HERITAGE INSTITUTE OF TECHNOLOGY

Final Year Project Synopsis

(Interim Viva)

Potato Disease Classification using Machine Learning

Prepared by:

Kinjal Kanjilal - 2161065

Debdipta Mitra - 2161093

Rupsha Sarkar - 2161067

Krishnarjun Nanda - 2161061

Mentor:

Jhalak Dutta (Department of CSE)

Abstract:

*This project focuses on early detection of potato diseases using a **Convolutional Neural Network (CNN)** model. Potatoes are a major global staple crop, but diseases like Early Blight and Late Blight significantly impact yield and quality, leading to financial losses for farmers. This model classifies potato leaf images into three categories: **Healthy, Early Blight, and Late Blight.** Using labelled images for training, the CNN achieves accurate classification to support proactive disease management. A **Web API** is also developed, allowing users to upload leaf images for instant classification, providing confidence scores to guide timely, effective treatment decisions.*

Introduction:

*Potatoes are a globally consumed staple, vital to food security and economic stability. However, diseases like Late Blight and Early Blight, caused by the pathogens **Phytophthora infestans and Alternaria solani,** result in severe yield losses by affecting the health of potato plants. Early identification and management of these diseases are essential to reduce financial losses and improve crop quality. Traditionally, disease detection relies on visual inspection by experts, but this approach can be impractical, especially on large or remote farms where there may be limited access to agricultural specialists. Recent advancements in image processing have introduced efficient, scalable disease monitoring through image analysis. This project aims to classify potato leaf images into three categories—**Late Blight, Early Blight, and Healthy**—using a **Convolutional Neural Network (CNN)** model. With early disease detection, farmers can make informed decisions and apply preventive measures in a timely manner. The model is integrated into a web application, allowing users to upload leaf images and receive instant classification results with confidence scores, making it accessible for routine use in agriculture. Considering that agriculture accounts for **18% of India's GDP** and supports about **60% of the population**, this CNN-based approach empowers farmers, especially in remote areas, to leverage technology for better crop management, sustainable practices, and enhanced productivity.*

Methodology:

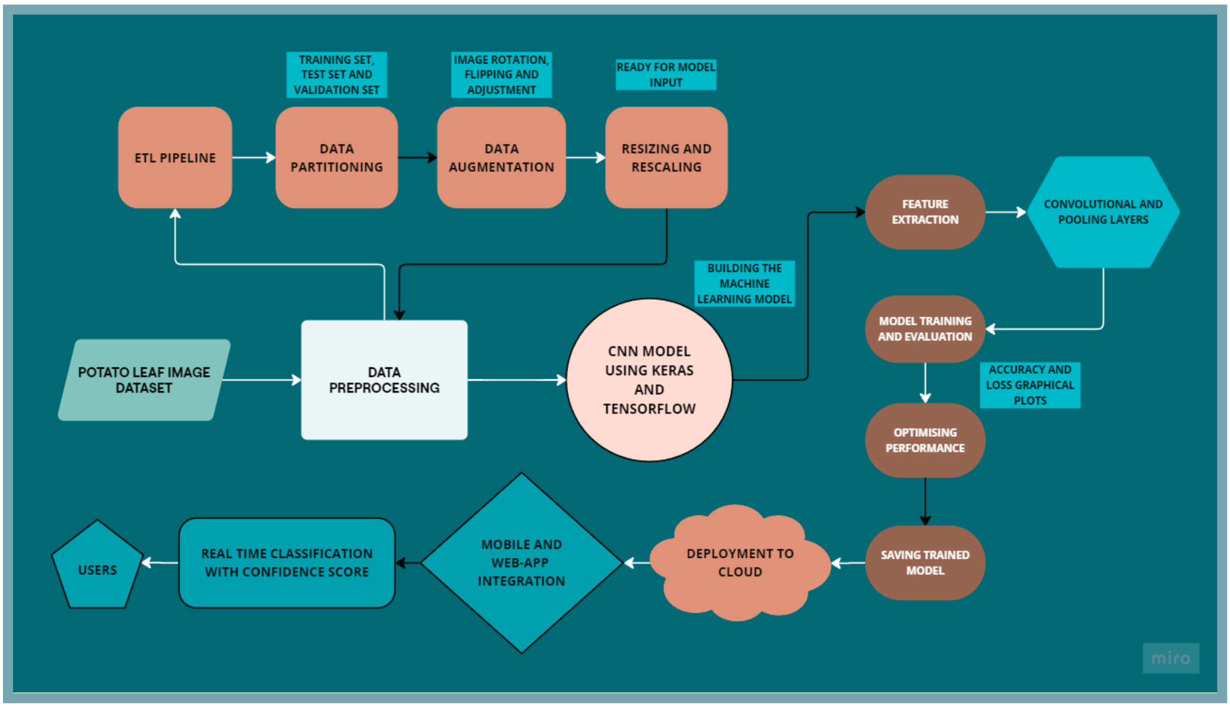
- *In this project, we developed a potato disease classification model using **Convolutional Neural Network (CNN)** technology to accurately identify and classify potato leaf images as Healthy, Early Blight, or Late Blight. This classification system enables early detection and monitoring of disease spread in crops, which is essential for effective management and prevention of large-scale agricultural losses.*

- *To build the model, we implemented a comprehensive **ETL (Extract, Transform, Load) pipeline**. This pipeline involved partitioning the dataset into training, validation, and testing sets, followed by data shuffling to prevent bias and enhance model generalizability. Each image was resized and rescaled to ensure uniform input dimensions and normalised pixel values for **optimal performance**. To increase data variability, we used data augmentation techniques, including random rotations, flips, and brightness adjustments. This preprocessing enhanced the model's robustness to diverse visual conditions, preparing it to perform well under real-world circumstances.*

- *Using **TensorFlow and Keras**, we structured the **CNN** with a **sequential model architecture**, incorporating multiple convolutional layers and max-pooling layers to progressively extract relevant features from the images. This structure allows the model to identify detailed patterns and distinguish between the disease stages effectively. For model evaluation, we used **Matplotlib** to plot training and validation accuracy and loss, enabling us to monitor the model's performance across epochs and prevent **overfitting**.*

- *After training, the model was compiled, and the final model was saved for further **integration into a web application**. This application allows users to upload potato leaf images for real-time classification, with confidence scores to guide their understanding of disease presence and severity. By making disease detection accessible to farmers, this project supports more timely interventions, promoting healthier crops and contributing to **sustainable agricultural practices.***

Methodology Flow Diagram:



Related Surveys :

➢ ***Detection of Plant Leaf Diseases Using CNNs***

Sladojevic et al. (2016) used a deep convolutional neural network to automatically detect 13 different plant diseases across various plant species. The model achieved an accuracy of 96.3%, demonstrating the potential of CNNs in identifying plant diseases from images. Their study highlighted the importance of sufficient training data for robust model performance.

➢ ***Transfer Learning for Plant Disease Classification***

Mohanty, Hughes, and Salathé (2016) explored transfer learning by using pre-trained models like AlexNet and GoogLeNet to classify plant leaf images into 38 different disease classes. The models achieved a test accuracy of up to 99.35%. This study *underscored the benefits of leveraging pre-trained networks to reduce training time and improve accuracy.*

- ***Improving Plant Disease Detection via Data Augmentation***

  Ferentinos (2018) investigated the impact of different data augmentation techniques, such as rotation, scaling, and flipping, on the performance of CNNs for plant disease detection. By augmenting the dataset, the model achieved an accuracy of 99.53% on a dataset containing 58 different plant diseases. This study demonstrated that data augmentation can significantly improve model robustness.

- ***Hyperparameter Optimization in CNN-based Plant Disease Models***

  Too et al. (2019) conducted a comprehensive analysis of different hyperparameters in CNNs, including learning rate, batch size, and the number of layers. They used a grid search to optimise these parameters, achieving an accuracy of 98.7% on a dataset with 38 plant diseases. The findings emphasised the need for proper hyperparameter tuning to enhance model performance.

- ***Plant Disease Detection using MobileNet***

  Howard et al. (2017) developed MobileNet, a lightweight deep learning model designed for mobile and embedded applications. Although not specifically focused on plant diseases, MobileNet has been adapted for plant disease classification tasks, achieving over 90% accuracy with significantly reduced computational requirements.

- ***A Comparative Study on CNN Architectures for Plant Disease Detection***

  Zhang et al. (2020) compared various CNN architectures such as VGG16, ResNet50, and DenseNet121 for plant disease detection. Their results showed that ResNet50 performed best with a test accuracy of 98.1%, while VGG16 and DenseNet121 followed closely behind. The study suggested that deeper networks like ResNet might offer an advantage when dealing with large datasets.

References:

[1]. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

[2]. Mirwaes Wahabzada, Anne-Katrin Mahlein, Christian Bauckhage, Ulrike Steiner, Erich Christian Oerke, Kristian Kersting, "'Plant Phenotyping using Probabilistic Topic Models: Uncovering the Hyperspectral Language of Plants," Scientific Reports 6, Nature, Article number: 22482, 2016.

[3]. Macedo-Cruz A, Pajares G, Santos M, Villegas-Romero I., "Digital image sensor-based assessment of the status of oat (Avena sativa L.) crops after frost damage," Sensors 11(6), 2011, pp. 6015–6036.

[4]. Yao Q, Guan Z, Zhou Y, Tang J, Hu Y, Yang B, "Application of support vector machine for detecting rice diseases using shape and color texture features," 2009 International Conference on Engineering Computation, IEEE, Hong Kong, 2009, pp. 79–83.

[5]. Phadikar S, Sil J, "'Rice disease identification using pattern recognition techniques," IEEE, Khulna, 2008, pp. 420_423.

[6]. Sharada Prasanna Mohanty, David Hughes, Marcel Salathe, "Using Deep Learning for ImageBased Plant Disease Detection," Computer Vision and Pattern Recognition, to be published.
