Optimization in Agriculture

Journal Homepage: sciencesforce.com/oia

Optimization Agri. Vol. 1 (2024) 105–114

Paper Type: Original Article

A Fresh Look at Tomato Leaf Disease Recognition using Vision Transformers

Walid Abdullah [1,*], Chomyong Kim [2], and Yunyoung Nam [3]

[1] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt; waleed@zu.edu.eg.

[2] Department of ICT Convergence, Soonchunhyang University, Asan, 31538, South Korea; monicakim89@sch.ac.kr.

[3] Department of Computer Science and Engineering, ICT Convergence Research Center, Soonchunhyang University, Asan, 31538, South Korea; ynam@sch.ac.kr.

Received: 12 Jan 2024 **Revised**: 20 Apr 2024 **Accepted**: 22 May 2024 **Published**: 25 May 2024

Abstract

Tomatoes is one of the major economically significant vegetables produced worldwide, contributing greatly to increased agricultural production and food security. However, tomato plants are unfortunately prone to a number of diseases, including several that target the leaves, which can significantly reduce crop productivity and quality. Recently, deep learning techniques have revolutionized the fields of computer vision and image analysis. By automatically learning hierarchical representations from raw pixel data. Transformer is a new deep learning technique that opens new possibilities for image understanding using self-attention mechanisms to capture global dependencies within input sequences. This approach is exemplified by the Vision Transformer (ViT). In this study, we utilize and evaluate the effectiveness of six variations of the Vision Transformer (ViT) architecture in the task of tomato leaf disease recognition. The variants include Mobile ViT, EANet, Swin ViT, ViT, Shift ViT, and Compact ViT. Utilizing a publicly available, multiple-source dataset of tomato leaf images containing various disease patterns. Performance for all models was evaluated and compared in classifying various types of tomato leaf diseases in terms of accuracy, loss, precision, recall, and F1-Score, and the results showed that. The CompactVit has achieved the best accuracy of 97% and precision of 97% and 96% for recall. While the mobile ViT has the lowest performance among all variations in tomato disease recognition, overall, ViT is showing its promise, and it can be utilized on a large scale for smart agriculture, which opens the door for further exploration of this area.

Keywords: Tomato Disease Recognition; Deep Learning; Convolution Neural Network; Vision Transformer.

1 | Introduction

Tomato is one of the most economically important vegetable crops globally, significantly improving food security and agricultural productivity [1]. Tomato farming, however, presents a variety of difficulties. One significant danger to yield and quality is foliar disease. For efficient disease control and sustainable agricultural productivity, early and accurate identification of these diseases is an essential task [2]. The visual examination process utilized in traditional disease detection methods often relies on human expertise, a process known to be time-consuming, subjective, and susceptible to errors [3]. As a solution, automated disease recognition

Corresponding Author: waleed@zu.edu.eg

https://doi.org/10.61356/j.oia.2024.1274

systems offer rapid and objective assessments of plant health. Recent advancements in artificial intelligence and computer vision, particularly the proliferation of deep learning (DL) techniques, have revolutionized the analysis of leaf images and pattern recognition, enabling the accurate detection and recognition of tomato leaf diseases [4-6].

Several techniques based on the DL approach have been developed in the literature to increase field crop persistence rates through early disease diagnosis and subsequent disease control [7]. DL is widely utilized for the detection and categorization of plant diseases, especially convolutional neural networks (CNN). It is a well-known and widely used technique for identifying, categorizing, and diagnosing tomato leaf diseases due to its impressive effectiveness when compared to other conventional techniques [8].

Transformer is a DL architecture proposed by Google researchers that was originally developed for natural language processing (NLP) [9]. Inspired by successful transformer structures in NLP, visual transformers (ViT) have demonstrated impressive performance on a range of computer vision problems [10]. ViT represents a significant step forward in the field of computer vision, offering a novel approach to image processing and analysis. It uses self-attention processes to capture both local and global dependencies in images, in contrast to typical convolutional neural networks (CNNs), which rely on local convolutional operations. As a result, they can acquire complex representations straight from image patches, which can perform very well on image classification tasks. ViT has been used in studies, either in combination with CNN or a pure transformer without reliance on CNNs, to classify and identify tomato diseases based on leaf images, performing very well on image classification tasks [11-13].

In this study, we propose to utilize and investigate the effectiveness of six variations of the ViT architecture, namely Mobile ViT, EANet, Swin ViT, ViT, Shift ViT, and CompactViT, in the task of tomato disease recognition. The models' performance was evaluated and compared in classifying various types of tomato leaf diseases in terms of accuracy, loss, precision, recall, and F1-Score. To identify the potential strengths and weaknesses of each variation. and address the efficiency of using ViT for extracting deep features from tomato leaf images. The experiments demonstrated that the CompactVit achieved the best accuracy of 97 and precision of 97 and 96 for recall. While the mobile ViT has the lowest performance among all variations in tomato disease recognition. overall, the following summarizes this study's primary contributions:

- Addressing the efficiency of ViT techniques in tomato disease recognition, using leaf images.
- We provide a comprehensive evaluation of six different ViT variants in the context of tomato disease recognition. By comparing their performance on a diverse dataset of tomato leaf images, we aim to offer insights into the relative effectiveness of each architecture.
- Our study serves as a benchmark for evaluating the suitability of ViT architectures for agricultural applications, specifically in the domain of crop disease detection. By establishing performance metrics and comparing them across multiple ViT variants.
- By identifying the most effective ViT variants for tomato disease recognition, we aim to contribute to the development of robust and efficient tools for the early detection and classification of plant diseases.

The rest of the paper is structured as follows: In Section 2, the most recent methods and relevant research in the identification and classification of tomato plant diseases are listed and reviewed including methods and findings. The material and techniques used in this paper, including the dataset and DL model, are covered in Section 3. The experimental Setup is shown in Section 4; The Experimental results and Discussion are presented in Section 5. Section 6 presents the implications of this study. The paper's conclusion and future directions are given in Section 7.

2 | Related Work

In this section, we review relevant literature on the utilization of DL techniques, including traditional CNN, transfer learning, and Vit techniques, in the field of detecting and classifying tomato diseases. By examining

previous studies, we aim to identify gaps and highlight key findings, methodologies, and advancements in previous studies.

Baser et al. [14], deployed an improved CNN model architecture to classify 10 different categories of tomato plant leaf diseases. PlantVillage dataset is a publicly available dataset with 16000 images of tomato leaves that was used to train the model, According to the experimental results, the model achieved an accuracy of 98.19%. Another study [15] used transfer learning techniques to identify tomato leaf diseases. The study utilized four pre-trained deep neural networks, including AlexNet, ResNet, VGG-16, and DenseNet for this task. The obtained results showed that, with the best accuracy of 99.9%, the DenseNet model using the RmsProp optimization strategy provides the most significant results.

Fuentes et al. [16] presented a DL-based approach for detecting diseases and pests in tomato plants using images captured by camera devices. The authors examine three popular object detection methods: Faster R-CNN, R-FCN, and SSD. They test these methods using powerful image analysis tools like VGG net and ResNet. they also propose a new way to label objects in images (both locally and globally) and to improve the data used for training. The results show how well these meta-architectures and feature extractors work.

Ullah et al. [17] developed a new hybrid method for detecting tomato plant diseases using deep learning. This "EffiMob-Net model" combines the strengths of two existing models, EfficientNetB3 and MobileNet. The outputs of the two models were combined to identify and categorize tomato leaf diseases and extract the significant features of leaf images accurately. The results demonstrated that this hybrid model achieved a 99.92% success rate in correctly identifying tomato leaf diseases. Another hybrid model to identify tomato leaf diseases was proposed by [18], to extract deep features. This study compacted three deep transfer layer models based on CNN, namely ResNet-18, ShuffleNet, and MobileNet. The extracted features from all models were merged, and then a hybrid feature selection method was utilized to produce a complete lower-dimensional feature collection. Furthermore, six classifiers are utilized to identify the tomato leaf diseases, namely Naïve Bayes, decision tree, K-nearest neighbor (KNN), quadratic discriminate analysis, the linear discriminate classifier, and support vector machine (SVM), and the experimental results showed that the KNN and SVM have achieved the greatest performance with an accuracy of 99.92% and 99.90%, respectively.

A multimodal hybrid DL approach using an attention-based dilated convolution feature extractor with logistic regression classification to detect and classify tomato leaf disease was proposed by [19]. This model takes advantage of CNN to extract the most relevant features from leaf images. And incorporates the Conditional Generative Adversarial Network (CGAN) model to generate a synthetic image to handle imbalances and noisy or wrongly labeled data to obtain good prediction results. The logistic regression (LR) classifier was used for the classification phase, and a well-known PlantVillage dataset was utilized to train and test the model. The experimental results demonstrate state-of-the-art performance by obtaining 100%, 100%, and 96.6% training, testing, and validation accuracy for multiclass on the Plant Village database of tomato leaf disease.

In [20], a lightweight attention-based CNN was proposed for tomato leaf disease classification. The model was developed by incorporating various attention modules, and the effectiveness of each module was evaluated with the model's computational complexity and performance. The performance of the models was evaluated in terms of precision, recall, and F1 score. According to the results, the lightweight model greatly lowered the complexity and number of network parameters. Moreover, while all attention modules improved CNN's performance, the self-attention (SA) mechanism and convolutional block attention module (CBAM) performed the best, with an average accuracy of 99.69% and 99.34%, respectively. Another tomato plant disease classification model is proposed by [21] by employing a new attention technique for the tomato leaf image disease identification, this technique integrates channel, pixel, and spatial attention, and this approach achieved impressive results: 99.88% accuracy on training data, 99.88% accuracy on validation data, and 99.83% accuracy on completely new test images.

While attention is applied in conjunction with convolutional networks, new research is being conducted to develop hybrid DL models based on ViT to model long-range features and allow the model to focus on relevant features. The authors in [22] proposed a novel approach for extracting deep features based on vision transformers and deep transfer learning techniques for plant disease classification, referred to as "TLMViT." Five pre-trained models are used individually with ViT to experiment with the proposed method. The results showed that the proposed model obtained 98.81% for VGG19, followed by the ViT model on the PlantVillage dataset. Additionally, pre-trained-based architecture is contrasted with TLMViT. According to the comparative result, TLMViT outperformed the transfer learning-based models by 1.11% in validation accuracy and 2.576% in validation loss. Other researchers applied a pure transformer directly to sequence image paths for image classification [13]. This study suggests a model-based method for differentiating between healthy and diseased plants utilizing a ViT model, which operates using the self-attention mechanism. The model is trained to detect 10 different tomato disease classes and compared with the Inception V3 DL pre-trained model. The results showed that model-based VIT yielded better performance than Inception V3, therefore motivating additional research in this research area.

3 | Visual Transformers for Tomato Leaf Disease Recognition

The Vision Transformer (ViT) has revolutionized computer vision by applying the transformer architecture which was initially developed for natural language processing, to image data. This approach involves dividing an image into patches and processing these patches as sequences of tokens, similar to words in a sentence. In this study, a set of Vision Transformer variations were utilized for tomato disease detection and identification:

- **Vision Transformer (ViT)** [10]: fundamentally changed the approach to image classification by treating images as sequences of patches. Each image is divided into fixed-size patches, which are then linearly embedded and fed into a transformer encoder. This model leverages self-attention mechanisms to capture global context, achieving state-of-the-art performance on large datasets. However, ViT requires substantial computational resources and extensive data for training, which can limit its usability in practical scenarios. architecture is presented in Figure 1.

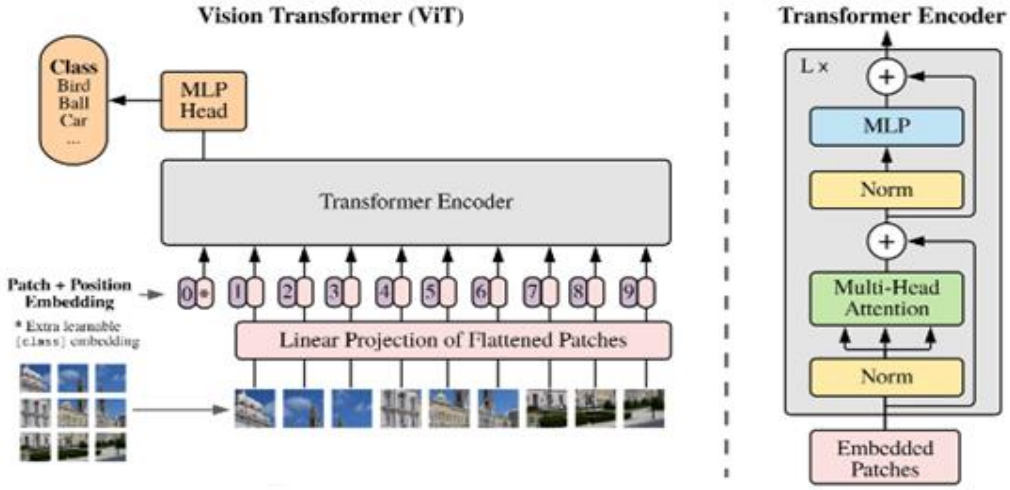


Figure 1. ViT architecture [10].

While the original ViT demonstrated exceptional performance on image classification tasks, it also highlighted significant challenges related to computational efficiency and the need for large-scale training datasets. In response, several variants of ViT have been developed to address these challenges and extend its applicability. These include Mobile ViT, EANet, Swin ViT, Shift ViT, and CompactViT, each introducing unique enhancements to the original architecture.

- **EANet** [23]: This model aims to optimize the attention mechanism used in ViT to enhance efficiency. it introduces more efficient techniques such as sparse attention and linear attention to reduce this

complexity. These improvements allow the model to maintain high performance while being more resource-efficient, making it suitable for large-scale and real-time applications.

- **MobileVit** [24]: adapts the Vision Transformer for mobile and edge devices, where computational and memory constraints are more pronounced. It combines the efficient convolutional operations of MobileNet with transformer blocks to handle both local and global features effectively. By integrating these two approaches, Mobile ViT reduces computational demands while preserving high performance.
- **Compact ViT** [25]: Techniques such as pruning, quantization, and the development of more efficient transformer blocks are employed to minimize the number of parameters and computational demand by focusing on reducing the size and complexity of the Vision Transformer.
- **Swin ViT** [26]**:** The Swin Transformer innovates by introducing a hierarchical architecture and a window-based self-attention mechanism. Instead of processing the entire image at once, it divides the image into non-overlapping windows and computes self-attention within each window. This hierarchical approach captures both local details and global context, enhancing the model's performance on tasks such as object detection and semantic segmentation.
- **Shift ViT** [27]: this variation incorporates shift operations into the transformer architecture to improve efficiency and performance. These operations involve shifting feature maps in specific patterns to better capture spatial dependencies inherent in image data. Shift ViT reduces computational costs and enhances the model's ability to recognize spatial relationships.

4 |Experimental Setup

4.1 |Dataset

A publicly accessible dataset with ten classifications that was gathered from multiple sources was used to train the suggested model, most of the images were primarily collected from a plant village database and the rest were collected from other various public sources [28]. The dataset consists of nine groups indicating various tomato leaf diseases, in addition to one class for healthy leaf images.

4.2 |Experimental Environment Setup

All experiments in this paper were conducted on the Kaggle environment with Nvidia Tesla P100 GPU and 16 GB of RAM using Python Version 3.7.6 and Keres Version 2.3.1. Furthermore, All DL Models were trained using Adam optimizer, with a learning rate of .0001, and using a batch size of 64 images.

4.3 |Evaluation Metrics

The models' efficacy was evaluated using adopting the standard classification metrics named, accuracy, loss, precision, recall, and F1-score. as shown in Eqs. (1-5), respectively.

- **Accuracy –** For measuring the overall correctness of predictions.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \quad (1)$$

- **Loss –** For measuring the discrepancy between predicted and actual values

$$\text{loss} = -\frac{1}{N}\sum_{i=1}^{n} y_i \log(\hat{y}_i) \quad (2)$$

- **Precision –** focuses on the precision of positive predictions, to assessing the precision of positive predictions among all predicted positives

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{3}$$

- **Recall -** emphasizes capturing all relevant instances, for quantifying the model's ability to correctly identify all relevant instances.

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{4}$$

- **F1 Score -** balances both precision and recall for comprehensive performance evaluation.

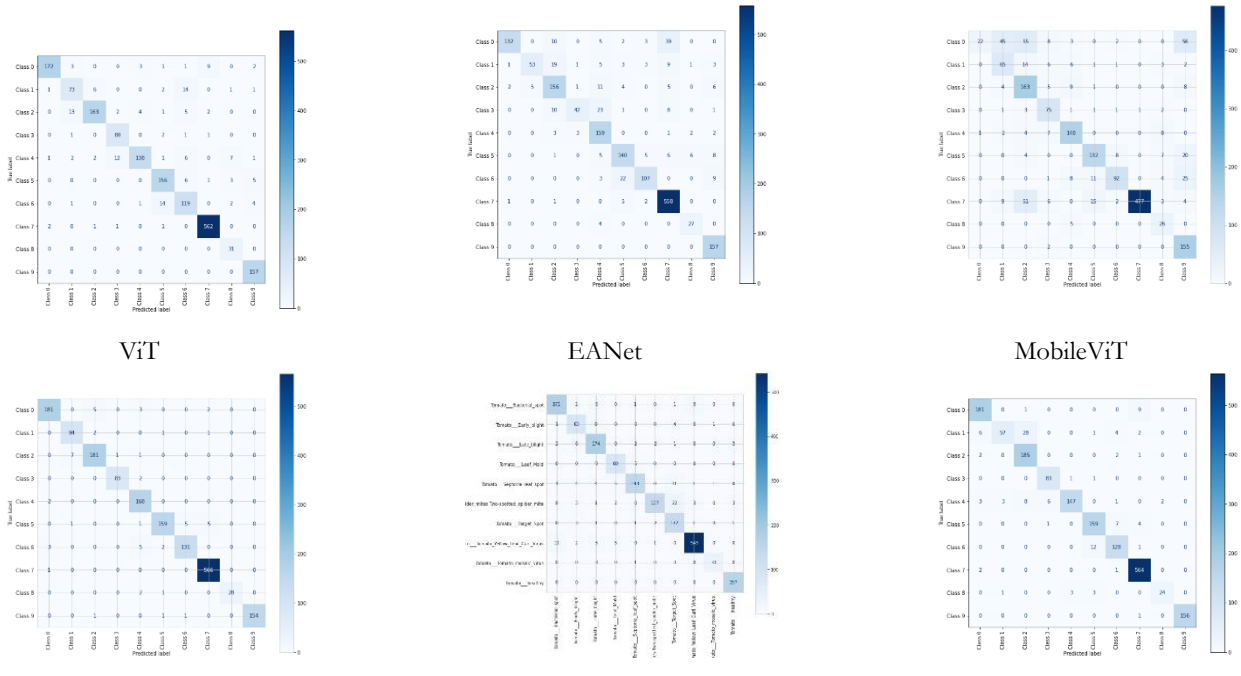
$$F1 \text{ Score} = 2 \times \frac{\text{recall} \times \text{Precision}}{\text{recall} + \text{Precision}} \tag{5}$$

5 | Experimental Results and Discussion

In this work, six different Variations of ViT models experimented to classify 10 images classes of tomato leaves images (9 different types of tomato diseases, and 1 healthy class) on the utilized dataset, all models are evaluated in terms of accuracy, loss, precision, recall, and F1- score, The performance of the six different models are shown in Table 1. Furthermore, Figure 2 shows the confusion matrices provided by various models on the test dataset. The ROC (Receiver Operating Characteristic) curves demonstrating the discriminative power of vision transformer models in tomato leaf disease recognition are presented in Figure 3, while Figure 4 shows the T-SNE (t-Distributed Stochastic Neighbor Embedding) graph to visualize feature representations Learned by vision transformer models.

Table 1. Performance comparison between the various vision Transformer models.

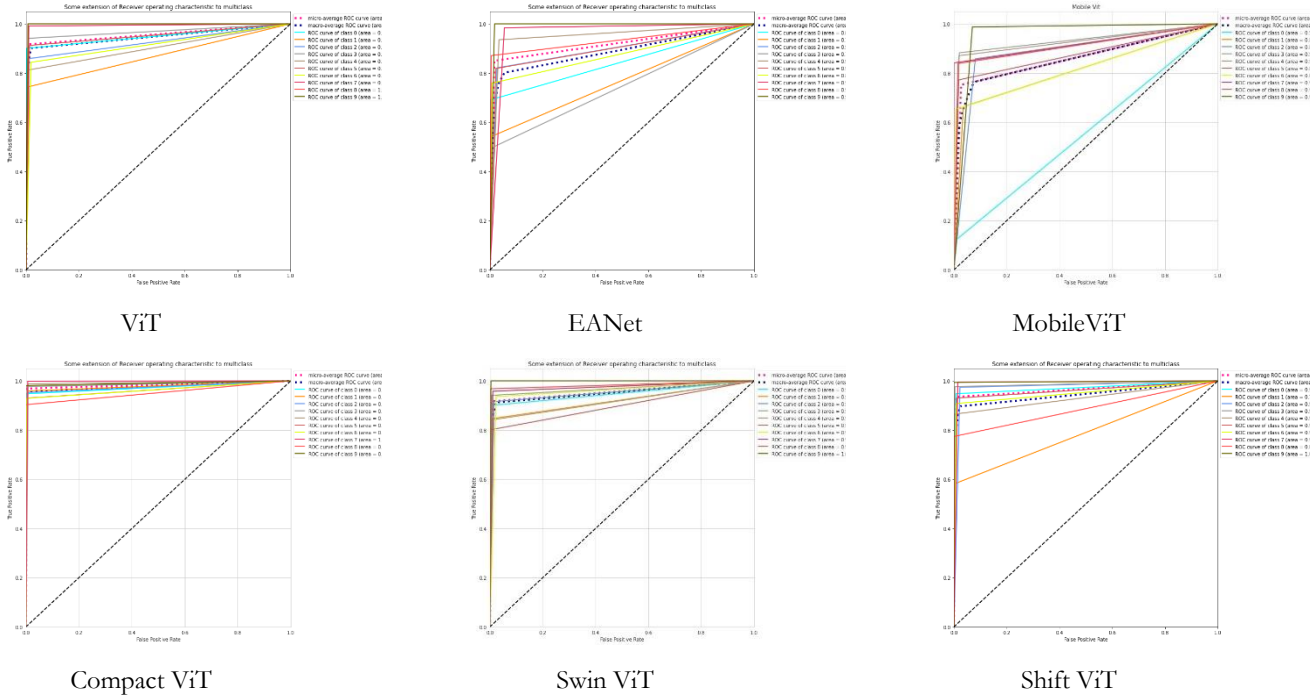
	Accuracy	Loss	Precision	Recall	F1-Score
Mobile ViT	0.75	1.37	0.73	0.75	0.69
EANet	0.85	0.53	0.85	0.79	0.80
Swin ViT	0.91	0.25	0.89	0.89	0.89
ViT	0.92	0.27	0.88	0.90	0.88
Shift ViT	0.94	0.25	0.93	0.89	0.91
CompactViT	0.97	0.11	0.97	0.96	0.96



ViT EANet MobileViT

Compact ViT Swin ViT Shift ViT

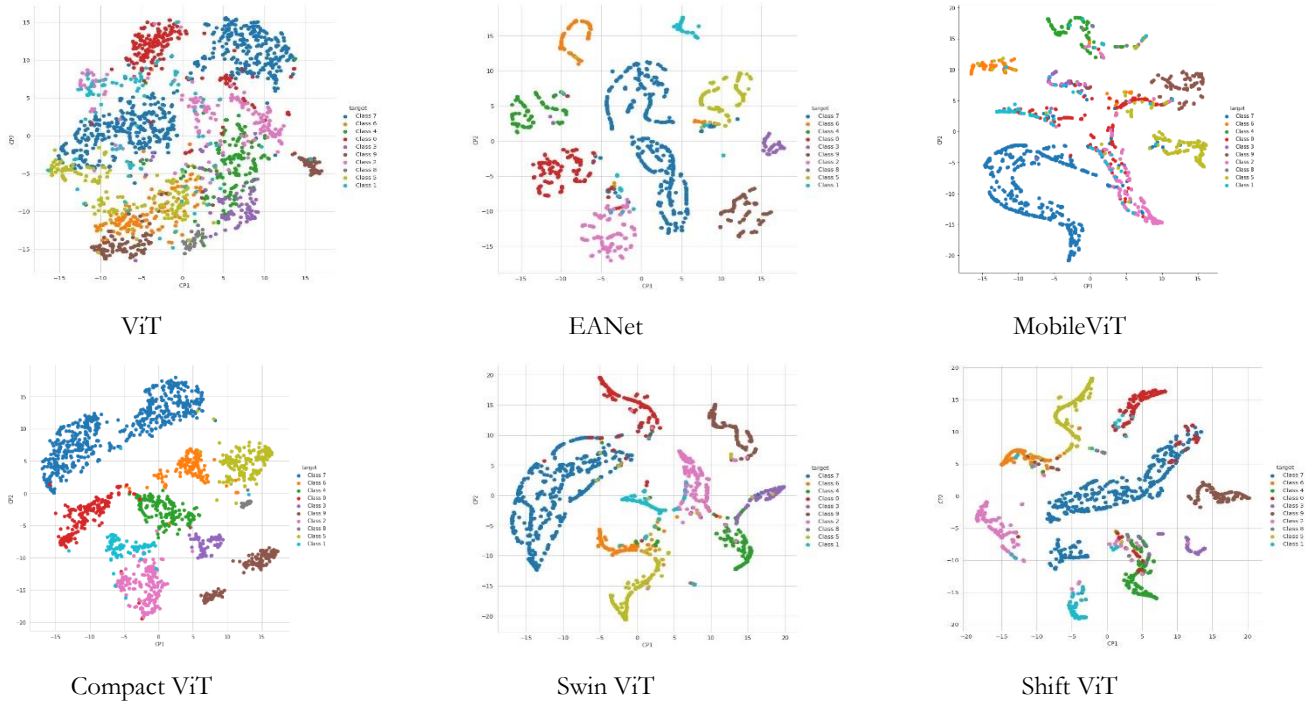
Figure 2. Confusion matrices provided by various Vision Transformer models in Tomato leaf disease recognition.



ViT EANet MobileViT

Compact ViT Swin ViT Shift ViT

Figure 3. ROC curves demonstrating the discriminative power of vision Transformer models in Tomato leaf disease recognition.



ViT EANet MobileViT

Compact ViT Swin ViT Shift ViT

Figure 4. T-SNE plots visualizing feature representations learned by vision Transformer models in Tomato leaf disease recognition.

As shown in the results, when comparing the performance of each model, it's evident that CompactViT outperforms the other models across all metrics. It achieves the highest accuracy of 97% and the highest F1-Score of 96%, indicating its effectiveness in accurately classifying tomato leaf diseases, Furthermore, Shift ViT also demonstrates strong performance, with an accuracy of 94% and an F1-Score of 91%. CompactViT, offers the second-highest accuracy and precision in disease classification tasks after CompactViT. While

Mobile ViT and EANet exhibit lower performance compared to the top-performing models, they still achieve respectable accuracy and F1-Score values. However, their precision and recall values are comparatively lower, indicating potential room for improvement in accurately identifying disease patterns. Overall, the experimental results highlight the effectiveness of Vision Transformer (ViT) architectures in tomato leaf disease classification. These models offer promising performance metrics, paving the way for their application in agricultural settings for automated disease detection and management.

6 | Implications

Economical crops such as tomatoes play a significant role in the global economy and productivity. Global vegetable consumption, especially tomato, continues to rise every year; however, plant diseases present a significant threat to yield productivity and quality. The significance of crop health and the efficient management of tomato diseases using new technologies is closely aligned with Egypt's Vision 2030 and the Sustainable Development Goals (SDGs) such as Food Security and Nutrition, Economic Growth, Environmental Sustainability, and Innovation and Technological Advancement [29], and which offer important insights into achieving the SDGs and advancing sustainable agricultural practices. Using new technologies in detecting plant diseases, such as DL techniques, has many benefits, such as early detection of diseases and precision and accuracy in plant disease identification and recognition.

7 | Conclusion

This study aims to address the efficiency of Vision Transformer in deep features extracting from tomato leaf images; furthermore, it provides a comprehensive evaluation of six variations of the ViT architecture for tomato leaf disease recognition, namely Mobile ViT, EANet, Swin ViT, ViT, Shift ViT, and CompactViT. Through rigorous experimentation and analysis, we have gained valuable insights into the capabilities of these ViT variants in accurately classifying various types of tomato diseases. Moreover, these variations' performance is compared to each other. Our findings highlight the effectiveness of ViT architectures, particularly CompactViT, in accurately classifying various types of tomato diseases, achieving the best performance among all variations with an accuracy and precision of 97% for both. This underscores the robustness and scalability of transformer-based approaches in image classification tasks, especially in the context of agricultural applications. Further exploration in this area holds great potential for innovation in precision agriculture. Future research directions could focus on optimizing transformer architectures for enhanced disease diagnosis performance. Additionally, there is scope for investigating the generalization of visual transformers to other crop species and expanding the scope of disease recognition to encompass newly discovered diseases and environmental factors.

Acknowledgments

This research was supported by the Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012724, HRD Program for Industrial Innovation) and the Soonchunhyang University Research Fund.

Author Contribution

All authors contributed equally to this work.

Funding

This research was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0012724, HRD Program for Industrial Innovation) and the Soonchunhyang University Research Fund.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] FAOSTAT, F., Agriculture organization of the united nations FAO statistical database. 2023.
[2] Abdullah, H.M., et al., Present and future scopes and challenges of plant pest and disease (P&D) monitoring: Remote sensing, image processing, and artificial intelligence perspectives. Remote Sensing Applications: Society and Environment, 2023: p. 100996.
[3] Sankaran, S., et al., A review of advanced techniques for detecting plant diseases. Computers and electronics in agriculture, 2010. 72(1): p. 1-13.
[4] Ngugi, L.C., M. Abelwahab, and M. Abo-Zahhad, Recent advances in image processing techniques for automated leaf pest and disease recognition–A review. Information processing in agriculture, 2021. 8(1): p. 27-51.
[5] Chowdhury, M.E., et al., Automatic and reliable leaf disease detection using deep learning techniques. AgriEngineering, 2021. 3(2): p. 294-312.
[6] Chowdhury, M., et al., Tomato leaf diseases detection using deep learning technique. Technology in Agriculture, 2021. 453.
[7] Zhao, S., et al., Tomato leaf disease diagnosis based on improved convolution neural network by attention module. Agriculture, 2021. 11(7): p. 651.
[8] Tan, L., J. Lu, and H. Jiang, Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. AgriEngineering, 2021. 3(3): p. 542-558.
[9] Vaswani, A., et al., Attention is all you need. Advances in neural information processing systems, 2017. 30.
[10] Dosovitskiy, A., et al., An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
[11] Tabbakh, A. and S.S. Barpanda, A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer "TLMViT" for Plant Disease Classification. IEEE Access, 2023. 11: p. 45377-45392.
[12] Yu, S., L. Xie, and Q. Huang, Inception convolutional vision transformers for plant disease identification. Internet of Things, 2023. 21: p. 100650.
[13] Barman, U., et al., ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture. Agronomy, 2024. 14(2): p. 327.
[14] Baser, P., J.R. Saini, and K. Kotecha, TomConv: An improved CNN model for diagnosis of diseases in tomato plant leaves. Procedia Computer Science, 2023. 218: p. 1825-1833.
[15] Bakır, H., Evaluating the impact of tuned pre-trained architectures' feature maps on deep learning model performance for tomato disease detection. Multimedia Tools and Applications, 2024. 83(6): p. 18147-18168.
[16] Fuentes, A., et al., A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors, 2017. 17(9): p. 2022.
[17] Ullah, Z., et al., EffiMob-Net: A deep learning-based hybrid model for detection and identification of tomato diseases using leaf images. Agriculture, 2023. 13(3): p. 737.
[18] Attallah, O., Tomato leaf disease classification via compact convolutional neural networks with transfer learning and feature selection. Horticulturae, 2023. 9(2): p. 149.
[19] Islam, M.S., et al., Multimodal hybrid deep learning approach to detect tomato leaf disease using attention based dilated convolution feature extractor with logistic regression classification. Sensors, 2022. 22(16): p. 6079.
[20] Bhujel, A., et al., A lightweight Attention-based convolutional neural networks for tomato leaf disease classification. Agriculture, 2022. 12(2): p. 228.
[21] C.K, S., J. C.D, and N. Patil, Tomato plant disease classification using Multilevel Feature Fusion with adaptive channel spatial and pixel attention mechanism. Expert Systems with Applications, 2023. 228: p. 120381.
[22] Tabbakh, A. and S.S. Barpanda, A Deep Features extraction model based on the Transfer learning model and vision transformer" TLMViT" for Plant Disease Classification. IEEE Access, 2023.

[23] Guo, M.-H., et al., Beyond self-attention: External attention using two linear layers for visual tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022. 45(5): p. 5436-5447.
[24] Mehta, S. and M. Rastegari, Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178, 2021.
[25] Hassani, A., et al., Escaping the big data paradigm with compact transformers. arXiv preprint arXiv:2104.05704, 2021.
[26] Liu, Z., et al. Swin transformer: Hierarchical vision transformer using shifted windows. in Proceedings of the IEEE/CVF international conference on computer vision. 2021.
[27] Wang, G., et al. When shift operation meets vision transformer: An extremely simple alternative to attention mechanism. in Proceedings of the AAAI Conference on Artificial Intelligence. 2022.
[28] Khan, Q., Tomato Disease Multiple Sources. CC0: Public Domain https://www. kaggle. com/datasets/cookiefinder/tomato-disease-multiple-sources, 2022.
[29] Nations, U., Transforming our world: The 2030 agenda for sustainable development. New York: United Nations, Department of Economic and Social Affairs, 2015. 1: p. 41.