Deep-Plant-Disease Dataset Is All You Need for Plant Disease Identification

Abel Yu Hao Chai
Swinburne University of Technology
Sarawak Campus
Kuching, Sarawak, Malaysia
abel_cyh94@hotmail.com

Kelly Li Zhen Jee
Swinburne University of Technology
Sarawak Campus
Kuching, Sarawak, Malaysia
kjee@swinburne.edu.my

Sue Han Lee
Swinburne University of Technology
Sarawak Campus
Kuching, Sarawak, Malaysia
shlee@swinburne.edu.my

Fei Siang Tay
Swinburne University of Technology
Sarawak Campus
Kuching, Sarawak, Malaysia
fstay@swinburne.edu.my

Jules Vandeputte
INRIA
Montpellier, France
jules.vandeputte@inria.fr

Hervé Goëau
AMAP, Univ Montpellier, IRD, CNRS,
INRAE, CIRAD
Montpellier, France
herve.goeau@cirad.fr

Pierre Bonnet
AMAP, Univ Montpellier, IRD, CNRS,
INRAE, CIRAD
Montpellier, France
pierre.bonnet@cirad.fr

Alexis Joly
INRIA
Montpellier, France
alexis.joly@inria.fr

Abstract

Deep learning models have emerged as a promising alternative to conventional approaches for plant disease identification, a critical challenge in agricultural production. However, the existing plant disease datasets are insufficient to address the complexities of real-world agricultural scenarios, such as multi crop disease, unseen, few-shot, and domain shift adaptation. Additionally, the lack of standardized evaluation protocols and benchmark datasets hinders the fair evaluation of models against these challenges. To bridge this gap, we introduce Deep-Plant-Disease, the largest and most diverse dataset with novel text data designed to enhance model generalization in multi crop disease identification. We revisit and reformulate the task by establishing a standardized evaluation framework that supports consistent benchmarking and guides future research. Through experiments, we further validate the robustness and adaptability of models trained on our dataset, highlighting their effective transferability to real-world agricultural challenges.

CCS Concepts

• **Computing methodologies** → **Object identification**; *Supervised learning*.

Keywords

Plant disease identification, fine-grained image-text pair dataset, vision model, vision language model

ACM Reference Format:
Abel Yu Hao Chai, Kelly Li Zhen Jee, Sue Han Lee, Fei Siang Tay, Jules Vandeputte, Hervé Goëau, Pierre Bonnet, and Alexis Joly. 2018. Deep-Plant-Disease Dataset Is All You Need for Plant Disease Identification. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Plant diseases are caused by fungi, bacteria or viruses that adversely affect agricultural productivity. Early and accurate identification of these diseases is an important task in preventing their outbreak and minimizing their impact. Traditionally, plant disease identification has relied on the assessment of plant pathologists through visual inspection and laboratory analysis. However, these methods are often time-consuming, resource-intensive and susceptible to human error. Consequently, automated plant disease identification using deep learning models has emerged as a promising solution. These models not only improve diagnostic efficiency, but also give non-experts, including farmers and the general public, access to plant disease identification. The identification involves simultaneously recognizing both the crop species and its associated disease. Numerous studies have shown that deep learning models are capable of learning feature representations of crop and diseases symptoms based on visual appearances [4, 6]. These models have been deployed in both single crop disease identification [12, 19], which focuses on detecting diseases affecting a specific crop, and multi crop disease identification [1, 3], which targets diseases across multiple crop species. In this study, we focus on multi crop disease identification, addressing the challenges of identifying and diagnosing diseases across diverse crops in real-world agricultural settings.

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

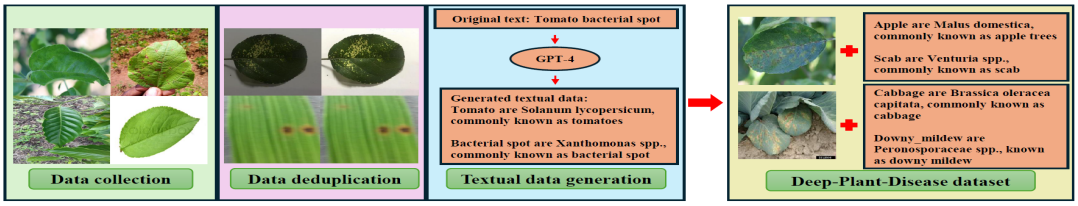

Figure 1: The curation process of our Deep-Plant-Disease dataset from 44 different public repositories. We performed data deduplication targeting samples with high visual similarity. Next, each sample was annotated with textual data using GPT-4.

In real-world applications, plant disease identification presents challenges, including generalization to unseen samples, few-shot, and domain adaptation. Recent studies have addressed these issues by reconceptualizing the task as a compositional task, where models are trained to learn and generalize both crop and disease concepts. This approach has shown promise, particularly in tackling the challenge of unseen plant disease identification [3, 4]. However, due to the limitations of existing datasets, these models can only be benchmarked on a single unseen crop-disease composition, resulting in performance that lacks convincing generalizability. There is also growing interest in exploiting text data to improve the performance of plant disease identification tasks [7, 30]. This is due to the fact that certain disease symptoms exhibit a high visual similarity, making it difficult to differentiate them using vision-based models alone. By incorporating textual information, such as metadata, symptom descriptions, or agronomic context, models can access complementary cues that guide them toward more accurate predictions. Although many recent works report promising results, they often rely on self-collected datasets and use varying evaluation metrics, making it difficult to benchmark state-of-the-art (SOTA) models consistently. As a result, the true performance of these models remains unclear. Therefore, this paper aims to establish a large-scale, standardized plant disease identification dataset, along with a unified evaluation framework to facilitate fair comparisons and improve model benchmarking in this field.

In this paper, we introduce Deep-Plant-Disease (DPD), the largest publicly available plant disease dataset, comprising 248,579 images across 55 crop species, 175 disease classes, and 333 unique crop-disease compositions. Drawing inspiration from [24], we incorporate a more plant biology-driven text to enhance the textual descriptions of crop disease symptoms. Specifically, we propose a novel scheme for generating textual descriptions for each crop and disease class, aimed at supporting the development of vision-language models. The methodology behind these text prompts will be detailed in Section 3.3. The dataset primarily focuses on leaf samples, as symptoms typically manifest on leaves in the early stages and such samples are the most accessible for data collection. The images were sourced from diverse settings, including both controlled laboratory environments and complex real-world conditions, ensuring broad applicability across various challenges. Additionally, we provide different testing dataset and a standardized evaluation protocol tailored to different environmental challenges. This enables researchers to rigorously assess model generalization and facilitates fair and consistent comparisons against SOTA methods. Our contributions are highlighted below:

(1) We curated the largest plant disease dataset with text descriptions known as Deep-Plant-Disease, comprising 248,579 images across 55 crop species, 175 disease classes, and 333 unique crop-disease compositions.
(2) We conducted comprehensive benchmarking across multiple downstream tasks in plant disease identification under diverse conditions that simulate different real-world challenges.
(3) We performed a comprehensive analysis using textual description evaluation and Grad-CAM visualization technique to demonstrate the robustness of representations learned from our proposed dataset, highlighting its effectiveness and reliability for real-world deployment.

2 Related Works

2.1 Plant Disease Datasets

Single vs Multi Crop Disease Datasets. Researchers have curated a variety of datasets to train models for the identification of plant diseases [10]. These datasets range from those focusing on a single crop to multi crop disease datasets. Table 1 summarizes some of the most recent large-scale datasets in this domain. Several datasets focus on specific crops, such as soybeans [17], cassava [18] or paddy [20]. While these datasets provide valuable insights into crop-specific disease manifestations, they are often limited in terms of scale or crop diversity. Such limitations reduce their effectiveness in evaluating a model's ability to generalize and remain robust across different crop species and diverse agricultural conditions. Various initiatives have been taken to collect more comprehensive datasets encompassing a wider range of crops. Among these, the PlantVillage (PV) dataset is widely used and remains the largest publicly available dataset for plant disease identification. However, the images in the PV dataset are often captured in controlled environments. Furthermore, the diversity of crops and diseases represented in the PV dataset remains limited, which restricts its applicability and hinders future advancements in plant disease identification.

From Controlled to Wild Setting. While PV datasets are widely used for plant disease identification, their images are primarily captured in laboratory settings. Consequently, models trained on PV data encounter significant domain shift challenges when applied in real-world environments. PlantDoc [23] and PlantWild [26] have been released to include crop disease samples from real-world environments. However, these datasets are still insufficient in terms of the number of crop and disease classes available. To fill this gap, we created the DPD dataset, which presents a more larger diversity of crops and diseases. The dataset includes 248,579 images of 55 crop species, 175 disease classes and 333 unique crop-disease compositions.

Table 1: Summary of publicly available and our proposed Deep-Plant-Disease (DPD) datasets

Dataset	I	C	D	CD	T	CI	DI	CDI	F	U	D
Soybean [17]	6,410	1	3	3			✓				
icassava [18]	22,031	1	5	5			✓				
Paddy Doctor [20]	16,225	1	13	13			✓				
PlantVillage [8]	54,309	14	26	38				✓			
PlantDoc [23]	2,598	13	17	27				✓			✓
CCDM [15]	137,000	16	48	60	✓	✓	✓	✓			✓
PDD271 [16]	220,592	47	121	271				✓			✓
PlantWild [26]	18,542	35	181	89	✓			✓			✓
Deep-Plant-Disease (DPD)	***248,579***	***55***	***175***	***333***	✓	✓	✓	✓	✓	✓	✓

I is the total number of images. C, D and CD are the total number of unique crop, disease and crop-disease pairs respectively. T is the available of text input data. CI, DI and CDI are the crop, disease and crop-disease identification tasks. F, U and D are the few shot, unseen and domain adaptation tasks.

Real-World Challenges. Existing plant disease datasets often lack the diversity and structure necessary to evaluate models under challenging real-world scenarios, such as generalization to unseen classes and few-shot learning. While some datasets, like CDDM [15] and PDD271 [16], cover a broader range of crops and diseases, their task formulations are not well-suited for evaluating model generalization to unseen crop or disease classes. This limitation arises primarily from the way crop-disease labels are constructed, typically as fixed pairs, resulting in limited flexibility for evaluating models in dynamic and practical settings, as highlighted in [3]. In contrast, our proposed dataset is explicitly designed to address these gaps. It supports more granular evaluations by enabling the identification of individual concepts (i.e., crop species or disease types independently) as well as crop-disease compositions. This structure provides greater flexibility and makes our dataset better suited for real-world plant disease identification tasks, supporting consistent and fair benchmarking across various learning challenges.

2.2 Multimodal Plant Disease Identification

Computer vision and pattern recognition, driven by deep learning, have revolutionized the application of technology in automatic plant disease identification. Specifically, deep learning approaches, primarily Convolutional Neural Networks (CNNs), have been extended from single crop to multi crop disease identification, leveraging powerful visual learning capabilities. Recent studies demonstrate that CNNs and Vision Transformers (ViTs) learn different disease patterns, and combining features from both architectures may offer complementary strengths [3, 4, 25].

However, plant disease identification remains challenging due to low inter-class variation and high intra-class variation, as different pathogens can cause similar symptoms across different plant species, or even within the same crop. Relying solely on visual cues may not be sufficient to accurately distinguish diseases. Recent research shows that integrating texture features and auxiliary information, such as textual descriptions, can enhance disease representation, even under zero-shot learning (ZSL) settings where the model has not previously encountered the disease [14]. As a result, a growing number of studies are developing multimodal methods that incorporate auxiliary information to enable ZSL. For instance, works such as [3, 9, 11, 28] incorporate semantic attributes, hierarchical class mappings, or natural language descriptions.

While these approaches show promising results, the impact of different textual prompts on disease representation has not been thoroughly studied, making it unclear which forms of text best complement visual inputs in plant disease identification. Motivated by these gaps, our work draws inspiration from BioCLIP [24] and integrates taxonomic descriptions to enrich the semantic understanding of crop disease symptoms. Through experiments, we demonstrate that our proposed textual descriptions significantly enhance the performance of vision-language models.

3 Dataset Overview

In this section, we provide a detailed overview of our plant disease dataset. A representative sample of the dataset is shown in Figure 1. Compared to previously curated datasets, our dataset exhibits greater diversity in both crop and disease types, and supports a wider range of downstream tasks. In addition, we systematically describe the dataset construction process, which includes data collection, filtering, automated annotation synchronization, text prompt generation and practical evaluation. Each step is designed to ensure the accuracy, robustness and applicability of the dataset to different challenges.

3.1 Data Collection and Filtering

The dataset proposed in this study is exclusively compiled from samples sourced from publicly available platforms such as Kaggle, Mendeley, and GitHub, as well as from existing research studies. Specifically, we aggregated samples from 44 distinct sources, as detailed in Table 3 in the supplementary materials. In total, the dataset comprises 370,342 images, representing 60 unique crop species and encompassing 221 different disease classes. The supplementary material and sample images for review are available at our GitHub repository[1]. The complete dataset will be available upon publication.

We first carried out a multi-stage filtering process. Specifically, we performed a full manual inspection of all samples. We eliminated duplicate or highly similar samples from different sources. we used the copy detection pipeline proposed by [21] on our dataset. Examples of samples detected as duplicates or nearly duplicates are shown in Figure 1. These redundant samples can originate from identical crops captured under different views. Such sample redundancies could increase the bias of the dataset and adversely affect the generalisation of the model. Therefore, we filtered out these redundant samples to improve the diversity of the dataset and reduce potential biases introduced during training. The resulting dataset consist of 248,579 images with 60 unique crop species and 212 different disease classes. To ensure dataset consistency, we restrict it to leaf images only. Leaves are not only the most abundant plant organ but also the most commonly used for plant disease identification. Additionally, other plant organs such as stems or fruits are unevenly represented across crop species, which could introduce bias and inconsistency in the learning process.

3.2 Automated Annotations Synchronization

We observed inconsistencies in the annotations across different sources. For instance, certain sources used scientific (Latin) names, while others use common names for the same class. As a result, a

[1]https://github.com/abelchai/Deep-Plant-Disease-Dataset-Is-All-You-Need-for-Plant-Disease-Identification

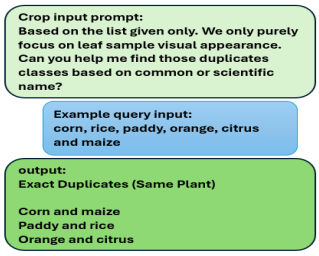

Figure 2: The example of utlitizing GPT-4 for label synchronization.

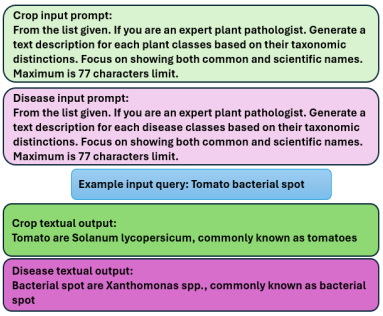

Figure 3: The example of utlitizing GPT-4 for botanical taxonomy textual data generation.

single class could be associated with multiple labels, introducing redundancy. These inconsistencies posed challenges for feature learning process, as the same crop might be interpreted as different classes, thereby degrading classification performance.

To standarized the annotations, we designed and proposed an automated annotations synchronization pipeline leveraging a Large Language Model (LLM). In particular, we employed GPT-4 to synchronise annotations or labels across all samples with our designed input prompt. An example illustrating the label synchronization process is presented in Fig. 2. By leveraging the extensive domain knowledge learnt in the LLM, our method was able to effectively synchronize annotations across datasets, significantly reducing the reliance on manual curation by domain experts or plant pathologists, while enhancing the consistency and reliability of the dataset for model training and evaluation.

3.3 Textual Data Generation

As discussed in Section 1, textual data can provide additional information for visual data to improve the feature extraction of different classes. Moreover, [14] demonstrated that incorporating textual information can significantly improve model performance, particularly in generalization to previously unseen tasks. At first, each crop and disease category was annotated with only simple class-level textual labels. However, [22] showed that structured textual prompts, rather than simple labels, can substantially enhance the feature learning process by providing more detailed contextual information.

Motivated by these findings, we aim to design textual prompts tailored to our tasks. Drawing inspiration from [24], which highlights the advantages of incorporating taxonomic and hierarchical knowledge into prompts for visual language models (VLMs), we hypothesize that structured textual prompts based on botanical taxonomy can enhance model performance across diverse tasks. Hence, we leverage the knowledge learnt from GPT-4, to automate the generation of botanical taxonomy textual description. Specifically, we design a task specific prompt to restrict GPT-4 to our predefined list and generate a botanical taxonomy textual description output for our dataset. The full input prompt and a detailed example of this process are illustrated in Figure 3. We also experimentally proven that our carefully designed textual prompt able to improve the performance of VLMs.

3.4 Benchmark Datasets and Evaluation Protocols

In this section, we describe the benchmark datasets utilized across multiple downstream tasks to evaluate the model's generalization capability in multi-crop disease identification, including zero-shot and few-shot scenarios. Additionally, we outline the benchmarking protocols designed to assess the model's ability to identify crops and diseases independently, from which the joint crop-disease performance is derived. This approach offers more granular insights into the model's effectiveness across both dimensions.

Benchmark Datasets. We strategically select 5 datasets as our downstream tasks, corresponding to different tasks or environmental conditions. Specifically, the selected datasets include PDD271 (PDD) [16], PlantDoc (PD) [23], PlantWildv3 (PWv3) [26], IDADP [29], and Herbarium. Details of the PDD, PD, PWv3, and IDADP datasets are presented in Table 1 in the supplementary material. The Herbarium dataset [13], sourced from the JSTOR Global Plants digital library, comprises 164 images with 91 crop species and 75 disease classes.

To evaluate the model under more challenging conditions, we additionally provide tailored training and testing splits of our dataset for unseen and few-shot plant disease identification tasks. For unseen task, we exclude 10 specific compositions from the training dataset as unseen compositions. These compositions are *Cherry aphis spp*, *Corn brown spot*, *Cotton bacterial blight*, *Cucumber downy mildew*, *Cucumber powdery mildrew*, *Grape black rot*, *Mango anthracnose*, *Pepper bell bacterial spot*, *Sugarcane rust* and *Tomato early blight*. For few show task, we define compositions with fewer than 50 samples in training set as few-shot classes, reflecting real-world situation where certain compositions are difficult to collect.

Standardized Evaluation Protocol. Traditional plant disease identification often evaluates performance on combined crop-disease classes (e.g., "tomato rust"). However, this masks how well models generalize across individual crops or diseases. To address this, we propose evaluating models both independently and jointly accuracies with crop acc $= \frac{1}{N}\sum_{i=1}^{N} 1(\hat{c}_i = c_i)$, disease acc $= \frac{1}{N}\sum_{i=1}^{N} 1(\hat{d}_i = d_i)$ and crop disease acc $= \frac{1}{N}\sum_{i=1}^{N} 1((\hat{c}_i, \hat{d}_i) = (c_i, d_i))$. Where $\hat{c}_i$ and $\hat{d}_i$ denoted as the predicted crop and disease labels. c_i and d_i denoted as the ground truth crop and disease labels. N is the total number of samples.

This protocol enables deeper insights into model generalization across both dimensions and is especially valuable in zero-shot or few-shot settings, where unseen crop-disease combinations must be inferred compositionally. Furthermore, to ensure a balanced evaluation across both seen and unseen or few-shot tasks, we report the harmonic mean accuracy, defined as $HM = \frac{2AB}{A+B}$, where A

and B represent the accuracies on seen and unseen (or few-shot) classes respectively, providing a more comprehensive measure of the model's generalization towards different tasks [27].

4 Experimental Results

We begin this section by evaluating the generalization ability of our proposed DPD dataset, comparing its performance against other large-scale benchmark datasets. To ensure a fair comparison, we use the same backbone model across all experiments, specifically, Vision Transformer (ViT_base_patch16_224). Subsequently, we study the effectiveness of our generated text description for each crop and disease labels with CLIP [22] and Bio-CLIP [24]. We select Bio-CLIP due to the model is trained with TREEOFLIFE-10M dataset which are closer to our crop disease domain and CLIP is the backbone they used. Next, we evaluate state-of-the-art (SOTA) models under unseen and few-shot real-world plant disease challenges to establish a standardized evaluation protocol and benchmarking results. To enhance model interpretability, we perform qualitative analysis using GradCAM visualization techniques. For reliability and reproducibility, all models are trained three times using ImageNet-pretrained weights, and their performance is reported as the average accuracies. We trained all models for a total of 60 epochs with learning rate of 0.001. We utilise SGD optimizer with momentum of 0.9 and weight decay of 0.00001.

4.1 Generalization Ability Evaluation

Table 2 presents a performance comparison between the proposed DPD dataset and several existing datasets using a linear probing approach to evaluate generalization capability across diverse downstream tasks. Notably, PDD and IDADP datasets represent multi crop and single crop disease identification tasks respectively. PD and PWv3 datasets also target multi crop disease identification, but incorporate samples from various plant organs such as fruits and stems. In addition, the Herb dataset is employed to evaluate cross domain generalization, testing the model's ability to transfer from fresh leaf samples to dried herbarium specimens. Based on the experimental results, the model pretrained on our DPD dataset outperformed those pretrained on other datasets, achieving a superior average accuracy of 72.54%. Notably, it outperformed the model pretrained on PlantNet330K by 3.16% in Top-1 Average (Avg) Accuracy, despite PlantNet330K covering a broader crop diversity with 1,081 unique crop species and focusing solely on multi crop identification without disease classes. These findings suggest that pretraining on our DPD dataset, which is enriched with disease-specific features, significantly enhances the model's ability to learn more generalizable representations for downstream tasks. Moreover, while the model pretrained on the PWv3 dataset, specifically designed for multi crop disease identification with integrated data from multiple plant organs, shows promising performance, it demonstrates limited effectiveness in identifying plant diseases in herbarium specimens, indicating poor generalization under domain shift conditions. Furthermore, a noteworthy finding is that the model pretrained on our DPD dataset achieves higher accuracy on PWv3 testing dataset compared to the model pretrained directly on the PWv3 dataset. Specifically, the model trained with our DPD dataset outperforms the PWv3-pretrained model by 3.53%. This highlights that the increased sample diversity and larger sample size in our DPD dataset significantly bolster the model's generalization capabilities across various downstream tasks.

Table 2: Generalization ability evaluation showing the Top 1 accuracy for each dataset, along with the overall average accuracy across all datasets.

Pretrained dataset	Top 1 Accuracy					
	PDD	IDADP	PD	PWv3	Herb	Avg
ImageNet-21k	84.09	99.17	54.24	65.10	16.18	63.76
ImageNet-1k (SSL)	82.92	99.45	53.81	61.39	27.94	65.10
PlantNet300k	88.07	99.31	52.97	71.24	**35.29**	69.38
PWv3	88.56	99.59	**58.05**	78.95	25.00	70.03
Deep-Plant-Disease	**88.61**	**99.72**	**58.05**	**82.48**	33.82	**72.54**

Avg is the average accuracies among all 5 downstream tasks.

Table 3: Textual description evaluation showing the plant (P) and disease (D) identification accuracies.

Model	P acc	D acc
CLIP (label)	24.94	0.45
CLIP (a photo of label)	27.41	0.37
CLIP (our proposed text description)	**33.19**	**5.50**
Bio-CLIP (label)	29.30	2.36
Bio-CLIP (a photo of label)	45.89	2.14
Bio-CLIP (our proposed text description)	**46.86**	**3.09**

P acc and D acc are plant and disease identification accuracies respectively.

Table 4: Unseen task on 10 unseen classes.

Model	Seen acc	Unseen acc	HM acc
ViT (ViT_base_patch16_224)	**78.53**	2.69	5.20
FF-ViT [5]	77.81	4.35	8.24
CL-ViT [3]	78.10	4.63	8.74
FF-CLIP [14]	76.14	**7.11**	**13.01**

Seen is the accuracy performance of classes that are available in the training dataset. HM is harmonic mean accuracy that balance between seen and unseen accuracies.

4.2 Textual Description Evaluation

A comparative experiment was conducted on CLIP [22] and Bio-CLIP [24] to evaluate the models' performance using three types of textual descriptions: class label-only (e.g., "rice", "tomato" for crops), label with an image prompt (e.g., "a photo of a tomato"), and our proposed descriptive text (e.g., "tomato are Solanum lycopersicum, commonly known as tomatoes"). We evaluate the effectiveness of these descriptions in zero-shot classification contexts following the protocol described in [22], in which images are classified based on their similarity to provided input text data. As shown in Table 3, the models using our proposed textual descriptions perform significantly better than those using other textual inputs, achieving the highest accuracy in the identification of both crops and diseases. These results indicate that our textual description based on botanical taxonomy facilitates a more meaningful alignment between textual and visual inputs.

4.3 Unseen and Few-Shot identification tasks

In this section, we present benchmarking results using various SOTA models for unseen and few-shot plant disease identification

tasks. The training and evaluation datasets used for each challenge are described in detail in Section 3.4.

For the unseen plant disease identification task, we evaluate performance using FF-ViT [5], CL-ViT [3] and FF-CLIP [14], which represent the latest SOTA in this area. It is important to note that prior studies on unseen plant disease identification [3, 5, 14] have assessed model generalization using only a single unseen class, namely *Pepper bell bacterial spot*, due to limitations in the dataset employed. In contrast, our DPD dataset includes a larger number of unseen classes (10 in total) with more images, offering a more comprehensive evaluation of model generalization. The results shown in Table 4 indicate that existing frameworks still have considerable room for improvement. Notably, there exists a trade-off between generalization and specialization—performance on seen tasks often inversely correlates with that on unseen tasks. This imbalance contributes to the persistent performance gap between seen and unseen tasks. This also highlights the challenge of achieving an optimal balance between these two objectives.

For few shot tasks, we evaluate the performance using conventional ViT (ViT_base_patch16_224) and CNN (Resnet152) models. In addition, we benchmark our results against the PlantAIM [4] model, which demonstrated SOTA performance on few show plant disease identification tasks, and DinoV1 [2], a self-supervised model with same ViT backbone. PlantAIM utilizes both ViT and CNN models to extract new features from their respective focus regions for plant disease identifications. As a result, PlantAIM achieves a HM accuracy of 66.24%, outperforming all models by at least 1.70%. This demonstrates the effectiveness of leveraging both ViT and CNN in few-shot plant disease identification tasks. These experiments highlight the effectiveness of our DPD dataset as a robust foundation for advancing research in plant disease identification, particularly under varied and challenging conditions.

4.4 Visualization Analysis

In this section, we perform a comprehensive Grad-CAM visualization analysis to examine the attention regions activated by different models trained on various pretraining datasets. A key observation is that the model trained with our DPD dataset is able to focus on crop disease-relevant features when identifying both crops and diseases. In contrast, models trained on ImageNet (ImageNet-21 or ImageNet-1k) or PlantNet300k tend to be distracted by background noise in herbarium (Herb) samples or disease images cluttered with noisy backgrounds (PD). Additionally, it is interesting to observe that the model pretrained on DPD captures distinct features within individual crop and disease domains. For example, when fine-tuned on the PDD dataset, the model selectively focuses on leaf characteristics such as stems and venation for plant identification, while concentrating on infected areas for disease identification, demonstrating the robustness of the features learned from DPD dataset. Besides, we also performed misclassification analysis in the supplementary materials.

5 Conclusion

In this study, we present the Deep-Plant-Disease dataset, which comprises 248,579 images covering 55 crop species, 175 disease categories and 333 unique crop-disease combinations. We outline the dataset curation pipeline, from data collection to textual data generation. To our knowledge, DPD is currently the largest publicly available dataset that integrates visual and botanical taxonomy textual data. Experimental results and visualisation analyses demonstrate that the models trained on our dataset have strong generalisation capabilities for multiple downstream tasks. Beyond the dataset itself, our contribution also includes a set of standardised and reproducible evaluation protocols designed to facilitate future research into different plant disease challenges, such as unseen and few-shot plant disease identifications.

Table 5: Few shot task on classes with less than 50 images.

Model	Seen acc	FS 50s acc	HM 50s acc
Resnet152	76.26	48.71	59.45
ViT (ViT_base_patch16_224)	78.11	54.98	64.54
DinoV1 [2]	69.14	29.52	41.37
PlantAIM [4]	**78.67**	**57.20**	**66.24**

Seen is the accuracy performance of classes that have more than 50 samples available in the training dataset. FS 50s is the accuracy performance of classes that have less than or equal to 50 samples in the training dataset. HM is harmonic mean accuracy that balance between both accuracies.

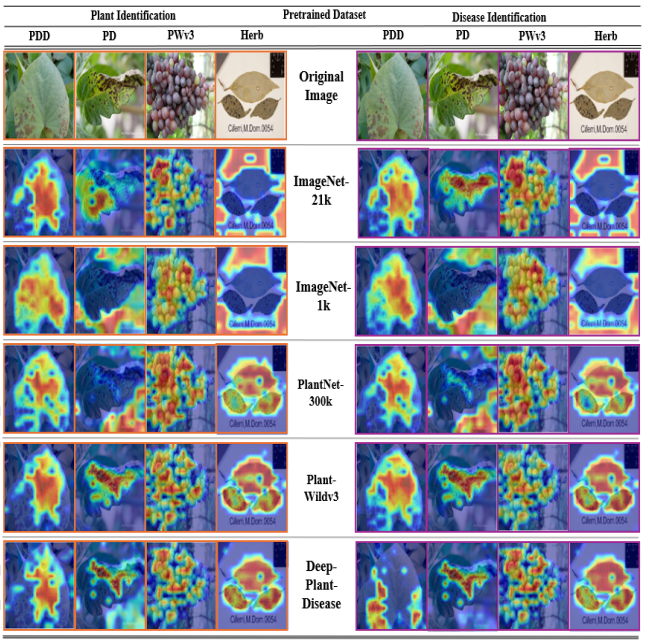

Figure 4: GradCAM visualizations across different samples.

Acknowledgments

This research is supported by the Fundamental Research Grant Scheme (FRGS) MoHE Grant (Ref: FRGS / 1 / 2021 / ICT02 / SWIN / 03 / 2), from the Ministry of Higher Education Malaysia and Swinburne Sarawak Research Supervision Grants (SSRSG) (Ref: SUTS / SoR / RMC / SSRGS / 2023). This work also benefited from a government grant managed by the Agence Nationale de la Recherche as part of the France 2030 programme, through the Pl@ntAgroEco project, under the reference "ANR-22-PEAE-0009". We also gratefully acknowledged the support of NEUON AI with the GPU workstation used for this research.

References

[1] Srabani Biswas, Ipsita Saha, and Abanti Deb. 2024. Plant disease identification using a novel time-effective CNN architecture. *Multimedia Tools and Applications* 83, 35 (2024), 82199–82221.

[2] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. 2021. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9650–9660.

[3] Abel Yu Hao Chai, Sue Han Lee, Fei Siang Tay, Pierre Bonnet, and Alexis Joly. 2024. Beyond supervision: Harnessing self-supervised learning in unseen plant disease recognition. *Neurocomputing* 610 (2024), 128608.

[4] Abel Yu Hao Chai, Sue Han Lee, Fei Siang Tay, Hervé Goëau, Pierre Bonnet, and Alexis Joly. 2025. PlantAIM: A New Baseline Model Integrating Global Attention and Local Features for Enhanced Plant Disease Identification. *Smart Agricultural Technology* (2025), 100813.

[5] Abel Yu Hao Chai, Sue Han Lee, Fei Siang Tay, Yi Lung Then, Hervé Goëau, Pierre Bonnet, and Alexis Joly. 2023. Pairwise feature learning for unseen plant disease recognition. In *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 306–310.

[6] Baofang Chang, Yuchao Wang, Xiaoyan Zhao, Guoqiang Li, and Peiyan Yuan. 2024. A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification. *Expert Systems with Applications* 237 (2024), 121638.

[7] Jiuqing Dong, Yifan Yao, Alvaro Fuentes, Yongchae Jeong, Sook Yoon, and Dong Sun Park. 2024. Visual information guided multi-modal model for plant disease anomaly detection. *Smart Agricultural Technology* 9 (2024), 100568.

[8] David Hughes, Marcel Salathé, et al. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060* (2015).

[9] Ruixiang Jiang, Lingbo Liu, and Changwen Chen. 2023. CLIP-Count: Towards Text-Guided Zero-Shot Object Counting. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. Association for Computing Machinery, New York, NY, USA, 4535–4545. doi:10.1145/3581783.3611789

[10] Alexis Joly, Pierre Bonnet, Antoine Affouard, Jean-Christophe Lombardo, and Hervé Goëau. 2017. Pl@ntNet - My Business. *Proceedings of the 25th ACM international conference on Multimedia* (2017). https://api.semanticscholar.org/CorpusID:34644257

[11] Pranav Kumar, Jimson Mathew, Rakesh Kumar Sanodiya, Thanush Setty, and Bhanu Prakash Bhaskarla. 2024. Zero shot plant disease classification with semantic attributes. *Artificial Intelligence Review* 57, 11 (2024), 305.

[12] Ghazanfar Latif, Sherif E Abdelhamid, Roxane Elias Mallouhy, Jaafar Alghazo, and Zafar Abbas Kazimi. 2022. Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model. *Plants* 11, 17 (2022), 2230.

[13] Sue Han Lee, Zhe Rui Liaw, Yu Hao Chai, Shien Lin Ng, Pierre Bonnet, Hervé Goëau, and Alexis Joly. 2024. Revolutionizing Plant Pathogen Conservation: The Past, Present, and Future of AI in Preserving Natural Ecosystems. *Biodiversity Information Science and Standards* 8 (2024), e133055.

[14] Jerad Zherui Liaw, Abel Yu Hao Chai, Sue Han Lee, Pierre Bonnet, and Alexis Joly. 2025. Can Language Improve Visual Features For Distinguishing Unseen Plant Diseases?. In *International Conference on Pattern Recognition*. Springer, 296–311.

[15] Xiang Liu, Zhaoxiang Liu, Huan Hu, Zezhou Chen, Kohou Wang, Kai Wang, and Shiguo Lian. 2024. A Multimodal Benchmark Dataset and Model for Crop Disease Diagnosis. In *European Conference on Computer Vision*. Springer, 157–170.

[16] Xinda Liu, Weiqing Min, Shuhuan Mei, Lili Wang, and Shuqiang Jiang. 2021. Plant disease recognition: A large-scale benchmark dataset and a visual region and loss reweighting approach. *IEEE Transactions on Image Processing* 30 (2021), 2003–2015.

[17] Maria Eloisa Mignoni, Aislan Honorato, Rafael Kunst, Rodrigo Righi, and Angélica Massuquetti. 2022. Soybean images dataset for caterpillar and Diabrotica speciosa pest detection and classification. *Data in Brief* 40 (2022), 107756.

[18] Ernest Mwebaze, Timnit Gebru, Andrea Frome, Solomon Nsumba, and Jeremy Tusubira. 2019. iCassava 2019 fine-grained visual categorization challenge. *arXiv preprint arXiv:1908.02900* (2019).

[19] Ishak Pacal. 2024. Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. *Expert Systems with Applications* 238 (2024), 122099.

[20] Petchiammal, Briskline Kiruba, Murugan, and Pandarasamy Arjunan. 2023. Paddy doctor: A visual image dataset for automated paddy disease classification and benchmarking. In *Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*. Association for Computing Machinery, 203–207.

[21] Ed Pizzi, Sreya Dutta Roy, Sugosh Nagavara Ravindra, Priya Goyal, and Matthijs Douze. 2022. A self-supervised descriptor for image copy detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14532–14542.

[22] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmLR, 8748–8763.

[23] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. 2020. PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. Association for Computing Machinery, 249–253.

[24] Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, et al. 2024. Bioclip: A vision foundation model for the tree of life. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 19412–19424.

[25] Poornima Singh Thakur, Shubhangi Chaturvedi, Pritee Khanna, Tanuja Sheorey, and Aparajita Ojha. 2023. Vision transformer meets convolutional neural network for plant disease classification. *Ecological Informatics* 77 (2023), 102245.

[26] Tianqi Wei, Zhi Chen, Zi Huang, and Xin Yu. 2024. Benchmarking in-the-wild multimodal disease recognition and a versatile baseline. In *Proceedings of the 32nd ACM International Conference on Multimedia*. Association for Computing Machinery, 1593–1601.

[27] Kai Yi and Mohamed Elhoseiny. 2021. Domain-Aware Continual Zero-Shot Learning. *ArXiv* abs/2112.12989 (2021). https://api.semanticscholar.org/CorpusID:245502766

[28] Jiarui Yu, Haoran Li, Y. Hao, Bin Zhu, Tong Xu, and Xiangnan He. 2023. CgT-GAN: CLIP-guided Text GAN for Image Captioning. *Proceedings of the 31st ACM International Conference on Multimedia* (2023). https://api.semanticscholar.org/CorpusID:261076397

[29] Yuan Yuan and Lei Chen. 2023. An image dataset for IDADP-grape disease identification. doi:10.11922/sciencedb.j00001.00311

[30] Yueyue Zhou, Hongping Yan, Kun Ding, Tingting Cai, and Yan Zhang. 2024. Few-Shot Image Classification of Crop Diseases Based on Vision–Language Models. *Sensors* 24, 18 (2024), 6109.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009