PANDIT DEENDAYAL ENERGY UNIVERSITY

SCHOOL OF TECHNOLOGY

Artificial Intelligence Lab

21CP313P

PROJECT REPORT

B.Tech. (Computer Science and Engineering)

Semester 6

Submitted To:

Dr. Pooja Shah

Submitted By:

HARSH SHAH

21BCP359

G11 Batch

INDEX

Plant Disease Classification

Abstract

This project aims to develop an AI system for automated plant disease classification. Leveraging deep learning techniques, the system will analyse images of plant leaves to accurately identify diseases. By training on a comprehensive dataset encompassing various plant species and diseases, the model will learn to differentiate between healthy and diseased plants with high precision. The proposed solution holds promise for early disease detection, enabling timely intervention and management practices to mitigate crop losses. Through its efficient and scalable approach, this AI system has the potential to revolutionize agricultural practices, fostering sustainable food production and enhancing global food security.

Domain introduction

The domain of plant disease classification stands as a vital frontier where agriculture intersects with cutting-edge artificial intelligence (AI) technologies. In the face of a burgeoning global population and the ever-growing demand for food, ensuring agricultural productivity and food security has become increasingly imperative. Traditional methods for detecting diseases in crops often rely on manual inspection by experts, a process that is labour-intensive, time-consuming, and prone to errors. However, recent advancements in computer vision and machine learning have paved the way for more efficient and accurate disease diagnosis in plants.

By leveraging AI techniques, particularly in image analysis, researchers and practitioners can develop systems capable of swiftly and accurately identifying various diseases afflicting plants solely based on images of their leaves. These AI-driven solutions have the potential to revolutionize agricultural practices by providing farmers with early detection tools that enable prompt intervention and targeted treatment, thus minimizing crop losses and optimizing yields. Moreover, the scalability and accessibility of these AI technologies mean that they can be deployed across diverse agricultural landscapes, benefitting farmers worldwide regardless of their location or resources.

In essence, the fusion of agriculture and AI in the realm of plant disease classification represents a promising paradigm shift towards more efficient, sustainable, and resilient food production systems. By harnessing the power of technology to tackle the challenges facing modern agriculture, we can pave the way for a future where food security is not just a goal but a reality for all.

About the Dataset

Title	New Plant Disease Dataset
Link	https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset
About	• This dataset is recreated using offline augmentation from the original dataset. • The images cover a wide range of plant species, including fruits, vegetables, and grains, afflicted by common diseases such as blight, rust, mold, and wilting. • Each image is meticulously labelled with the corresponding disease type or marked as healthy, enabling supervised learning approaches for disease detection and classification tasks.
Size	Total Images: 87K RGB Train: 70K Valid: 17K No. of Classes: 38
Features	• High-resolution images (minimum 256x256 pixels) captured under diverse lighting and environmental conditions. • Labelled annotations indicating the presence of specific diseases or the healthy status of the plants. • Images covering a wide range of plant species, including fruits, vegetables, and grains, from multiple geographical regions. • A balanced distribution of diseased and healthy samples, ensuring unbiased model training and evaluation.

Implementation

1. Data Pre-processing:

- Pre-processed the collected images to prepare them for model training.
- Applied techniques such as resizing, normalization, and augmentation to enhance the dataset's diversity and quality.

2. Model Development:

- **Model Architecture:**
    - ➢ Utilized a Convolutional Neural Network (CNN) architecture for its efficacy in extracting spatial features from images, crucial for accurate classification tasks.

- **Building Blocks:**
    - Incorporated key components including Convolutional Layers, Pooling Layers (e.g., Max Pooling), Activation Functions (ReLU-Internal layers, Softmax-Output layer), Flatten Layer, and Fully-Connected Layers.
    - Also used Dropout Layers to prevent overfitting.
    - Each layer played a specific role in processing and transforming the input data, ultimately leading to precise disease classification.

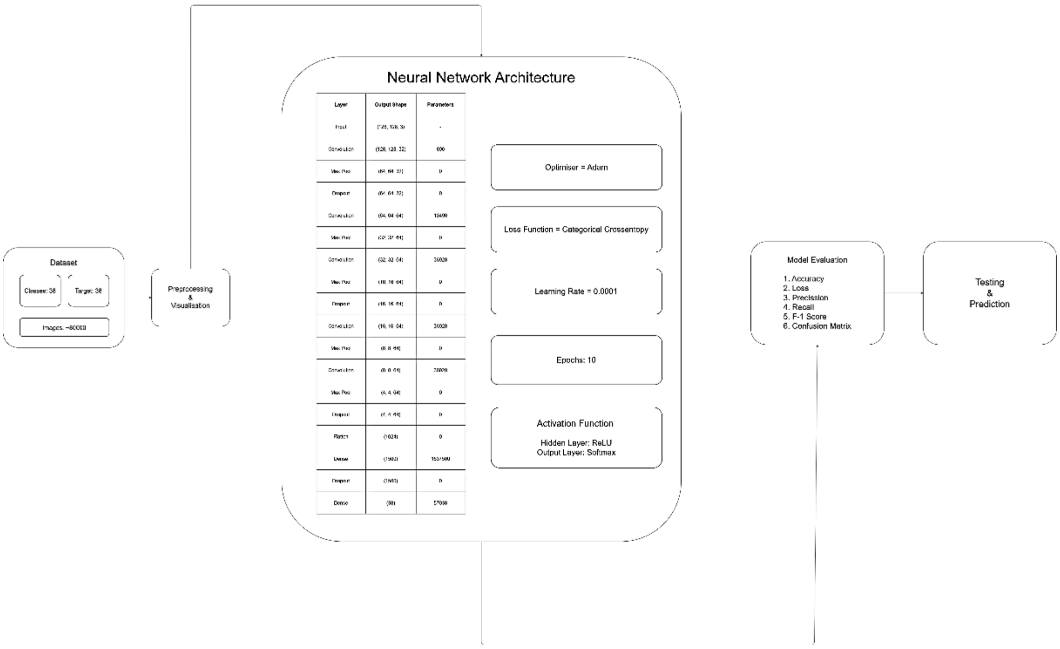
3. **Training Process:**

- **Data Splitting:**
    - Divided the pre-processed data into distinct sets for training, validation, and testing.
    - Utilized the training set for model training, the validation set for monitoring training progress, and the test set for final model evaluation.
- **Model Learning:**
    - Implemented an iterative training process where the model adjusted its internal weights based on the training data and the categorical cross-entropy loss function.
    - Leveraged the Adam optimizer to guide weight adjustments, aiming to minimize the loss function and enhance classification accuracy.

4. **Evaluation and Deployment:**

- **Model Evaluation:**
    - Conducted comprehensive evaluation post-training by assessing the model's performance on the held-out test set.
    - Utilized standard evaluation metrics such as accuracy, precision, recall, and F1-score to measure classification effectiveness.
- **Deployment:**
    - Upon achieving satisfactory performance, deployed the trained model for real-world application.
    - Integrated the model into user-friendly platforms such as mobile applications or web services, enabling convenient and on-demand plant disease prediction capabilities.

Flow Chart



Results and discussion

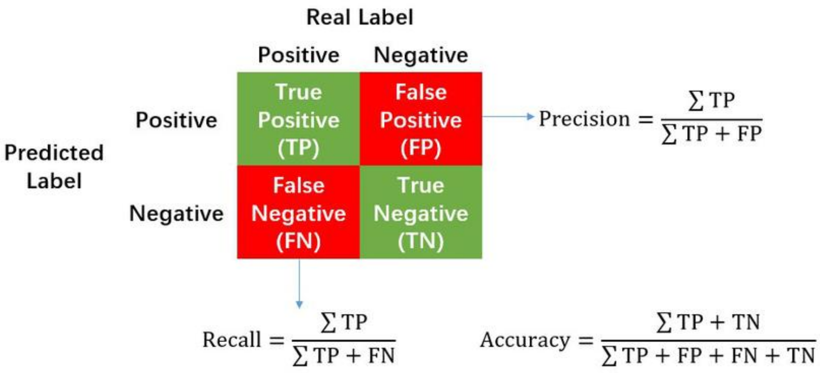
- **Accuracy Metrics:** Upon evaluating the trained model on the test set, achieved a training accuracy of **94.5%** and validation accuracy of **92.18%**. This indicates the proportion of correctly classified plant leaves out of the total tested samples.

```
💡 Click here to ask Blackbox to help you code faster
# Training Accuracy
train_loss, train_acc = cnn.evaluate(training_set)
print('Training accuracy:', train_acc)
                                                                  Python
2197/2197 [==============================] - 54s 24ms/step - loss: 0.1681 - accuracy: 0.9456
Training accuracy: 0.945600688457489
```

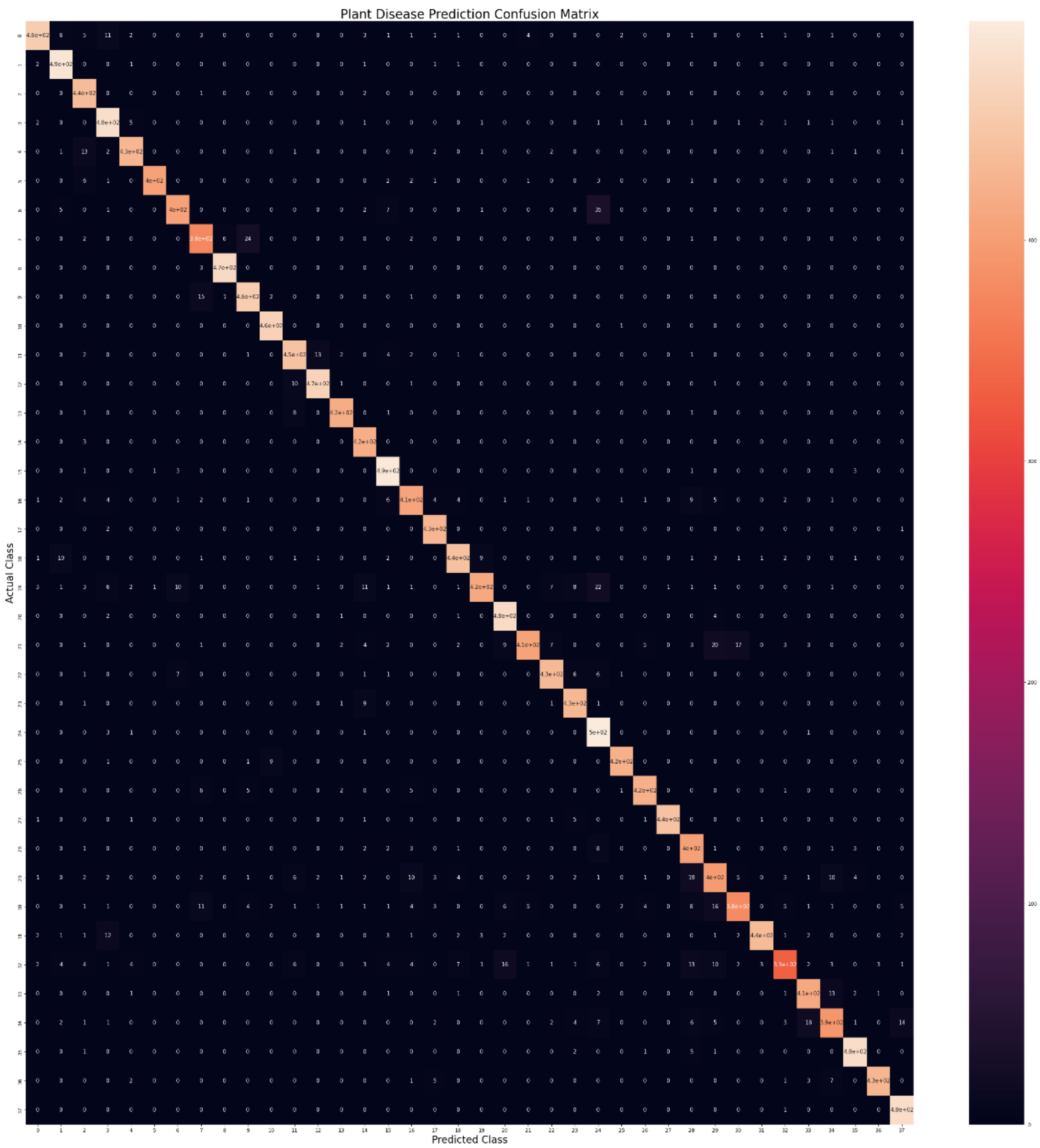
```
💡 Click here to ask Blackbox to help you code faster
# Validation Accuracy
val_loss, val_acc = cnn.evaluate(validation_set)
print('Validation accuracy:', val_acc)
                                                                  Python
550/550 [==============================] - 14s 25ms/step - loss: 0.2447 - accuracy: 0.9218
Validation accuracy: 0.9218074083328247
```

- **Precision and Recall:** Precision measures the proportion of true positive predictions out of all positive predictions, while recall measures the proportion of true positive predictions out of all actual positives. Achieved weighted average precision value of 0.93 and weighted average recall value of 0.92, highlighting the model's ability to make accurate predictions while minimizing false positives and false negatives.

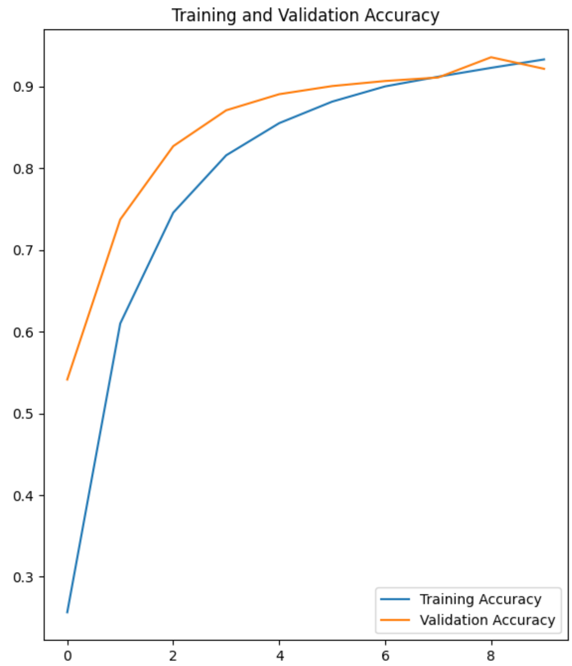
- **F1-Score:** Calculated the harmonic mean of precision and recall to obtain the F1-score, which serves as a balanced measure of the model's overall performance. Achieved an weighted average F1-score of 0.92, indicating a robust balance between precision and recall.

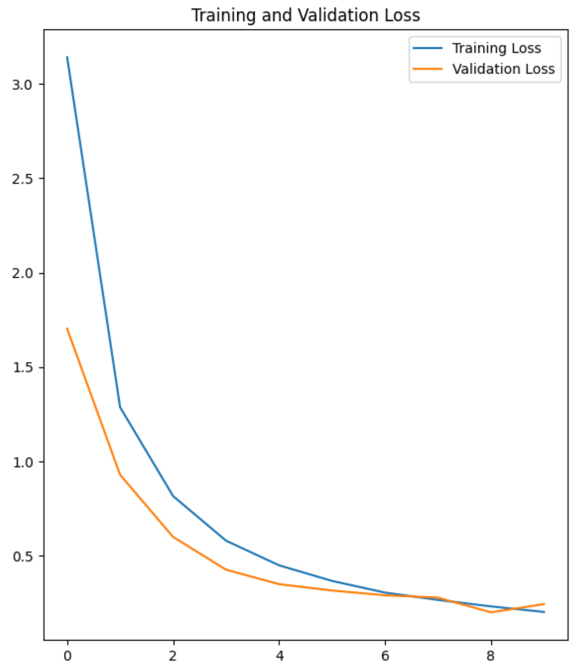


Confusion Matrix



Visualisation: Training and Validation - Accuracy and Loss





Future scope

In the realm of plant disease classification, future research holds promise for significant advancements. Enhancing model architectures beyond traditional CNNs to incorporate attention mechanisms, graph neural networks, or transformers could lead to improved accuracy and robustness. Integrating multi-modal data, such as infrared or hyperspectral imaging, alongside RGB images, offers a more comprehensive view of plant health. Transfer learning and domain adaptation techniques can optimize model generalization across diverse geographical regions and plant species. Active learning and semi-supervised approaches enable efficient utilization of labelled and unlabelled data for model training.

Developing explainable AI methods enhances trust and understanding among stakeholders. Real-time monitoring through IoT integration facilitates timely intervention in agricultural fields. Collaboration and data sharing initiatives promote knowledge exchange and reproducibility. Finally, user-centric design principles ensure that AI solutions meet the specific needs of farmers and agricultural practitioners. By exploring these avenues, we can advance the field, leading to more effective and sustainable solutions for plant disease management and crop protection.