

Configuration Manual

MSc Research Project
MSc Data Analytics

Jennifer Daniel
Student ID: X23268123

School of Computing
National College of Ireland

Supervisor: Teerath Kumar Menghwar

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: ...Jennifer Daniel.....

Student ID: ...x23268123.....

Programme: ...Msc Data Analytics..... **Year:**2025.....

Module:Research Practicum.....

Lecturer: Teerath Kumar Menghwar

Submission

Due Date:11-08-2025.....

Project Title: Audio Deepfake Detection with Hybrid CNN and Vision Transformer Models

Word Count:497..... **Page Count:**4.....

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:Jennifer Daniel.....

Date:11-08-2025.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
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.	<input type="checkbox"/>

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

Jennifer Daniel
Student ID: x23268123

1 Introduction

This configuration manual contains the configuration and procedure documentations for setting up and carrying out the MSc research project, "Audio Deepfake Detection with Hybrid CNN and Vision Transformer Models". The study is purely focused on the detection of synthetic speech by AI systems in clean environments as well as noisy environments. Models implemented are CNN, CRNN, PaSST (Patchout Spectrogram Transformer) and MobileViT. The datasets from sources include FakeAVCeleb for both the real and synthesized audio, as well as a Hospital Noise dataset for simulating real-world noisy conditions. The manual consists of hardware and software requirements, information for the dataset obtainment, how to configure the environment, how to execute the model and how to verify its correctness.

2 System Requirements

2.1 Hardware Requirements

- **Processor:** Intel Core i7 / AMD Ryzen 7 or higher
- **GPU:** NVIDIA RTX 3060 or higher (8 GB VRAM recommended)
- **RAM:** 16 GB minimum (32 GB recommended)
- **Storage:** Minimum 50 GB free space

2.2 Software Requirements

- **Operating System:** Windows 10/11, Ubuntu 20.04+, or Google Colab environment
- **Python Version:** Python 3.9+
-

3 Dataset Information

3.1 FakeAVCeleb

Source: <https://www.kaggle.com/datasets/noxolosizakele/deepfake-detection-in-audio?select=Audio+Deepfake+Detection>

The FakeAVCeleb dataset allows both authentic and synthetic audio samples. It includes a variety of speakers, accents, and styles to allow for thorough training of models. All files are in .wav format, and have been standardised to 5 seconds, and at a 16 kHz sample rate.

3.2 Hospital Noise Dataset

Source: <https://www.kaggle.com/datasets/nafin59/hospital-ambient-noise>

The Hospital Noise dataset contains authentic environment noise recordings from a health care setting, consisting of conversation, machines beeping, and ambient sound, which can be used to simulate a noisy environment when testing models.

4 Project Structure

```
audio-deepfake-detection/  
|— data/  
|   |— FakeAVCeleb/  
|   |— HospitalNoise/  
|— notebooks/  
|   |— CNN_and_CRNN.ipynb  
|   |— PaSST.ipynb  
|   |— MobileViT.ipynb  
|— utils/  
|   |— preprocessing.ipynb  
|   |— augmentation.ipynb  
|— results/  
|   |— accuracy_plots.png  
|   |— f1score_plots.png  
|   |— comparison_plots.png  
|— requirements.txt  
|— README.md
```

5 Configuration Settings

- **Audio length:** 5 seconds (clip/pad)
- **Sampling rate:** 16 kHz
- **Feature extraction:** Mel-spectrogram & MFCCs

- **Augmentation:**
 - RandAugment (CNN, CRNN)
 - Background noise addition (Hospital Noise dataset)
- **Batch size:** 32
- **Learning rate:** 0.001

6 Module Execution Guide

7.1 Running in Jupyter Notebook (Local)

1. **Activate your virtual environment** (if applicable):

```
source env/bin/activate    # Linux/Mac
env\Scripts\activate      # Windows
```

2. **Launch Jupyter Notebook:**

```
jupyter notebook
```

3. **Go to the notebooks/ folder in the Jupyter interface.**
4. **Click on the appropriate model:**
 - **CNN_and_CRNN.ipynb** → CNN and CRNN training (includes RandAugment & noisy data testing)
 - **PaSST.ipynb** → For the Patchout Spectrogram Transformer model
 - **MobileViT.ipynb** → For the Mobile Vision Transformer model

7.2 Running in Google Colab

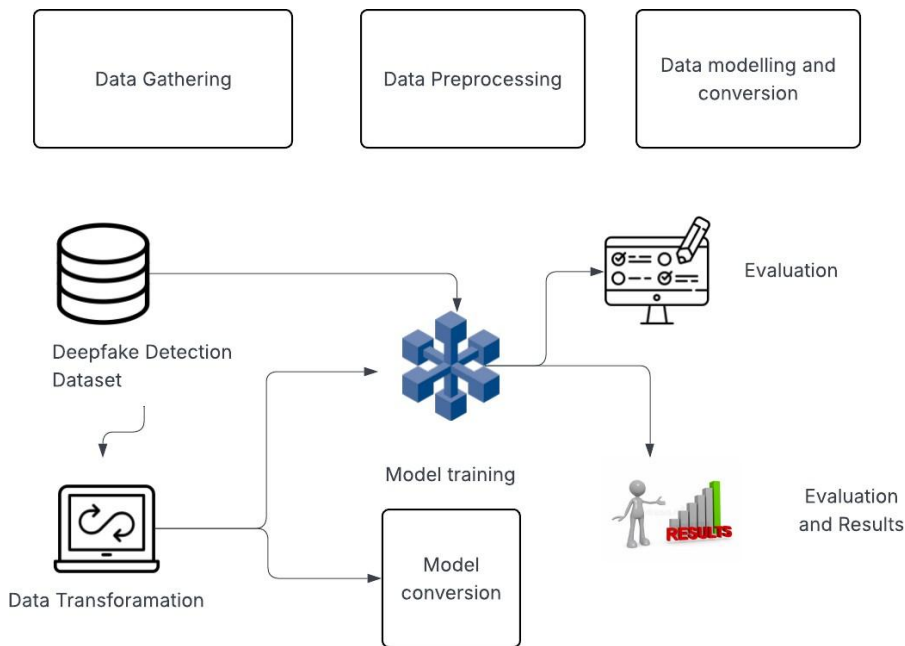
1. Upload .ipynb file to Google Colab (<https://colab.research.google.com>).
2. Upload or mount your datasets:
 - Mount Google Drive:

```
from google.colab import drive
drive.mount('/content/drive')
```

Place datasets in the mounted path.

3. Run all cells sequentially .

7 Flowchart



8 Expected Outputs & Verification

- **Metrics Generated:** Accuracy, Precision, Recall, F1-score
- **Example:**

Model	Augment	Dataset	Accuracy	Precision	Recall	F1-Score
CNN	NoAug	Clean	0.9889	0.9889	0.9889	0.9889
CNN	NoAug	Noisy	0.6643	0.7810	0.6643	0.5604
CRNN	NoAug	Clean	0.9830	0.9830	0.9830	0.9829
CRNN	NoAug	Noisy	0.6518	0.7758	0.6518	0.5350
CNN	RandAug	Clean	0.9885	0.9885	0.9885	0.9885
CNN	RandAug	Noisy	0.6603	0.7793	0.6603	0.5524
CRNN	RandAug	Clean	0.9890	0.9891	0.9890	0.9890
CRNN	RandAug	Noisy	0.6578	0.7782	0.6578	0.5474
MobileViT	--	Clean	0.9860	0.9863	0.9860	0.9859
MobileViT	--	Noisy	0.6843	0.7669	0.6843	0.6031
PaSST	--	Clean	0.9765	0.9766	0.9765	0.9764
PaSST	--	Noisy	0.6573	0.7199	0.6573	0.5553