A picture containing graphical user interface

Description automatically generated **Configuration Manual**

MSc Research Proposal

Research in Computing

Akshen Doke

Student ID: x18191592

School of Computing

National College of Ireland

Supervisor: - Prof. Rashmi Gupta

A picture containing text, clipart

Description automatically generated

**National College of Ireland**

**MSc Project Configuration Manual**

**School of Computing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Name:** | Akshen Doke | | |
| **Student ID:** | x18191592 | | |
| **Programme:** | MSc in Data Analytics | **Year:** | 2019-2020 |
| **Module:** | Research Project | | |
| **Supervisor:** | Prof. Rashmi Gupta | | |
| **Submission Due Date:** | 17th December 2020. | | |
| **Project Title:** | Image Classification: Optimizing and Benchmarking contemporary deep neural networks over chest x-ray images. | | |
| **Word Count:**4413 | **Page Count** 20 | | |
|  |  | | |

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:** | ……………………………………………………………………………………………………………… |
| **Date:** | 17h December 2020 |

**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): |  |

1. **Hardware and Software Requirements**

For this project, all compute intensive tasks like modelling, data visualization and prediction was done on a cloud service called Google Colab which was accessed using a MacBook Air. Only the data downloaded from data source was converted from jpeg to png and was renamed on local device (MacBook Air) before uploading it to the cloud using bash program.

**Cloud Setup Option**

|  |  |
| --- | --- |
| Processor | page13image60933248page13image38216000  On-demand |
| Graphic Card | page13image60947072page13image60947456  TPU and GPU option available  page13image60948032page13image60937856 |
| RAM | Min 8Gb-Max 32GB |
| HDD | 12GB free space |

Bash scripts for data format changing and renaming.

Graphical user interface, text, application

Description automatically generated

1. **Google Collaboratory (Colab) Setup**

Since this research was carried out using Google Colab’s Cloud infrastructure, we need to first upload our dataset to Google drive which can be connected to our notebook (code platform of colab) were we are going to code and use the data.

Graphical user interface, application

Description automatically generated

We create three folders, one in which we are going to store our training data, second our test data and third for the models on which the training is going to be happening.

As mentioned in (Google, n.d.) Google Colab is an Infrastructure and Software as a Service free to use provided by Google for tasks related to machine learning, data analytics and artificial intelligence in python and its related libraries.

To mount the drive to our notebook we use the code given below

Graphical user interface, application

Description automatically generated

After our drive is mounted successfully we can set paths for our train and test files, also import the required libraries and functions for our project.

Text

Description automatically generated

To get maximum speed and utilization of our notebook we change our runtime to GPU from None, this will make our program execution faster while we train and run our predictions on the dataset.

Graphical user interface, application

Description automatically generated

**3. Data Preparation and Visualization**

Next, we set paths for our train and test datasets in the respect variable names.

Text

Description automatically generated

Now we need to calculate the overall count of each set of images and represent it visually for that we use python based library called matplotlib

Text

Description automatically generated

The output of Data spread which we get is

Chart, pie chart

Description automatically generated

As we can see the count of covid is relatively low, in order to balance this out we would be using data augmentation techniques while training our model.

1. **Implementation of Base Models**

Since we are going to make use of ensemble methods for prediction of our model, we would be training around 7 models using which we would be performing the ensemble based prediction.