POTATO DISEASE CLASSIFICATION

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

Branch

SUBMITTED BY

DIVYANSHU KUMAR (2104095)

ANURADHA BHARTI (2104072)

GURU NANAK DEV ENGINEERING COLLEGE,

LUDHIANA

Table of contents:

Introduction

The potato disease classification project using deep learning involves training a neural network on a dataset of images of healthy and diseased potato plants to accurately classify different diseases affecting potato crops, by learning and recognizing patterns and features in the images.

Here, we'll develop an end-to-end Deep Learning project in the field of Agriculture. We'll create an Image Classification Model that will categorize potato disease using simple CNN. We'll start by gathering data from different sites and then we'll prepare the data for model building and then we'll display the output in a webpage and any app prototype. This webpage will be developed on React Native and the App on React JS. Also we'll be using TF Serving and FASTAPI for building the local server for transferring the real life image to predict the output.

Figure 1

Rationale

Farmers who grow potatoes suffer from serious financial standpoint losses each year which cause several diseases that affect potato plants. The diseases Early Blight and Late Blight are the most frequent. Early blight is caused by fungus and late blight is caused by specific micro-organisms and if farmers detect this disease early and apply appropriate treatment then it can save a lot of waste and prevent economical loss. The treatments for early blight and late blight are a little different so it's important that you accurately identify what kind of disease is there in that potato plant. Behind the scene, we are going to use Convolutional Neural Network – Deep Learning to diagnose plant diseases.

Objectives:

1. Identify diseases in the early stages to prevent their spread and minimize damage.
2. Accurately classify different potato diseases, such as late blight, early blight, or black scurf, to ensure the correct treatment or preventive measures are applied.
3. Provide farmers with a cost-effective tool for disease diagnosis, reducing dependency on expensive lab tests or expert consultations.
4. Reduce loss caused by diseases by ensuring proper and timely management

Literature Review:

Potato is one of the world's most important food crops, playing a crucial role in global food security. However, potato crops are susceptible to various diseases, including late blight, early blight, black scurf, and bacterial wilt, which can significantly reduce yield and quality. Accurate and timely classification of potato diseases is essential for effective management and mitigation strategies. Advances in machine learning (ML), computer vision, and remote sensing have provided innovative tools for disease classification. This literature review explores the state-of-the-art techniques, datasets, and methodologies in potato disease classification.

Traditional Approaches Traditional disease classification relied heavily on visual inspection by agricultural experts. This approach has several limitations, including subjectivity, high labour costs, and the inability to scale across large agricultural areas. Microscopic and biochemical analyses have also been used, but these methods are time-consuming and require specialized equipment.

Deep Learning (DL) Models The advent of deep learning has revolutionized potato disease classification. Convolutional Neural Networks (CNNs) have become the most commonly used architecture due to their ability to automatically learn hierarchical features from image data. Studies have demonstrated that CNNs can achieve high accuracy in classifying diseases such as late blight and early blight.

Feasibility Study:

Conducting a feasibility study is essential to determine the practicality of implementing potato disease classification systems in real-world settings. Key aspects to consider include:

- **Technical Feasibility**: Assessing the availability of computational resources, such as GPUs, and evaluating the scalability of the chosen ML or DL models.
- **Economic Feasibility**: Estimating costs associated with data collection, model training, and deployment, and comparing them against potential benefits such as increased crop yield and reduced pesticide usage.
- **Operational Feasibility**: Ensuring that the system can be easily integrated into existing agricultural workflows and is user-friendly for farmers and field workers.
- **Environmental Feasibility**: Evaluating the system's effectiveness under diverse environmental conditions, such as different climates, soil types, and disease prevalence.

Methodology/ Planning of work:

Extensive Exploratory Data Analysis (EDA): Perform a thorough analysis of the dataset to gain valuable insights and understand the characteristics of potato diseases. Feature Selection: Utilize correlation analysis to uncover the interrelationships among variables, helping to select the most important features for the classification task. Model Robustness: Employ advanced techniques like Variance Inflation Factor (VIF) to mitigate multicollinearity and enhance the robustness of the model.

1. Problem Definition

Identify the specific diseases of interest (e.g., early blight, late blight, bacterial wilt).

Determine the desired outcomes (e.g., disease type, severity level).

Specify constraints such as computational resources, time, and accuracy requirements.

2. Data Collection

Sources of Data:

Collect potato leaf images from open-source datasets (e.g., PlantVillage).

Capture real-world images using smartphones, cameras, or drones in various lighting and environmental conditions.

3. Data Preprocessing

Image Augmentation:

Enhance the dataset diversity using techniques like flipping, rotation, zooming, and adding noise.

4. Feature Extraction

Deep Learning Features:

Use convolutional neural networks (CNNs) to learn relevant features directly from the data.

5. Model Development

Model Selection:

Train custom CNN architectures if sufficient data is available.

Model Training:

Use a training-validation split to evaluate performance during training.

Loss Function:

Use categorical cross-entropy or binary cross-entropy, depending on the output format.

Optimizer:

Choose optimizers like Adam or SGD for model training.

6. Model Evaluation

Metrics:

Evaluate the model using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

Confusion Matrix:

Analyse the confusion matrix to understand misclassifications.

Cross-validation:

Perform k-fold cross-validation to ensure robustness.

Planning of Work

Phase	Tasks	Duration	Deliverables
Phase 1: Data	Collect and preprocess data	2–3 weeks	Annotated and augmented dataset
Phase 2: Model	Develop and train model	4–5 weeks	Trained and validated model
Phase 3: Test	Evaluate on test data	1–2 weeks	Performance metrics and insights
Phase 4: Deploy	Implement and deploy solution	2–3 weeks	User-accessible model (app or interface)
Phase 5: Monitor	Gather feedback and update	Ongoing	Updated model versions

Facilities required for proposed work:

Software Requirements

Machine learning and deep learning libraries (e.g., TensorFlow, PyTorch, OpenCV).

Data preprocessing tools like pandas, NumPy, and scikit-learn.

Tools for building a mobile app (e.g., Flutter, React Native) to make the classifier user-friendly.

Operating System

: Windows 10 home and above version

Language:

Python

Source Code Editor:

Google Collab, Visual studio code, Jupiter Notebook Graphics card 1080

Hardware Requirements

Processor:

11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40GHz 2.42 GHz

RAM:

4GB & Above

Hard disk:

512 GB SSD

Expected outcomes:

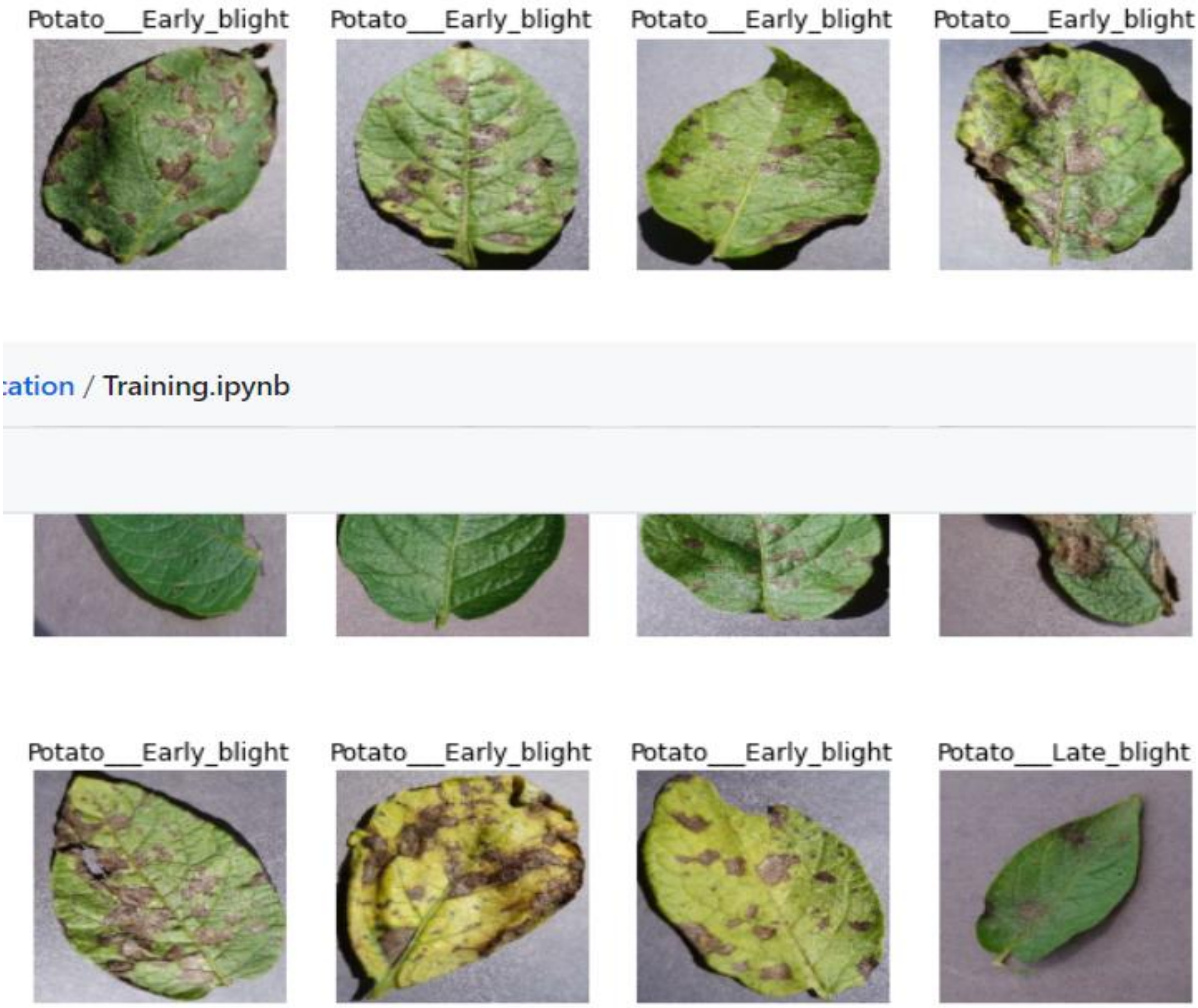

Figure 2.

References:

www.google.com

www.scribe.com

www.codebasic.com

www.nareshit.com

www.w3school.com

www.geekforgeeks.com

www.javatpoint.com

www.youtube.come