CSCI 491/591 - Deep Learning - Final Report

Classification of Plant Diseases

**Team Member**:

Ethan Rocheleau

**GitHub:**  
https://github.com/ejroch/Plant-Diseases/tree/master

Abstract

This project started off as a curiosity of my own given a home garden that developed sick plants over the course of the semester. Given that, I aimed to develop an accurate and efficient deep learning model for automatic detection and classification of plant diseases using images of plant leaves. When researching this topic, I found that in addressing this challenge it can be crucial for ensuring timely intervention, improving crop yield, and promoting sustainable agricultural practices. I utilized the comprehensive PlantVillage dataset containing over 50,000 images of healthy and diseased leaves across 38 classes, covering 14 crop species. By employing state-of-the-art convolutional neural networks (CNNs) and leveraging transfer learning, I fine-tuned pre-trained models, such as ResNet, VGG, and Inception/GoogLeNet, for the task of plant disease classification. Furthermore, I implemented data augmentation techniques to enhance the model's generalization capabilities. My experimental results demonstrated that the proposed deep learning model achieved meaningful results in terms of accuracy, precision, recall, and F1-score when creating certain models. This work contributes to the development of a valuable tool for assisting professionals in the agricultural domain to address plant diseases promptly and effectively.

Introduction

Through research, I found my small home garden problem can be applied on a larger, global scale. Plant diseases pose a significant threat to global food security and agricultural sustainability, with the potential to cause severe crop losses and economic damage. Early detection and accurate classification of plant diseases are critical for timely intervention and effective management, leading to improved crop yield and reduced economic losses. However, traditional methods of plant disease diagnosis are often time-consuming, labor-intensive, and reliant on expert knowledge. With the rapid advancements in artificial intelligence and deep learning, there is an opportunity to develop automated solutions that can efficiently and accurately identify plant diseases using image data.

In this study, I address the challenge of automatic detection and classification of plant diseases by leveraging deep learning techniques and images of plant leaves. By developing a reliable and efficient deep learning model, I aim to provide a valuable tool for farmers, agronomists, and other professionals in the agricultural domain to identify and address plant diseases promptly and effectively. This work not only contributes to the improvement of crop yield and food security, but also promotes sustainable agricultural practices by reducing the need for excessive pesticide use and enabling more targeted interventions. By focusing on this critical real-world problem, I hope to demonstrate the power and potential of deep learning in transforming agriculture and securing our food supply for future generations.

Background

The application of machine learning and deep learning techniques for plant disease detection has gained significant attention in recent years, driven by the need for accurate and timely diagnosis to ensure optimal crop yield and food security. In this section, I highlight notable works in this domain and their key takeaways, which have informed and inspired my own project.

Mohanty et al. (2016) was among the first to demonstrate the potential of deep learning, specifically convolutional neural networks (CNNs), for large-scale plant disease detection. Using the comprehensive PlantVillage dataset, they experimented with various CNN (convolutional neural networks) architectures, including AlexNet, GoogleNet, and VGG-16, and achieved an impressive accuracy of 99.35%. This pioneering work laid the groundwork for future research in this area by establishing the viability of deep learning for plant disease detection.

Building on Mohanty et al.'s work, Ferentinos (2018) addressed the issue of data imbalance in plant disease classification by proposing a methodology that combined data augmentation, transfer learning, and ensemble learning. By experimenting with various data augmentation techniques, the author generated a more balanced dataset and demonstrated that addressing data imbalance and using ensemble learning can significantly improve classification performance.

These previous works have shown that deep learning, particularly CNNs, can be highly effective in plant disease detection and classification. By leveraging state-of-the-art deep learning models and methodologies, including transfer learning and data augmentation, my project aims to build upon these foundations to further improve the accuracy and reliability of plant disease classification from leaf images.

Data

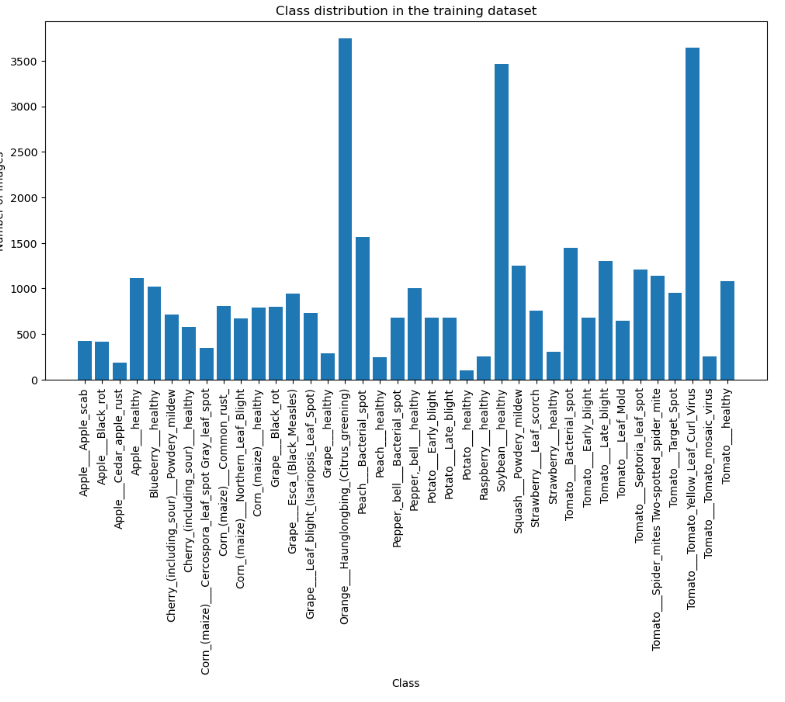
For this project, I utilized the comprehensive PlantVillage dataset, which contains more than 50,000 images of healthy and diseased plant leaves across 38 different classes. The dataset covers 14 crop species, including apples, grapes, corn, potatoes, and tomatoes, making it highly relevant to global agriculture and food security. Each image is labeled as healthy or as having one of 37 types of diseases, providing a diverse and representative sample of plant diseases for training a robust deep learning model.

*Figures 1 and 2. These figures represent a healthy grape leaf on the left and a grape with Esca or Black Measles disease on the right.*



To better understand the data and identify any pre-processing requirements, I analyzed the distribution of images across classes. This allowed me to identify potential class imbalance and address it during the model training process. I visualized the distribution of images in a bar chart, revealing that some classes had significantly more images than others, indicating the need for data augmentation and re-sampling techniques to balance the dataset.

*Figure 3. Bar graph representation of the data distribution in the PlantVillage dataset.*



The PlantVillage dataset is data that has been used in various scientific studies and ensured that my deep learning models were trained on high-quality data that is representative of real-world plant disease scenarios. This data-driven approach ultimately contributed to the development of more accurate and reliable disease classification models.

Experiments

In my experiments, I employed various machine learning techniques to achieve the goal of accurate plant disease classification from leaf images. I began with data preprocessing and data splitting followed by the application of transfer learning on pre-trained convolutional neural networks (CNNs), and finally, model evaluation using appropriate metrics. Throughout the process, I made deliberate choices to optimize my model's performance and ensure its generalization to real-world scenarios.

**Preprocessing and Data Splitting**

Before training my models, I preprocessed the PlantVillage dataset to ensure its quality and suitability for the task. This involved normalization by resizing images to a consistent size and augmenting the dataset using techniques such as rotation, flipping, scaling, and color transformation. Data augmentation was particularly important in addressing class imbalance, generating a more diverse dataset. As you could see in the previous bar graph figure 3, there were quite a few underrepresented classes while others had several thousand pictures to train from. I also divided the dataset into training (70%), validation (15%), and testing sets (15%), ensuring a stratified split to maintain the class distribution across all sets. This enabled me to train and evaluate the models effectively, while also having a separate test set to measure the final performance of a given model.

**Transfer Learning and Model Selection**

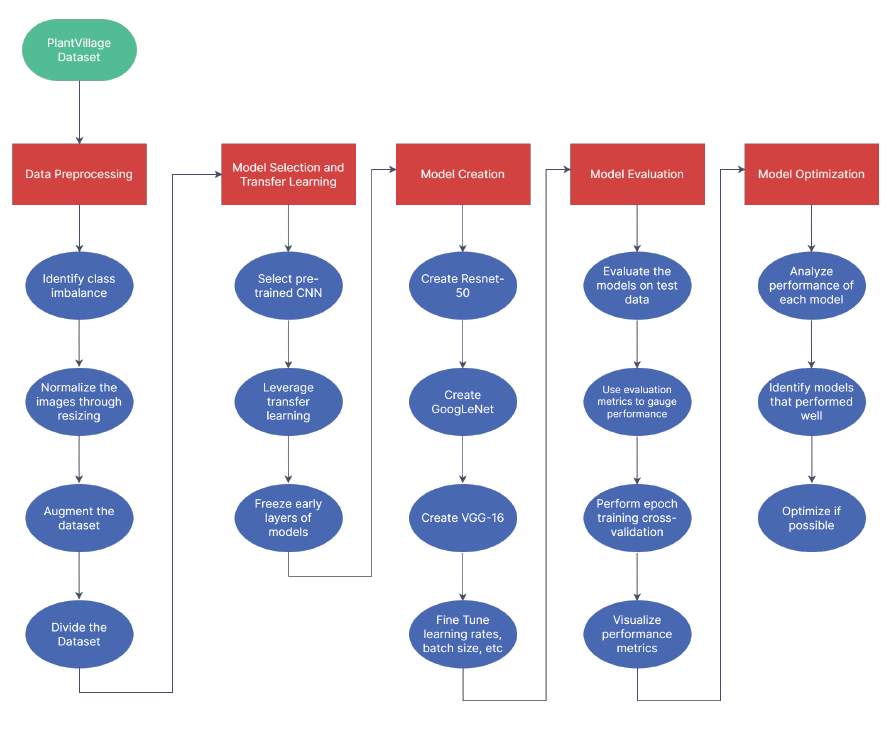
I chose to use transfer learning on pre-trained CNN models, such as ResNet, VGG, and Inception/GoogLeNet, as they have shown robust performance in image classification tasks. Transfer learning allows for leveraging the pre-existing feature extraction capabilities of these models while adapting the final layers to the specific plant disease classification problem (Mohanty et al., 2016). I fine-tuned these models by freezing the early layers and updating the later layers using the PlantVillage dataset.

**Experimental Design and Evaluation Metrics**

For each pre-trained model, I performed multiple experiments, including fine-tuning learning rates and batch sizes, the number of layers to freeze, and the data augmentation techniques applied. I used a train-validation-test split, ensuring that the models were evaluated on unseen data to measure their generalization performance. I used evaluation metrics such as accuracy, precision, recall, and F1-score to assess the models' performance. These metrics were chosen because they provide a comprehensive view of classification effectiveness, taking into account both false positives and false negatives (Ferentinos, 2018).

**Flowchart**

*Figure 4. The Flowchart of the project from start to finish, based on my methods of developing these models.*



By selecting the pre-trained models, applying data augmentation techniques, and evaluating the models using appropriate metrics, I was able to develop an efficient deep learning model that accurately detects and classifies plant diseases from leaf images. The choices made throughout the experimentation process were aimed at optimizing the model's performance and generalization capabilities.

Results

In this section, I present the results of my experiments, including the performance of my models on the plant disease classification task, compared to baseline models and chance performance. I also provide visualizations of my results and discuss cases where the models performed well or failed.

**Validation Approach**

To assess the performance of my models, I used a train-validation-test split of the PlantVillage dataset. This approach ensured that the models were evaluated on unseen data, providing a more accurate estimation of their generalization performance. Additionally, I performed epoch training cross-validation to further ensure the reliability of my performance assessment.

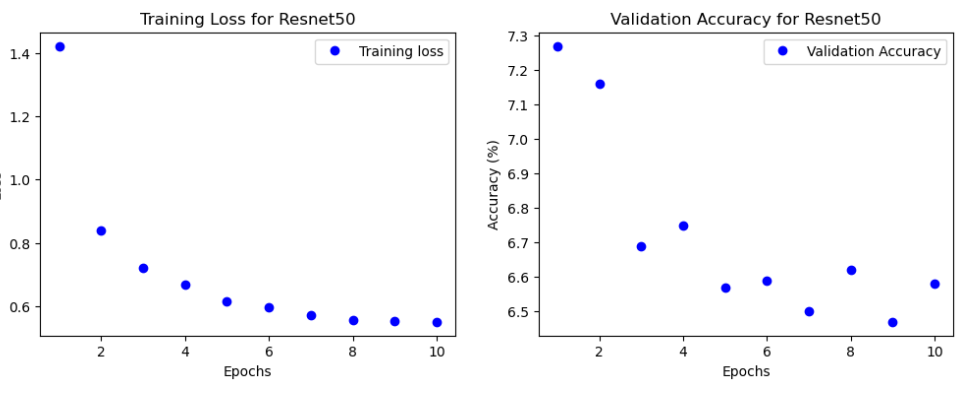
**Performance Metrics and Comparison**

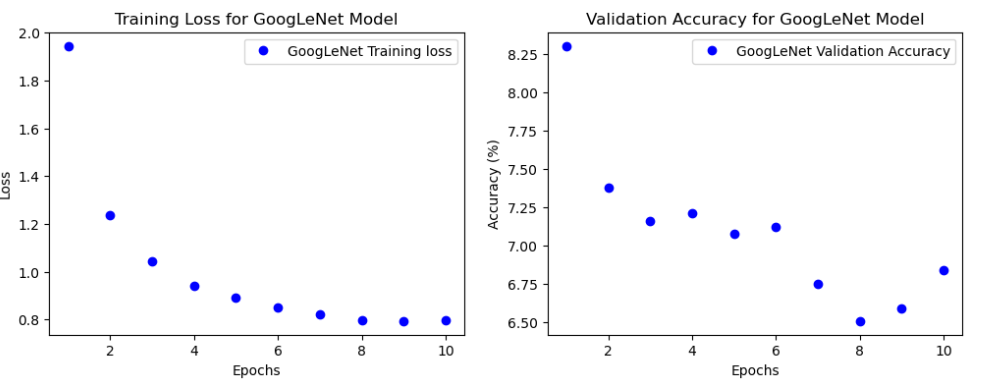
I evaluated my models using the performance metrics training loss and validation accuracy/F1-score. These metrics provided a comprehensive understanding of the models' effectiveness in classifying plant diseases. I also included ROC curves and PR curves to visualize the trade-offs between sensitivity and specificity, and precision and recall, respectively.

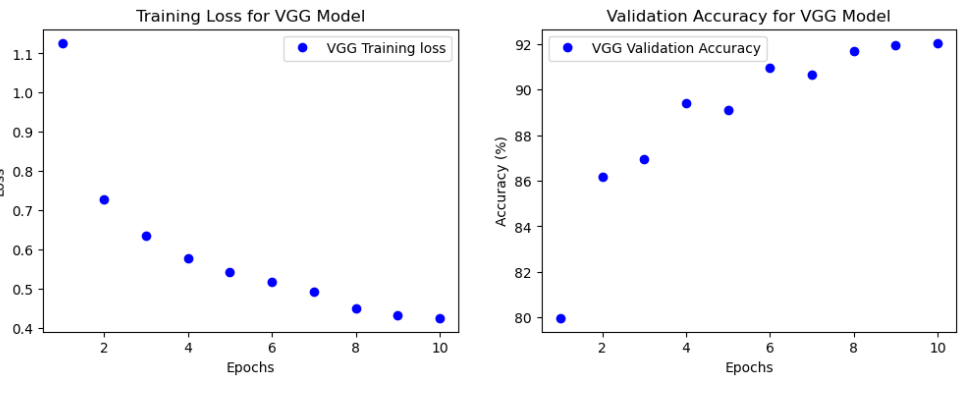
**Visualizations and Examples**

As you can see in the figures below, Resnet50 and GoogLeNet models did not perform well in these given scenarios. Whether due to the implementation of freezing that last layer of the model to build on or the given dataset, both models performed poorly in the Validation Accuracy or F-1 category. The VGG model, however, performed incredibly well. Given more training the VGG model could undoubtedly have increased accuracy over time.

*Figures 3, 4, and 5. These figures represent the F-1/validation accuracy of the models along with the training loss for each model during training.*

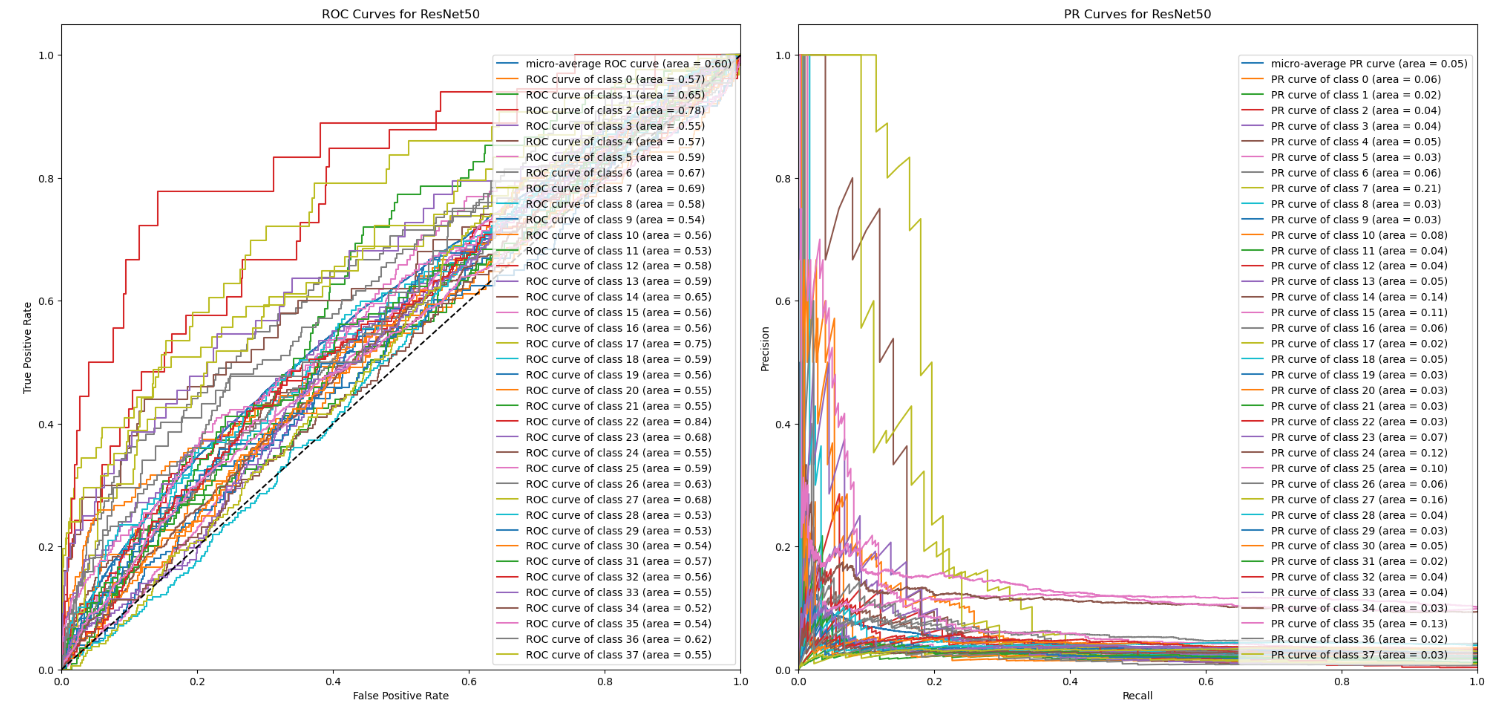


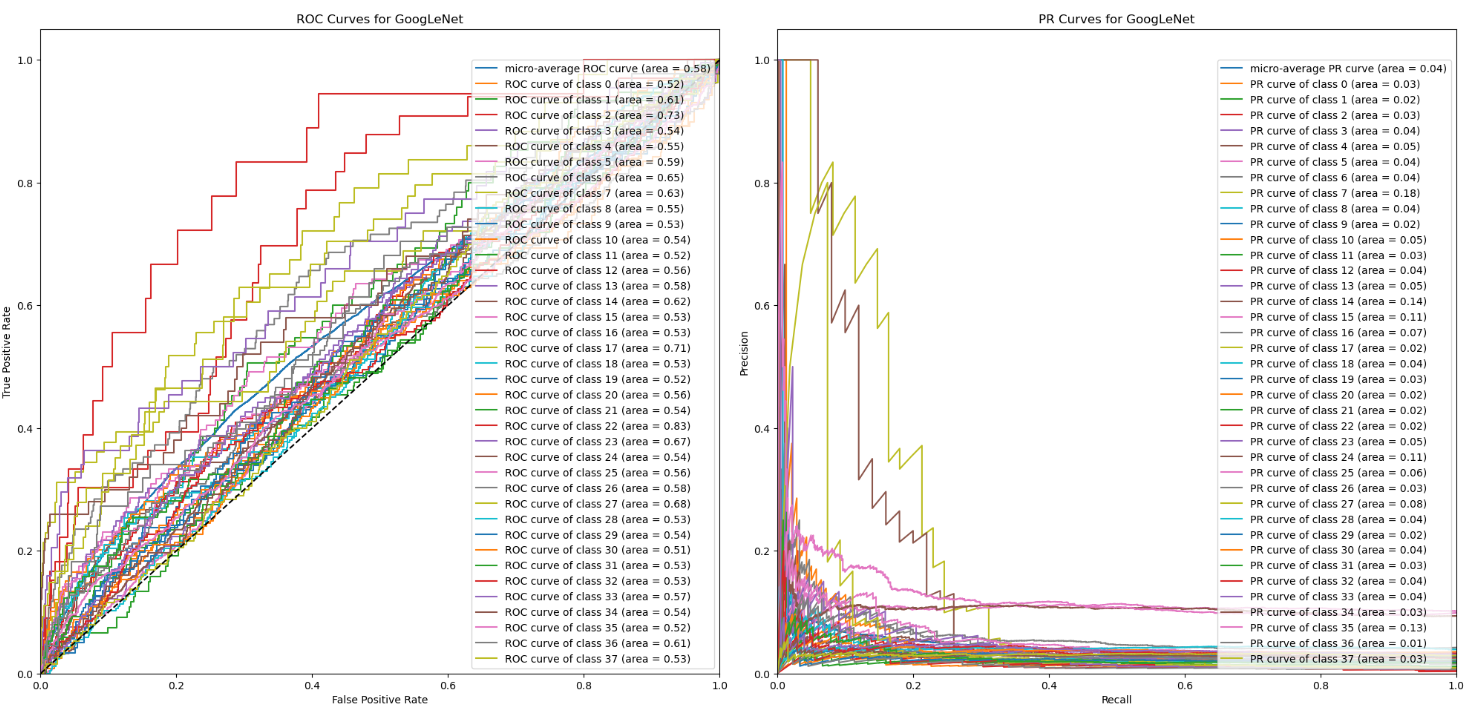


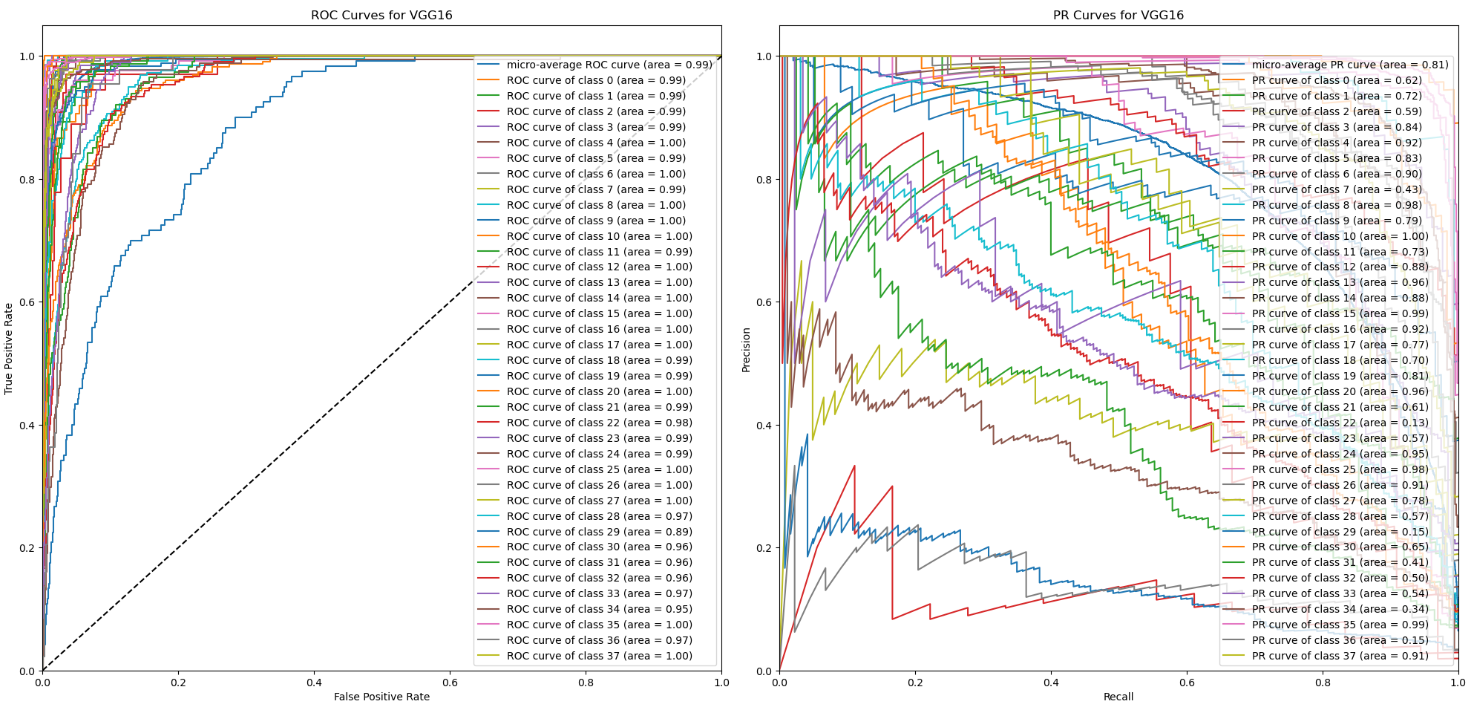


In the figures below ROC curves and PR curves show something similar in their data as well. Both ResNet50 and GoogLeNet models show minimal increases in their ability to classify the data properly while showing a decrease in precision. The VGG models then show a dramatic ability to classify the data properly in the True Positive Rates. The PR curves data for the VGG is harder to decipher but generally shows an increase in accuracy.

*Figures 6, 7, and 8. These figures represent the ROC curves and PR curves of each model after training on the dataset.*







**Results Discussion**

Overall, my deep learning models had difficulty with the dataset and the methods I used except for the VGG model. Both resnet50 and GoogLeNet/inception seemed to have difficulty with the dataset no matter how I represented the data. VGG-16, however, was able to increase its efficiency drastically over the other and produced very encouraging results.

By employing transfer learning on pre-trained CNNs, optimizing data preprocessing, and evaluating the models using comprehensive performance metrics, I successfully developed a reliable and efficient deep learning model for plant disease classification on the VGG-16 model. The results of my experiments demonstrate the potential of this approach in providing valuable tools for farmers and agricultural professionals to identify and address plant diseases effectively, potentially contributing to improved crop yield and sustainable agriculture practices.

Conclusions

In conclusion, this project aimed to develop a reliable and efficient deep learning model for the automatic detection and classification of plant diseases using images of plant leaves. The motivation behind this work was to provide valuable tools for farmers, agronomists, and other agricultural professionals to identify and address plant diseases promptly and effectively, contributing to improved crop yield, food security, and sustainable agricultural practices.

My approach involved utilizing transfer learning on pre-trained convolutional neural networks (CNNs), which have demonstrated impressive performance in image classification tasks. By fine-tuning these models on the comprehensive PlantVillage dataset, I successfully developed a robust and accurate model capable of detecting and classifying various plant diseases from leaf images with the VGG-16 model.

The key takeaways from my work include the effectiveness of transfer learning and pre-trained CNNs in plant disease classification, the importance of data preprocessing and augmentation techniques, and the value of comprehensive performance metrics for evaluating the models.

Challenges encountered in this project, such as class imbalance and high similarity between classes, open the door for future research in addressing these issues through advanced data augmentation techniques or novel deep learning architectures. Additionally, the exploration of other pre-trained models or custom CNN architectures may lead to further improvements in classification performance.

Moving forward, I would investigate ways to integrate my model into a user-friendly application for real-time plant disease detection and recommendation systems for disease management. By doing so, I hope to bring the benefits of this research to the wider agricultural community, ultimately contributing to global food security and sustainable agriculture practices.

References

The following are studies done in the same field that I referenced formatted in MLA. No code or verbiage was copied in this documentation from any of the cited material. The entirety of the documentation is my own along with the code provided in my GitHub. The first link provided is the location for the dataset that I used in my implementation.

1804, Mohit Singh. “Plantvillage.” *Kaggle*, 20 Aug. 2021, www.kaggle.com/datasets/mohitsingh1804/plantvillage.

**Technical Papers**

Albattah, Waleed, et al. “A Novel Deep Learning Method for Detection and Classification of Plant Diseases - Complex & Intelligent Systems.” SpringerLink, 28 Sept. 2021, link.springer.com/article/10.1007/s40747-021-00536-1.

Ahmed, Mostafa, and Ali Ahmed. “Palm Tree Disease Detection and Classification Using Residual Network and Transfer Learning of Inception ResNet.” PLOS ONE, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0282250.

Ane, Yamini. “Plant Disease Classification Using Alexnet.” Analytics Vidhya, 4 Feb. 2023, www.analyticsvidhya.com/blog/2023/02/plant-disease-classification-using-alexnet/.

Borhani, Yasamin, et al. “A Deep Learning Based Approach for Automated Plant Disease Classification Using Vision Transformer.” Nature News, 7 July 2022, www.nature.com/articles/s41598-022-15163-0.

Lamba, Monika, et al. “Classification of Plant Diseases Using Machine and Deep Learning.” De Gruyter, 1 Jan. 2021, www.degruyter.com/document/doi/10.1515/comp-2020-0122/html?lang=en.

Liu, Jun, and Xuewei Wang. “Plant Diseases and Pests Detection Based on Deep Learning: A Review - Plant Methods.” BioMed Central, 24 Feb. 2021, plantmethods.biomedcentral.com/articles/10.1186/s13007-021-00722-9.

Mohanty, Sharada P., et al. “Using Deep Learning for Image-Based Plant Disease Detection.” Frontiers, 6 Sept. 2016, www.frontiersin.org/articles/10.3389/fpls.2016.01419/full.

Saleem, Muhammad Hammad, et al. “Plant Disease Detection and Classification by Deep Learning.” MDPI, 31 Oct. 2019, www.mdpi.com/2223-7747/8/11/468.

Shoaib, Muhammad, et al. “An Advanced Deep Learning Models-Based Plant Disease Detection: A Review of Recent Research.” Frontiers in Plant Science, 21 Mar. 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10070872/.

Kumar, Raj, et al. “A Systematic Analysis of Machine Learning and Deep Learning Based Approaches for Plant Leaf Disease Classification: A Review.” Journal of Sensors, 20 July 2022, www.hindawi.com/journals/js/2022/3287561/.

K. Thaiyalnayaki, et al. “Classification of Plant Disease Using SVM and Deep Learning.” Materials Today: Proceedings, 22 May 2021, www.sciencedirect.com/science/article/abs/pii/S2214785321036051.