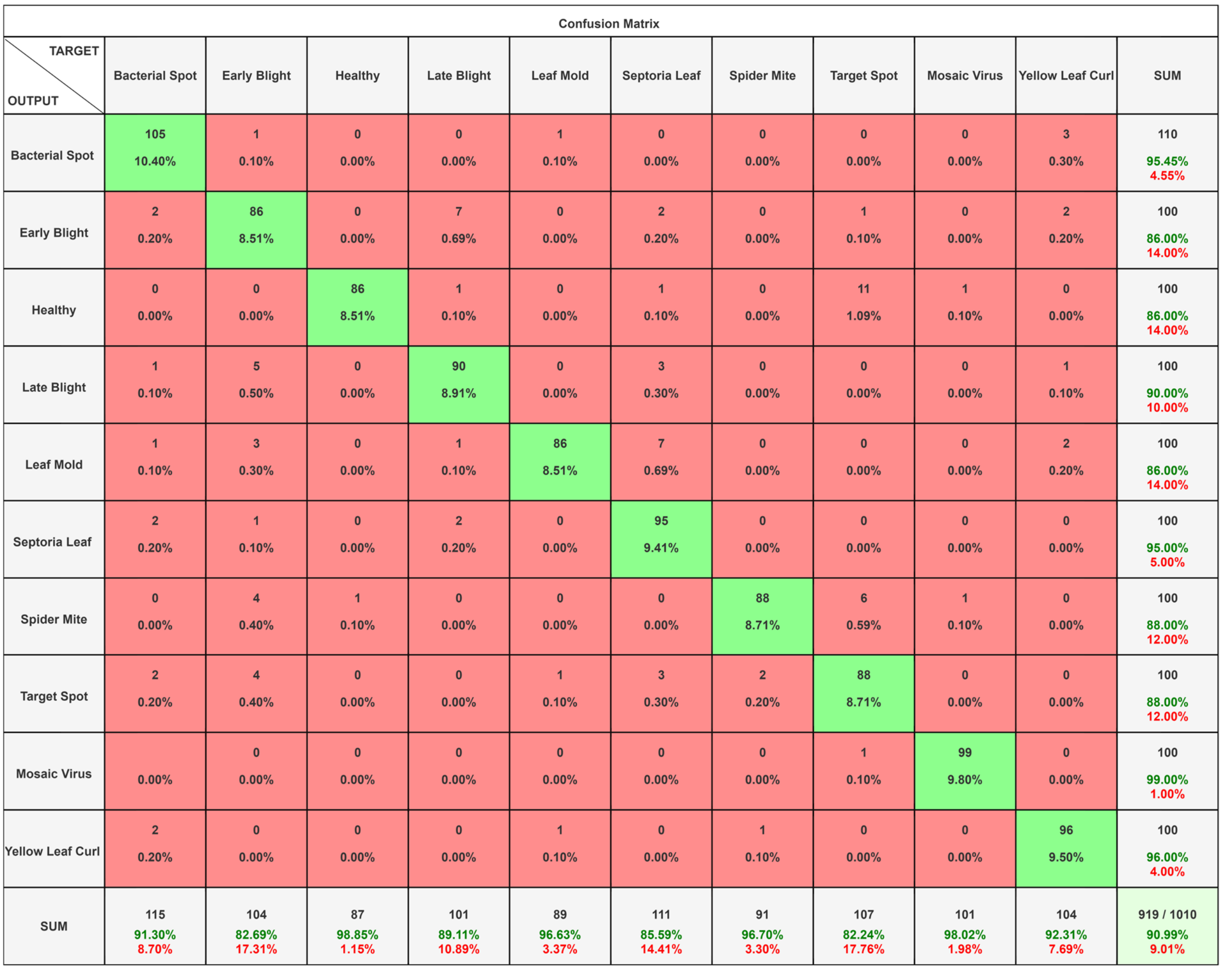
Furthermore, the validation accuracy and validation loss are presented in Table 6. Validation accuracy represents the model's performance on a separate dataset not used during training; here, ViT maintains a higher accuracy of 95.76% compared to Inception V3's 88.98%. Similarly, ViT achieves a lower validation loss of 0.15% in contrast to Inception V3's 1.22%.

The results in Table 6 highlight the superior training and validation performance of the ViT model over Inception V3 in terms of accuracy and loss metrics. These findings underscore the effectiveness of ViT in achieving higher accuracy and better optimization during the training process, making it a more robust model for tomato disease detection. ViT surpassed Inception V3 in terms of producing better accuracy at a lower loss rate, according to the results in Table 7. Therefore, we created a confusion matrix for the ViT model and computed a classification report.

Table 8 presents the comprehensive classification report for the ViT model, which stands out as the best-performing classification algorithm in our study. The model showcases impressive precision, recall, and F1 scores across various classes representing different tomato diseases. Figure 6 shows the confusion matrix of the ViT model for tomato leaf disease detection in our experiment.

Table 8. Classification report on the best classification algorithm (ViT).

Model	Class	Precision	Recall	F1 Score	Testing Accuracy
ViT	Tomato Bacterial Spot Disease	0.95	0.91	0.93	90.99%
	Tomato Early Blight	0.86	0.82	0.84	
	Tomato Healthy	0.86	0.82	0.91	
	Tomato Late Blight	0.90	0.89	0.89	
	Tomato Leaf Mold	0.86	0.96	0.91	
	Tomato Septoria Leaf Spot	0.95	0.85	0.90	
	Tomato Spider Mite	0.88	0.96	0.92	
	Tomato Target Spot	0.88	0.82	0.85	
	Tomato Mosaic Virus	0.99	0.98	0.98	
	Tomato Yellow Leaf Curl Virus	0.96	0.92	0.94	



Confusion Matrix

TARGET OUTPUT	Bacterial Spot	Early Blight	Healthy	Late Blight	Leaf Mold	Septoria Leaf	Spider Mite	Target Spot	Mosaic Virus	Yellow Leaf Curl	SUM
Bacterial Spot	105 10.40%	1 0.10%	0 0.00%	0 0.00%	1 0.10%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	3 0.30%	110 95.45% 4.55%
Early Blight	2 0.20%	86 8.51%	0 0.00%	7 0.69%	0 0.00%	2 0.20%	0 0.00%	1 0.10%	0 0.00%	2 0.20%	100 86.00% 14.00%
Healthy	0 0.00%	0 0.00%	86 8.51%	1 0.10%	0 0.00%	1 0.10%	0 0.00%	11 1.09%	1 0.10%	0 0.00%	100 86.00% 14.00%
Late Blight	1 0.10%	5 0.50%	0 0.00%	90 8.91%	0 0.00%	3 0.30%	0 0.00%	0 0.00%	0 0.00%	1 0.10%	100 90.00% 10.00%
Leaf Mold	1 0.10%	3 0.30%	0 0.00%	1 0.10%	86 8.51%	7 0.69%	0 0.00%	0 0.00%	0 0.00%	2 0.20%	100 86.00% 14.00%
Septoria Leaf	2 0.20%	1 0.10%	0 0.00%	2 0.20%	0 0.00%	95 9.41%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	100 95.00% 5.00%
Spider Mite	0 0.00%	4 0.40%	1 0.10%	0 0.00%	0 0.00%	0 0.00%	88 8.71%	6 0.59%	1 0.10%	0 0.00%	100 88.00% 12.00%
Target Spot	2 0.20%	4 0.40%	0 0.00%	0 0.00%	1 0.10%	3 0.30%	2 0.20%	88 8.71%	0 0.00%	0 0.00%	100 88.00% 12.00%
Mosaic Virus	0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	1 0.10%	99 9.80%	0 0.00%	100 99.00% 1.00%
Yellow Leaf Curl	2 0.20%	0 0.00%	0 0.00%	0 0.00%	1 0.10%	0 0.00%	1 0.10%	0 0.00%	0 0.00%	96 9.50%	100 96.00% 4.00%
SUM	115 91.30% 8.70%	104 82.69% 17.31%	87 98.85% 1.15%	101 89.11% 10.89%	89 96.63% 3.37%	111 85.59% 14.41%	91 96.70% 3.30%	107 82.24% 17.76%	101 98.02% 1.98%	104 92.31% 7.69%	919 / 1010 90.99% 9.01%

Figure 6. Confusion matrix of ViT for tomato leaf disease detection.

For the detection of Tomato Bacterial Spot Disease, the ViT model achieves a precision of 0.95, indicating a high accuracy of positive predictions. The recall of 0.91 reflects the model's ability to capture a significant portion of actual positive instances. The corresponding F1 score, a harmonic mean of precision and recall, is calculated at 0.93, underlining a robust balance between precision and recall for this class. The overall testing accuracy for Tomato Bacterial Spot Disease is reported as 90.99%, emphasizing the model's efficacy in accurately classifying instances of this particular disease.

A similar high-level performance is observed for other tomato diseases, with varying precision, recall, and F1 scores. For instance, Tomato Mosaic Virus detection reaches exceptional precision, recall, and F1 scores of 0.99, 0.98, and 0.98, respectively, showcasing the model's excellence in identifying this specific disease. Additionally, Tomato Late Blight, Tomato Septoria Leaf Spot, and Tomato Spider Mite demonstrate commendable performance, with balanced precision, recall, and F1 scores.

This detailed classification report (Figure 6) provides insights into the ViT model's effectiveness in distinguishing between different tomato diseases, highlighting its ability to deliver accurate and reliable predictions across various classes. The robust evaluation metrics presented in Table 8 underscore the model's potential for practical deployment in real-world scenarios, particularly in the context of tomato disease detection and classification.

In the comparative analysis (Table 9), we studied various authors who experimented with tomato disease classification using different algorithms and datasets. Here are some observations:

- Abbas et al. [24] achieved the maximum accuracy of 97.11% using the DenseNet121 algorithm on the Plant Village Synthetic dataset.
- Agarwal et al. [4] achieved an accuracy of 91.20% using the VGG16 algorithm on the Plant Village tomato dataset.
- Hossain et al. [14] used the Multi-Axis Vision Transformer algorithm and obtained an accuracy of 93.00% on the Plant Village dataset.
- The proposed work achieved an accuracy of 90.99% using the ViT algorithm in independent testing.

Table 9. Comparative analysis of tomato disease identification.

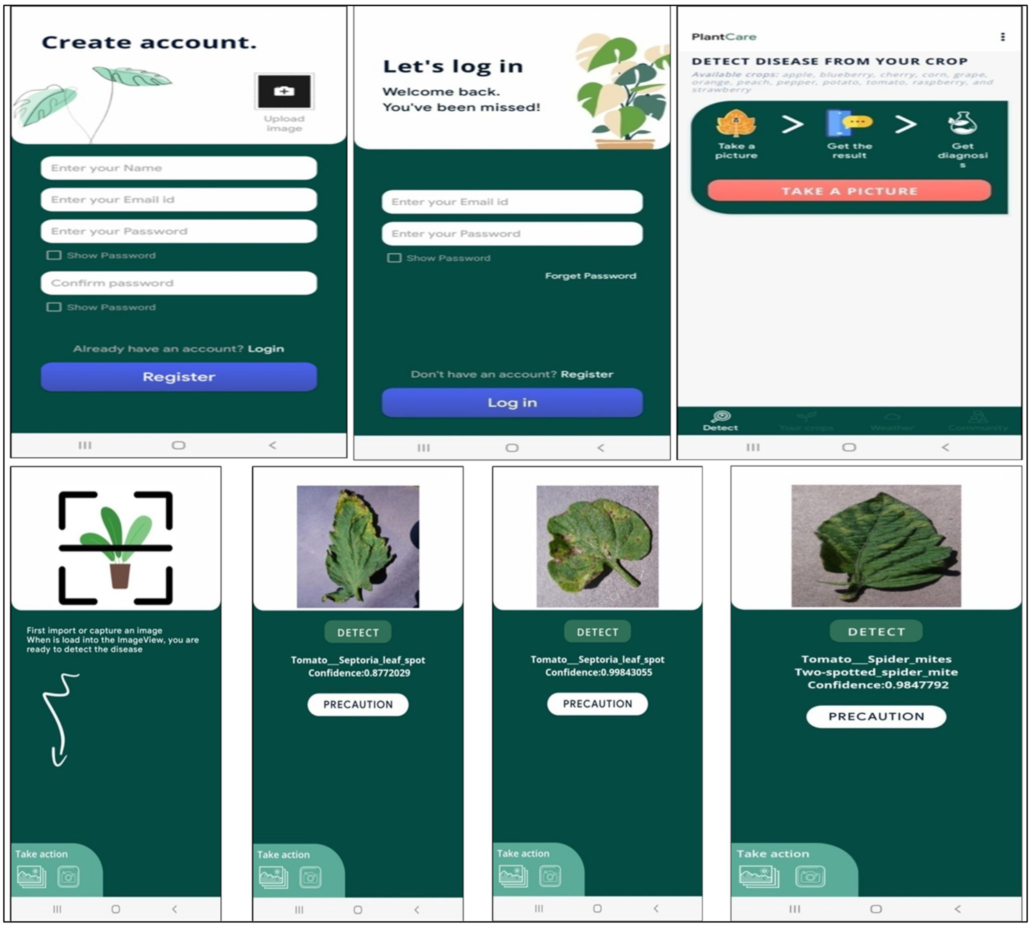
Sl No.	Author	Tomato Dataset	No. of Classes	Algorithm	Testing Accuracy
1	Abbas et al. [24]	Plant Village Synthetic	10	DenseNet121	97.11%
2	Agarwal et al. [4]	Plant Village	10	VGG16	91.20%
3	Hossain et al. [14]	Plant Village	10	Multi-Axis Vision Transformer	93.00%
4	Proposed Work	Plant Village	10	ViT	90.99%

The selection of the Vision Transformer (ViT) model was based on its proven efficacy in various image recognition tasks and its ability to capture global dependencies in data, particularly in the context of plant disease classification [25]. The comparison with Inception V3 was chosen to benchmark ViT against a well-established convolutional neural network (CNN) architecture. Inception V3 serves as a reference model widely used in plant disease classification tasks [26]. The goal was to assess the relative performance of ViT against a recognized baseline, providing insights into the strengths and weaknesses of each model.

ViT demonstrated impressive performance in handling diverse datasets, making it a suitable candidate for our study. The Vision Transformer (ViT) model demonstrated a testing accuracy of 90.99%, marginally lower than some counterparts. The comparison mandates consideration of factors such as model complexity and dataset characteristics. The 88% validation accuracy of Inception V3 highlights the inherent variability in model performance. However, interpreting these results necessitates attention to additional factors like dataset size, data preprocessing techniques, and experimental setup. A holistic evaluation is essential, emphasizing the multifaceted nature of assessing model efficacy beyond raw accuracy numbers.

4. Model Deployment

TensorFlow Lite is a collection of technologies that makes it possible for developers to run their models on mobile, embedded, and IoT devices, enabling on-device machine learning. Following the model construction, the ViT model was exported as .tflite files based on the model accuracy. For the deployment phase, we utilized JAVA to develop the integrated models. The .tflite files were subsequently imported into the tool for seamless integration. Post-integration, the tool was prepared and set for real-time testing (Figure 7).



Welcome back.
You've been missed!

Available crops: apple, blueberry, cherry, corn, grape, orange, peach, pepper, potato, tomato, raspberry, and strawberry

First import or capture an image
When is load into the ImageView, you are ready to detect the disease

Tomato__Septoria_leaf_spot
Confidence:0.8772029

Tomato__Septoria_leaf_spot
Confidence:0.99843055

Tomato__Spider_mites
Two-spotted_spider_mite
Confidence:0.9847792

Figure 7. Android App-based model deployment and testing for tomato leaf disease detection.

5. Conclusions

In this study, the vision transformer (ViT) model was used to learn and categorize nine different tomato leaf disease classes and a healthy class using the self-attention paradigm. We also developed an Android smartphone app for model deployment, based on the ViT model. The findings from this work are a viable direction for the application of ViT models to address agricultural challenges. The underlying ViT model is favored over Inception V3 for its excellent accuracy, as demonstrated in Table 7. This demonstrates how successfully the convolutional-free ViT model understands an image as a sequence of patches and processes it using a standard transformer encoder to learn patterns between infected and healthy images by utilizing the self-attention mechanism. In this regard, it is concluded that the use of vision transformers in agricultural applications for picture classification could redefine the way vision tasks are approached by bypassing traditional CNN-based models. We also intend to employ the vision transformer classifier as the backbone of an object detection architecture to locate and identify weeds and plants in future research.

Author Contributions: Conceptualization, U.B. and M.J.S.; methodology, U.B., P.S. and M.R.; software, V.D., S.L. and V.S.; validation, U.B., S.L. and V.S.; formal analysis, U.B., P.S. and M.R.; investigation, U.B. and M.J.S.; resources, U.B., V.S. and M.J.S.; data curation, M.R., V.D. and S.L.; writing—original draft preparation, U.B., P.S., M.R., V.D. and S.L.; writing—review and editing, M.J.S.; visualization, U.B. and M.J.S.; supervision, U.B. and M.J.S.; project administration, U.B. and M.J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The required data is presented in the manuscript. The full data will be shared upon request.

Acknowledgments: The authors acknowledge the Faculty of Computer Technology, AdtU, and the Department of IT, Gauhati University, for providing resources for research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Saeed, A.; Abdel-Aziz, A.A.; Mossad, A.; Abdelhamid, M.A.; Alkhaled, A.Y.; Mayhoub, M. Smart Detection of Tomato Leaf Diseases Using Transfer Learning-Based Convolutional Neural Networks. *Agriculture* **2023**, *13*, 139. [CrossRef]
2. Barman, U.; Choudhury, R.D.; Uddin, I. Predication of Soil pH Using K Mean Segmentation and HSV Color Image Processing. In Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 13–15 March 2019; pp. 31–36.
3. Barman, U.; Choudhury, R.D.; Sahu, D.; Barman, G.G. Comparison of Convolution Neural Networks for Smartphone Image Based Real Time Classification of Citrus Leaf Disease. *Comput. Electron. Agric.* **2020**, *177*, 105661. [CrossRef]
4. Agarwal, M.; Singh, A.; Arjaria, S.; Sinha, A.; Gupta, S. ToLeD: Tomato Leaf Disease Detection Using Convolution Neural Network. *Procedia Comput. Sci.* **2020**, *167*, 293–301. [CrossRef]
5. Ahmed, S.; Hasan, M.B.; Ahmed, T.; Sony, M.R.K.; Kabir, M.H. Less Is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification. *IEEE Access* **2022**, *10*, 68868–68884. [CrossRef]
6. Tarek, H.; Aly, H.; Eisa, S.; Abul-Soud, M. Optimized Deep Learning Algorithms for Tomato Leaf Disease Detection with Hardware Deployment. *Electronics* **2022**, *11*, 140. [CrossRef]
7. Zhao, S.; Peng, Y.; Liu, J.; Wu, S. Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module. *Agriculture* **2021**, *11*, 651. [CrossRef]
8. Karthik, R.; Hariharan, M.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R. Attention Embedded Residual CNN for Disease Detection in Tomato Leaves. *Appl. Soft Comput.* **2020**, *86*, 105933.
9. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
10. Maurício, J.; Domingues, I.; Bernardino, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Appl. Sci.* **2023**, *13*, 5521. [CrossRef]
11. Li, Z.; Tao, W.; Liu, J.; Zhu, F.; Du, G.; Ji, G. Tomato Leaf Disease Recognition via Optimizing Deep Learning Methods Considering Global Pixel Value Distribution. *Horticulturae* **2023**, *9*, 1034. [CrossRef]
12. Dhanya, V.G.; Subeesh, A.; Kushwaha, N.L.; Vishwakarma, D.K.; Kumar, T.N.; Ritika, G.; Singh, A.N. Deep Learning Based Computer Vision Approaches for Smart Agricultural Applications. *Artif. Intell. Agric.* **2022**. [CrossRef]
13. Thakur, P.S.; Chaturvedi, S.; Khanna, P.; Sheorey, T.; Ojha, A. Vision Transformer Meets Convolutional Neural Network for Plant Disease Classification. *Ecol. Inform.* **2023**, *77*, 102245. [CrossRef]
14. Hossain, S.; Tanzim Reza, M.; Chakrabarty, A.; Jung, Y.J. Aggregating Different Scales of Attention on Feature Variants for Tomato Leaf Disease Diagnosis from Image Data: A Transformer Driven Study. *Sensors* **2023**, *23*, 3751. [CrossRef]
15. Zhang, D.; Huang, Y.; Wu, C.; Ma, M. Detecting Tomato Disease Types and Degrees Using Multi-Branch and Destruction Learning. *Comput. Electron. Agric.* **2023**, *213*, 108244. [CrossRef]
16. Liu, J.; Wang, X. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Front. Plant Sci.* **2020**, *11*, 898. [CrossRef]
17. Ouhami, M.; Es-Saady, Y.; Hajji, M.E.; Hafiane, A.; Canals, R.; Yassa, M.E. Deep Transfer Learning Models for Tomato Disease Detection. In *Image and Signal Processing*; El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12119, pp. 65–73, ISBN 978-3-030-51934-6.
18. Trivedi, N.K.; Gautam, V.; Anand, A.; Aljahdali, H.M.; Villar, S.G.; Anand, D.; Goyal, N.; Kadry, S. Early Detection and Classification of Tomato Leaf Disease Using High-Performance Deep Neural Network. *Sensors* **2021**, *21*, 7987. [CrossRef]
19. Guerrero-Ibañez, A.; Reyes-Muñoz, A. Monitoring Tomato Leaf Disease through Convolutional Neural Networks. *Electronics* **2023**, *12*, 229. [CrossRef]
20. Wu, S.; Sun, Y.; Huang, H. Multi-Granularity Feature Extraction Based on Vision Transformer for Tomato Leaf Disease Recognition. In Proceedings of the 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 10–12 December 2021; pp. 387–390.
21. Tabbakh, A.; Barpanda, S.S. A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer "TLMViT" for Plant Disease Classification. *IEEE Access* **2023**. [CrossRef]
22. Shewale, M.V.; Daruwala, R.D. High Performance Deep Learning Architecture for Early Detection and Classification of Plant Leaf Disease. *J. Agric. Food Res.* **2023**, 100675. [CrossRef]
23. Geetharamani, G.; Pandian, A. Identification of Plant Leaf Diseases Using a Nine-Layer Deep Convolutional Neural Network. *Comput. Electr. Eng.* **2019**, *76*, 323–338.
24. Abbas, A.; Jain, S.; Gour, M.; Vankudothu, S. Tomato Plant Disease Detection Using Transfer Learning with C-GAN Synthetic Images. *Comput. Electron. Agric.* **2021**, *187*, 106279. [CrossRef]

25. Thakur, P.S.; Khanna, P.; Sheorey, T.; Ojha, A. Vision Transformer for Plant Disease Detection: PlantViT. In *Computer Vision and Image Processing*; Raman, B., Murala, S., Chowdhury, A., Dhall, A., Goyal, P., Eds.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2022; Volume 1567, pp. 501–511, ISBN 978-3-031-11345-1.
26. Maeda-Gutiérrez, V.; Galván-Tejada, C.E.; Zanella-Calzada, L.A.; Celaya-Padilla, J.M.; Galván-Tejada, J.I.; Gamboa-Rosales, H.; Luna-Garcia, H.; Magallanes-Quintanar, R.; Guerrero Mendez, C.A.; Olvera-Olvera, C.A. Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases. *Appl. Sci.* **2020**, *10*, 1245. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.