BAHRIA UNIVERSITY

INTRODUTION TO DATA SCIENCE LAB

Potato Leaf Disease Classification Using CNNs

PROJECT REPORT

SUBMITTED BY:

MATEEH ULLAH ABBASI
01-135221-066

BAHRIA UNIVERSITY
ISLAMABAD

Table of Contents

Potato Leaf Disease Classification Using Convolutional Neural Networks (CNNs)

1. Introduction & Problem Statement

Potatoes are one of the most widely consumed crops around the world and are a critical part of the agricultural economy in many regions. However, potato crops are susceptible to a variety of diseases such as Early Blight and Late Blight, which can cause significant yield losses and economic damage to farmers.

Manual disease detection by farmers or experts is time-consuming, expensive, and prone to human error. Hence, there is a strong need for an automated, reliable, and scalable solution. This project aims to address this issue by leveraging Convolutional Neural Networks (CNNs) to accurately classify potato leaf diseases based on image data.

Automating disease identification through deep learning not only increases efficiency and accuracy but can also be integrated into mobile or drone-based applications, offering real-time support to farmers and agronomists.

2. Objectives

Primary Objective:

To build a deep learning model using CNNs that can classify potato leaf images into three categories: Healthy, Early Blight, and Late Blight.

Secondary Objectives:

- To preprocess and normalize the image dataset for optimal model performance.
- To evaluate the model's accuracy using performance metrics such as precision, recall, F1-score, and accuracy.
- To use data augmentation techniques to increase the robustness of the model.
- To visualize training results and evaluate the model using test data.

3. Dataset Description

- **Source:** PlantVillage Dataset on Kaggle
- **Size:** 2,150 images across 3 categories (Healthy, Early Blight, Late Blight)
- **Format:** RGB images (.jpg format), size 256x256 pixels

Preprocessing Steps:

- Resizing all images to a uniform size (256x256).
- Normalizing pixel values to range [0,1].
- Augmenting the dataset using rotation, flipping, and zoom to improve generalization.
- Splitting into training , validation , and test sets.

4. Methodology

This project employs a Convolutional Neural Network (CNN) for multi-class image classification. CNNs are well-suited for image-based tasks due to their ability to learn spatial hierarchies of features.

Model Details:

- Sequential CNN architecture
- Convolutional layers + MaxPooling
- Dropout for regularization
- Dense output layer with Softmax activation

Tools & Libraries:

- Python
- TensorFlow & Keras
- NumPy, Matplotlib, Seaborn
- Google Colab for development

Why CNN?

CNNs are the state-of-the-art for image classification problems due to their ability to extract spatial features automatically, without manual feature engineering.

5. Expected Outcomes

- A trained CNN model that can classify potato diseases with at least 90% accuracy on test data.
- A confusion matrix and classification report for performance analysis.
- A graphical analysis of training vs validation loss and accuracy over epochs.
- Insights on overfitting/underfitting using visualization tools.

6. Timeline

Week	Task
1	Project proposal writing, dataset selection
2	Data loading and preprocessing
3	Model architecture design
3	Model training and fine-tuning
4	Evaluation and visualization of results
4	Writing final report and documentation

7. References

1. Arjun Tejaswi, PlantVillage Dataset – Kaggle: **https://www.kaggle.com/datasets/emmarex/plantdisease**
2. Chollet, François. *Deep Learning with Python*, 2nd Edition.
3. TensorFlow Documentation: https://www.tensorflow.org