National College of Ireland

Configuration Manual

Multimodal Depression Detection using Audio and Visual Features
MSc Artificial Intelligence

Sachleen Singh Chani
Student ID: x22244778

School of Computing
National College of Ireland

Supervisor: Dr. Muslim Jameel Syed

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sachleen Singh Chani
Student ID:	x22244778
Programme:	MSc Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr. Muslim Jameel Syed
Submission Due Date:	12/08/2024
Project Title:	Configuration Manual
Word Count:	647
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:	Sachleen Singh Chani
Date:	12th August 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):	

Contents

Configuration Manual

Sachleen Singh Chani
x22244778

1 Introduction

This configuration manual gives the outline for the steps needed to setup, configure and run the proposed deep learning multimodal model for depression detection. This can be run either on the local machine or a cloud environment like google Colab or Kaggle. For this project, python language has be utilized to preprocess and train the model, and the code is executed in Jupyter notebooks.

The research conducted is for the multimodal binary classification for depression detection. The two modalities used are video and audio features. The proposed model for this research is Bidirectional Long Short-Term memory model.

2 System Configuration

2.1 Hardware

2.1.1 Local Machine

- CPU: i5 6200U 2.30GHz
- GPU: Nvidia GForce GTX 940M
- RAM: 16GB DDR3

2.1.2 Google Colab

- TPU
- L4 GPU

2.2 Dependencies

The following are the necessary dependencies for this research:

- OS: Windows 11 64-bit
- VS Code: version 1.89.0
- VS Code Jupyter Extension: v2024.7.0
- Python: version 3.11.6

- Python Libraries:
  - pip 24.1.1
  - pandas 2.2.2
  - matplotlib version 3.9.0
  - scikit-learn 1.5.0
  - seaborn 0.13.2
  - tensorflow 2.16.1

3 Installation and Setup

3.1 VS Code Installation

Visit the Visual Studio Code download page on their official website and download the executable file for windows. After the installation, install Jupyter notebook extension and Python extension inside the VS Code IDE.

Optional: After the VS code has been installed, a virtual environment (.venv) can be created to maintain the version compatibility between the various libraries. Visit the vs code webpage for complete instructions. Following command can be used to create the environment:

```
python –m venv myvenv
```

3.2 Python Installation

Visit the Python download page and download and run the python executable. Follow the on-screen instructions to finish the installation.

3.3 Python libraries installation

In the VS Code terminal, run the following command:

```
pip install <library–name>
```

3.4 Google Colab

- Upload the project files to either google drive or directly to Colab
- Make sure the select the hardware accelerator (L4 or TPU) from the Colab's settings.

4 Dataset

The data used for this research is part of the Distress Analysis Interview Corpus (DAIC) (Gratch et al.; 2014). This data was collected to create a computer agent (Ellie) that could interview people and from the visual and audio cues identify the indicators of mental illness. The interviews were conducted by the animated virtual interviewer called Ellie, which was controlled by a human interviewer from another room. This process was conducted by University of South California.

Due to the sensitive nature of the data and consent constrains, it can not be shared. Please visit the website (https://dcapswoz.ict.usc.edu/) for more information.

5 Running the Project

To run the project, create a folder and extract the code base into the folder.

5.1 Creating the datasets

First the labels dataset needs to compiled. This is done using the 'labels.ipynb'. This code will output the PHQ8 labels for each participant.

Execute the jupyter notebook named 'dataset.ipynb'. This file will perform the data preprocessing and will create two csv files for the audio and video features.

5.2 Multimodal Model

Execute the jupyter notebook named 'audio+video.ipynb'. This code will read the two csv files generated earlier, construct the multimodal model, and train the model. Further, the 'Evaluation' section inside the notebook will give the evaluation results.

5.3 Additional Code (Single Modality)

For the experimentation on the single modality, the jupyter notebook likes are present in a separate folder named 'Additional Code (Single Modality)'.

This folder contains:

Dataset files:

- dataset_video.ipynb
- dataset_audio.ipynb

Model training files:

- video.ipynb
- audio.ipynb

Executing the dataset files above in the given order, will create the individual datasets for single modal models. Further executing the 'video.ipynb' and 'audio.ipynb' will train the chosen Machine learning models and output the evaluation results.

References

Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S., Traum, D., Rizzo, S. and Morency, L.-P. (2014). The distress analysis interview corpus of human and computer interviews, *in* N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk and S. Piperidis (eds), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 3123–3128.
URL: *http://www.lrec-conf.org/proceedings/lrec2014/pdf/508_Paper.pdf*