Report: Vine Fungi Detection Using Vision Transformers

Introduction

This report presents the development and evaluation of a binary classification model for detecting the presence of fungi in vine microscopy images. The project uses a Vision Transformer (ViT) architecture fine-tuned on a specialized microscopy image dataset to distinguish between healthy and fungi-infected vine samples.

Dataset Overview

The model was trained on "An Eye on the Vine" dataset, which contains fluorescence microscopy images of vine wood samples. For this specific task, two subsets were used:

- **No Fungi (B2)**: 312 images of healthy vine samples
- **Fungi (B3)**: 569 images showing fungi presence

All images in the dataset are uniformly sized at 256×256 pixels and stored in RGB format, making them suitable for direct processing with the ViT model. The class distribution reveals a slight imbalance, with approximately 65% of images containing fungi.

Methodology

Data Preparation

The dataset was processed using a custom FungiDataset class to handle the TIF image format and converted to the HuggingFace dataset format for easier manipulation. Images were preprocessed using the ViT feature extractor to ensure compatibility with the model's input requirements.

Model Architecture

The implementation leveraged transfer learning by fine-tuning a pre-trained Vision Transformer (google/vit-base-patch16-224-in21k). This architecture, which treats images as sequences of patches, has demonstrated strong performance on image classification tasks without the inductive biases present in convolutional neural networks.

Training Approach

The model was trained with the following parameters:

- Training/test split: 80%/20%
- Batch size: 16
- Learning rate: 2e-4
- Number of epochs: 4
- Evaluation strategy: Per epoch
- Optimizer: AdamW (default in Transformers library)

Results and Analysis

The model demonstrated excellent performance in distinguishing between images with and without fungi. The evaluation metrics improved consistently across training epochs:

Epoch	Training Loss	Validation Loss	Accuracy	F1 Score	F1 No Fungi	F1 Fungi
1	0.2784	0.2333	0.9096	0.9101	0.8519	0.9350
2	0.2574	0.1986	0.9209	0.9213	0.8704	0.9431
3	0.0900	0.1438	0.9435	0.9432	0.9038	0.9600
4	0.0609	0.1490	0.9492	0.9495	0.9174	0.9633

The final model achieved:

- **Accuracy**: 94.35%

- **Overall F1 Score**: 94.32%
- **Class-specific F1 Scores**:
    - No Fungi: 90.38%
    - Fungi: 96.00%
- **Precision**:
    - No Fungi: 92.16%
    - Fungi: 95.24%
- **Recall**:
    - No Fungi: 88.68%
    - Fungi: 96.77%

Performance Analysis

1. **Class Performance Disparity**: The model performed slightly better on the "Fungi" class (F1: 96.00%) compared to the "No Fungi" class (F1: 90.38%). This difference may be attributed to:

    - The imbalanced dataset with more "Fungi" examples
    - Potentially more distinctive visual patterns in fungi-infected samples
2. **Precision vs. Recall**: For the "No Fungi" class, precision (92.16%) was higher than recall (88.68%), indicating that the model was more conservative in predicting this class but more accurate when it did. This suggests some fungi-free samples were incorrectly classified as containing fungi.

3. **Training Convergence**: The substantial decrease in training loss (from 0.2784 to 0.0609) indicates good model convergence. However, the slight increase in validation loss in the final epoch (from 0.1438 to 0.1490) may suggest the onset of overfitting, despite continued improvements in accuracy and F1 scores.

Conclusions

The Vision Transformer-based model demonstrates strong performance for automated detection of fungi in vine microscopy images. With an overall accuracy of 94.35% and F1 score of 94.32%, the model provides a reliable tool for identifying fungi presence, which could significantly aid in early detection of vine diseases.

The slightly lower performance on the "No Fungi" class suggests that further improvements might be achieved by:

1. Addressing the class imbalance through techniques such as weighted sampling or loss functions
2. Expanding the dataset with more "No Fungi" examples
3. Exploring data augmentation to increase the diversity of training samples

This successful implementation showcases the effectiveness of Vision Transformers for specialized microscopy image classification tasks, potentially offering a valuable tool for vineyard disease management and research.