”Disease Detection in Plants”

Satyadev Subudhi
Roll No: 112101058
Indian Institute of Technology, Palakkad
112101058@smail.iitpkd.ac.in

Smart AgriTech Centre for Advanced Research and Development

This report submitted in partial fulfilment of the requirements for the Internship program, Smart AgriTech Centre for Advanced Research and Development. This project has been carried out with the guidance of Dr Satyajit Das and Dr Bhavesh Ram.

July 13, 2024

Abstract

This project aims to develop a robust model for classifying and detecting plant diseases using images. Leveraging the Plant Village Dataset, the project employs advanced image processing and deep learning techniques, specifically using MobileNetV2, ResNet, and VGG architectures. The study focuses on achieving high accuracy in disease classification to enhance crop management and agricultural productivity by enabling early intervention and treatment of plant diseases. MobileNetV2 emerged as the most effective model, showcasing superior performance in terms of accuracy.

Contents

1. Introduction

1.1 Background

Agriculture is crucial for global food security and economic stability, but it faces significant challenges from plant diseases, which can severely impact crop yields and quality. Traditional methods of plant disease detection involve manual inspection by experts, which is labor-intensive, time-consuming, and prone to human error. The advent of deep learning and image processing technologies presents an opportunity to automate and enhance the accuracy of plant disease detection.

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized image classification tasks, showing remarkable capability in identifying patterns and features in images. By applying these techniques to plant disease detection, we can potentially transform agricultural practices, leading to more efficient and effective disease management.

1.2 Objective

The primary objective of this project is to develop an accurate and efficient model for classifying and detecting various plant diseases from images. This is achieved using the Plant Village Dataset, which contains a diverse collection of images representing different plant species and disease conditions. For practical purposes, a subset of the dataset, focusing on five specific classes including potato and pepper plants, was selected.

The project explores three state-of-the-art deep learning architectures—MobileNetV2, ResNet, and VGG—to determine the most effective model for this task. The models are trained, validated, and tested to evaluate their performance in classifying plant diseases accurately.

By developing a reliable plant disease classification model, the project aims to support farmers in managing crops more effectively, thereby enhancing agricultural productivity and sustainability. Early and accurate disease detection will enable timely interventions, reducing crop losses and contributing to global food security.

2. Methodology

2.1 Research Design

This project employs a quantitative research design to develop and evaluate deep learning models for the classification and detection of plant diseases from images. The research process involves multiple stages, including data collection, data preparation, model development, model evaluation, and results visualization. The primary objective is to identify the most accurate model among MobileNetV2, ResNet, and VGG architectures for plant disease classification.

2.2 Data Collection

The dataset used for this project is sourced from the Plant Village Dataset, available on Kaggle. The dataset comprises 20,639 images across 15 different classes of leaves. For this study, a subset containing images from five specific classes, including potato and pepper plants, was selected to expedite the testing phase while maintaining a diverse representation of plant types.

2.3 Data Preparation

2.3.1 Loading the Dataset:

The images were loaded using TensorFlow's image dataset loading utilities. The dataset was shuffled and resized to a uniform size suitable for model training.

2.3.2 Data Augmentation:

To enhance the robustness of the model, data augmentation techniques such as rotation and flipping were applied. This helps in preventing overfitting by increasing the diversity of the training data.

2.3.3 Dataset Splitting:

The dataset was divided into training, validation, and testing sets with a split ratio of 0.7, 0.2, and 0.1 respectively.

2.4 Model Development

Three advanced deep learning models were developed and evaluated: MobileNetV2, ResNet, and VGG. These models were chosen for their unique characteristics and proven performance in image classification tasks.

2.4.1 MobileNetV2:

MobileNetV2 is a lightweight model designed for mobile and embedded vision applications, known for its efficiency and real-time performance capabilities.

2.4.2 ResNet:

ResNet (Residual Network) is renowned for its deep architecture and the introduction of residual learning, which facilitates training very deep networks.

2.4.3 VGG:

VGG is known for its simplicity and uniform architecture, which has been widely used for image classification tasks.

2.5 Model Training

Each model was trained using the training dataset with hyperparameters such as batch size = 32 and epochs = 20. The Adam optimizer and categorical cross-entropy loss function were used to compile the models.

2.6 Model Evaluation

The models were evaluated on the validation and testing datasets to determine their accuracy and effectiveness. Metrics such as precision, recall, and F1 score were calculated to provide a comprehensive assessment of model performance.

2.7 Results Visualization

The training and validation accuracy and loss were visualized using matplotlib to compare the performance of the models over the training epochs. Additionally, sample predictions on test images were visualized to illustrate the model's ability to classify plant diseases correctly.

3. Conclusion

3.1 Summary of Findings

This project successfully developed a robust model for classifying and detecting plant diseases from images using advanced deep learning techniques. By leveraging the Plant Village Dataset and employing three state-of-the-art deep learning architectures—MobileNetV2, ResNet, and VGG—we were able to create models that effectively identify and categorize plant diseases. Among these, MobileNetV2 demonstrated superior performance in terms of accuracy, making it the most suitable model for real-time applications on devices with limited computational power.

3.2 Future Work

While the current project achieved promising results, several areas for future work can further enhance the model's performance and applicability-

3.2.1 Utilizing the Full Dataset:

Future iterations should consider training the models on the entire Plant Village Dataset, which includes 20,639 images across 15 classes, to improve the generalizability and robustness of the model.

3.2.2 Addressing Data Imbalance:

Techniques such as oversampling, undersampling, or generating synthetic data using methods like SMOTE (Synthetic Minority Over-sampling Technique) can be employed to balance the dataset and improve the model's performance on minority classes.

3.2.3 Exploring Additional Models:

Investigating other deep learning architectures, such as EfficientNet or custom-built CNNs, may yield better performance and further improvements in accuracy and efficiency.

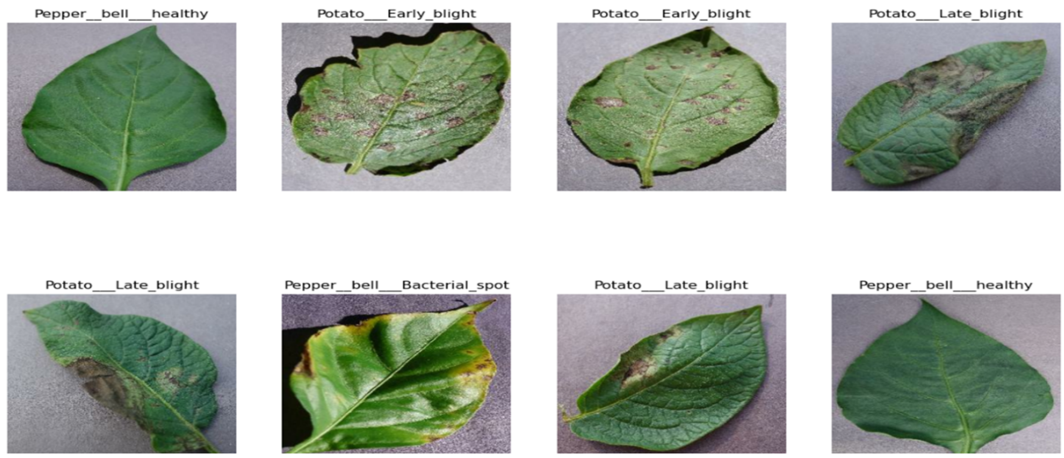
3.2.4 Enhancing Model Architecture:

Adding more convolutional layers, using dropout layers to reduce overfitting, and experimenting with different hyperparameters can lead to a more optimized model.

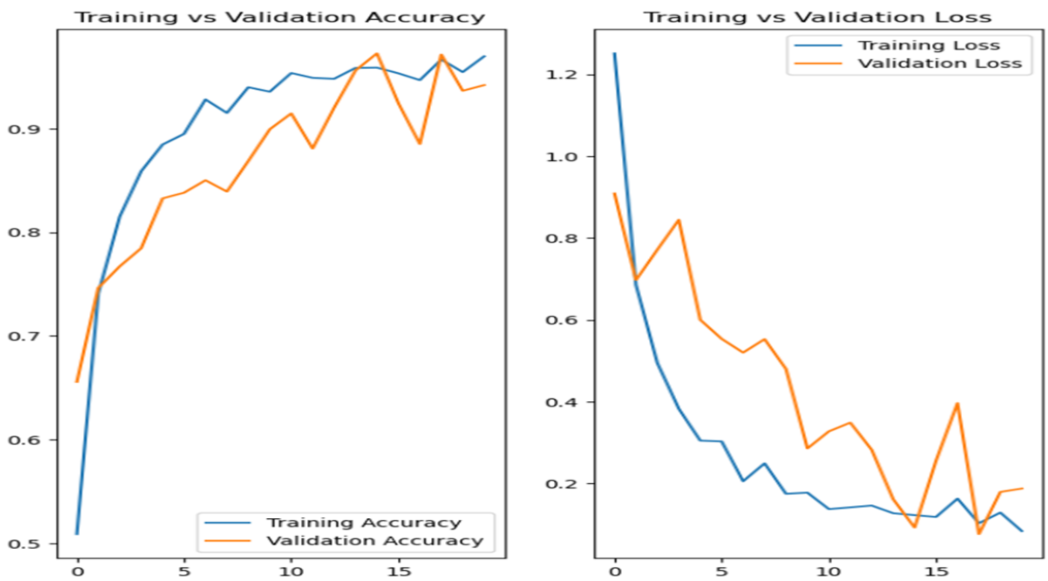
4. List of Figures

4.1 Visualisation of Data:

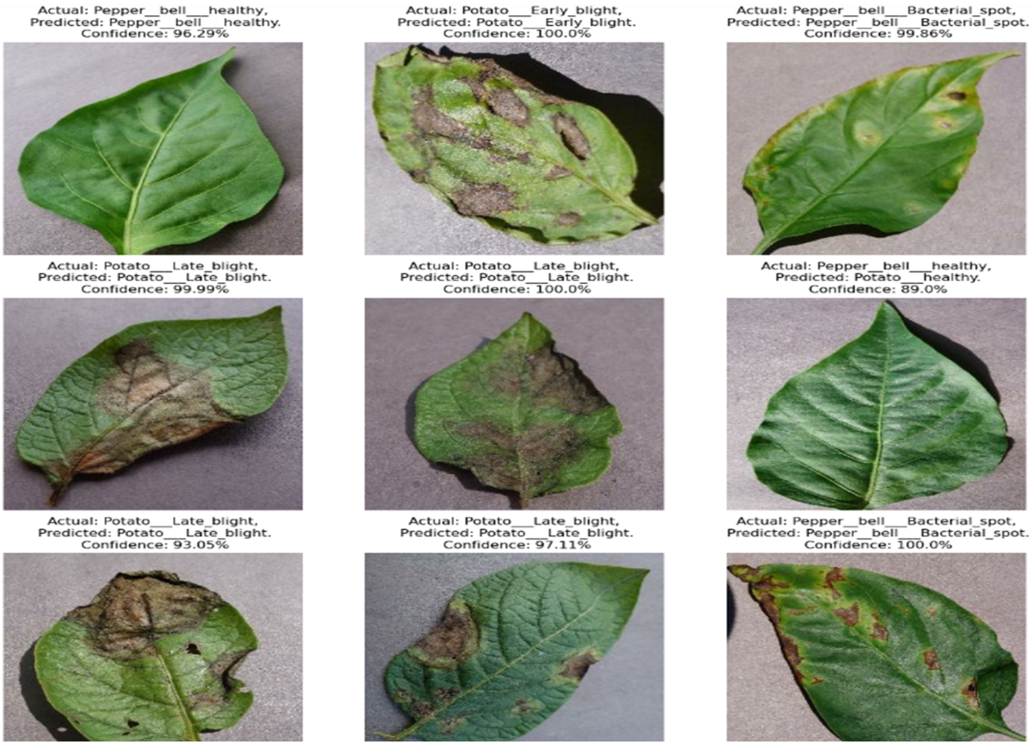
```
In [76]: # vizualization of dataset
plt.figure(figsize=(15, 15))
for image_batch, labels_batch in dataset.take(1):
    for i in range(12):
        ax = plt.subplot(3, 4, i + 1)
        plt.imshow(image_batch[i].numpy().astype("uint8"))
        plt.title(class_names[labels_batch[i]])
        plt.axis("off")
```



4.2 Training vs Validation Accuracy and Loss:



4.3 Testing model on custom input:



A. Annexure

A.1 Link of code:

Plant disease classification (Kaggle notebook)

A.2 Link of dataset:

[Plant Village Disease Classification](https://www.kaggle.com/datasets/emmarex/plantdisease)