Guava Disease Classification Using Transfer Learning with MobileNetV2

Project Overview

This project implements a machine learning pipeline to classify guava leaves into three categories: **Anthracnose (disease)**, **Fruit Fly (disease)**, and **Healthy Guava**, using transfer learning with MobileNetV2. The goal is to create an efficient and accurate model that can help detect guava diseases early, potentially aiding in better crop management and reducing agricultural losses.

The model leverages **TensorFlow** and **Keras**, applying **transfer learning** to adapt the pre-trained MobileNetV2 for this multi-class classification problem.

Dataset

The dataset consists of images categorized into three classes:

- **Anthracnose**: Guava leaves affected by fungal infection.
- **Fruit Fly**: Leaves with visible damage caused by fruit flies.
- **Healthy Guava**: Leaves without signs of disease.

The dataset is split into three subsets:

1. **Training Set**: 2647 images
2. **Validation Set**: 755 images
3. **Test Set**: 382 images

All images are resized to 224x224 pixels to match the input requirements of MobileNetV2.

Methodology

1. Data Augmentation

To make the model more robust and avoid overfitting, data augmentation techniques like rotation, zoom, shear, and brightness adjustments are applied to the training data. The pixel values are normalized to a range of `[0, 1]` to match the requirements of MobileNetV2.

2. Transfer Learning with MobileNetV2

We utilized **MobileNetV2**, a lightweight and efficient convolutional neural network pre-trained on the ImageNet dataset.

- The pre-trained layers of MobileNetV2 are frozen to preserve the learned features.
- A custom classification head is added, including:
    - **Global Average Pooling** to reduce dimensions.
    - **Dense Layers** for learning disease-specific features.
    - **Dropout** to prevent overfitting.
    - A final **Softmax Layer** for multi-class classification.

3. Model Optimization

The model is compiled with:

- **Adam Optimizer**: Adaptive learning rate for faster convergence.
- **Categorical Crossentropy Loss**: Suitable for multi-class classification.
- Metrics: **Accuracy** to measure performance.

Two callbacks were employed:

- **ReduceLROnPlateau**: Dynamically reduces the learning rate when the validation loss plateaus.

- **Early Stopping**: Stops training when validation loss stops improving, restoring the best model weights.

4. Model Training

The model was trained for **20 epochs**, with early stopping ensuring optimal convergence.

Performance Metrics

Accuracy

- Training Accuracy: 96.88%
- Validation Accuracy: 98.64%
- Test Accuracy: 98.01%

Confusion Matrix

A confusion matrix demonstrates the model's ability to correctly classify each category:

Classification Report

Class	Precision	Recall	F1 - Score
Anthracnose	0.99	0.99	0.99
Fruit Fly	0.96	0.98	0.97
Healthy Guava	0.98	0.95	0.96

Overall Metrics:

- Accuracy: 98%
- Macro Average F1-Score: 98%
- Weighted Average F1-Score: 98%

Sample Images