Final Report: Plant Disease Detection Using Swin Transformer

1. Introduction

Crop diseases are a big challenge in the field of agriculture, often resulting in reduced crop yields along with substantial economic losses. Age-old methods of detecting plant diseases relied on the manual inspection of trained professionals. These methods were slow and prone to errors, as they depended on human judgment. But with the advancement of deep learning and computer vision, automated plant disease detection has become a real alternative. This project aims to use a detection model based on a transformer architecture from deep learning in detecting plant diseases from leaf images with real-time performance.

2. Problem Statement

This project aims to create a deep learning model that classifies a range of plant diseases accurately. The data used for the training and testing of the model comes from the PlantVillage dataset. Some characteristics of the model include the following:

- Manage 38 separate classes of plant diseases
- Excel at standard image classification metrics
- Be deployable via a web application grounded in Streamlit

3. Dataset Description

- Dataset Name: PlantVillage
- Source: https://github.com/spMohanty/PlantVillage-Dataset
- Total Classes: 38
- Format: RGB images organized into class-named folders

Preprocessing Steps:

- Resize images to 224x224 pixels
- Normalize using ImageNet statistics
- Apply data augmentation (random horizontal flip, rotation, and brightness jitter) during training

4. Model Architecture

Base model: Swin Transformer (swin_base_patch4_window7_224)

• Library: timm (PyTorch Image Models)

• Size of Input: 224x224

• Size of Output: 38 classes

• The weights from the pretrained model should be initialized using ImageNet.

In contrast to conventional CNNs, the Swin Transformer utilizes a hierarchical structure with shifted windows that result in efficient attention computation. This architecture enables it to capture an image's local and global features equally well.

5. Training Details

- **Epochs:** 5
- **Batch Size:** 32
- **Optimizer:** AdamW
- **Loss Function:** CrossEntropyLoss
- **Learning Rate:** 3e-5

Results by Epoch:

- Epoch 1: Loss = 215.99
- Epoch 2: Loss = 39.84
- Epoch 3: Loss = 20.41
- Epoch 4: Loss = 20.55
- Epoch 5: Loss = 15.95

6. Evaluation Metrics

We assessed on the complete dataset with the model we trained.

Accuracy: Determined by the division of correct predictions by the total number of samples.

• Computed per class and averaged: Precision, Recall, F1-score.

• Confusion Matrix: Used to see if the system is confounding one class for another

7. Application Deployment

A user interface for detecting diseases was developed using Streamlit. The following are features the app provides:

• Leaf image uploader

• Softmax confidence scores for the top 3 predicted diseases

• Immediate feedback after the image is uploaded

To execute the application: streamlit run app/streamlit_app.py

8. Discussion and Conclusion

The PlantVillage dataset showed that the Swin Transformer could achieve state-of-the-art results in very few training epochs and with minimal preprocessing of the data. Classical convolutional networks may take up to several hours to train even when provided with top-tier hardware. In contrast, the Swin Transformer can yield intelligible results in an evening's time.

Strengths:

• GPU training is fast (less than 10 minutes)

• Predictions are made in real time

• Providing probabilities for the top three classes gives users much more contextual information.

Limitations:

The dataset is collected in a laboratory setting; accuracy in real-world applications may differ.

• No accounting for the presence of multiple leaves or for intricate backgrounds.

9. Future Work

• Implement Grad-CAM for model interpretability

• Deploy to mobile using ONNX/TFLite

• Acquire and analyze images from the field to validate in real-world conditions.

App extension for multi-class or multi-label predictions support.

10. References

- Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale"
- Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows"
- PlantVillage Dataset
- TIMM Library
- Streamlit