Early Detection and Diagnosis of Plant Diseases via Machine Learning

Aayush Sangani, Bijesh Shrestha, Chad Hucey, Daniel Fox, Wen Chiang Ivan Lim

1 Introduction

Research into the domain of plant disease detection using computer vision capabilities has piqued the interest of researchers from both the academic and industrial sides alike. We are growing accustomed to having various machine-learning applications that help us with difficult tasks, such as plant disease identification with varied levels of accuracy and usability. This project is driven by the potential of convolutional neural networks (CNNs) to significantly advance the identification process of plant ailments with a higher accuracy and efficiency that were previously unattainable. Our thesis explores advanced disease identification methods in plants through an active literature review and translating these techniques into a practical, user-friendly tool. We aim to expand on an existing researchers' codebase to develop an open-source web application that can serve both as a sophisticated plant disease detection system and as an educational platform for future developers and researchers in the field.

In this experiment, we make the following contributions:

- **Tool Function:** Our tool leverages ML and deep-learning algorithms to accurately classify plant disease from images that can be used for diagnosis and remediation suggestions.
- **Impact:** Our tool will be a step forward in improving global agriculture efforts, informing hobbyists, and inspiring aspiring researchers through access to a web platform that can be used as a product or boilerplate code repository to develop the tool further.

2 One-sentence description

Our project will develop an open-source web application that leverages advanced Convolutional Neural Networks to provide accurate and efficient plant disease detection and diagnosis, facilitating agricultural productivity, and serving as an educational tool for technology and agriculture enthusiasts.

3 Prior Works

- **Seed Paper:** Yao et al. [4] reviewed deep learning techniques for plant species identification and disease detection via leaf images, outlining four strategies: multi-model, multi-label, multi-output, and multi-task. They proposed an advanced method, Generalised Stacking Multi-output CNNs (GSMo-CNN), using stacked multi-output layers and interlayer connections for enhanced prediction accuracy.
- **Available Major Applications**

  There are several commercially available apps that can identify infected plants. Some of these apps include: "PlantSaver – Plants Identifier", "Agrio," and "Plant Disease Identifier," which are all available on the Apple App Store. Each app allows the user to take a picture of a plant, and the app will determine whether or not the plant is healthy. It is important to note that each app requires a monthly subscription for uninterrupted usage. Additionally, "Agrio" is the only service that offers a web-based application, which requires a monthly subscription.
- **Our Contribution**

  For this project, this team will develop a free-to-use guide for beginners to the field of Machine Learning who want to create their own Plant Disease Classifier. This team will outline the steps involved, including preprocessing techniques, model architectures, and evaluation metrics needed to create a robust model.
- **Impact: Difference, Difficulty and Challenges**

  The impact of this project lies in its ability to serve as a guide for future works. Future researchers can use this project as a reference and a starting point for their research.

  The challenges of developing this tool lies in democratizing access to a researcher's complex codebase of a deep learning model designed to classify samples across two dimensions (plant species and leaf disease). Our efforts focus on reviewing how we may further optimize the model's performance and improve its documentation to lower the entry barriers for replication and further research by others, including junior researchers.

4 Audience

The audience for this project can be broadly categorized into several groups:

Farmers and hobbyists are the primary beneficiaries of this project as they can use the developed system for the early detection of plant diseases. This enables them to take timely preventive measures, such as applying pesticides or adjusting irrigation, to protect their crops and prevent significant yield losses. **Aspiring researchers** can benefit from this project through the published guide and repository to learn and expand upon the work.

5 Approach

5.1 Data

- Primary: PlantDoc is a dataset for visual plant disease detection, consisting of 2,598 data points across

13 plant species and up to 17 disease classes, annotated from internet-scraped images to facilitate computer vision approaches in early plant disease identification [3].

- Secondary: In addition to PlantDoc, the paper also referenced the Plant Village dataset [2] and the Plant Leaves dataset [1]. The former consists of 54,305 images of either healthy or diseased leaves across 14 plant species (Apple, Blueberry, Cherry, Corn, Grape, Orange, Peach, Bell Pepper, Potato, Raspberry, Soybean, Squash, Strawberry, and Tomato), and 22 disease classes, while the latter consists of 4,502 images of healthy or diseased leaves across 22 plant species-disease classes.

5.2 ML pipeline and Architecture

The deep learning model we are replicating uses a CNN (Convolutional Neural Network) architecture structure, as shown in Figure 1. Specifically, the GSMo-CNN model introduces a novel approach by using the probabilities of plant species and diseases, from initial predictions through a cross-connection strategy between multi-output layers to enhance the final disease and plant species predictions, respectively.

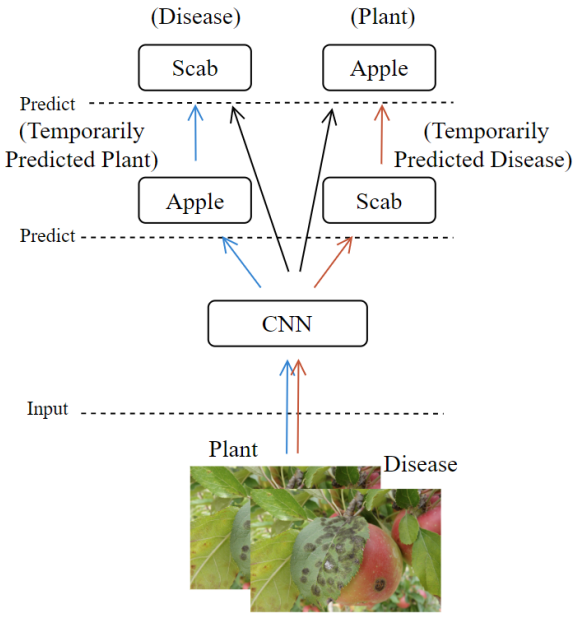

Figure 1: Model proposed by Yao et. al. , source: [4]

5.3 Web-app and its use

Building a web-based application for plant disease detection can greatly enhance its accessibility and usability for the intended audience. It can serve as a versatile tool for farmers, researchers, extension workers, and agri-tech companies, facilitating communication, collaboration, and knowledge exchange within the agricultural community. By leveraging the power of technology and data-driven insights, the application can empower users to make informed decisions and take proactive measures to safeguard crop health and enhance agricultural productivity.

6 Resources Needed

- High-Performance Computing Resource: Each team member will need active access to WPI's Turing network.

7 Define Success

For this work to be useful to the intended audience, we need:

- Web application: Deliver a sophisticated yet intuitive web application that consumers can use with varying levels of technical expertise. The application should offer a user-friendly interface, detailed diagnostic information, and (if time permits) guidance on potential remedial actions.
- The open-sourced and public repository needs to follow industry-standard knowledge management. In addition, the repository needs to be easy to implement for replication or further development.

References

[1] S. S. Chouhan. A Database of Leaf Images: Practice towards Plant Conservation with Plant Pathology, June 2019.

[2] D. P. Hughes and M. Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics, Apr. 2016. arXiv:1511.08060 [cs].

[3] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra. PlantDoc: A Dataset for Visual Plant Disease Detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 249–253, Hyderabad India, Jan. 2020. ACM.

[4] J. Yao, S. N. Tran, S. Garg, and S. Sawyer. Deep learning for plant identification and disease classification from leaf images: Multi-prediction approaches. *ACM Comput. Surv.*, jan 2024. Just Accepted.