Final Project for CIS*6050: Neural Networks
Project Title: Symptom Driven Plant Diseases Classification

Poojan Vadaliya
Mathematics and Statistics
University of Guelph
Guelph, Canada
pvadaliy@uoguelph.ca

Mohammad Sakibul Islam
Mathematics and Statistics
University of Guelph
Guelph, Canada
mislam22@uoguelph.ca

Gurleen Pelia
Mathematics and Statistics
University of Guelph
Guelph, Canada
gpelia@uoguelph.ca

Abstract

This project introduces a strategy for diagnosing plant diseases, which goes beyond traditional image-based classification methods by incorporating predictive symptom analysis. Unlike previous approaches that solely rely on images, our method combines image analysis with textual symptom descriptions generated by a language model, Gemini-vision-pro which provided a more comprehensive understanding of the disease. Through training on diverse datasets, a multi-modal approach was developed that is capable of classifying diseases, with the aim of enhancing interpretability and practical utility. Four Convolutional Neural Network models which includes VGG16, Densenet InceptionV3 and a bespoke CNN model are compared along with leveraging fusion techniques to integrate information of image and text features. Densenet seemed to be the most reliable model with a training accuracy of 95.5% and a validation accuracy of 85% among all. This demonstrated the ability to not only classify the disease but also comprehend the significance of the prediction made by the model to classify the disease. This approach not only enhances disease diagnosis but also holds immense potential for agriculture by empowering farmers with an effective tool for disease detection and sustainable crop management.

Introduction

i. Motivation

Agriculture is crucial for sustaining life, providing food security and supporting economic development. As plants are one major part of it, ‘Plant Health’ is essential for achieving optimal crop yields, maintaining ecosystem balance and promoting environmental sustainability. In order to effectively control diseases, accurate and timely identification of diseases is crucial. The motivation of our research is to contribute in advancing the field of automated plant disease diagnosis and inform the development of more efficient and accessible tools for plant health monitoring and management. Moreover, exploration of classification using a combined approach of image and text processing in the context of plant diseases would not only contribute to finding a way to fulfill the requirements but also employ to expand our knowledge to new horizons. Ultimately, the project is driven by the desire to cope with advanced technology, specifically Convolutional Neural Networks(CNN) and Large Language Models(LLM), and to address the key challenge in agriculture, which is accurate identification of plant diseases based on symptoms recognition which states how the disease has been classified.

ii. Significance of Project

Over the past years, the prevalence and severity of plant diseases have escalated due to factors such as changes in cultivation methods, and ineffective plant protection measures (Alsakar et al., 2023). As a consequence, there is a pressing need for automated techniques to identify diseases in a precise and efficient manner. These automated methods alleviate the need for manual inspection and monitoring, thereby reducing labor-intensive tasks.

iii. Potential Applications

The ability to diagnose symptoms impacting crop health during cultivation would be tremendously useful, allowing timely action to disrupt disease spread and effectively halt the progression of the disease. Our project offers potential applications in early disease detection, automated monitoring systems, and informed decision-making, assisting to effectively reduce economic losses and agricultural sustainability.

Problem Statement

The importance of addressing the problem of plant disease diagnosis cannot be overstated. Crop losses due to diseases can have a disastrous impact on economic stability, and food production, especially in areas where agriculture plays a major role. Crop disease diagnosis automation in the field of plant health is the need of the hour which leads us to work upon this project that aims to automate disease diagnosis by identifying symptoms leveraging visual and textual features. Traditional plant diseases identification methods predominantly rely on image-based classification, neglecting the predictive power of textual feature analysis.

Given a set of plant images, the objective is to develop an automated system for plant disease diagnosis. This system should leverage visual features extracted from images to generate image summary of plant diseases. The system must then classify diseas based on the combined analysis of visual and textual features.

The primary objective of our project is to enhance the reliability and robustness of plant disease by extending the work by considering textual description of the image along with image of the plant. Subsequently, we will compare the accuracy of our classification approach by applying multiple CNN models which were readily available along with a newly created custom CNN model to check the classification power of each model.

i. Challenges in Classification

Several challenges exist in plant disease identification research, including the scarcity and limited diversity of datasets, which hinder effective disease recognition due to the substantial image variety required for deep learning models (Barbedo, 2018).A wide range of symptoms, including as discoloration,wilting, lesions, and abnormalities, can be present in plant diseases. The appearance of these symptoms might vary which makes it difficult to diagnose illnesses solely based on visual characteristics. Moreover, plant disease symptoms exhibits variability due to weather condition, geographic location, and crop development stages (Li et al., 2021). Additionally, multiple diseases may coexist on the same plant (Bhagat and Kumar, 2022), leading to difficulties in distinguishing between different diseases. Furthermore, similar symptoms across various diseases make the classification more complicated. Reliable illness diagnosis depends on the system's ability to withstand changes in image quality, and disease severity. The reliance on image-based data introduces complexities such as variations in lighting, image quality, and background clutter (Barbedo, 2016), which can significantly impact the accuracy of symptom extraction and classification. Addressing these challenges necessitates the development of different approach to effectively handle these factors for plant disease classification.

Related Work

Plant diseases pose a significant danger to global agricultural production and food security. Traditionally, this required manual work of experts by visual inspection which was time consuming, subjective and may require specialized knowledge. Previous research relied on automating plant disease diagnosis through machine learning methods, specifically image classification algorithms.

The study titled "Research on plant disease identification based on CNN, Sun et al. (2022) utilize Convolutional Neural Networks (CNNs), including EfficientNet, DenseNet and

ResNet models, to create an automated system for precise identification of plant diseases from images. Another research study Hussain et al. (2022) focused on cucumber leaf disease recognition, employing a deep learning framework featuring VGG and Inception V3 models, trained on a privately collected dataset consisting of a sample of cucumber plants, yielding an accuracy rate of 96.5%. The study Atila et al. (2021) proposed EfficientNet for plant leaf disease classification, employing transfer learning on the PlantVillage dataset to compare its performance with other deep learning models.

Alsakar et al. (2023) study provides insights into the current trends and challenges in utilizing machine learning and deep learning methods for plant disease detection and classification. The study Islam et al. (2022) showed a approach leveraging parallel convolutional neural network with various activation layers and normalization techniques in context of tomato leaf diseases classification in order to achieve high accuracy.

Moreover, the study by Wan and Shao (2023) proposes a disease classification model leveraging both images and textual descriptions through an adaptive multi-modal attention mechanism and demonstrates significant improvements in classification accuracy over single-modal approaches. another study Hasan et al. (2024) introduces a multimodal framework leveraging ResNet50, Vision Transformers, and Generative Pre-trained Distilled-GPT2 models to improve classification accuracy of biomedical images with multiple labels. Kalakuntla et al. (2023) research investigates multimodal language models, integrating text and images through fusion techniques such as early and late fusion, along with attention mechanisms, to enhance comprehension and exhibit superior performance.

Our project's topic is positioned within the framework of earlier studies on machine learning—specifically, image classification algorithms—as a means of automatically diagnosing plant diseases. Our initiative provides a supplementary viewpoint to current research

Class	Symptoms
Leaf blight	Dark green to black lesions on leaves
Leaf curl	Leaf curling, puckering, with pale green or yellow leaves.
Septoria leaf spot	small, circular, dark brown spots with gray centers on lower leaves.
Verticulium wilt	Yellowing, wilting, stunting, brown vascular discoloration.
Healthy	vibrant green and free from discoloration, spots, or curling.

Table 1: Symptoms of different classes in Tomato Plants

by putting forth a fusion method of combining visual and textual features for plant disease classification, which may improve the reliability and applicability of disease diagnosis systems. As in the past there were work done on the classification part in the field of plant disease identification but here extending to it we have added the possibility to also populate the results with image summary which gives the reasoning of identification of the plant disease apart from classifying the plant disease by applying multi-modal fusion.

Methodology

i. Dataset Information

We tried to accommodate the dataset requirement by acquiring data from multiple sources as per the availability and the need for the project. The data was combined from "Plant Village"(Hughes, David and Salathé, Marcel and others, 2015) from "Kaggle" and "Dataset for Crop Pest and Disease Detection" of "Mendeley Data"(Mensah, Kwabena Patrick and others, 2023), containing images depicting crop parts with both healthy and infected areas.

We are investigating tomato plant diseases classification, encompassing five distinct classes: healthy, Leaf blight, Leaf curl, Septoria leaf spot, and Verticillium wilt.

ii. Data Pre-Processing

Since the dataset mainly comprises images of plant leaves, augmenting these images will be crucial. We've implemented augmentation methods such as image scaling, rotation, and flipping to enhance the diversity of training examples and ensured that the model is well-prepared to handle a broad spectrum of variations effectively just for the convolutional neural network to capture more variety of features. We have also ensured that different disease classes have a balanced distribution of samples, which improves the model's ability to generalize to all classes equally.

For comparing the image based training with the proposed approach , image data of around 5000 images were used to train the custom neural network model whereas for the multi-modal fusion only 500 images were utilised to train the model due to constraints raised during symptom generation where 100 images were considered for each class.

iii. Model Architecture

Here we have tried to implement the process of distillation (Sun et al., 2021) where we use the concept of teacher student model in which Gemini-vision-pro is considered as teacher model that prepares the data or rationale database for the student model which in our case is the custom multi-fusion model that would be trained and fine-tuned as per the task specific requirements which would be plant disease detection.

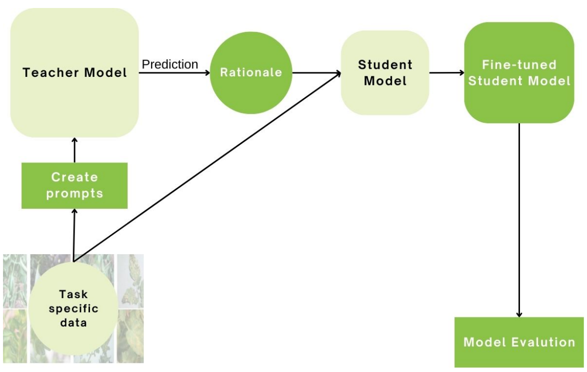

Figure 1: Abstract View of project architecture

Gemini-vision-pro has been utilised to extract image summary which contains the symptom description of the disease in the image of plant leaf. These symptoms serve as a textual summary of plant leaf images. Once we obtain the summary for each image, we'll establish a rationale database containing the image, symptoms, and the label. This database will then be fed into a secondary model, built using Multi-modal fusion technique.

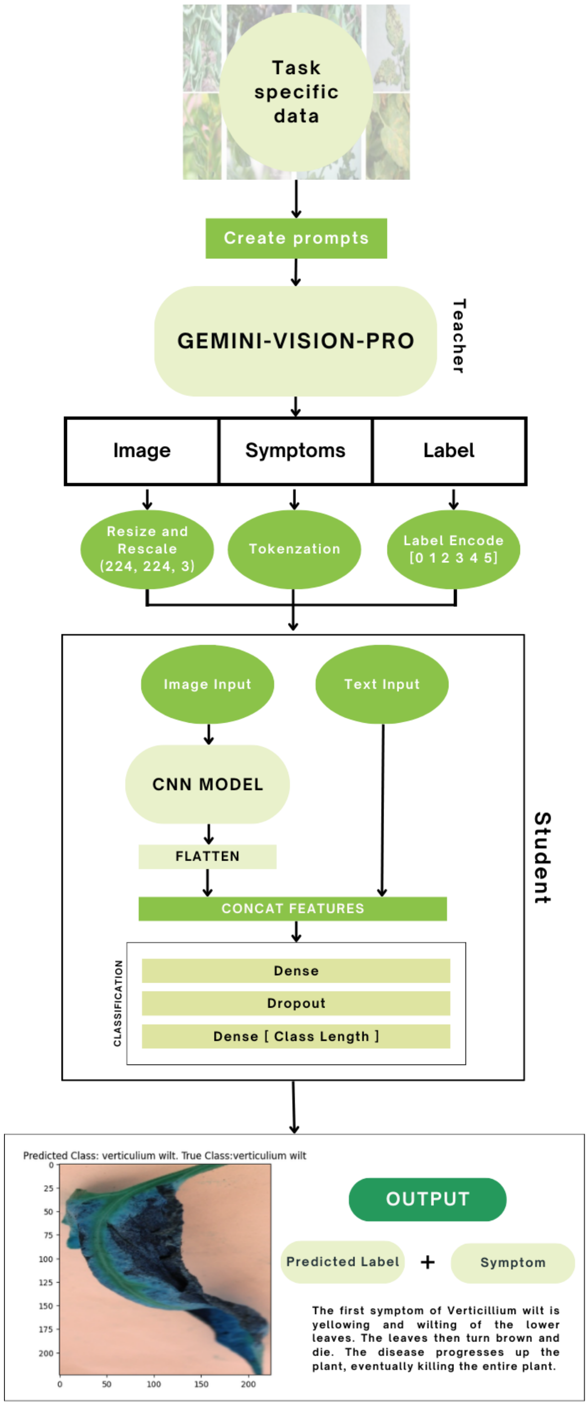

Figure 2: Flow of our project architecture

Within the Multi-modal fusion framework, the initial step involves standardizing the input images through a pre-processing pipeline. This involves resizing and re-scaling each image to a fixed size of 224x224 pixels and ensuring pixel values fall within the range of 0 to 1, facilitating uniformity and efficient processing.

For image classification part CNN was utilised over vision large language model as LLM seemed to be constrained with the hallucination problem which lowered the accuracy of the disease classification and also recognising the past research work done on CNN along with its application CNN seemed to be the best option for image feature analysis. Four models have been utilised for comparison where pretrained neural network models and custom made CNN was used for image feature extraction where for pretrained networks VGG16, DenseNet and InceptionV3 have been used and also building a custom CNN which comprised of six 2D-convolutional layers paired with max-pooling layers, contributing as training parameters, followed by a flattening layer followed by two dense layers are applied. These layers collectively capture intricate spatial patterns and reduce the spatial dimensions of feature maps, enabling effective feature extraction.

Integrating the CNN models into the multimodal architecture involves passing the image input through the CNN to extract image features, while simultaneously concatenating textual symptom descriptions with the flattened CNN output. This fusion of image and text representations enabled the model to leverage both modalities for disease classification, enhancing its predictive capabilities.

Further processing occurs through fully connected layers, transforming the concatenated features into a higher-dimensional space. A dropout layer is introduced to mitigate overfitting by randomly deactivating neurons during training. Finally, a soft-max activation function at the output layer predicts the probability distribution across disease classes, facilitating multi-class classification.

During model compilation, the Adam optimizer is utilized alongside sparse categorical cross-entropy loss, suited for integer-encoded labels in multi-class classification tasks. Accuracy is chosen as the evaluation metric to monitor the model's performance during training, providing insights into its effectiveness in disease diagnosis.

To evaluate the effectiveness of the approach multiple CNN models were compared to check the classification accuracy along with the summary output of how accurately the summary were produced aligning to the test image using cosine similarity.

Evaluation

A. Outcome of Project

i. Classification of Disease:

We have classified the plant diseases portrayed in images based on the symptoms and then compared this approach with the image-based plant diseases classification.

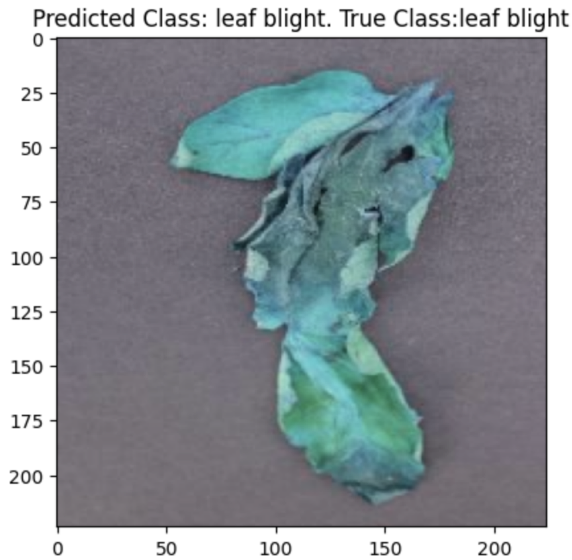

Figure 3: Classification Prediction

ii. Symptom prediction:

The symptom description of the disease depicted in the plant leaf image has been generated using Gemini-Vision-Pro, a sophisticated language model introduced by Google. This model possesses the capability to recognize both image and text input, allowing it to generate accurate and detailed descriptions of the disease symptoms based on the visual input of the plant leaf.

Generated Symptoms: The tomato plant leaf image shows symptoms of leaf blight, a fungal disease caused by Phytophthora infestans. The symptoms of leaf blight include brown spots on the leaves, which can eventually lead to the leaves wilting and dying. The disease can also cause the fruit to rot. Leaf blight is a common problem in tomato plants, and it can be difficult to control. However, there are a number of things that can be done to reduce the risk of infection, such as using resistant tomato varieties, avoiding overhead watering, and keeping the plants well-spaced.

Figure 4: Generated summary output for the predicted class.

B. Results

For comparing the generated symptoms for various plant diseases with original reference symptom texts, a pre-trained sentence embedding model was employed. After aggregating symptoms by disease label, cosine similarity scores are calculated between generated symptoms and reference texts for each disease class. Cosine similarity is chosen because it measures the cosine of the angle between the vectors of generated symptoms and reference texts.So cosine similarity is a useful metric for assessing the degree and direction of similarity because this enables us to assess how well the generated symptoms align with the actual reference symptoms. After calculating the average cosine similarity for each of the class, it is revealed that higher similarity scores are observed for the generated symptoms of leaf blight, septoria, and healthy plants, while lower scores are observed for verticulium wilt and leaf curl.

Class	Average Cosine Similarity
leaf blight	0.868
verticulium wilt	0.416
septoria	0.864
healthy	0.876
leaf curl	0.692

Figure 5: Average Cosine Similarity for each class

Considering the data distribution among the training and the validation set 80% of total image data was considered for training set and 20% for validation in all cases.

Four different models have been explored, which includes VGG16, DenseNet, InceptionV3 and custom CNN model in the proposed architecture. After analyzing the accuracy curves for training and validation dataset for all models, it is evident that DenseNet model provides better performance in classification among all the considered models with a training accuracy of 95.5% and validation accuracy of 85% just after training the model for 10 epochs.

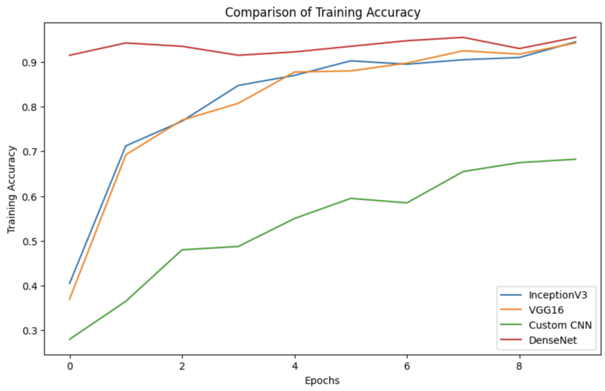

Figure 6: Comparison of Training Accuracy

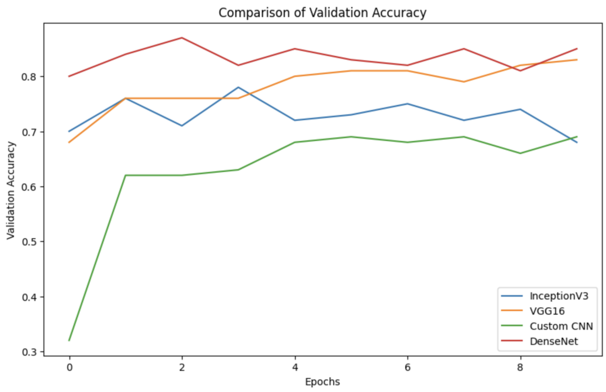

Figure 7: Comparison of Validation Accuracy

To evaluate the classification result of the model, different evaluation metrics including precision, recall and f1-score are utilized. Precision evaluates the accuracy of the model's positive predictions for each diseases class by calculating the ratio of correctly predicted positive instances to all instances predicted as positive within that class. Among of the four investigated models, classification results and ROC curve of the best-performing DenseNet model in the proposed multimodal architecture are presented in the report. The model achieves a precision of 0.96 for leaf blight, which means that it correctly identifies the presence of leaf blight in plants 96% of the time.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall evaluates the model's ability to accurately identify actual positive occurrences for each illness class by measuring the proportion of correctly predicted positive cases to all actual positive cases in that class. The recall for verticullium wilt is 0.86 which indicates the model accurately predicts 86% of actual cases of verticullium wilt.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

The proposed model shows strong ability to accurately identify healthy, leaf blight, and septoria with F1-scores of 0.98, 0.93, and 0.84 respectively by maintaining balance between precision and recall.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

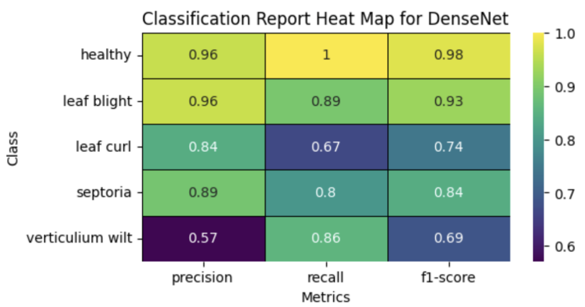

Figure 8: Classification Report for best performing model

Examining the ROC-AUC curve score results, we can see that all class AUC scores are above 90%. This indicates that the model effectively learns image features, enabling it to distinguish between classes and achieve accurate disease classification rather than random assignment of the classified results.

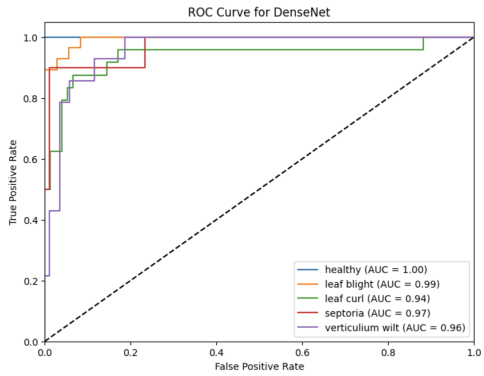

Figure 9: ROC Curve for for best performing model

C. Comparison with Image Based Classification

In contrast to the image-based classification approach, it was discovered that after 50 epochs, the custom neural network achieved a training accuracy of 92% and a validation accuracy of 91%. However, by utilizing only one-tenth of the total data employed for image classification, the multimodal approach surpassed the predominant technique. The top-performing CNN model in the proposed approach where Densenet was used for image feature extraction attained a training accuracy of 95.5% and a validation accuracy of 85% after being trained for just 10 epochs.

Figure 10: Training and Validation Accuracy and Loss for Image based Classification

Conclusions

Incorporating both image data and symptom descriptions generated by Gemini, a multimodal system was developed, with CNN chosen over LLM for classification to address the issue of hallucination and ensure accurate disease retrieval from the image. This system takes both text and image inputs to extract features and assist in classifying disease types and provides description of the diagnosed disease along with the classification. Even with only 500 images the top performing model, Densenet in our proposed multimodal architecture was capable to show training accuracy of 95.5% and validation accuracy of 85% which showed the effectiveness of the ability to classify the disease using both image and text data compared to just the image data in the traditional approach. This might also be enhanced more by adding greater number of images or exploring the utilization of attention mechanism or various fine tuning methods to make it even more robust.

In future research, it can be crucial to explore the influence of external factors like weather conditions, soil properties, and geographical location on plant disease occurrence. Expanding the scope of disease classification models to encompass a broader range of crop types such as wheat, rice, and maize will provide a more comprehensive understanding of disease dynamics in agricultural settings.

References

Yasmin M. Alsakar, Nehal A. Sakr, and Mohammed Elmogy. 2023. Plant disease detection and classification using machine learning and deep learning techniques: Current trends and challenges. In *World Conference on Internet of Things: Applications & Future*.

Umit Atila, Murat Ucar, Kemal Akyol, and Emine Uçar. 2021. Plant leaf disease classification using efficientnet deep learning model. *Ecological Informatics*, 61:101182. Published in March.

Jayme Garcia Arnal Barbedo. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, 144:52–60.

Jayme Garcia Arnal Barbedo. 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture*, 153:46–53.

M. Bhagat and D. Kumar. 2022. A comprehensive survey on leaf disease identification and classification. *Multimedia Tools and Applications*, pages 1–29.

Md. Rakibul Hasan, Md Rafsan Jani, and Md Mahmudur Rahman. 2024. Image and text feature based multimodal learning for multi-label classification of radiology images in biomedical literature. *ACL Anthology*. Computer Science Department, Morgan State University, Baltimore, Maryland, U.S.A.

Hughes, David and Salathé, Marcel and others. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv*.

N. Hussain, M.A. Khan, U. Tariq, S. Kadry, M.A.E. Yar, A.M. Mostafa, A.A. Alnuaim, and S. Ahmad. 2022. Multiclass cucumber leaf diseases recognition using best feature selection. *Comput Mater Continua*, 70:3281–3294.

M.P. Islam, K. Hatou, T. Aihara, S. Seno, S. Kirino, and S. Okamoto. 2022. Performance prediction of tomato leaf disease by a series of parallel convolutional neural networks. *Automation in Agriculture*, 4:100054. Author links open overlay panel.

Sai Teja Kalakuntla, Lankapothu Sai Teja Reddy, Lankapothu Pavan Kumar Reddy, and Navya Manjari Uppaluri. 2023. Multimodal llms: Combining text and images for enhanced understanding. *Journal of Multimodal Learning*, 1(1):1–15.

L. Li, S. Zhang, and B. Wang. 2021. Plant disease detection and classification by deep learning—a review. *IEEE Access*, 9:56683–56698.

Mensah, Kwabena Patrick and others. 2023. Dataset for crop pest and disease detection. Mendeley Data. V1, doi: 10.17632/bwh3zbpkpv.1.

Siqi Sun, Samuel Humeau, Prafulla Dhariwal, Trang Pham, and Quoc V Le. 2021. Distilling step-by-step: Outperforming larger language models with less training data and smaller model sizes. *Google Research Blog*.

Xuewei Sun, Guohou Li, Peixin Qu, Xiwang Xie, Xipeng Pan, and Weidong Zhang. 2022. Research on plant disease identification based on cnn. *Cognitive Robotics*, 2:155–163.

Zhengyu Wan and Xinhui Shao. 2023. Disease classification model based on multi-modal feature fusion. *Journal of Medical Imaging and Health Informatics*, XX(XX):XX–XX.