Disease Detection in Plants Using ML Models on Plant Leaf Dataset

Submitted as a requirement of the course
EE769 Introduction to Machine Learning
Indian Institute of Technology Bombay 2024

Ayush Joshi
210100036
Electrical Engineering
Indian Institute of Technology Bombay

Mayank Gupta
210101002
Electrical Engineering
Indian Institute of Technology Bombay

Vansh Aggarwal
210070091
Electrical Engineering
Indian Institute of Technology Bombay

***Abstract*— This project explores the application of machine learning techniques for the detection of diseases in plants, aiming to mitigate the challenges posed by plant diseases to global food security and agricultural sustainability. Working on a dataset of plant leaf images curated for this study, we evaluate several machine learning models, including convolutional neural networks, batch gradient descent, support vector machines and random forest for disease detection in plants. Our results demonstrate promising performance, indicating the potential of machine learning in early disease diagnosis and prevention. The models could spot plant diseases with high accuracy. This means we might be able to catch plant diseases earlier and stop them from spreading. We talk about what all this means for farmers and scientists. We discuss the implications of our findings for agriculture and propose future research directions to further enhance the effectiveness of machine learning in plant disease management.**

I. Introduction

Imagine a world where plants could tell us when they're feeling under the weather. Sounds like something out of a science fiction novel, right? Well, with the power of machine learning, this futuristic scenario is becoming increasingly plausible.

Plants are susceptible to diseases that can wreak havoc on agricultural productivity and global food security. From fungal infections to viral outbreaks, these diseases pose a significant threat to our ability to sustainably feed the growing global population. Traditional methods of disease detection in plants have relied on visual inspection by trained experts, a time-consuming and often subjective process. However, with the emergence of machine learning techniques, there lies a promising avenue for revolutionizing plant disease management.

In this project, we explore the application of machine learning in the early detection of diseases in plants. Through empirical experimentation and analysis, we seek to develop robust machine learning models tailored for the task of disease detection in plants. Employing curated, we aim to demonstrate the efficacy of our approach in accurately identifying and classifying plant diseases. By harnessing the power of machine learning, we aspire to empower farmers and stakeholders with innovative tools for early diagnosis and targeted intervention, thereby mitigating the impact of plant diseases on food production and security.

II. Datasets

The dataset we have used is from https://github.com/spMohanty/PlantVillage-Dataset. It has images of leaves of various healthy and diseased plants, in colour and greyscale. It has a total of 39 categories, e.g. healthy apple, scab affected apple, healthy cherry, powdery mildew affected cherry, etc.

III. DATA ANALYSIS

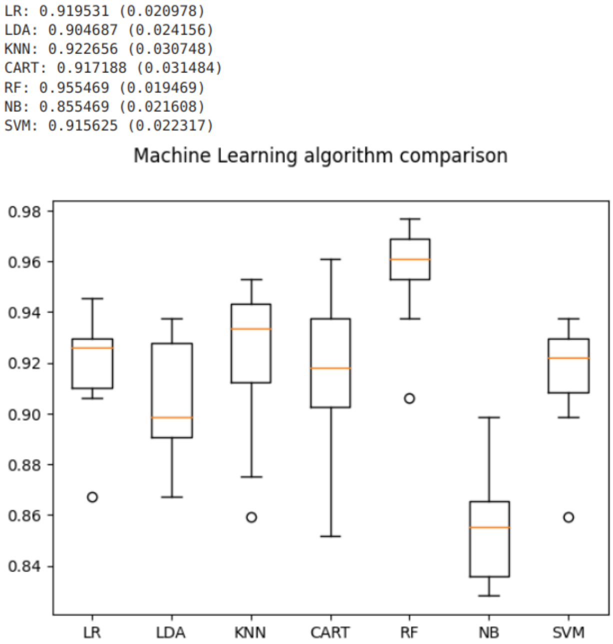
- Preprocessing the data, we resize, crop and convert the images to tensors.
- We split the datapoints in our dataset into 32311 for training, 13848 for validation and 21994 for testing.
- We use cross-entropy loss as loss criterion and Adam as our optimiser.

IV. RESULTS & INFERENCE

.

A. *Model training and selection –*

- Using PyTorch, we train a CNN using batch gradient descent over 5 epochs with a batch size of 64.
- This gives us test accuracy of 88.53%.
- We also train the model using SVM, over 20 epochs, which gives an accuracy of 91.64%.
- Training using Naïve Bayes (NB) gives an accuracy of 85.70%.
- Training using Random Forest gives an accuracy of 95.70%.
- Training using Classification and Regression Trees (CART) gives an accuracy of 90.62%.
- Training using K-Nearest Neighbour (KNN) gives an accuracy of 92.03%.
- Training using Linear Discriminant Analysis (LDA) gives an accuracy of 90.78%.
- Training using LR gives an accuracy of 91.64%.
- The accuracy data is represented in the following graph



B. *Key Learnings*

- This project gave us an opportunity to use the concepts learned in the course on a real-world problem.
- We got to learn and use new models as well along with the ones taught in class.
- We also learnt that different models work best for different types of data, e.g. RF gives best accuracy for disease prediction for this dataset, whereas NB performs relatively worse than the others.
- This project helped us apply machine learning and neural networks to images, instead of raw data as we had done in previous assignments.

CONCLUSION

Multiple organisations today focus on delivering optimal plant pesticides and medication through robots instead of farmers. This technology can help these robots find diseased crops more easily, thus saving essential chemicals and manpower. This can help robots decide optimal amount of raw materials to provide to plants to prevent them from catching diseases by predicting future outcomes. Using this, we can prevent more plants from catching diseases by detecting diseases prematurely, if trained with the right dataset.

Machine learning can be used to help farmers understand when pesticides need to be sprayed by using such models, proving to be of high importance in a country like India. This project helped us apply our knowledge obtained in class practically in a universally useful project.

Acknowledgment

We would like to thank Prof. Amit Sethi for teaching us this course, without which we would not have been able to complete this project.

References

[1] Dataset used: https://github.com/spMohanty/PlantVillage-Dataset

[2] Models : scikit-learn.org

[3] Plant Disease Detection using Image Processing and Machine Learning, Pranesh Kulkarni Et. Al. [2106.10698] Plant Disease Detection Using Image Processing and Machine Learning (arxiv.org)

[4] Performance analysis of AI- based solutions for crop disease identification, detection and classification. Divyanshu Tirkey Et. Al. Performance analysis of AI-based solutions for crop disease identification, detection, and classification - ScienceDirect