Configuration Manual

MSc Research Project
MSc Data Analytics

Gopi Krishna Inampudi
Student ID: X20182708

School of Computing
National College of Ireland

Supervisor: Taimur Hafeez

Student Name:	Gopi Krishna Inampudi		
Student ID:	X20182708		
Programme:	MSc Data Analytics	**Year:**	2022…………………………..
Module:	Research Project		
Lecturer:	Taimur Hafeez		
Submission Due Date:	15-08-22		
Project Title:	Using Chest X-ray Images to diagnose and distinguish COVID-19 Pneumonia, Viral Pneumonia and Lung Opacity		
Word Count:	970	**Page Count:** 9	

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.

<u>ALL</u> 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: ……

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

Gopi Krishna Inampudi
Student ID: X20182708

1 Introduction

Complete instructions for installing the necessary software and gathering the various parts and pieces of hardware for putting together the system from the blueprints are provided in this manual. Using a rational approach, the Configuration Manual will provide a duplicate of the research work that has already been done. Here, the interface is coupled with all of the code's evaluations.

The Configuration Manual will contain 6 parts, they are as follows:
1. Environmental Setup
2. Libraries
3. Dataset Description
4. Interface
5. Evaluations
6. Code Description

2 Environment Configuration

2.1 Hardware Configuration

RAM: 8GB
SSD: 256 GB
Processor: intel core i5
210 Mhz

2.2 Software Requirement

Windows 10
Python 3.8.9

3   Packages and Libraries

Pandas

For the purpose of manipulating and analysing data, the Python programming language has a software package called pandas. It includes specific data structures and procedures for working with time series and mathematical tables. Pandas was used to convert our dataset to data frame.

NumPy

Large, multi-dimensional arrays and matrices are supported by NumPy, a library for the Python programming language, along with a substantial number of high-level mathematical operations that may be performed on these arrays. NumPy was used to convert our dataset to NumPy array in order to perform some pre-processing.

Matplotlib

The Python programming language and the NumPy library for numerical mathematics are supported by the Matplotlib graphing tool. With its object-oriented API, plots may be easily integrated into software. Matplotlib was used for plotting graphs for our exploratory data analysis

Seaborn

A matplotlib-based Python data visualization library is called Seaborn. It offers a sophisticated drawing tool for creating eye-catching and educational statistical visuals. Seaborn: Seaborn was used for plotting graphs of our EDA as well.

Glob

All pathnames that fit a given pattern are located by the glob module. Glob was used to add the file path of our dataset and load them to pandas as data frame.

Scikit Learn

It offers a suite of powerful applications for statistics and machine learning. Scikit Learn: The Scikit learn library was used to run our classification reports, confusion matrix, split the dataset into training and testing.

TensorFlow

Google's TensorFlow is a free package optimized for deep learning use cases. TensorFlow was used to train our deep learning model.

Random

The Python Random library is a built-in function for producing random numbers. We used the python random library to generate random samples of the dataset.

4   Dataset

Research teams from Qatar University in Doha and the University of Dhaka in Bangladesh, along with collaborators from Pakistan and Malaysia, assembled the data. 3616 COVID-19 positive cases, 10,192 normal, 6012 non-COVID lung infection, and 1345 viral pneumonia images are included in the collection as well each classes masked images. The dataset is freely available on Kaggle

https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

5   Evaluation and Code

The Jupyter notebook with all the artefact have been attached with this document. Anaconda3 2021.11 and python 3 (64-bit) were used to implement the study project, both of which were published by Anaconda, Inc. of version 2021.11.

```
In [5]: train_set , test_set = train_test_split(data,test_size=0.2,random_state=17)
        train_set.shape,test_set.shape

Out[5]: ((16932, 2), (4233, 2))
```

Fig 1 training and test split

An overview of the models' test and train data is shown in Fig 1. The data is split 80:20 between the training and testing

```
In [14]: print(classification_report(test_data.labels,y_pred))

                 precision    recall  f1-score   support

             0       0.95      0.77      0.85       724
             1       0.78      0.93      0.85      1197
             2       0.93      0.88      0.90      2061
             3       0.79      0.85      0.82       251

      accuracy                           0.87      4233
     macro avg       0.86      0.86      0.85      4233
  weighted avg       0.88      0.87      0.87      4233
```
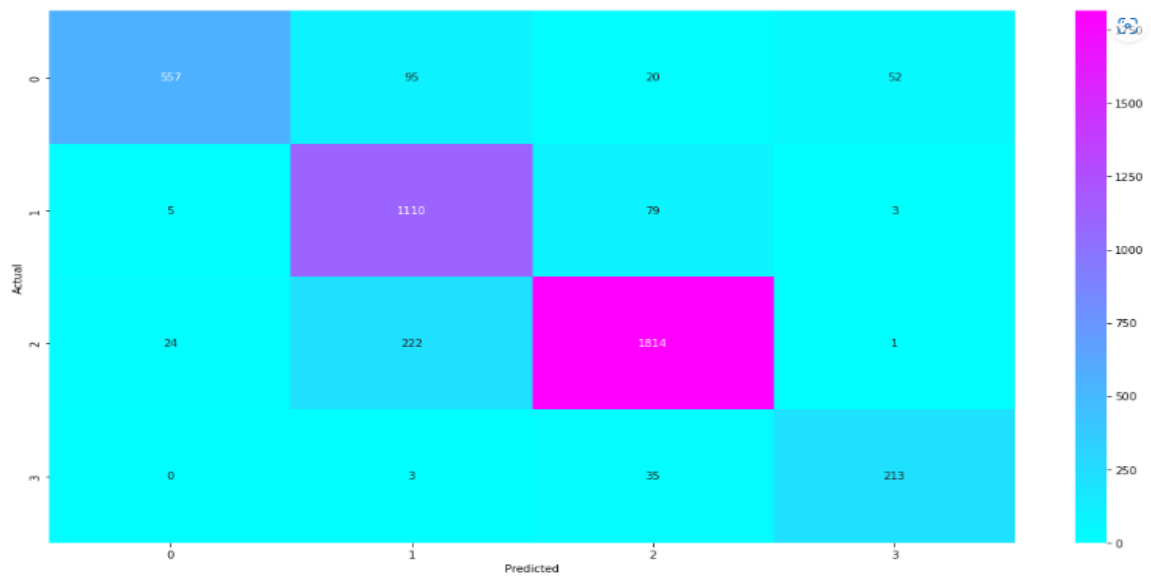
environments.

Fig 2 Densenet 121 classification report

Fig 3 confusion matrix

```
[25]: print(classification_report(test_data.labels,y_pred))

               precision    recall  f1-score   support

           0       0.87      0.83      0.85       724
           1       0.74      0.83      0.78      1197
           2       0.89      0.82      0.86      2061
           3       0.71      0.87      0.78       251

    accuracy                           0.83      4233
   macro avg       0.80      0.84      0.82      4233
weighted avg       0.83      0.83      0.83      4233
```
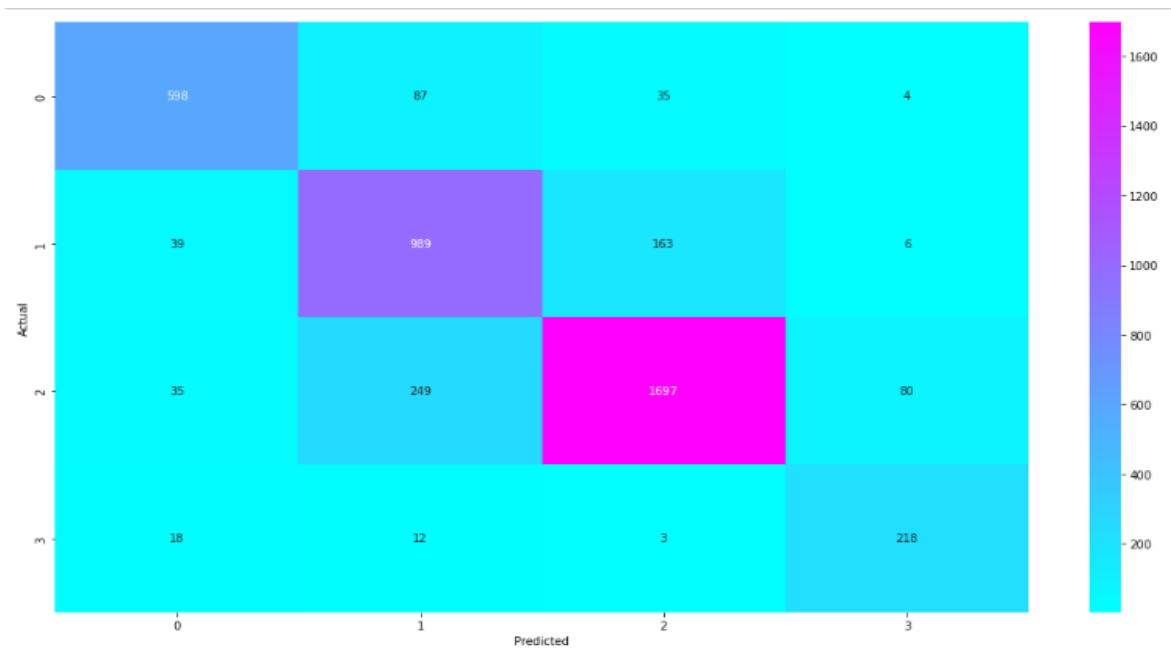
Fig 4 InceptionV3 classification report

Fig 5 InceptionV3 confusion matrix

```
In [36]: print(classification_report(test_data.labels,y_pred))
              precision    recall  f1-score   support

           0       0.98      0.73      0.83       724
           1       0.92      0.80      0.86      1197
           2       0.84      0.97      0.90      2061
           3       0.90      0.95      0.93       251

    accuracy                           0.88      4233
   macro avg       0.91      0.86      0.88      4233
weighted avg       0.89      0.88      0.88      4233
```
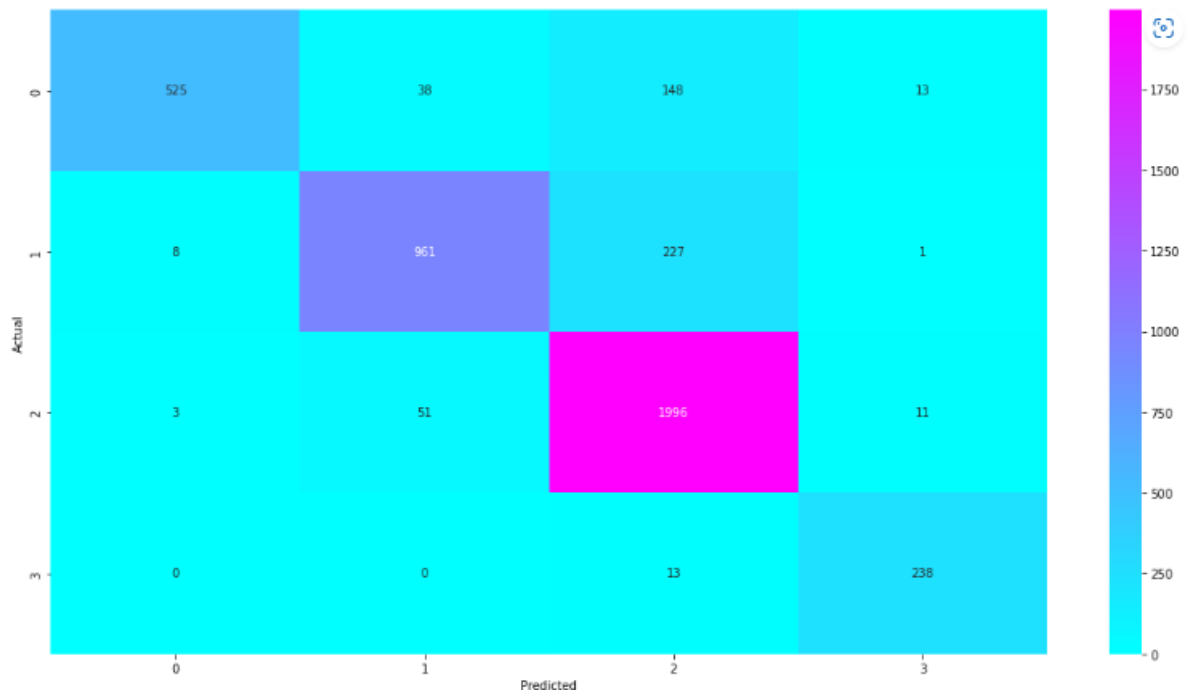
Fig 6 Xception classification report

Fig 7 Xception confusion matrix

While comparing how accurate each model were, we found out that the Xception model performed better than the other models. All models were ran with 5 epoch, with the Xception model having an accuracy of 88% compared to the Densenet 121 and InceptionV3 model with 86% and 84% respectively. The precision for the Xception model was also higher than other models for each of the classes. The InceptionV3 model performed poorly to other models in terms of precision, recall, F1- score and confusion matrix.

Table 1 Models Comparison

Model	Model Accuracy	Precision	Recall	F1-Score
Xception	88%	91%	86%	88%
Densenet 121	86%	86%	86%	85%
InceptionV3	84%	80%	84%	82%

References

Kaggle. *COVID-19 Radiography Database*. Kaggle.com. Retrieved 14 August 2022, from
https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

https://www.almabetter.com/blogs/introduction-to-python-pandas-for-beginners