Edge AI Prototype – Project Report

Using TensorFlow Lite for On■Device Image Classification

Author: Nolin Masai Wabuti

1. Project Overview

The purpose of this project is to create a lightweight Edge AI image classification system using TensorFlow and TensorFlow Lite. The model classifies fruit images and is optimized for deployment on devices such as the Raspberry Pi, allowing fast, offline, and private inference.

2. Dataset Description

The Fruits_1 dataset from Kaggle contains categorized fruit images. The dataset was resized to 128x128 pixels and normalized. An 80/20 split was used for training and validation.

3. Model Architecture

A compact Convolutional Neural Network (CNN) was used with three convolutional blocks, max pooling, flattening, and dense layers. The model has ~826k parameters and is optimized for real■time edge deployment.

4. Training Results

The model trained for 3 epochs and achieved outstanding performance:
• **Validation Accuracy:** 1.0000
• **Validation Loss:** 0.0000 These results indicate excellent generalization under the provided dataset.

5. TensorFlow Lite Conversion

The Keras model was converted to TensorFlow Lite using optimization settings, producing a lightweight .tflite model ideal for resource■constrained devices.

6. Benefits of Edge AI

• **Real■time inference:** No delay caused by cloud communication. • **Privacy:** Images never leave the device. • **Offline capability:** Works without internet. • **Lower cost:** No cloud compute or bandwidth fees. These advantages make Edge AI crucial in smart agriculture, robotics, healthcare devices, and industrial quality control.

7. Deployment Steps (Raspberry Pi)

1. Install TensorFlow Lite runtime. 2. Load the .tflite model. 3. Preprocess the input image. 4. Run inference. 5. Display or use the prediction in real■time applications.

8. Conclusion

This project successfully demonstrates an end■to■end pipeline for training, converting, and deploying an Edge AI model. The optimized TensorFlow Lite model enables fast and efficient on■device inference, making it suitable for real■world applications.