National
College *of*
Ireland

Configuration Manual

MSc Research Project
Data Analytics

Darshan Vetal
Student ID: 23215194

School of Computing
National College of Ireland

Supervisor: Dr. Teerath Kumar

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Darshan Vetal
Student ID:	23215194
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr. Teerath Kumar
Submission Due Date:	20/12/2024
Project Title:	Configuration Manual
Word Count:	XXX
Page Count:	5

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Darshan Vetal
Date:	11th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Darshan Vetal
23215194

1 Introduction

This configuration manual provides detailed instructions to replicate the research process, including hardware and software setup, data directory organization, and the execution of the implemented models. The goal is to ensure ease of reproducibility for future researchers.

2 Pre-Requisites

Hardware Requirements:

1. A machine with a GPU-enabled environment is recommended for efficient computation. Minimum specifications:

   (a) OS: windows 10/11 (if running code locally)

   (b) CPU: Intel i7/i9 (10 cores or up) or AMD Ryzen 7 (10 cores or up).

   (c) GPU: NVIDIA Tesla T4 (Google Colab GPU).

   (d) RAM: 16GB or higher.

   (e) Disk Space: 30 GB

2. Cloud Integration: Data and models are stored in a Google Cloud Storage (GCS) bucket. Access credentials for GCS are required to download the datasets and saved model files.

3 Software Requirements

1. Python Environment: Python Version: 3.10 or higher. Pycharm or Anaconda Navigator is optional but can be used for managing environments.

2. Python Libraries: Install the following libraries using pip install or via Anaconda: numpy, pandas, scipy, matplotlib, seaborn, nibabel, sklearn, imblearn, torch, torchvision, flask, lazypredict, pytorch-grad-cam

3. Development Environments: Google Colab for model training and experimentation. Local Flask Application for deploying the trained model for predictions.
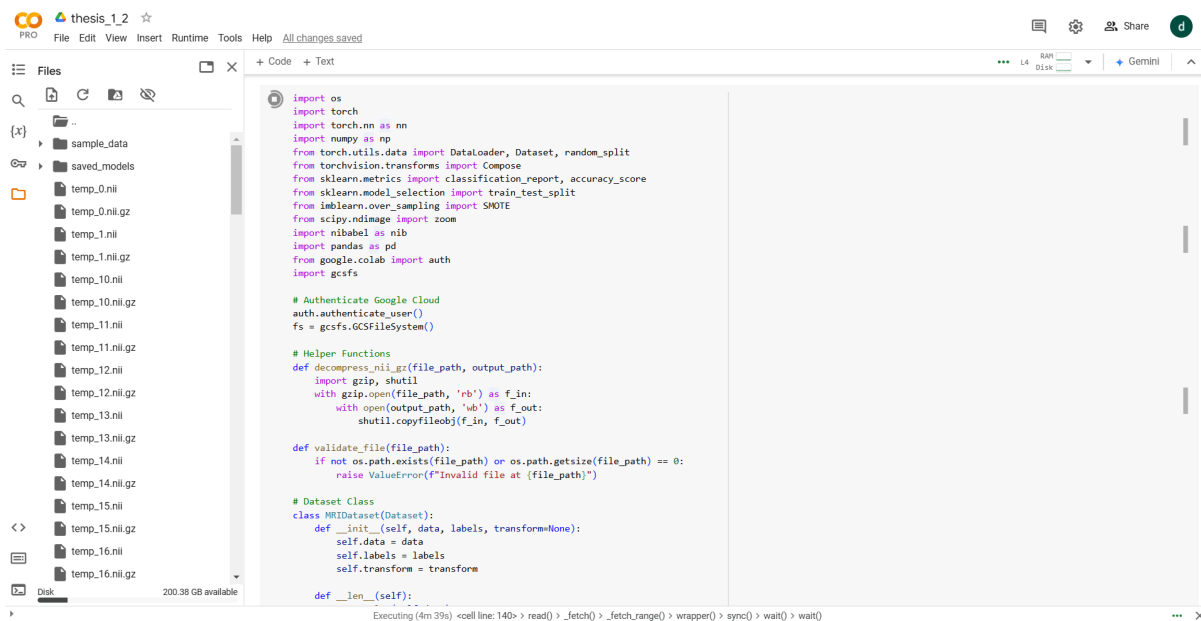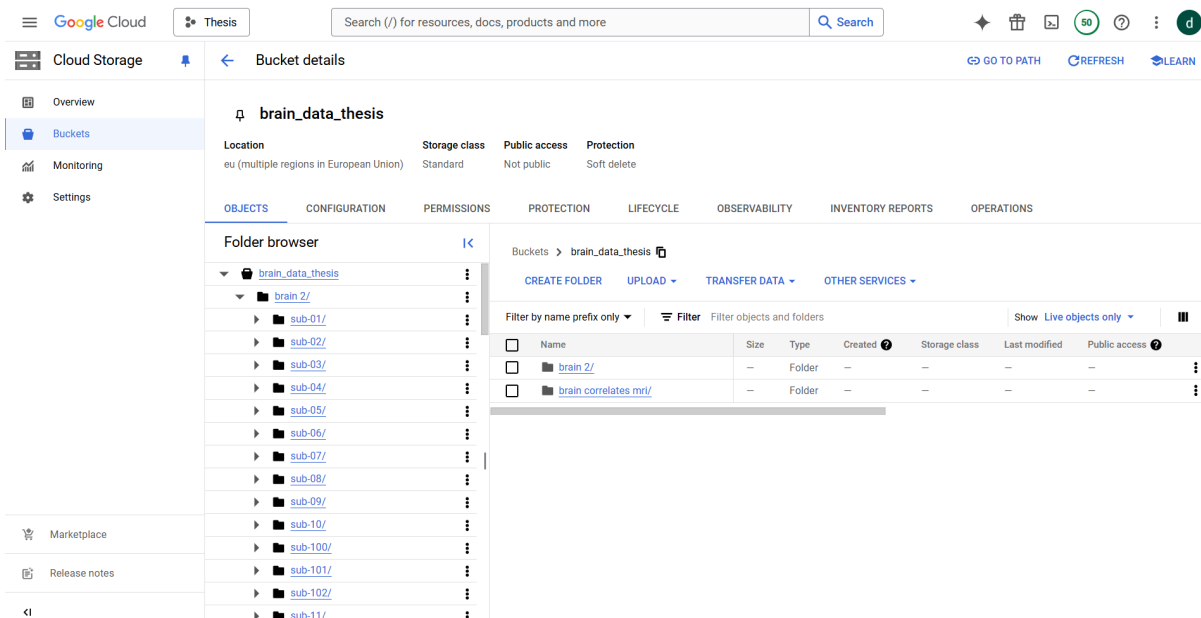
Figure 1: Google Collab Environvent

Figure 2: GCS Bucket

4 Setting Up the Environment

1. Google Colab Setup

   (a) Open Google Colab.

   (b) Upload or connect your notebook files to Colab.

   (c) Install the required Python libraries:
       !pip install -r requirements.txt

   (d) Mount Google Drive for storing datasets or output
       from google.colab import drive
       drive.mount('/content/drive')

   (e) Connect to a GPU runtime:
       Go to Runtime - Change runtime type and select GPU under Hardware Accelerator.
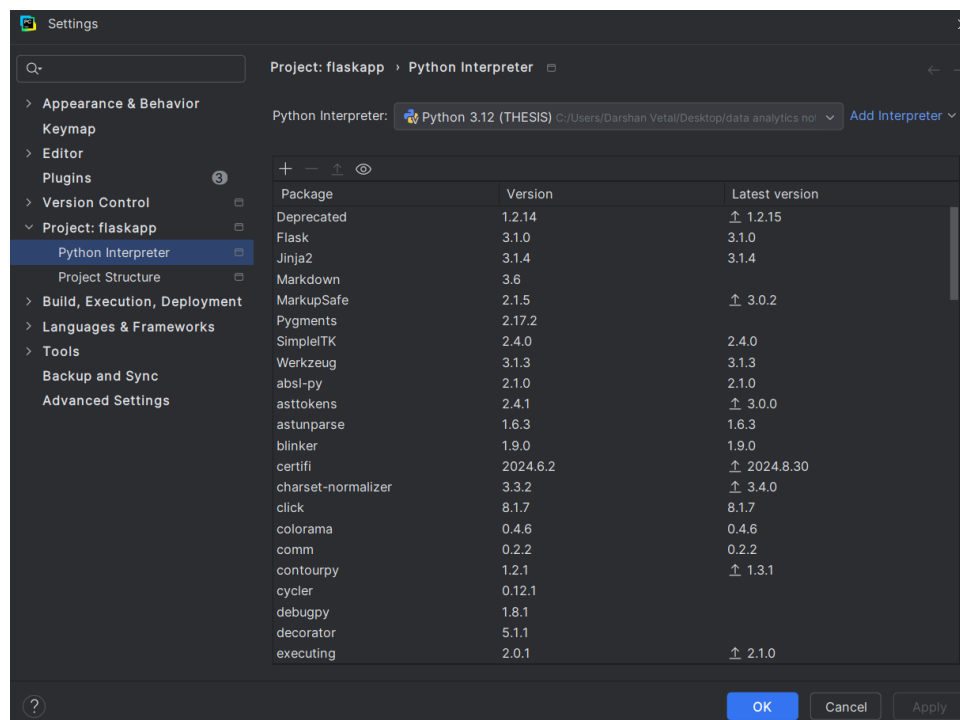
2. PyCharm with Virtual Environment

Figure 3: PyCharm Environment

   (a) Install PyCharm Community or Professional edition from JetBrains.

   (b) Create a new project or open the existing project.

   (c) Set up a virtual environment:

       i. Go to File - Settings - Project: Project Name - Python Interpreter.
       ii. Click on the gear icon and select Add....
       iii. Choose New Virtual Environment and specify the location.

(d) Install project dependencies:
    pip install -r requirements.txt

(e) Configure PyCharm for Flask app development:
    i. Go to Run - Edit Configurations....
    ii. Add a new Flask Server configuration.
    iii. Set the FLASK_APP variable to your main Flask script (e.g., app.py).

5   Directory Structure

Datasets:

1. Dataset 1: Schizophrenia OpenNeuro (2022)

2. Dataset 2: Healthy vs. Schizophrenia OpenNeuro (2018)

Directory Structure is as follows:

```
Thesis/
├── brain correlates mri/        # Dataset 1 (Schizophrenia: Hallucinators vs. Non-hallucinators)
├── brain 2/                     # Dataset 2 (Healthy vs. Schizophrenia)
├── saved_models/                # Trained models
├── notebooks/                   # Jupyter notebooks for data preprocessing, training, and evaluation
│   ├── thesis.ipynb             # CNN / CNN + SMOTE / CNN + ADASYN
│   ├── lazy_predict_code.ipynb  # Lazy Predict models
├── main.py                      # Flask application to run
├── dataset.py                   # Load Dataset
├── util.py                      # Defined Utilities
├── model.py                     # Load Model and Gradcams
├── requirements.txt             # Python dependencies
├── README.md                    # Project documentation
```

Figure 4: Folder Structure

6   Flask App Setup

To run the MRI classifier Flask based web application follow these replication steps:

1. Clone the repository:
   git clone https://github.com/x23215194/Thesis.git
   cd Thesis

2. Install dependencies:
   pip install -r requirements.txt

3. Run the Flask app locally:
   python main.py
   App runs by default on: http://127.0.0.1:5000/

References

OpenNeuro (2018). Working memory in healthy and schizophrenic individuals. Dataset available at `https://openneuro.org/datasets/ds000115/versions/00001`.

OpenNeuro (2022). Brain correlates of speech perception in schizophrenia patients with and without auditory hallucinations. Dataset available at `https://openneuro.org/datasets/ds004302/versions/1.0.1`.