MegaPlant

Preprint, compiled December 7, 2025

Chris Andrei Irag [1*], **Ashley Pramio** [1], **Monique Antoinette Mendoza** [1], **and Rod Vincent Dela Vega** [1]

[1]University of Science and Technology of Southern Philippines

Abstract

We introduce MegaPlant, a consolidated leaf-image dataset designed to support plant disease classification models that generalize across diverse environmental conditions, from controlled laboratory settings to highly variable in-field scenarios. MegaPlant integrates multiple publicly available datasets and standardizes them into a unified taxonomy of healthy and diseased leaf categories, enabling robust training across modalities. In addition, we propose a compartmentalized decision-making framework tailored for fully autonomous, in-field agents such as UAVs and mobile scouting robots. The framework separates disease detection from symptom identification, reducing single-point failure risks and improving reliability in real-world deployments. This modular structure also enhances interpretability, allowing practitioners to diagnose which stage of the pipeline misperformed when errors occur. Together, MegaPlant and our decision framework facilitate more dependable, scalable, and transparent plant disease surveillance systems suited for modern precision agriculture.

***K*eywords** Plant pathology, Image classification, Deep learning, Convolutional neural networks

1 Background

Plant diseases are abnormal changes in appearance and behaviour that progresses over time, unlike plant injury that occurs immediately (DeBusk, 2019).

These are caused by pathogens such as viruses, bacteria, fungus, oomycetes (fungus-like micro-organisms), parasitic nematodes (worm-like micro-organisms), and parasitic plants. Pathogens and pests (PPs) account for about 20% and at least 10% of harvest yield loss in major crops [1], [2].

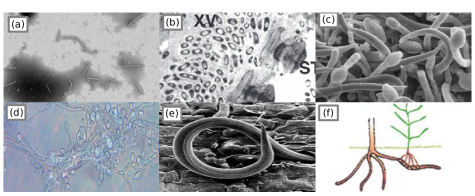

Figure 1: morphology of various pathogens. (a) Pepper mild mottle virus [3]; (b) Cabbage afflicted with black rot marked XV [4]; (c) fungus on an infected host flower [5]; (d) oomycete “potato blight” *P. infestans* [6]; (e) parasitic nematode [7]; (f) parasitic plant “witchweed” *Striga* [8].

Although most diseases are caused by pathogens or biotic factors, some are a result of direct injury or abiotic factors, also called environmental factors. These factors are drought, winter, disruptive human activities, etc. Diseases caused by abiotic factors are easier to diagnose but harder to control [8].

Considering that these pathogens are micro-organisms and invisible to the human eye, the method for identifying if a plant is unhealthy or infected is by identifying the symptoms and signs visually. However, identifying the exact disease-causing agent will require certain procedures often done by professional plant pathologists [2], UNH Extension, 2015.

1.1 Symptoms

To diagnose a plant immediately is by looking at the symptoms, these symptoms are reactions of the plant to the pathogen, not necessarily a sign of the particular pathogen itself. Signs of a plant disease are physical evidence of the causal agent or pathogen, signs are not symptoms (Penn State Extension, 2017).

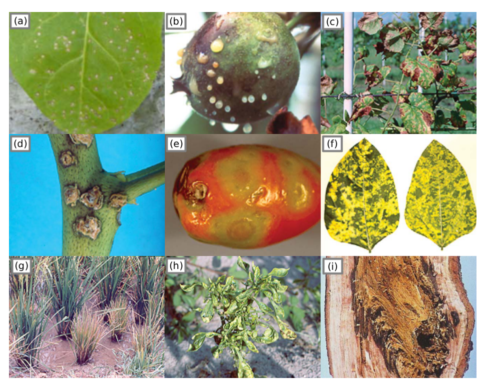

Figure 2: Plants showing 9 of many symptoms observable by the human eye. (a) localized lesions caused by virus [3], 4; (b) an apple exuding bacterial ooze, (c) scorched leaf symptoms caused by bacteria, (d) canker on a leaf stem, (e) tomato ring malformations, (f) mosaic pattern symptom (g) dwarfism/stunted growth and bronzing caused by virus, (h) malformation and mosaic pattern symptoms, (i) tree bark afflicted with rot [8].

Symptoms caused by abiotic factors are referred to as disorders. These symptoms are usually uniform, affecting large or evenly distributed areas of vegetation. In contrast, diseases caused by pathogens are often non-uniform and appear as scattered or irregular patches in the field (UNH Extension, 2015).

With deep learning models that focus on visual inspection for plant disease detection, it may be more practical to determine if a plant is unhealthy or otherwise, rather than detecting specific diseases or disorders. For example, a deep learning model may classify a plant as unhealthy due to bronzing during autumn, but may be harder to determine what the disease or disorder is, without additional temporal or environmental context.

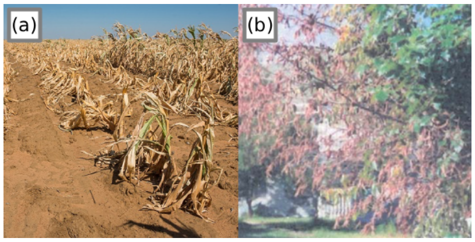

Figure 3: abiotic versus biotic induced symptoms. (a) Total loss of a corn field due to drought (The Independent, 2019); (b) Dehydrated tree due to fungi inhibiting water passage to the tree branches [8].

Figure 3 shows the same symptoms in two different plants but caused by different factors. This shows that environmental context will be needed when diagnosing plant diseases. As such, this case study will focus only on identifying whether a plant is healthy or unhealthy based on the symptoms observed by the human eye.

1.2 Signs

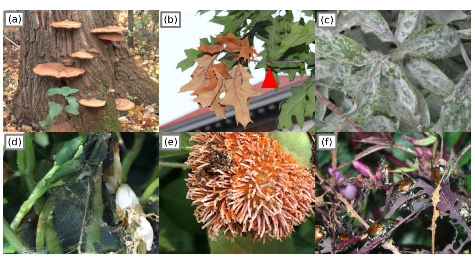

Figure 4: Signs of the PPs inflicting disease. (a) fungus growing as conks, (b) flagging, the insect feeding on the branch is highlighted, (c) powdery mildew appearance is composed of a fungus' hyphae, (d) web-like hyphae of a fungus, (e) rusting and spores of a fungus, (f) japanese beetles feeding Beckerman & Creswell, 2022.

Signs are physical evidence of the pathogen or pest, the real cause of the plant disease. Knowing the signs is key information to generating actions or solutions for PP management.

Observing the signs is a sure enough method to indicate that the plant might be unhealthy. However, that will take a more complex deep learning model, to consider another piece of information, for example, the bugs scattered on the branches, or fungus hyphae fully covering the subject leaf or plant. This study's scope will only incorporate symptoms observed on the leaves and some simpler signs, like fungus mildew.

1.3 Detection Methods

Plant disease detection can be performed using a range of approaches, from traditional manual observation to advanced computational techniques. Each method varies in accuracy, cost, scalability, and practicality depending on the use case.

1.3.1 Manual and Laboratory Methods

Visual inspection is the most common and oldest method, where farmers or plant pathologists examine the visible symptoms on leaves, stems, or fruits such as spots, blight, or discoloration. Although simple and fast, this method is subjective and heavily reliant on human expertise and environmental conditions (Penn State Extension, 2017).

In laboratory diagnostics, several scientific tests are employed to accurately identify pathogens:

- **Microscopy** – Used to observe fungal spores or bacterial colonies.
- **Culture tests** – Pathogens are isolated and grown in nutrient media for species identification.
- **Serological tests (e.g., ELISA)** – Use antibodies to detect specific proteins associated with viruses or bacteria.
- **Molecular techniques (PCR, qPCR, LAMP)** – Detect pathogen DNA or RNA, offering high sensitivity and specificity [9], [10].

While these approaches are precise, they require laboratory equipment, trained personnel, and are not suitable for large-scale or real-time monitoring.

1.3.2 Image Processing

Before the advent of deep learning, plant disease detection often relied on handcrafted features derived from image processing.

Key features such as color, texture, and shape were extracted using algorithms like:

- Gray-Level Co-occurrence Matrix (GLCM) for texture analysis
- Local Binary Patterns (LBP) for surface variation
- Color histograms in RGB or HSV space for spotting discoloration

These features were then classified using traditional machine learning algorithms such as:

- Support Vector Machines (SVMs)
- k-Nearest Neighbors (k-NN)
- Random Forests
- Naïve Bayes classifiers

For example, [11] demonstrated that SVM models using color and texture features achieved high accuracy in detecting citrus diseases. However, the performance of these systems is limited by the need for manual feature engineering, and they often fail to generalize well to diverse environmental conditions.

1.3.3 Spectral and Sensor-Based Method

More recently, spectral imaging technologies such as hyperspectral, multispectral, and thermal imaging have been applied for early plant disease detection. These methods capture light reflectance across multiple wavelengths, including the visible, near-infrared (NIR), and thermal infrared regions. Diseased plants exhibit distinct reflectance patterns, enabling early detection even before visible symptoms appear [12].

Such approaches are commonly integrated into precision agriculture systems, where drones or UAVs collect large-scale field data. Despite their promise, these systems are often expensive, complex to analyze, and require specialized sensors, limiting their accessibility to smallholder farmers.

2 Related Work

The related work and literature that will be discussed will be: (1) the recent advances and innovations in deep learning for plant leaf disease detection. (2) their downsides and limitations. (3) Existings datasets and their pros and cons.

2.1 Recent Advances

Recent developments in deep learning have significantly improved plant leaf disease detection. Traditional CNNs such as VGG16 and ResNet are still widely used, but newer approaches focus on improving model accuracy, generalization, and field performance. One major advancement is the integration of attention mechanisms and Transformer-based models, which provide stronger feature extraction and robustness in real agricultural environments [13]. These models help address challenges like varying lighting and complex backgrounds, which often reduce CNN performance in real-field conditions [14]. Another important trend is the development of lightweight and mobile-friendly architectures optimized for edge computing. Recent reviews highlight that compact CNN models can achieve high accuracy while being efficient enough for deployment on smartphones or IoT devices used in farms [15]. This allows farmers to detect diseases in real time without needing high-end hardware. Researchers have also focused on multi-crop and large-scale datasets, enabling models to recognize multiple plant diseases across different species rather than being limited to one crop at a time [16]. This improves the practicality of deep learning models for real agricultural use.

2.2 Current limitations

Although deep learning has significantly advanced plant disease detection, current innovations still face several unresolved limitations that are discussed in other research papers like [17]. Many studies rely heavily on controlled or laboratory style datasets with uniform backgrounds, making models difficult to generalize to real-world field conditions when lighting, background, or leaf appearance changes. Our approach depends heavily on visual symptoms captured from lab images, field photos and stock images, yet its performance varies across these sources due to domain shift.

Existing models from other papers also struggle with domain shift, meaning a model trained on one environment or imaging condition often performs poorly when tested in another, this reflects a common limitation noted in [18]. Our paper only focuses on identifying whether a plant is healthy or unhealthy based on the symptoms observed by the human eye, but we cannot reliably distinguish abiotic disorders from biotic diseases.

Studies like [17] and [18] show that accuracy can drop significantly when their models are evaluated outside their original training domain, indicating that many architectures are overfitted to specific datasets rather than truly learning robust, generalizable features. This becomes a serious problem for real-world deployment, because farmers and agricultural stakeholders often need a single system that can handle multiple crops, varying environments and different camera sources. Deploying smartphones or IoT devices used in farms [15] can actually help farmers to detect diseases in real time without needing high-end hardware.

Although our dataset includes multiple image sources, it remains limited in scale compared to the large, balanced datasets required for strong deep learning generalization. Many reviews, such as [15] similarly highlight that insufficient data diversity leads to reduced robustness in real-world agricultural settings.

In addition to data-related issues, computational demands also represent a major limitation. Many high-performing deep learning models require powerful GPUs, large memory capacity and stable internet connectivity, resources that are often unavailable to farmers, smallholder communities and agricultural workers in developing regions. [19] emphasize that without lightweight and hardware-efficient solutions, the gap between technological innovation and real-world adoption remains wide.

In conclusion, these limitations show that while our approach demonstrates the potential of deep learning in multi-condition environments, the system still requires more diverse data, improved robustness to domain variability and integration of richer visual information to achieve reliable, field-ready performance.

2.3 Known datasets

[20] proposed plant disease datasets have little diversity due to the laboratory conditions the image datasets were taken in, particularly datasets like the PlantVillage dataset by [21]. Their PlantDoc dataset improves upon this limitation but doesn't account for images with uniform background.

3 MegaPlant Dataset

MegaPlant integrates leaf-image subsets from PlantDoc, PlantVillage, and DiaMOS [22] to produce a model robust across varied imaging conditions. Only leaf images are included. PlantDoc and PlantVillage were obtained from their Kaggle derivatives due to issues accessing the original GitHub repositories.

- (a) PlantDoc: nirmalsankalana/plantdoc-dataset
- (b) PlantVillage: abdallahalidev/plantvillage-dataset
- (c) The DiaMOS dataset was retrieved directly from its official repository.

Table 1: Dataset Image counts

Dataset	Reported size	Retrieved images
DiaMOS	3,901	3,006
PlantVillage	54,305	54,306
PlantDoc	2,922	2,598

We obtained different image counts across repositories due to duplicates, corrupted files, dataset inconsistencies, and untracked modifications present in derivative versions of the datasets. To ensure consistency, we applied the constraint that only leaf images were included. Accordingly, from DiaMOS, we retrieved images only from the `leaves/` directory; from PlantDoc, only the `train/` and `test/` folders; and from PlantVillage, only the `color/` directory containing colored leaf images.

All images were consolidated and mapped into two primary classes: healthy (0) and unhealthy (1). The unhealthy category contains 12 symptom subclasses: blight, greening, malformation, powdery mildew, feeding, mold, mosaic, rot, rust, scab, scorch, and spot. We define the subclass criteria for dataset integration as follows:

1. If a folder is labeled with a subclass, all images within are assigned to that class.
2. Significant changes in leaf shape are classified as malformation.
3. Changes in leaf color (hue) are classified as greening.
4. Damage caused by insect feeding is classified as feeding.

An exception is Esca (Black Measles), which, despite its unique pathology, is labeled under the spot subclass. The final dataset was divided into train (70%), validation (20%), and test (10%) splits to enable reliable model evaluation and reduce bias.

4 MODELING

Before training any model, we resize all input images to 32×32 pixels. Standardizing the input size allows the network to accommodate images of varying original dimensions without requiring architectural changes. After resizing, each RGB image is converted into a tensor of shape 3×32×32, which serves as the model's input.

This simple model is designed to be used on edge devices that often lack the computational power that more complex models need.

Throughout our experiments, we use a simple convolutional neural network architecture. Figure 5 details the layer configuration and corresponding dimensions. The output layer is adapted to each task: for binary disease detection, the output dimension is 1; for multi-class symptom classification, it equals the number of symptom classes; and for the combined task, the model outputs the number of symptom classes plus one additional healthy class.

The network was trained using Stochastic Gradient Descent for the optimization algorithm with a learning rate of 0.01 with 50

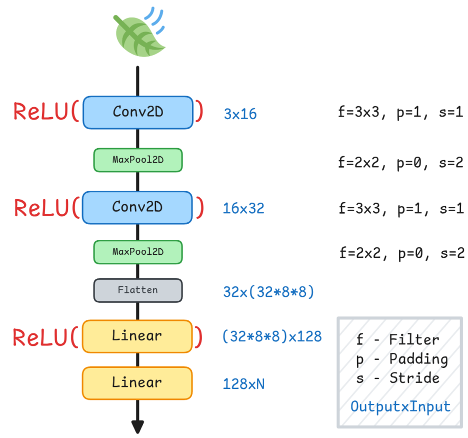

Figure 5: A simple Convolutional Neural Network architecture (SimpleCNN)

epochs. We use Mean Square Error and Cross Entropy as the loss function when doing binary classification or multi-class classification.

$$y = \begin{cases} 1, & \text{if z } \geq i \\ 0, & \text{else} \end{cases} \tag{1}$$

During inference, we classify the logits as either healthy or unhealthy using a 0.5 classification threshold.

We evaluate two approaches for disease detection and symptom identification:

1. Single-model approach: A multiclass classifier that predicts the healthy class and the 12 symptom subclasses.
2. Two-stage approach: A binary classifier first determines whether a leaf is diseased, followed by a multi-class classifier that identifies the specific symptom for diseased samples.

The single-model approach is simpler but suffers from a single point of failure. Any misclassification immediately affects all downstream predictions. In contrast, the two-stage approach compartmentalizes decision-making, allowing flexibility of complexity in downstream models or system architectures, and improving robustness in practical deployments such as UAV-based plant disease surveillance. In many real-world applications, reliably detecting whether a plant is diseased is more critical than precisely identifying the causal agent, which is often better handled by agronomists or plant pathologists.

This structure also supports additional downstream tasks, such as plant species identification or causal agent analysis, and enables clearer interpretability when the pipeline misperforms by isolating which stage produced the error. We additionally compare the proposed approaches against state-of-the-art object detection models to contextualize performance in broader plant disease detection pipelines.

Figure 6 describes a potential application of the modular framework where an unmanned aerial vehicle (UAV) is delegated the

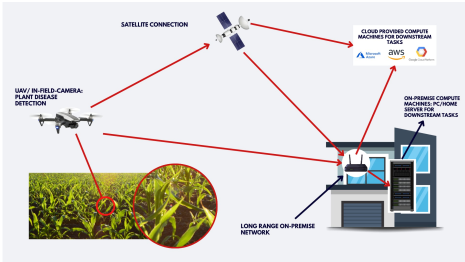

Figure 6: Imagined application of a modular decision framework in a farming business

Table 2: Perfomance of Disease Detection with SimpleCNN

Dataset	F1 Score	Accuracy
DiaMOS	0.9918	0.9837
PlantVillage	0.9901	0.9856
PlantDoc	0.8954	0.8392

task of detecting plant diseases. If it detects any leaves with a disease, it sends the picture to a computer machine where downstream tasks such as symptom or disease identification may be performed. This approach lessens the demand for computing power on edge devices, while also compartmentalizing the decision pipeline. Similar solutions like [23] only perform disease detection, without symptom identification.

4.1 *System Configuration*

All experiments were conducted using [24] on an NVIDIA 1050 Ti GPU (driver version 535.247.01) with CUDA 12.2. The machine that was used had 32 GB of RAM and an AMD Ryzen 5 5600 6-Core Processor.

5 Results and Discussion

We perform four main tasks and evaluate their performance on the MegaPlant dataset. Disease detction for binary classification, symptom identification for multi-class classification, a single model approach for multi-class classification of both symptoms and a healthy class, then a double model approach which separates model responsibility into two models for disease detection and symptom identification tasks.

5.1 *Disease Detection*

Our goal is to produce a model that serves as a foundation for downstream tasks, specifically for the two-stage approach we've discussed. We perform binary disease detection on the three datasets after training using the MegaPlant dataset.

SimpleCNN achieves an accuracy of 98.16% on the test set, an F1 score of 0.9876, with a parameter count of 267,489. Table 2 also shows the performance on each foundational dataset. The scores across seem balanced and indicates decent generalization of the model.

Table 3: Perfomance of Symptom Identification with SimpleCNN

Dataset	F1 Score	Accuracy
DiaMOS	0.8919	0.8853
PlantVillage	0.9302	0.9305
PlantDoc	0.7692	0.7569

Table 4: Perfomance of combined task classification with single model approach

Dataset	F1 Score	Accuracy	Binary F1 Score	Binary Accuracy
DiaMOS	0.8705	0.8666	0.9942	0.9886
PlantVillage	0.9475	0.9478	0.9934	0.9905
PlantDoc	0.7897	0.7754	0.9477	0.9242

5.2 *Symptom Identification*

Considering that we begin with a foundational model for disease detection, we then introduce a downstream model that performs symptom identification. In practice, once a leaf image is classified as unhealthy, it is passed to this secondary model to determine the specific symptoms present. This is in line with the two-stage approach.

SimpleCNN achieves an accuracy of 86.15% on the test set, an F1 score of 0.8614, with a parameter count of 268,908. Table 3 also shows the performance on each foundational dataset. The scores across seem to favor PlantVillage considering that it composes the biggest portion of the MegaPlant dataset.

The original authors of [20] used a variation of a ResNet model [25] to perform symptom identification with a performance 70.53% accuracy and 0.70 F1 score. Our model performs better with a superior 75.69% accuracy and 0.7692 F1 score. Our model generalized better when incorporating other datasets to train a model.

5.3 *Single model approach*

This single model approach combines both disease detection and symptom identification into one model.

SimpleCNN achieves an accuracy of 88.07% on the test set, an F1 score of 0.8808, with a parameter count of 269,037. Table 3 also shows the performance on each foundational dataset. The scores across seem to favor PlantVillage considering that it composes the biggest portion of the MegaPlant dataset.

Our model slightly performs better than the best model produced by the authors of DiaMOS [26] with an absolute difference of 0.61%. Indicating slight increase in model performance when trained on the MegaPlant dataset.

SimpleCNN performed worse but on par with pretrained models like GoogleNET that have near perfect performance of 99% accuracy on the PlantVillage dataset [27] [17].

5.4 *Double model approach*

This double model approach simulates a compartmentalized approach of inference, where disease detection and symptom

Table 5: Perfomance of combined task classification with double model approach

Dataset	F1 Score	Accuracy	Binary F1 Score	Binary Accuracy
DiaMOS	0.8726	0.8713	0.9918	0.9837
PlantVillage	0.9360	0.9360	0.9900	o.9856
PlantDoc	0.8726	0.8713	0.9918	0.9837

identification are done by two different models, but in this case, are trained on the same data.

5.5 *Limitations*

Considering that the framework has yet to be practically validated in real-world deployment scenarios, its performance under unconstrained environmental conditions remains uncertain. Factors such as varying illumination, background clutter, camera quality, and occlusions may significantly affect the reliability of the system when compared to controlled experimental settings. Furthermore, the symptom identification module may fail in cases where multiple disease symptoms are simultaneously present on a single leaf. Since overlapping visual features can confuse the classifier, this limitation could lead to incorrect or incomplete diagnoses, which is particularly critical for real-world agricultural applications where co-infections are common.

6 Conclusion

We introduced the basics of plant pathology to give an idea of where our case study might be situated in the research field. We discussed what symptoms are and how our case study approaches the problem of detecting plant diseases by detecting the symptoms visually. We identified traditional and alternative methods of plant disease detection. These methods often required feature engineering steps or expensive requirement

The weaknesses and strengths of innovations in plant disease detection using deep learning were discussed and informed us of how we might tackle the problem of detecting plant diseases using deep learning, particularly on datasets of leaf images with varying conditions such as laboratory, on field conditions, and stock images.

The MegaPlant dataset is hosted in a HuggingFace dataset repository can be retrieved from this link: https://huggingface.co/datasets/chrisandrei/MegaPlant. All relevant Python code, Jupyter notebooks, notes, and references can be found in this git repository hosted on GitHub: https://github.com/iragca/DS413-final-project.

Glossary

abiotic physical components of the environment that affect living organisms. - Wikipedia. 1, 2

biotic living components of the environment, living organisms. - Wikipedia. 1, 2

canker Sunken necrotic patch of bark. - Beckerman Creswell, n.d., 9. 1

morphology in biology, is the study of the size, shape, and structure of animals, plants, and microorganisms and of the relationships of their constituent parts. - Britannica. 1

Acronyms

PP Pathogens and pests. 2

PPs Pathogens and pests. 1, 2

Acknowledgements

This work benefited from the assistance of large language models for tasks such as literature discovery, writing support, and code autocompletion. All research ideas and methodological decisions were conceived and developed by the authors, with inspiration drawn from relevant prior work.

References

[1] Serge Savary, Laetitia Willocquet, Sarah Jane Pethybridge, Paul Esker, Neil McRoberts, and Andy Nelson. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3(3): 430–439, 2 2019. ISSN 2397-334X. doi: 10.1038/s41559-018-0793-y. URL http://dx.doi.org/10.1038/s41559-018-0793-y.

[2] Richard N. Strange and Peter R. Scott. Plant Disease: A Threat to Global Food Security. *Annual Review of Phytopathology*, 43(1):83–116, 9 2005. ISSN 1545-2107. doi: 10.1146/annurev.phyto.43.113004.133839. URL http://dx.doi.org/10.1146/annurev.phyto.43.113004.133839.

[3] Philippe Colson, Hervé Richet, Christelle Desnues, Fanny Balique, Valérie Moal, Jean-Jacques Grob, Philippe Berbis, Hervé Lecoq, Jean-Robert Harlé, Yvon Berland, and Didier Raoult. Pepper Mild Mottle Virus, a Plant Virus Associated with Specific Immune Responses, Fever, Abdominal Pains, and Pruritus in Humans. *PLoS ONE*, 5(4):e10041, 4 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0010041. URL http://dx.doi.org/10.1371/journal.pone.0010041.

[4] M. Dow, S.-Q. An, and A. O'Connell. *Bacterial Diseases*, pages 59–68. Elsevier, 2017. ISBN 9780123948083. doi: 10.1016/b978-0-12-394807-6.00051-4. URL http://dx.doi.org/10.1016/B978-0-12-394807-6.00051-4.

[5] Olienaide Ribeiro de Oliveira Pinto, Celli Rodrigues Muniz, José Emilson Cardoso, Francisca Samara Assunção de Oliveira, and Joilson Silva Lima. Morphological analyses of Pseudoidium anacardii infecting brazilian cashew plants. *Summa Phytopathologica*, 42(3): 257–260, 9 2016. ISSN 0100-5405. doi: 10.1590/0100-5405/2101. URL http://dx.doi.org/10.1590/0100-5405/2101.

[6] Waqas Raza, Muhammad Usman Ghazanfar, Muhammad Asif, Ikram-ul Haq, Muhammad Zakria, and Laith Khalil Tawfeeq Al-Ani. Morphological Characterization of Phytophthora infestans and its Growth on Different Growth Media. *Sarhad Journal of Agriculture*, 38(4), 2022. ISSN 2224-5383. doi: 10.17582/journal.sja/2022/38.4.1189.1202. URL http://dx.doi.org/10.17582/journal.sja/2022/38.4.1189.1202.

[7] Misgana Mitiku. Plant-Parasitic Nematodes and their Management: A Review. *Agricultural Research & Technology: Open Access Journal*, 16(2), 5 2018. ISSN 2471-6774. doi: 10.19080/artoaj.2018.16.55580. URL http://dx.doi.org/10.19080/ARTOAJ.2018.16.55580.

[8] G.N. Agrios. *Plant Pathogens and Disease: General Introduction*, pages 613–646. Elsevier, 2009. ISBN 9780123739445. doi: 10.1016/b978-012373944-5.00344-8. URL http://dx.doi.org/10.1016/B978-012373944-5.00344-8.

[9] ELAINE WARD, SIMON J FOSTER, BART A FRAAIJE, and H ALASTAIR MCCARTNEY. Plant pathogen diagnostics: immunological and nucleic acid-based approaches. *Annals of Applied Biology*, 145(1):1–16, 8 2004. ISSN 0003-4746. doi: 10.1111/j.1744-7348.2004.tb00354.x. URL http://dx.doi.org/10.1111/j.1744-7348.2004.tb00354.x.

[10] Norman W. Schaad, Reid. D. Frederick, Joe Shaw, William L. Schneider, Robert Hickson, Michael D. Petrillo, and Douglas G. Luster. A<scp>DVANCES IN</scp> M<scp>OLECULAR</scp>-B<scp>ASED</scp> D<scp>IAGNOSTICS IN</scp> M<scp>EETING</scp> C<scp>ROP</scp> B<scp>IOSECURITY AND</scp> P<scp>HYTOSANITARY</scp> I<scp>SSUES</scp>. *Annual Review of Phytopathology*, 41(1): 305–324, 9 2003. ISSN 1545-2107. doi: 10.1146/annurev.phyto.41.052002.095435. URL http://dx.doi.org/10.1146/annurev.phyto.41.052002.095435.

[11] R. Pydipati, T.F. Burks, and W.S. Lee. Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture*, 52(1–2):49–59, 6 2006. ISSN 0168-1699. doi: 10.1016/j.compag.2006.01.004. URL http://dx.doi.org/10.1016/j.compag.2006.01.004.

[12] Anne-Katrin Mahlein. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease*, 100(2):241–251, 2 2016. ISSN 1943-7692. doi: 10.1094/pdis-03-15-0340-fe. URL http://dx.doi.org/10.1094/PDIS-03-15-0340-FE.

[13] Asadulla Y. Ashurov, Mehdhar S. A. M. Al-Gaashani, Nagwan A. Samee, Reem Alkanhel, Ghada Atteia, Hanaa A. Abdallah, and Mohammed Saleh Ali Muthanna. Enhancing plant disease detection through deep learning: a Depthwise CNN with squeeze and excitation integration and residual skip connections. *Frontiers in Plant Science*, 15, 1 2025. ISSN 1664-462X. doi: 10.3389/fpls.2024.1505857. URL http://dx.doi.org/10.3389/fpls.2024.1505857.

[14] Thandiwe Nyawose, Rito Clifford Maswanganyi, and Philani Khumalo. A Review on the Detection of Plant Disease Using Machine Learning and Deep Learning Approaches. *Journal of Imaging*, 11(10):326, 9 2025. ISSN 2313-433X. doi: 10.3390/jimaging11100326. URL http://dx.doi.org/10.3390/jimaging11100326.

[15] Abhishek Upadhyay, Narendra Singh Chandel, Krishna Pratap Singh, Subir Kumar Chakraborty, Balaji M. Nandede, Mohit Kumar, A. Subeesh, Konga Upendar, Ali Salem, and Ahmed Elbeltagi. Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*, 58(3), 1 2025. ISSN 1573-7462. doi: 10.1007/s10462-024-11100-x. URL http://dx.doi.org/10.1007/s10462-024-11100-x.

[16] Mahmoud Khaled Elfouly, Amr M. AbdelAziz, Wael Hassan Gomaa, and Mohammed Abdalla. A deep learning-based framework for large-scale plant disease detection using big data analytics in precision agriculture. *Journal of Big Data*, 12(1), 8 2025. ISSN 2196-1115. doi: 10.1186/s40537-025-01265-9. URL http://dx.doi.org/10.1186/s40537-025-01265-9.

[17] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, 7, 9 2016. ISSN 1664-462X. doi: 10.3389/fpls.2016.01419. URL http://dx.doi.org/10.3389/fpls.2016.01419.

[18] Konstantinos P. Ferentinos. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145:311–318, 2 2018. ISSN 0168-1699. doi: 10.1016/j.compag.2018.01.009. URL http://dx.doi.org/10.1016/j.compag.2018.01.009.

[19] Sana Parez, Naqqash Dilshad, Turki M. Alanazi, and Jong Weon Lee. Towards Sustainable Agricultural Systems: A Lightweight Deep Learning Model for Plant Disease Detection. *Computer Systems Science and Engineering*, 47 (1):515–536, 2023. ISSN 0267-6192. doi: 10.32604/csse.2023.037992. URL http://dx.doi.org/10.32604/csse.2023.037992.

[20] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. Plantdoc: A Dataset for Visual Plant Disease Detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, pages 249–253. ACM, 1 2020. doi: 10.1145/3371158.3371196. URL http://dx.doi.org/10.1145/3371158.3371196.

[21] David. P. Hughes and Marcel Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics. 2015. doi: 10.48550/ARXIV.1511.08060. URL https://arxiv.org/abs/1511.08060.

[22] Gianni Fenu and Francesca Maridina Malloci. Diamos Plant: A Dataset for Diagnosis and Monitoring Plant Disease. *Agronomy*, 11(11):2107, oct 21 2021. ISSN 2073-4395. doi: 10.3390/agronomy11112107. URL http://dx.doi.org/10.3390/agronomy11112107.

[23] Reazul Hasan Prince, Abdul Al Mamun, Hasibul Islam Peyal, Shafiun Miraz, Md. Nahiduzzaman, Amith Khandakar, and Mohamed Arselene Ayari. Csxai: a lightweight 2d CNN-SVM model for detection and classification of various crop diseases with explainable AI visualization. *Frontiers in Plant Science*, 15, 7 2024. ISSN 1664-462X. doi: 10.3389/fpls.2024.1412988. URL http://dx.doi.org/10.3389/fpls.2024.1412988.

[24] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An Imperative Style, High-Performance Deep Learning Library, 2019. URL https://arxiv.org/abs/1912.01703.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, 2015. URL https://arxiv.org/abs/1512.03385.

[26] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. 2019. doi: 10.48550/ARXIV.1905.11946. URL https://arxiv.org/abs/1905.11946.

[27] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE, 6 2015. doi: 10.1109/cvpr.2015.7298594. URL http://dx.doi.org/10.1109/CVPR.2015.7298594.