Healthy vs Rotten Fruit & Vegetable Classification Project

1. Dataset Overview

The dataset used in this project is titled 'Fruit and Vegetable Disease (Healthy vs Rotten)'.

It contains 28 classes, representing both healthy and rotten categories for 14 different types of fruits and vegetables.

The data is sourced from Kaggle and consists of high-quality images in JPG, PNG, and WEBP formats.

Dataset Size: ~4.7 GB

Total Classes: 28

Use Cases: Image classification, deep learning, transfer learning

This dataset is well-suited for training models for agricultural quality control and food safety detection.

2. Data Augmentation

Data augmentation is a technique used to artificially expand the size of a training dataset by applying transformations like rotation, scaling, flipping, brightness changes, and more. This increases model robustness and helps prevent overfitting.

In this project, data augmentation was initially planned to be used, but it was skipped because the images in the dataset were already preprocessed and cropped. Despite skipping augmentation, model accuracy was not significantly impacted.

However, skipping augmentation may lead to slightly longer training times.

Example transformations include:

- Rotation

- Width/Height Shift

- Horizontal Flip

- Brightness Adjustment

- Zoom

These techniques improve generalization by introducing synthetic variation to the training data, allowing the

model to perform better on unseen inputs.

3. Visualization of Augmented Images

You can visualize data augmentation by using the provided Python script (data_augmentation_example.py).

Just replace the placeholder path with an actual image file and run the script to see 10 transformed images.

This helps you understand how diverse your training data becomes with augmentation.

Tool used: TensorFlow Keras' ImageDataGenerator.