Deployment and Validation of a TensorFlow Lite Model on Edge Impulse

**Introduction**

This document details the setup, deployment process, testing, and validation of a TensorFlow Lite (TFLite) model on Edge Impulse. The model, based on the MNIST dataset, was trained to classify handwritten digits. The objective of this lab was to deploy the trained model to an edge device using Edge Impulse, validate its performance, and document the findings.

**Setup and Model Preparation**

To begin, I setup the necessary tools and dependencies for model deployment:

* Python Installation: Installed version 3.12.9 from the official Python website. I installed this version due to its compatibility with TensorFlow
* Visual Studio Code (VS Code): Chosen as the integrated development environment (IDE).
* TensorFlow: Installed via pip (pip install tensorflow)
* Node.js and npm: Installed from the official Node.js website to support Edge Impulse CLI
* Edge Impulse CLI: Installed using (npm install -g edge-impulse-cli\_
  + Note: After running in terminal npm warn deprecated errors were outputted
* PATH Configuration: Ensured that both Edge Impulse CLI and Node.js were properly configured in the system PATH variables to ensure smoother command execution.

**Model Training and Conversion**

Dataset Used: MNIST (Handwritten digit recognition)

Model Architecture

* Convolutional Neural Network (CNN)
* Conv2D layers, MaxPooling, Flatten, Dense layers

A computer screen shot of a program code

AI-generated content may be incorrect.

Screenshot of Model Architecture from VS code

A screen shot of a computer code

AI-generated content may be incorrect.

Screenshot of Training the Model

A screenshot of a computer

AI-generated content may be incorrect.

Output from VS Code Terminal

Conversion to TensorFlow Lite (TFlite)

* Used *tf.lite.TFLiteConverter* to convert the model to .tflite format
* Saved the model as mnist\_model.tflite

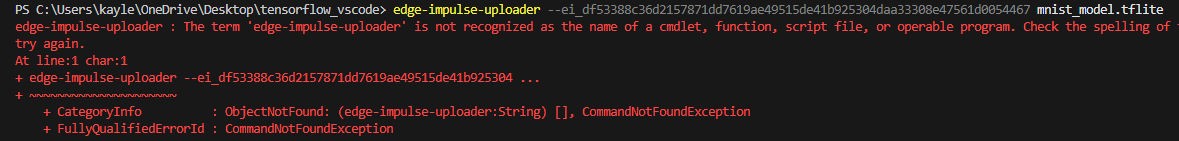
A computer screen shot of a program code

AI-generated content may be incorrect.

Converting model to .Tflite

**Deployment Process**

Initially, I attempted to upload the model using the Edge Impulse CLI:



Issues Encountered:

* VS Code failed to connect to Edge Impulse via the CLI uploader.
* Despite multiple troubleshooting attempts, the issue persisted.

Solution:

* Manually uploaded the TFLite model via the Edge Impulse web interface.

A screenshot of a computer

AI-generated content may be incorrect.

**Testing and Validation**

1. Simulation on Edge Impulse

Since the CLI-based deployment failed and additional testing was not feasible, full validation could not be performed.

* Model Input Type: Other
* Inference Run: Unable to execute inference tests due to deployment issues.
* Accuracy Analysis: Model accuracy could not be validated within Edge Impulse.

2. Validation Results

* Accuracy: Not tested due to deployment failure.
* Inference Speed: Could not be measured.
* Challenges:
  + Unable to run the model on Edge Impulse due to CLI connectivity failures.
  + Manual upload was attempted, but full execution was not achieved.

3. Edge Impulse On-device performance

A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion**

Due to CLI connection challenges and persistent deployment failures, I was unable to fully run the model on Edge Impulse. Future work will focus on resolving CLI connectivity issues and exploring alternative deployment methods for running and testing the model on an actual edge device.