Supplementary Materials: 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

1 Supplementary Materials

1.1 Benchmark Datasets

In this section, we provide a detailed description of the benchmark datasets utilized to assess the generalization capabilities of models pretrained on different datasets. Table 1 summarizes each dataset, including training images, testing images, number of unique crop classes, disease classes and crop-disease classes. Besides, Figure 1, 2, 3, 4 and 5 also show some sample images from PDD, IDADP, PD, PWv3 and Herb dataset respectively. These datasets encompass a diverse set of tasks and environmental conditions. Specifically, PDD [16] focuses on multi crop disease identification using only leaf images. PD [34] and PWv3 [41] evaluate multi crop disease classification across various plant organs, including leaves, fruits, and stems. IDADP [46] targets single crop disease classification, while Herb [15] consists exclusively of dried herbarium specimens.

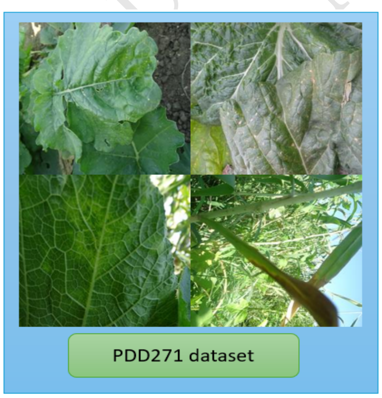

Figure 1: The sample images from PDD dataset which focus on multi crop disease identification for solely leaf samples.

Table 1: The train and test splitting of datasets for generalization experiments.

Dataset	Train Images	Test Images	Number of Crop	Disease	Crop-Disease
PDD	8,128	2,037	47	121	271
IDADP	2,893	726	1	7	7
PD	2,316	236	13	17	27
PWv3	7,916	2,295	35	71	115
Herb	73	62	62	61	62
PN-300K	243,916	31,113	1081	-	-
DPD (ours)	198,711	49,867	55	175	333

Crops and Diseases are the total number of unique crop and disease classes respectively. Images are the total number of samples for the dataset.

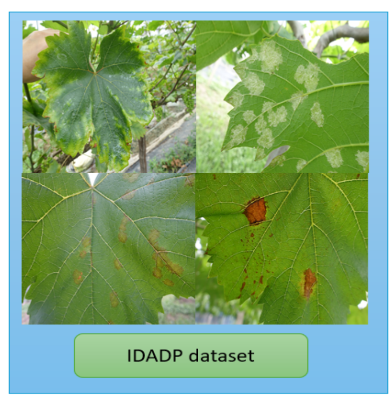

Figure 2: The sample images from IDADP dataset which focus on single crop disease identification for grape species.

Figure 3: The sample images from PD dataset which focus on multi crop disease identification under real-world environment.

Figure 4: The sample images from PW dataset which focus on multi crop disease identification under real-world environment.

Figure 5: The sample images from Herb dataset which focus on multi crop disease identification for dried herbarium specimens.

1.2 Deep-Plant-Disease Composition

There are a total of 44 datasets to formed the proposed dataset Deep-Plant-Disease (DPD) as summarized in Table 2. These datasets are found from public repository such as Kaggle, Mendeley Data, and GitHub. The datasets are used solely for research purposes and in accordance with the appropriate licenses and citation requirements. Manual inspection were performed during selecting samples into the DPD dataset. The distribution of crop, disease and crop-disease classes are also shown in Figure 6, Figure 7 and Figure 8 respectively.

1.3 Plant and Disease Textual Description Generation

All crop and disease labels in our DPD dataset are accompanied by textual descriptions. The generation of these botanical taxonomy textual descriptions is inspired by prior study in [35], which demonstrated that incorporating taxonomic information can significantly enhance model performance in both unseen and few shot identification tasks. The example of crop or disease textual description are shown in Table 3 and 4. The complete list will be publicly available via our GitHub repository[1] upon publication.

1.4 Finetuning on Benchmark Tasks

Table 6 presents a comparative analysis of the feature representations learned from different pretraining datasets, evaluated across multiple benchmark datasets. We adopt the fine-tuning protocol applied in [6, 14], wherein the entire model, including feature extractor is fine-tuned using a learning rate of 0.001. The training and testing split including unique crop or disease classes are shown in Table 5. Experimental results demonstrate that models pretrained on the DPD dataset outperform those pretrained on other datasets by at least 1.09% in Top-1 average (Avg) accuracy for plant disease

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

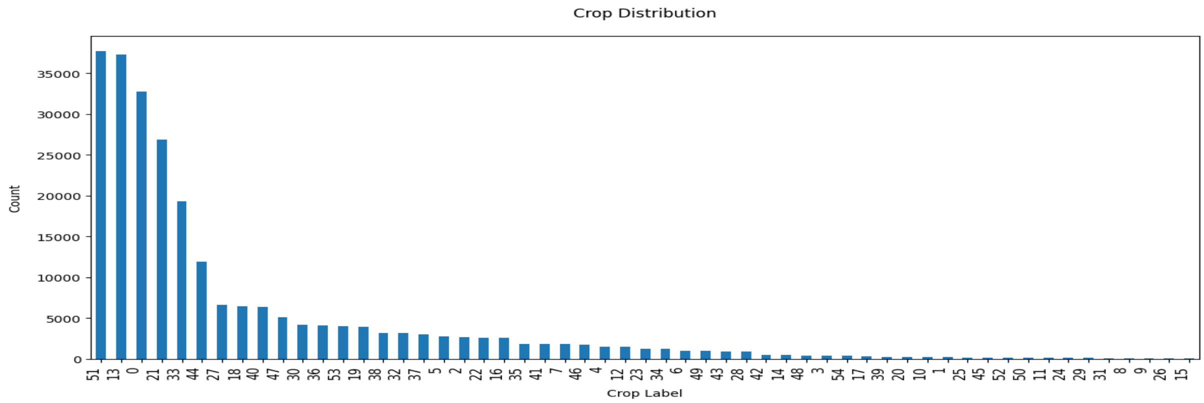

Figure 6: The distribution of crop classes for our proposed DPD dataset.

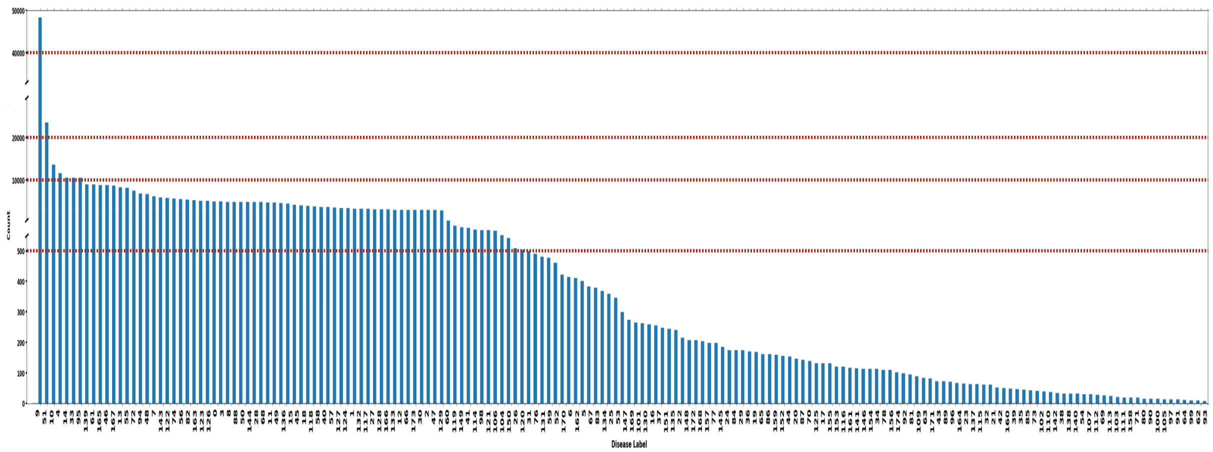

Figure 7: The distribution of disease classes for our proposed DPD dataset.

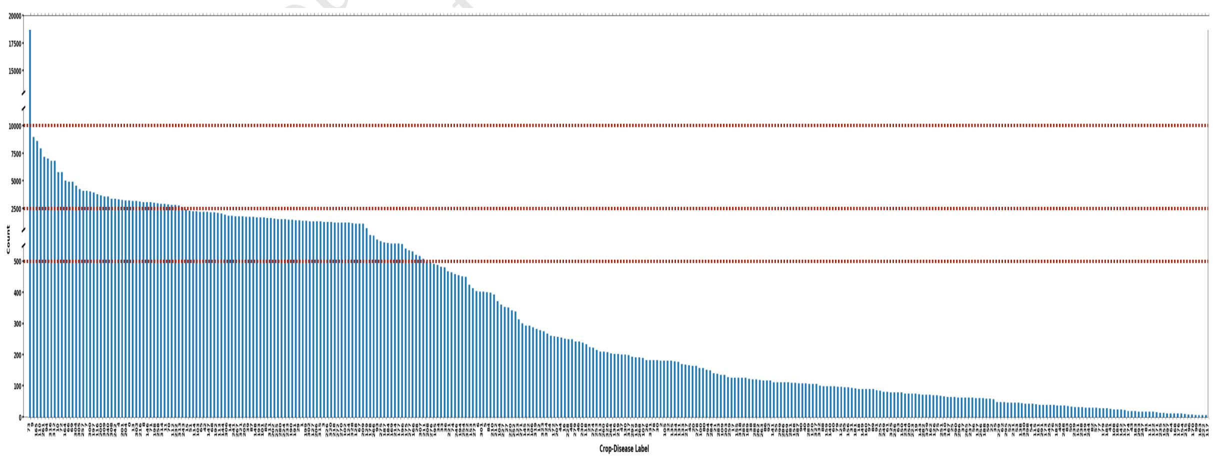

Figure 8: The distribution of crop-disease classes for our proposed DPD dataset.

Table 2: Datasets Licensing

Dataset	Link	License*
PlantWildv2 [42]	https://tqwei05.github.io/PlantWild/	1
Cassava Leaf Disease Image Dataset [27]	https://scholarsphere.psu.edu/resources/215d1acd-2c1e-440b-a27a-03d212761ef7	2
MangoLeafBS Dataset [1]	https://data.mendeley.com/datasets/hxsnvwty3r/1	3
Cucumber Disease Recognition Dataset [36]	https://data.mendeley.com/datasets/y6d3z6f8z9/1	4
Crop Pest and Disease Detection [18]	https://data.mendeley.com/datasets/bwh3zbpkpv/1	4
Coffee Crop [5]	https://data.mendeley.com/datasets/vfxf4trtcg/5	4
ESCA Dataset [3]	https://data.mendeley.com/datasets/89cnxc58kj/1	4
FlowerNet [31]	https://data.mendeley.com/datasets/7z67nyc57w/2	4
Guave Dataset [26]	https://data.mendeley.com/datasets/x84p2g3k6z/1	4
Groundnut Leaf Dataset [2]	https://data.mendeley.com/datasets/22p2vcbxfk/3	4
Images of Soybean Leaves [19]	https://data.mendeley.com/datasets/bycbh73438/1	4
Sugarcane Leaf Disease Dataset [7]	https://data.meley.com/datasets/9424sndekmnrk/1	4
Sun Flower Fruits and Leaves Dataset [29]	https://data.mendeley.com/datasets/b83hmrzth8/1	4
Tea Sickness Dataset [12]	https://data.mendeley.com/datasets/j32xdt2ff5/2	4
VegNet [30]	https://data.mendeley.com/datasets/t5sssfgn2v/3	4
Banana Leaf Disease Images [11]	https://data.mendeley.com/datasets/rjykr62kdh/1	4
BananaLSD Dataset [4]	https://data.mendeley.com/datasets/9tb7k297ff/1	4
Rice Leaf Disease Image Samples [32]	https://data.mendeley.com/datasets/fwcj7stb8r/1	4
Mango Pest Classification [13]	https://data.mendeley.com/datasets/94jf97jzc8/1	4
Rice Leaf Diseases [33]	https://archive.ics.uci.edu/dataset/486/rice+leaf+diseases	4
BDPapayaLeaf [22]	https://data.mendeley.com/datasets/p997fvf526/2	4
Blackgram PLant Leaf Disease Dataset [37]	https://data.mendeley.com/datasets/zfcv9fmrgv/3	4
DiaMOS [10]	https://zenodo.org/records/5557313	4
FieldPlant [21]	https://universe.roboflow.com/plant-disease-detection/fieldplant/dataset/11	4
Sugarcane Leaf Image Dataset [39]	https://data.mendeley.com/datasets/9twjtv92vk/1	4
PlantVillage [20]	https://github.com/spMohanty/PlantVillage-Dataset	5
Maize_TZ_Image_Dataset [17]	https://dataverse.harvard.edu/file.xhtml?fileId=6420463&version=6.0	6
Bean Leaf Dataset	https://www.kaggle.com/datasets/prakharrastogi534/bean-leaf-dataset	6
Cotton Plant Disease [8]	https://www.kaggle.com/datasets/dhamur/cotton-plant-disease?select=Cotton+leaves	7
Potato Disease Leaf Dataset (PLD) [28]	https://www.kaggle.com/datasets/rizwan123456789/potato-disease-leaf-datasetpld	7
Rice Diseases Image Dataset	https://www.kaggle.com/datasets/minhhuy2810/rice-diseases-image-dataset/data	8
Paddy Doctor [25]	https://ieee-dataport.org/documents/paddy-doctor-visual-image-dataset-automated-paddy-disease-classification-and-benchmarking	9
CNN_olive_dataset	https://github.com/sinanuguz/CNN_olive_dataset	4
Leaf Spot Attention Network [45]	https://github.com/cvmllab/Leaf-Spot-Attention-Network	4
Coffee Dataset [9]	https://github.com/esgario/lara2018	8
Plant Pathology 2020 - FGVC7 [38]	https://www.kaggle.com/competitions/plant-pathology-2020-fgvc7/data	9
Cassava Disease Classification [23]	https://www.kaggle.com/competitions/cassava-disease/data	10
PlantDiseaseNet [40]	https://github.com/mturkoglu23/PlantDiseaseNet	10
CDDM Dataset [44]	https://github.com/UnicomAI/UnicomBenchmark/tree/main/CDDMBench	10
OSF Dataset [43]	https://osf.io/p67rz/?view_only=	6
Date Palm Data	https://www.kaggle.com/datasets/hadjerhamaidi/date-palm-data	4
Plant Pathology Dataset [38]	https://www.kaggle.com/competitions/plant-pathology-2020-fgvc7/data	10
Coffee Plant Disease [24]	https://data.mendeley.com/datasets/c5yvn32dzg/2	4
CustomisedPD	https://drive.google.com/file/d/1HhtA939IwSjrN2XKRyeTgMQnTaY4zniA/view	4

License*
1: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International
2: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
3: Creative Commons Attribution-NonCommercial 3.0 Unported
4: Creative Commons Attribution 4.0 International
5: Creative Commons Attribution-ShareAlike 3.0 Unported
6: Creative Commons CC0 1.0 Universal Public Domain Dedication
7: Database: Open Database, Content: Database Content (ODbL v1.0)
8: MIT License
9: Apache License 2.0
10: Cite

Table 3: Crop descriptions

Crop	Description
apple	apple are Malus domestica, commonly known as apple trees
apricot	apricot are Prunus armeniaca, commonly known as apricot trees
basil	basil are Ocimum basilicum, commonly known as sweet basil
blueberry	blueberry are Vaccinium corymbosum, commonly known as blueberries
coriander	coriander are Coriandrum sativum, commonly known as cilantro or coriander
cotton	cotton are Gossypium hirsutum, commonly known as upland cotton plants
eggplant	eggplant are Solanum melongena, commonly known as eggplants
ginger	ginger are Zingiber officinale, commonly known as ginger
lettuce	lettuce are Lactuca sativa, commonly known as lettuce
olive	olive are Olea europaea, commonly known as olive trees
paddy	paddy are Oryza sativa, commonly known as rice plants
pear	pear are Pyrus communis, commonly known as pear trees
pumpkin	pumpkin are Cucurbita pepo, commonly known as pumpkins
raspberry	raspberry are Rubus idaeus, commonly known as raspberries
soybean	soybean are Glycine max, commonly known as soybeans
sugarcane	sugarcane are Saccharum officinarum, commonly known as sugarcane
sunflower	sunflower are Helianthus annuus, commonly known as sunflowers
tea	tea are Camellia sinensis, commonly known as tea plants
walnut	walnut are Juglans regia, commonly known as English walnut
zucchini	zucchini are Cucurbita pepo, commonly known as zucchini

Table 4: Disease descriptions

Disease	Description
algal_leaf	algal_leaf are Cephaleuros spp., known as algal leaf spot
alternaria_blotch	alternaria_blotch are Alternaria spp., commonly known as leaf blotch
black_rot	black_rot are Xanthomonas campestris, commonly known as black rot
black_stem_borer	black_stem_borer are Xylosandrus compactus, known as black stem borer
brown_leaf_spot	brown_leaf_spot are Phoma or Alternaria spp., known as brown leaf spot
cedar_apple_rust	cedar_apple_rust are Gymnosporangium spp., commonly known as cedar apple rust
crinckle	crinckle are Viral symptom, commonly known as leaf crinkle
dappula_tertia	dappula_tertia are Dappula tertia, known as oil palm leaf-eating caterpillar
eriosoma_lanigerum	eriosoma_lanigerum are Eriosoma lanigerum, commonly known as woolly aphid
fusarium_wilt	fusarium_wilt are Fusarium oxysporum, known as Fusarium wilt
frog_eye_leaf_spot	frog_eye_leaf_spot are Botryosphaeria obtusa, known as frogeye leaf spot
greening	greening are Candidatus Liberibacter spp., known as citrus greening
gummy_stem_blight	gummy_stem_blight are Didymella bryoniae, known as gummy stem blight
icerya_seychellarum	icerya_seychellarum are Icerya seychellarum, known as seychelles scale
leaf_blight	leaf_blight are Necrotic spread, known as leaf blight
mosaic	mosaic are Viral disease, known as mosaic
powdery_mildew	powdery_mildew are Erysiphales fungi, known as powdery mildew
purple_discoloration	purple_discoloration are Stress signs, known as purple discoloration
septoria_leaf_spot	septoria_leaf_spot are Septoria spp., known as leaf spot
yellow_mosaic_virus	yellow_mosaic_virus are Yellow mosaic virus, known as yellow mosaic

Table 5: The train and test splitting of datasets for different plant tasks

Task	Training images	Training crop	Training disease	Seen testing images	Seen crop	Seen disease	Unseen/Few shot testing images	Testing crop	Testing disease
Unseen	192943	55	175	48,419	55	175	1,448	9	10
Few shot	198711	55	175	49,325	51	134	542	27	61

Table 6: The performance of finetuning models pretrained with different datasets on various downstream tasks for plant disease identification.

Pretrained Dataset	Crop Disease Identification Top 1 Accuracy					
	PDD	IDADP	PD	PWv3	Herb	Avg
ImageNet-21k	88.12	99.40	59.60	76.44	24.19	69.55
ImageNet-1k	84.83	99.63	49.72	65.91	14.52	62.92
PlantNet300K	89.49	99.72	57.20	77.04	**38.71**	72.43
PWv3	90.28	99.45	58.90	78.43	30.65	71.54
DPD	**90.72**	**99.86**	**60.17**	**81.39**	35.48	**73.52**

Table 7: The performance of finetuning models pretrained with different datasets on various downstream tasks for crop identification task.

Pretrained Dataset	Crop Identification Top 1Accuracy					
	PDD	IDADP	PD	PWv3	Herb	Avg
ImageNet-21k	96.58	100.0	91.38	87.64	27.42	80.60
ImageNet-1k	93.44	100.0	82.63	79.27	16.13	74.29
PlantNet300K	97.10	100.0	89.41	88.24	43.55	83.66
PWv3	**97.40**	100.0	**92.80**	88.24	41.94	84.08
DPD	**97.20**	100.0	91.95	**90.63**	**45.16**	**84.99**

Table 8: The performance of finetuning models pretrained with different datasets on various downstream tasks for disease identification task.

Pretrained Dataset	Disease Identification Top 1 Accuracy					
	PDD	IDADP	PD	PWv3	Herb	Avg
ImageNet-21k	89.49	99.40	64.12	83.27	29.03	73.06
ImageNet-1k	87.11	99.63	59.46	74.52	17.74	67.69
PlantNet300K	90.82	99.72	63.14	83.79	**48.39**	77.17
PWv3	91.46	99.45	63.56	84.97	38.71	75.63
DPD	**91.95**	**99.86**	**65.25**	**87.02**	40.32	**76.88**

identification. These results highlight the effectiveness of the model pretrained with our DPD dataset in providing a more robust and transferable initialization for diverse downstream tasks.

The performance of the finetuning models of different pretraining datasets are also evaluated on individual crop and disease identification tasks as shown in Table 7 and Table 8 respectively. Experimental results indicate that models pretrained on the DPD dataset consistently outperformed those trained on other datasets, achieving average accuracies of 84.99% for crop identification and 76.88% for disease identification. Overall, disease identification proved to be more challenging than crop identification, as evidenced by consistently lower accuracy across all models. This performance gap can be attributed to the fact that disease-related features or symptoms often appear in variable locations and scales on leaf surfaces, making it more difficult for models to accurately learn and extract consistent disease features. We also present the confusion matrixes of our DPD dataset on crop and disease identitication tasks in Figure 9 and Figure 10.

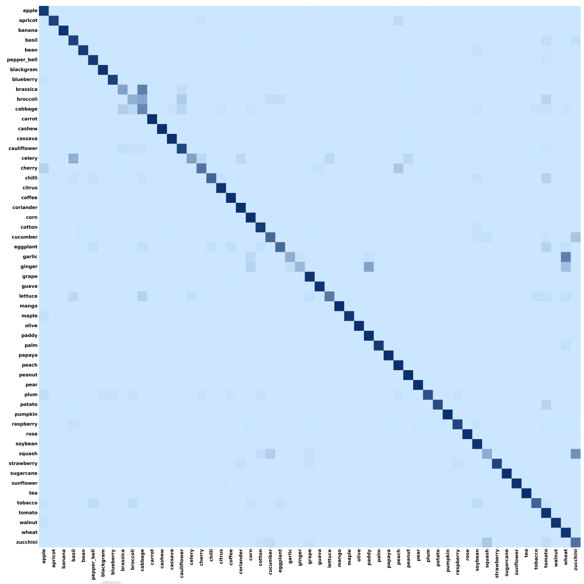

Figure 9: The confusion matrix for crop identification on DPD dataset.

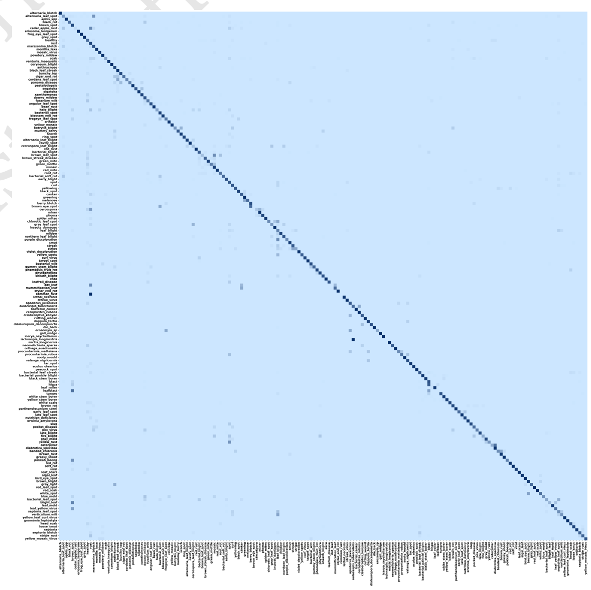

Figure 10: The confusion matrix for disease identification on DPD dataset.

1.5 Misclassification Analysis

To advance future research in multi-crop disease identification and to assess the generalization capacity of our pretrained model, we conducted a detailed misclassification analysis to uncover the underlying factors that limit model performance. Specifically, we examined instances where the model failed to produce correct predictions, aiming to understand the root causes of these errors. Figure 11 presents Grad-CAM visualizations of attention maps

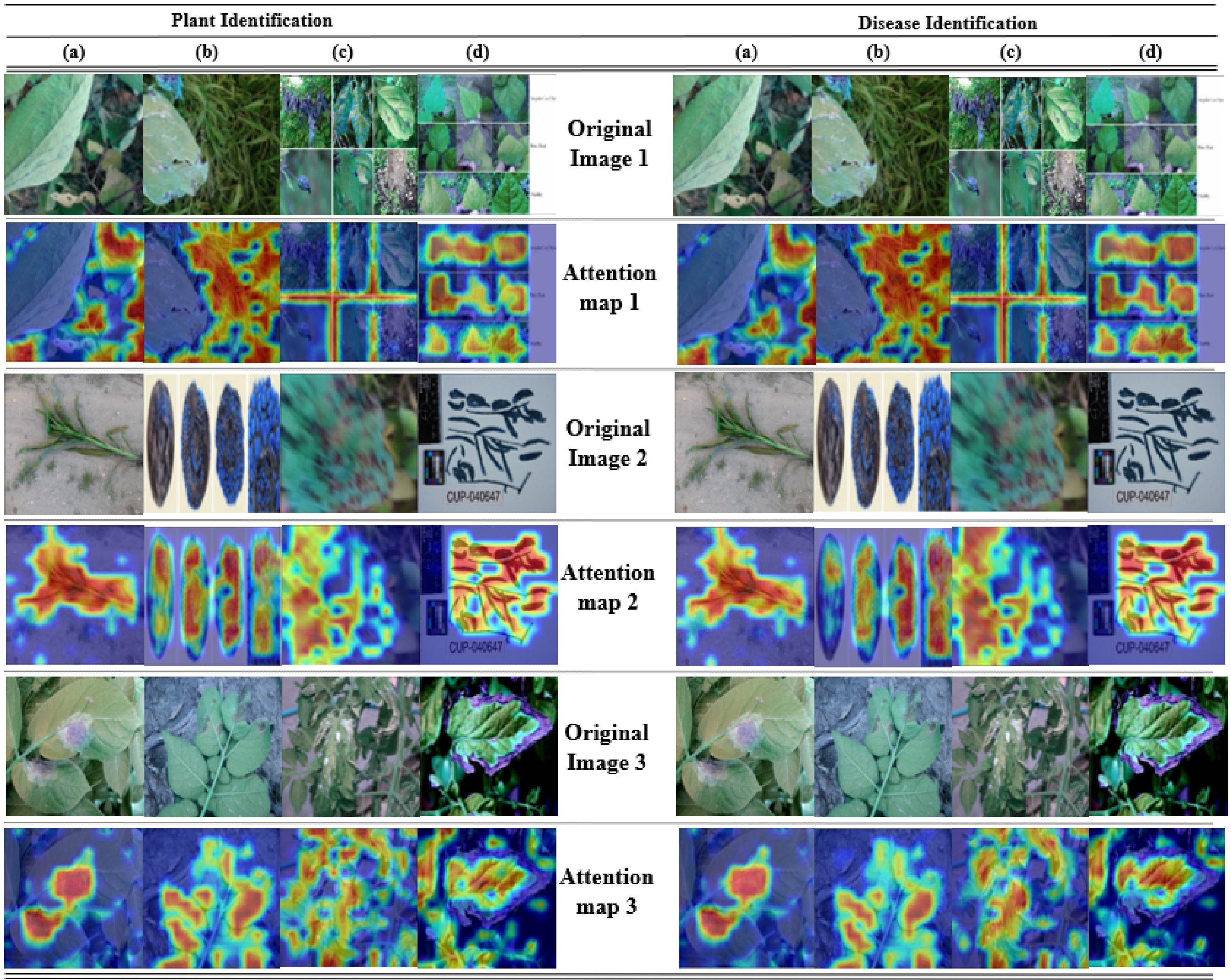

Figure 11: The misclassification analysis for model pretrained with our DPD dataset

produced by models pretrained on the DPD dataset, highlighting representative examples of misclassified samples. Based on this analysis, we identified three key factors that consistently contribute to misclassification.

First, the presence of multiple overlapping or background leaves within an image, illustrated in Original Image 1 and Attention Map 1, introduces significant noise. For instance, in Images 1(a) and 1(b), the model erroneously focuses on background leaves rather than the intended foreground leaf. Furthermore, as seen in Images 1(c) and 1(d), some samples retrieved from online repositories are collages composed of multiple leaves within a single image. This added visual complexity further impairs the model's ability to accurately isolate and classify the target leaf sample. Future work could explore multi-crop disease localization methods to better identify affected regions and improve classification performance.

Second, the leaf samples are either too distant or blurry (Original Image 2 and Attention Map 2). For example, in Images 2(a) and 2(d), the leaves appear too far from the camera, while in Images 2(b) and 2(c), the samples are noticeably blurry. These issues lead to a degradation in visual quality, thereby hindering the model's ability to accurately extract crop- or disease-related features. However, this decline in performance is understandable, as even human experts would face difficulties diagnosing diseases from such low-quality images.

Third, the presence of low inter-class variation and high intra-class variation presents a significant challenge in multi-disease classification. For instance, Images 3(a) and 3(b) correspond to samples of *Potato early blight* and *Potato late blight* respectively. Although they belong to different disease classes, their symptoms share highly similar visual characteristics, such as greyish discoloration in the affected regions. Conversely, Images 3(c) and 3(d) both represent

Tomato bacterial spot, yet they display noticeably different visual appearances, which in turn pose challenges for accurate identification. This inter-class similarity coupled with intra-class variability constitutes a major challenge for the model's ability to learn robust and consistent disease features. To address this, future work could explore incorporating additional data modalities, such as environmental or temporal information, or advanced feature learning techniques to enhance model discriminability.

References

[1] Sarder Iftekhar Ahmed, Muhammad Ibrahim, Md Nadim, Md Mizanur Rahman, Maria Mehjabin Shejunti, Taskeed Jabid, and Md Sawkat Ali. 2023. MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves. *Data in Brief* 47 (2023), 108941.

[2] MP Aishwarya and A Padmanabha Reddy. 2023. Dataset of groundnut plant leaf images for classification and detection. *Data in Brief* 48 (2023), 109185.

[3] M Alessandrini, R Calero Fuentes Rivera, L Falaschetti, D Pau, V Tomaselli, and C Turchetti. 2021. A grapevine leaves dataset for early detection and classification of esca disease in vineyards through machine learning. *Data in Brief* 35 (2021), 106809.

[4] Shifat E Arman, Md Abdullahil Baki Bhuiyan, Hasan Muhammad Abdullah, Shariful Islam, Tahsin Tanha Chowdhury, and Md Arban Hossain. 2023. BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning. *Data in Brief* 50 (2023), 109608.

[5] Alvaro Leandro Cavalcante Carneiro, Lucas de Brito Silva, and Marisa Silveira Almeida Renaud Faulin. 2021. Artificial intelligence for detection and quantification of rust and leaf miner in coffee crop. *arXiv preprint arXiv:2103.11241* (2021).

[6] 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.

[7] Swapnil Dadabhau Daphal and Sanjay M Koli. 2023. Enhancing sugarcane disease classification with ensemble deep learning: A comparative study with transfer learning techniques. *Heliyon* 9, 8 (2023).

[8] Dhamodharan. 2023. Cotton plant disease. doi:10.34740/KAGGLE/DSV/5127834

[9] José GM Esgario, Renato A Krohling, and José A Ventura. 2020. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture* 169 (2020), 105162.

[10] Gianni Fenu and Francesca Maridina Malloci. 2021. DiaMOS plant: A dataset for diagnosis and monitoring plant disease. *Agronomy* 11, 11 (2021), 2107.

[11] Y Hailu. [n. d.]. Banana leaf disease images. Mendeley Data 1 (2021).

[12] Gibson Kimutai and Anna Förster. 2022. Tea sickness dataset. *Mendeley Data* 2 (2022).

[13] K Kusrini, S Suputa, A Setyanto, IMA Agastya, H Priantoro, K Chandramouli, and E Izquierdo. 2020. Dataset for pest classification in Mango farms from Indonesia. *Mendeley Data* (2020).

[14] Sue Han Lee, Hervé Goëau, Pierre Bonnet, and Alexis Joly. 2020. New perspectives on plant disease characterization based on deep learning. *Computers and Electronics in Agriculture* 170 (2020), 105220.

[15] 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.

[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. doi:10.1109/TIP.2021.3049334

[17] Neema Mduma, Hudson Laizer, Loyani Loyani, Mbwana Macheli, Zablon Msengi, Alice Karama, Irine Msaki, Sophia Sanga, and Kennedy Jomanga. 2022. The Nelson Mandela African Institution of Science and Technology Maize dataset. doi:10.7910/DVN/GDON8Q

[18] Patrick Kwabena Mensah, Vivian Akoto-Adjepong, Kwabena Adu, Mighty Abra Ayidzoe, Elvis Asare Bediako, Owusu Nyarko-Boateng, Samuel Boateng, Esther Fobi Donkor, Faiza Umar Bawah, Nicodemus Songose Awarayi, et al. 2023. CCMT: Dataset for crop pest and disease detection. *Data in Brief* 49 (2023), 109306.

[19] 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.

[20] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. 2016. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science* 7 (Sep 2016). doi:10.3389/fpls.2016.01419

[21] Emmanuel Moupojou, Appolinaire Tagne, Florent Retraint, Anicet Tadonkemwa, Dongmo Wilfried, Hyppolite Tapamo, and Marcellin Nkenlifack. 2023. FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning. *IEEE Access* 11 (2023), 35398–35410.

[22] Sumaya Mustofa, Md Taimur Ahad, Yousuf Rayhan Emon, and Arpita Sarker. 2024. BDPapayaLeaf: A dataset of Papaya leaf for disease detection, classification, and analysis. *Data in Brief* 57 (2024), 110910.

[23] 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).

[24] Jorge Parraga-Alava, Kevin Cusme, Angélica Loor, and Esneider Santander. 2019. RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition. *Data in brief* 25 (2019), 104414.

[25] 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.

[26] Aditya Rajbongshi, Sadia Sazzad, Rashiduzzaman Shakil, Bonna Akter, and Umme Sara. 2022. A comprehensive guava leaves and fruits dataset for guava disease recognition. *Data in Brief* 42 (2022), 108174.

[27] Amanda Ramcharan, Kelsee Baranowski, Peter McCloskey, Babuali Ahmed, James Legg, and David P Hughes. 2017. Deep learning for image-based cassava disease detection. *Frontiers in plant science* 8 (2017), 1852.

[28] Javed Rashid, Imran Khan, Ghulam Ali, Sultan H Almotiri, Mohammed A Al-Ghamdi, and Khalid Masood. 2021. Multi-level deep learning model for potato leaf disease recognition. *Electronics* 10, 17 (2021), 2064.

[29] Umme Sara, Aditya Rajbongshi, Rashiduzzaman Shakil, Bonna Akter, Sadia Sazzad, and Mohammad Shorif Uddin. 2022. An extensive sunflower dataset representation for successful identification and classification of sunflower diseases. *Data in brief* 42 (2022), 108043.

[30] Umme Sara, Aditya Rajbongshi, Rashiduzzaman Shakil, Bonna Akter, and Mohammad Shorif Uddin. 2022. VegNet: An organized dataset of cauliflower disease for a sustainable agro-based automation system. *Data in Brief* 43 (2022), 108422.

[31] Sadia Sazzad, Aditya Rajbongshi, Rashiduzzaman Shakil, Bonna Akter, and M Shamim Kaiser. 2022. RoseNet: Rose leave dataset for the development of an automation system to recognize the diseases of rose. *Data in Brief* 44 (2022), 108497.

[32] Prabira Kumar Sethy. 2020. Rice leaf disease image samples. *Mendeley Data* 1 (2020), 2020.

[33] Harshadkumar; Shah, Jitesh; Prajapati and Vipul Dabhi. 2017. Rice Leaf Diseases. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5R013.

[34] 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* (Hyderabad, India) *(CoDS COMAD 2020)*. Association for Computing Machinery, New York, NY, USA, 249–253. doi:10.1145/3371158.3371196

[35] 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.

[36] Nusrat Sultana, Sumaita Binte Shorif, Morium Akter, and Mohammad Shorif Uddin. 2023. A dataset for successful recognition of cucumber diseases. *Data in Brief* 49 (2023), 109320.

[37] Srinivas Talasila, Kirti Rawal, Gaurav Sethi, Sanjay Mss, et al. 2022. Black gram Plant Leaf Disease (BPLD) dataset for recognition and classification of diseases using computer-vision algorithms. *Data in Brief* 45 (2022), 108725.

[38] Ranjita Thapa, Noah Snavely, Serge Belongie, and Awais Khan. 2020. The plant pathology 2020 challenge dataset to classify foliar disease of apples. *arXiv preprint arXiv:2004.11958* (2020).

[39] Sandip Thite, Yogesh Suryawanshi, Kailas Patil, and Prawit Chumchu. 2023. Sugarcane Leaf Image Dataset. *Mendeley Data* 1 (2023).

[40] Muammer Turkoglu, Berrin Yanikoğlu, and Davut Hanbay. 2022. PlantDiseaseNet: Convolutional neural network ensemble for plant disease and pest detection. *Signal, Image and Video Processing* 16, 2 (2022), 301–309.

[41] 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*. 1593–1601.

[42] Tianqi Wei, Zhi Chen, Zi Huang, and Xin Yu. 2024. Benchmarking In-the-Wild Multimodal Plant Disease Recognition and A Versatile Baseline. In *ACM International Conference of Multimedia*.

[43] Tyr Wiesner-Hanks, Ethan L Stewart, Nicholas Kaczmar, Chad DeChant, Harvey Wu, Rebecca J Nelson, Hod Lipson, and Michael A Gore. 2018. Image set for deep learning: field images of maize annotated with disease symptoms. *BMC research notes* 11, 1 (2018), 1–3.

[44] Liu Xiang, Liu Zhaoxiang, Hu Huan, Chen Zezhou, Wang Kohou, Wang Kai, and Lian Shiguo. 2025. A Multimodal Benchmark Dataset and Model for Crop Disease Diagnosis. In *Computer Vision – ECCV 2024*. Springer Nature Switzerland, Cham, 157–170.

[45] Hee-Jin Yu and Chang-Hwan Son. 2020. Leaf spot attention network for apple leaf disease identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 52–53.

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