Pooling-Based Vision Transformer (PiT) for Enhanced Plant Disease Classification

Siva Adduri g00095118

Mohamed Alkhaja Alawadhi b00094286

Hamza Khan b00092525

Arya Sankhe b00097405

1 Introduction

Plant disease detection is a critical area of research in agriculture, as plant diseases can significantly impact crop yields and food security. According to the Food and Agriculture Organization (FAO), plant diseases and pests cause a loss of 20-40% in global food production annually [1]. Traditional methods of disease detection are often labour-intensive, time-consuming, and prone to human error. The integration of computer vision and deep learning techniques offers a promising solution to these challenges by providing rapid, accurate, and automated disease detection systems. This project aims to leverage the capabilities of pooling-based Vision Transformers (PiTs) to enhance the accuracy and efficiency of plant disease detection.

Pooling-based Vision Transformers (PiTs) are particularly advantageous for this application due to their ability to capture both local and global features of images, which is crucial for identifying subtle disease symptoms in plants. Unlike traditional Convolutional Neural Networks (CNNs), which may lose important spatial information during pooling, PiTs maintain this information, leading to more accurate disease classification [2]. Additionally, PiTs are computationally efficient, making them suitable for real-time applications in large-scale agricultural settings. By comparing PiTs with other state-of-the-art models such as Vision Transformers (ViTs), EfficientNet, and TrincNet, we will provide a comprehensive evaluation of their performance in plant disease detection.

The relevance of our project is highlighted by the growing need for sustainable agricultural practices and the increasing global population, which is projected to reach 9.7 billion by 2050. Early and accurate detection of plant diseases can significantly reduce crop losses and improve food security. Studies have shown that advanced deep learning models, including PiTs, can achieve high accuracy in plant disease detection, with some models reaching over 99% accuracy [3]. Our project will contribute to the development of robust and scalable solutions for plant disease management, ultimately supporting sustainable agriculture and food security.

2 Objectives

Primary Objective

To evaluate the effectiveness of Pooling-based Vision Transformer (PiT) architecture for plant disease detection using the Plant Village dataset, and to conduct a comprehensive comparative analysis against traditional Vision Transformer (ViT) models in the context of agricultural applications.

Specific Objectives

1. To implement and optimize the PiT architecture for plant disease detection tasks using the Plant Village dataset
2. To compare the performance metrics (accuracy, precision, recall, and F1-score) of PiT against traditional ViT models
3. To analyze the computational efficiency of PiT in terms of FLOPs and processing speed compared to existing ViT implementations
4. To evaluate the model's robustness against input variations and challenging conditions specific to agricultural imaging

Scope

- Dataset: Plant Village dataset comprising 70,000 images across 9 plant species
- Model Architecture: Implementation and optimization of PiT architecture
- Comparative Analysis: Benchmarking against traditional ViT models and state-of-the-art agricultural vision models
- Performance Metrics: Accuracy, precision, recall, F1-score, computational efficiency, and robustness measures

Limitations

1. Dataset Constraints:
    - Limited to controlled environment images from the Plant Village dataset
    - May not fully represent real-world agricultural conditions
    - Restricted to 9 plant species
2. Technical Constraints:
    - Focus solely on PiT and ViT architectures, excluding other potential deep learning approaches
    - Computational resource limitations in training and optimization
    - Model evaluation limited to available computational infrastructure
3. Application Constraints:
    - Results may not generalize to other agricultural datasets or conditions
    - Performance analysis limited to disease detection tasks
    - Real-time implementation considerations not addressed

3 Literature review

1. Rethinking Spatial Dimensions of Vision Transformers [5]

This study introduces the Pooling-based Vision Transformer (PiT) which uses pooling layers to overcome the limitation of the original Vision Transformers (ViT). It includes pooling layers, which allows PiT to achieve computational efficiency, reduction in FLOPs, with enhancement in processing speed and generalization. PiT is also much more robust against input changes, such as occlusions or adversarial attacks; hence, it finds broader applications in further complicated computer vision tasks. PiT achieves an optimal balance between the trade-off of computational efficiency and generalization by combining depth-wise convolution with reduction in spatial sizes, yielding state-of-the-art performance compared to ViT on limited training data or when the input is challenging.

2. Visual Intelligence in Precision Agriculture: A Deep Dive into Plant Disease Detection Using Efficient Vision Transformers [7]

This study introduces GreenViT, which is a Vision Transformer optimized for plant disease detection. GreenViT segments an image of a plant and processes them through several transformer layers to classify their labels as either 'Healthy' or 'Infected'. It resulted in high scores for accuracy-100%, 98%, and 99% on different datasets (including Planet Village dataset) with a reduction in the count of parameters from 85 million to 21.65 million. GreenViT, however, outperformed other state of the art models with a very low False Alarm Rate and proved that the Vision Transformers are apt for precision agriculture prone to accurate disease identification with minimum computational resources.

3. Explainable Vision Transformer Enabled Convolutional Neural Network for Plant Disease Identification: PlantXViT [8]

PlantXViT fuses the powers of CNNs and Vision Transformers for enhanced accuracy and interpretability in plant disease identification/detection. The overall performance was high accuracy in detecting plant diseases from various datasets, ranging from 93.55% to 98.86%, while using CNNs for extracting local features and ViTs for capturing global dependencies. The approach is explainable with Grad-CAM and LIME, enabling the model to place emphasis on image regions that create certain predictions by offering

insight into the decision-making process. Explainability within this hybrid architecture, along with strong performance, proves the capability of discovering diseases that are both accurate and interpretable, informing actionable insights on the management of agricultural diseases.

4. TrIncNet: A Lightweight Vision Transformer Network for Identification of Plant Diseases [4]

TrIncNet is a lightweight, Vision-Transformer-based deep learning model for plant disease identification. It achieves a testing accuracy of 99.93% in PlantVillage and 96.93% in Maize datasets. TrIncNet is lightweight with only 6.94 million trainable parameters compared to the other ViT and CNN architectures. Besides, it proposes an Inception module within the Trans-Inception block, replacing the traditional MLP module for enhancing efficiency in processing. The lightweight design of TrIncNet, along with its high accuracy, makes it quite suitable under resource-constrained conditions that enable efficient and effective plant disease classification accordingly.

5. A Systematic Review of Different Categories of Plant Diseases Detection Using Deep Learning-Based Approaches [6]

The review describes the deep learning models used to classify diseases in each of the plant specimens of pepper, potato, and tomato by comparing several architectures such as DenseNet, NasNetLarge, and MobileNetV2 on the PlantVillage dataset. Then, preprocessing techniques, morphological features extraction, and some performance metrics-accuracy, precision, recall, F1 score-have been applied to measure the performance of these models. Therefore, DenseNet201 topped the rest with an accuracy of 98.67% validation and higher precisions and recalls, hence higher F1-scores. Another hybrid model comprising EfficientNetB7 and ResNet152V2 also gave outstanding performance but could not provide results as effectively as DenseNet201. This review has showcased the strengths of CNN architecture-based deep learning approaches to attain high-accuracy-based plant disease detection.

4 Datasets

We plan to use the Plant Village dataset from Kaggle linked below:

https://www.kaggle.com/datasets/tushar5harma/plant-village-dataset-updated

This dataset consists of 70,000 high-quality images of diseased and healthy plant leaves from 9 different species, which demonstrates diversity and volume. This diversity will further help the model to generalize across different plant conditions. Furthermore, this dataset has been used by most of the research papers, which highlights the quality of this dataset. Given the quality of the data and the enormous variety of data, we believe this dataset is the right fit for testing the PiT model.

References

[1] M. Dang *et al.*, "Computer Vision for Plant Disease Recognition: A Comprehensive Review," *Bot. Rev.*, vol. 90, no. 3, pp. 251–311, Sep. 2024, doi: 10.1007/s12229-024-09299-z.

[2] L. Kouadio *et al.*, "A Review on UAV-Based Applications for Plant Disease Detection and Monitoring," *Remote Sensing*, vol. 15, no. 17, p. 4273, Aug. 2023, doi: 10.3390/rs15174273.

[3] J. Chen, J. Chen, D. Zhang, Y. A. Nanehkaran, and Y. Sun, "A cognitive vision method for the detection of plant disease images," *Machine Vision and Applications*, vol. 32, no. 1, p. 31, Jan. 2021, doi: 10.1007/s00138-020-01150-w.

[4] Gole, P., Bedi, P., Marwaha, S., Haque, Md. A., & Deb, C. K. (2023). TrIncNet: A lightweight vision transformer network for identification of plant diseases. *Frontiers in Plant Science*, *14*, 1221557. https://doi.org/10.3389/fpls.2023.1221557

[5] Heo, B., Yun, S., Han, D., Chun, S., Choe, J., & Oh, S. J. (2021). *Rethinking Spatial Dimensions of Vision Transformers* (No. arXiv:2103.16302). arXiv. http://arxiv.org/abs/2103.16302

[6] Kumar, Y., Singh, R., Moudgil, M. R., & Kamini. (2023). A Systematic Review of Different Categories of Plant Disease Detection Using Deep Learning-Based Approaches. *Archives of Computational Methods in Engineering*, *30*(8), 4757–4779. https://doi.org/10.1007/s11831-023-09958-1

[7] Parez, S., Dilshad, N., Alghamdi, N. S., Alanazi, T. M., & Lee, J. W. (2023). Visual Intelligence in Precision Agriculture: Exploring Plant Disease Detection via Efficient Vision Transformers. *Sensors*, *23*(15), 6949. https://doi.org/10.3390/s23156949

[8] Thakur, P. S., Khanna, P., Sheorey, T., & Ojha, A. (2022). *Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT* (No. arXiv:2207.07919). arXiv. http://arxiv.org/abs/2207.07919