



Plant Wild Multimodal Plant Disease Recognition

Lecturer : Dr. Hajer Saleh

Prepared By: Hussein Mudeheg

1. Abstract

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. Reference Models

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

2.1 Benchmarking In-the-Wild Multimodal Plant Disease Recognition and A Versatile Baseline (Wei et al., 2024)

The research emphasizes that the plant disease classification has reached a sufficient level under the laboratory conditions, however, they observed that the same is not applied in-the-wild and real live plant images, because, similar appearance may be demonstrated across different diseases. Therefore, they proposed to include visual and text-based descriptions to provide rich information for each disease along with the large number of disease classes' dataset which support to improve the recognition in real live conditions and recognize the disease in few shots. The research proposes a Multimodal Versatile Plant Disease Recognition (MVPDR) and used CLIP based Vision Language Model to extract visual features from training images and then groups the features within each class to generate multiple prototypes

2.1.1 Methodology

The PlantWild dataset was constructed as large scale for wild plant disease recognition with around 89 classes categorized as 33 healthy plant classes and 56 disease plant classes. The images were collected from multiple online search engines to ensure the diversity and real world variability. The images were filtering procedure used image exemplars of University of Minnesota (UMN) as annotation guidance where each image was reviewed at least by two expert annotators and in case of disagreement a third expert conduct the final review

2.1.2 Preprocessing

The preprocessing techniques applied in the selected study are as follow

- Real-world image cleaning
Irrelevant and low quality images were removed to ensure data set reliability
- Image resizing and normalization
The images were resized to be suitable for deep learning models, and Pixel normalization was applied to scandalized to improve training stability
- Text prompt engineering
Descriptive textual prompt was generated for each class of the disease using GPT 3.5
- Feature normalization
Visual and text features were scaled to standard format to ensure they work consistently
- Prototype generation per disease class

Create a typical example for each disease to act as a benchmark to notify the difference between one disease and another

2.1.3 DataSet

PlantWild Dataset was used with the detailed below characteristics

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

2.1.4 Result

- MVPDR demonstrated strong real world performance
It confirm that it is effective under complex environment with natural background and lightning variation
- Improved robustness in few-shot scenarios
The model show ability to generalize with limited labeled samples utilizing the textual prompt and prototype generation
- Outperformed traditional CNN models under complex conditions
Was able to distinguish visually similar diseases with fine grained differences

2.2 A comparative study of fine-tuning deep learning models for plant disease identification (Too et al., 2019)

This study put more focus on evaluating the effectiveness of different deep learning architecture for plant disease classification. Unlike the above paper the emphasized the real world multimodal learning, this research concentrate on benchmarking several pre trained convolutional neural network (CNN) modals through learning and ultimately evaluate their accuracy and computational efficiency

2.2.1 Methodology

The study applied transfer learning by fine tune the deep learning models which trained on large scale dataset such as ImageNet, The models were fine-tuned to replace the final classification layers and retrain them on the plant disease dataset. Performance comparison was conducted based on classification accuracy and validation metrics. The evaluated model include:

- VGG16
- VGG19
- ResNet50
- InceptionV3
- DenseNet121

2.2.2 Preprocessing

The preprocessing focused on optimizing CNN training performance under controlled dataset conditions. The preprocessing steps included:

- Image resizing
All images were resized to match the required input size of each CNN architecture
- Data normalization
Pixel values were normalized to improve training stability and convergence
- Data augmentation
Some techniques such as rotation, flipping, and scaling were applied to increase dataset variability and reduce overfitting.

2.2.3 Data Set

The study used **PlantVillage** dataset which has the following characteristics

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

The PlantVillage dataset does not contain real world background complexity, which makes classification less challenging

2.2.4 Result

- DenseNet121 achieved the highest classification accuracy among the evaluated models
- Transfer learning significantly improved performance compared to training from scratch
- The models demonstrated very high accuracy under laboratory conditions

The study did not evaluate performance in real world environments with complex backgrounds. The final results confirmed that deep CNN architectures are highly effective when trained and evaluated on clean, controlled datasets.

2.3 Using Deep Learning for Image-Based Plant Disease Detection (Mohanty et al., 2016)

This was one of the earliest studies applying deep learning for automated plant disease detection. The research demonstrated that deep convolutional neural networks can achieve high accuracy in classifying plant diseases from leaf images. The study mainly investigated whether deep learning models could outperform traditional machine learning approaches in agricultural disease detection.

2.3.1 Methodology

The authors implemented deep CNN architectures, and compared their performance in plant disease classification, using AlexNet and GoogLeNet. The model was trained using supervised learning and evaluated on labeled plant disease and Similar to Too et al. (2019), this study relied entirely on visual image features and did not incorporate textual or multimodal information.

2.3.2 Preprocessing

- Image resizing

Images were resized to fit the input requirements of the CNN models

- Normalization

Pixel intensity values were normalized for stable training

- Dataset splitting

The dataset was divided into training and testing sets to evaluate generalization performance

2.3.3 Data Set

The study also, used **PlantVillage** dataset were which has the following characteristics

- 54,306 images
- 38 classes of disease
- Laboratory controlled conditions
- Uniform background

2.3.4 Result

- The deep CNN models achieved very high classification accuracy which considered as state-of-the-art at the time
- GoogLeNet slightly outperformed comparing to AlexNet in both performance and efficiency
- The study demonstrated the strong potential of deep learning for plant disease detection.

2.4 Overall Comparison

1.Mohanty et al. (2016) established the feasibility of CNN based on plant disease detection under controlled conditions

2.Too et al. (2019) improved performance further through fine tuning and comparative evaluation of modern CNN architectures.

Wei et al. (2024) extended the research into real world multimodal recognition using PlantWild and vision language modeling, addressing few shot learning and work under environmental complexity.

Ref	Models	Datasets	The Model	Results
Wei et al., 2024	MVPDR, CLIP	PlantWild	Multimodal Vision Language Model	Strong real world performance
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)