

Plant Wild Multimodal Plant Disease Recognition

Lecturer : Dr. Hajer Saleh

Prepared By: Hussein Mudeheg

1. Overview

Plant disease recognition has become an important topic in the research area of Artificial Intelligence and deep learning due to its impact on the agriculture and food security. Early experiments studies such as Mohanty et al (2016) has demonstrated that deep convolutional neural networks (CNN) like AlexNet and GoogleNet could support to achieve high accuracy when trained on lab-controlled datasets such as PlantVillage. Similarly, Too et al (2019) has compared multiple CNN architecture including VGG16, ResNet50, and DenseNet121, which report an accuracy rate that exceed 99% under controlled conditions

Although, the results were very promising, the studies relied mainly on the PlantVillage dataset, which contain images captured in clean laboratory environment with unified background. The models performed very well in these settings, the ability to generalize to real world plants condition remained limited

2. Methodology Analysis

This section covers the summary of three experimental studies focusing on plant disease recognition model and their evaluation

2.1 Referenced Models

The selected experimental studies has implemented the following Models

1) Too Et al. 2019

This model were implemented using transfer learning techniques

- VGG16
- ResNet50
- DenseNet121

2) Mohanty et al., 2016

These deep learning CNN were trained for planet disease classification using supervised learning

- AlexNet
- GoogleNet

3) Wei et al., 2024

This Model utilized both visual and textual information to improve the recognition in the real live conditions

- MVPDR (Multimodal Versatile Plant Disease Recognition)
- CLIP-based Vision–Language Model

2.2 Preprocessing

The preprocessing techniques applied in the selected studies are as follow

Too et al., 2019

- Image resizing (256 × 256 pixels)
- Normalization
- Data augmentation (rotation, flipping)

Mohanty et al., 2016

- Image resizing
- RGB and grayscale comparison
- Background removal experiments
- Dataset splitting

Wei et al., 2024

- Real-world image cleaning
- Image resizing and normalization
- Text prompt engineering
- Feature normalization
- Prototype generation per disease class

2.3 DataSet

2.3.1 **PlantVillage** Dataset were used in Too et al., 2019 and Mohanty et al., 2016 with the following characteristics, and has a high performance in lab settings but reduced generalization in real world scenarios

- 54,306 images
- 38 classes
- Laboratory controlled conditions
- Clean backgrounds

2.3.2 **PlantWild Dataset** was used in Wei et al., 2024 with the detailed below characteristics

- 18,542 images
- 89 disease classes
- Real world environments of agricultural
- Support Textual descriptions

2.4 Results

Too et al., 2019

- The best model was DenseNet121
- Accuracy percentage is 99.75%
- Show high performance under laboratory conditions

Mohanty et al., 2016

- Shows the best accuracy of 99.35%
- The performance was decreased on the real world images

Wei et al., 2024

- MVPDR demonstrated strong real-world performance
- Improved robustness in few-shot scenarios
- Outperformed traditional CNN models under complex conditions

Ref	Models	Datasets	The Model	Results
Too et al., 2019	VGG16, ResNet50, DenseNet121	PlantVillage	Transfer Learning CNN	99.75% accuracy
Mohanty et al., 2016	AlexNet, GoogleNet	PlantVillage	Deep CNN Classification	99.35% (lab data)
Wei et al., 2024	MVPDR, CLIP	PlantWild	Multimodal Vision Language Model	Strong real world performance