

# Configuration Manual

MSc Research Project  
MSC Data Analytics

Aniket Rane  
Student ID: x23299720

School of Computing  
National College of Ireland

Supervisor: Prof. Shubham Subhnil

National College of Ireland  
Project Submission Sheet  
School of Computing



|                             |                       |
|-----------------------------|-----------------------|
| <b>Student Name:</b>        | Aniket Rane           |
| <b>Student ID:</b>          | x23299720             |
| <b>Programme:</b>           | MSC Data Analytics    |
| <b>Year:</b>                | 2025                  |
| <b>Module:</b>              | MSc Research Project  |
| <b>Supervisor:</b>          | Prof. Shubham Subhnil |
| <b>Submission Due Date:</b> | 28/01/2026            |
| <b>Project Title:</b>       | Configuration Manual  |
| <b>Word Count:</b>          | 678                   |
| <b>Page Count:</b>          | 3                     |

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.

|                   |                   |
|-------------------|-------------------|
| <b>Signature:</b> |                   |
| <b>Date:</b>      | 22nd January 2026 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

|                                                                                                                                                                                            |                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| Attach a completed copy of this sheet to each project (including multiple copies).                                                                                                         | <input type="checkbox"/> |
| <b>Attach a Moodle submission receipt of the online project submission</b> , to each project (including multiple copies).                                                                  | <input type="checkbox"/> |
| <b>You must ensure that you retain a HARD COPY of the project</b> , 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.

| <b>Office Use Only</b>           |  |
|----------------------------------|--|
| Signature:                       |  |
| Date:                            |  |
| Penalty Applied (if applicable): |  |

# Configuration Manual

Aniket Rane  
x23299720

## Abstract

This configuration manual gives detailed information about the required tools, libraries, and settings for ensuring reproducibility of the results. It outlines the software and hardware requirements in a structured and easily understandable manner.

## 1 System Requirements

The experiments and codes were run on a personal Windows laptop with the following hardware and software configurations

### 1.1 Hardware Requirements

- Operating System: Windows or Linux
- Processor: Intel Core i5 10th Gen or Higher (Tested on Intel Core Ultra 7 155H)
- Cores: Minimum 6 core CPU for good performance (Tested on 16 Core CPU)
- RAM: Minimum 16GB RAM
- Storage: Atleast 20GB of free disk space for storing dataset and model files
- GPU: Minimum 8GB GPU for good performance (Tested on NVIDIA GeForce RTX 4060 Laptop GPU with 8GB VRAM)

### 1.2 Software Requirements

- Python version 3.10 or higher (Tested on 3.12.3)
- Jupyter notebook or Google Colab required to run the code file
- CUDA: Minimum CUDA version 11.8 or higher (Tested on CUDA version 12.1)

## 2 Extracting Files from Zip

- Download the provided .zip file
- Extract the files present in the zip folder
- Place the extracted files at the desired folder path

- The zip file also contains a word file with the link to the dataset. This link can be used to download the dataset as the dataset size is 6.07GB in total.

## 3 Environment Setup

In order to run the code file, an anaconda environment needs to be setup with all the required libraries and python packages along with the dependencies. For doing this, an environment creation file named **cancer-classification.yml** containing all the required packages has been provided in the .zip file. Just simply run the following commands given below in anaconda prompt after extracting the .yml file from the zip. Before running the command for environment creation ensure that the working directory for anaconda is set to the folder location in which you have placed the cancer-classification.yml file.

**Note:** Expect the environment setup to take upto 15 to 30 minutes in total.

```
conda env create -f cancer-classification.yml
```

After creating the environment, activate it using the command given below:

```
conda activate cancer-classification
```

## 4 Running the code

After setting up the environment and activating it, open Jupyter notebook to run the code file.

### 4.1 Dataset directory

After opening Jupyter notebook, ensure that the working directory is set to the location where you have saved the provided dataset files downloaded from the given link. As given in the code line below, you can replace the path with the path where you have stored the dataset files.

```
cd "Path where the dataset is stored"
```

Also the 'ls' command could be used to ensure you are in the right directory. Once the path following cd is correctly replaced, simply run all cells.

### 4.2 Checking if GPU is being used by torch

In order to ensure that the models are being trained on the GPU, run the commands given below:

```
print(torch.cuda.is_available())
print(torch.cuda.get_device_name(0) if torch.cuda.is_available() else "CPU only")
print(torch.version.cuda)
```

After running the above check, expect to see an output similar to the one given below:

---

```
True
NVIDIA GeForce RTX 4060 Laptop GPU
12.1
```

### 4.3 Setting the seed

To ensure reproducibility, random seed was set using the following commands:

```
seed = 42
set_seed(seed, reproducible=True)
torch.manual_seed(seed)
np.random.seed(seed)
random.seed(seed)
```

### 4.4 Tensorboard logging

Tensorboard logging was utilized to store the various charts under Exploratory data analysis (EDA) and model training metrics (losses, accuracy, etc) under conventional training. The Tensorboard could be accessed right from the Jupyter notebook by using the following commands:

```
%load_ext tensorboard

%tensorboard --logdir=runs/dataset_eda
-- To view the charts for EDA

%tensorboard --logdir=runs/cancer_resnet34
-- To view the logs for ResNet34

%tensorboard --logdir=runs/cancer_efficientnet_b2
-- To view the logs for EfficientNetB2

%tensorboard --logdir=runs/cancer_vit_b16
-- To view the logs for ViT_Base_Patch16_224

%tensorboard --logdir=runs/cancer_deit_small
-- To view the logs for DeiT_Small_Patch16_224
```

**Note:** The command to access Tensorboard logs has been hashed out in the main code file. To access the tensorboard logs remove the hash and use any of the codes provided above to access the desired log.

These are the instructions that need to be followed in order to run the code file.

## References