

# **Configuration Manual**

MSc Research Project Data Analytics

Mansi Atul Chowkkar Student ID: x18134599

School of Computing National College of Ireland

Supervisor: Cristina Muntean

#### National College of Ireland

#### **MSc Project Submission Sheet**



#### **School of Computing**

Student Name: Mansi Chowkkar

**Student ID:** X18134599

**Programme:** MSc. DATA ANALYTICS

Module: MSc. RESEARCH PROJECT

Lecturer: Cristina Muntean

Submission Due 12/12/2019 Date:

- **Project Title:** Breast Cancer Detection from Histopathological Images using Deep Learning and Transfer Learning.
- **Word Count:** 721

Page Count: 13

Year: 2019-2020

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):	

# **1** Hardware/Software Requirements

#### **1.1 Hardware Requirements**

The below hardware would be ideal for running the experiment smoothly.

OS	Windows 10
RAM	Minimum 8GB (2.14 GB from Colaboratory)
Hard Disk Space	Minimum 100GB (100 GB drive space)

### **1.2 Software Requirements**

Programming Language Tools	Google Colaboratory (Cloud based Jupyter
	notebook environment), Python version 3,
	Microsoft Excel, Overleaf
Web Browser	Google Chrome or Mozilla Firefox
Email	Access to a Gmail account

# 1 Google Colaboratory Environment Setup

This section will explain how to setup Google Colaboratory environment in order to perform this experiment. Screenshots are included in order to facilitate a better understanding. A new Gmail address as a mansithesis.unet@gmail.com account is created in order to access Google Colaboratory.

1. Sign in with your Gmail account:

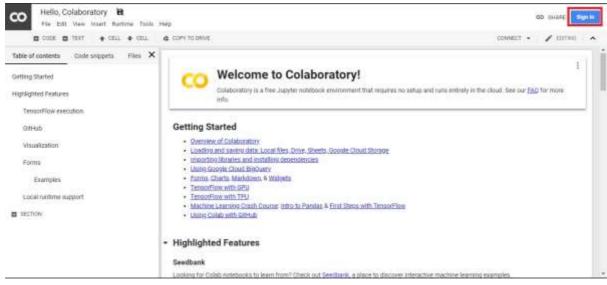


Figure 1: Sign in to Google Colaboratory

2. Once in the notebook, all required libraries are imported. The libraries required for each algorithm are written at the start of the coding of the algorithm.

# 2 Data preparation for Experiments

This section will explain how to upload data on Google Drive and access it in Collaboratory notebook.

1. Upload the image data folder on Google Drive:

Histopathological image dataset of the Breast Biopsy is taken from the Kaggle as shown in the figure 2. Images of all magnifications are mixed in one folder for executing Experiment 1 and image data folder is uploaded on Google drive as shown in figure 3.

MSc Research Project Rep	ort Te 🗶 k bread cano	c dataget from break	+				
€ → C ∎ kaggle	com/kritika397/breast-car	cer dataset from break	his				÷ 🗊
III Appen 🔹 SSIS to SQL Set		alt_ Ø Discover and us		Projects -	Ove., 👻 NCLCloud :	iervices 🧰 Non-structu	re Data 🗰 34
	breast cancer d	ataset from bi	reakhis	+	+	+ -	+ +
	Anukritika • upda	ted 7 months ago (Ven	sian 1)	*	+	4	
	Data Tasks Kernels	(7) Discussion Act	tivity Metad	lata		Download (4 GB)	New Notebook
	h Usability 4.4	😅 License Data fi	ilea © Origina	Author	∿ Ta	gs cancer	
	Data (4 G8)						×
10	lata Sources				About this file		
	e 🖾 toldt			1	No description yet		
	> 🖂 test						
	🛩 🗁 train						
	~ 🗅 100X						
	→ B_100X						
	SOB_B_A	14-22549AB-100-001.j	ping				
	IN SOB B.A.	14-22549AB-100-002.	png	_			

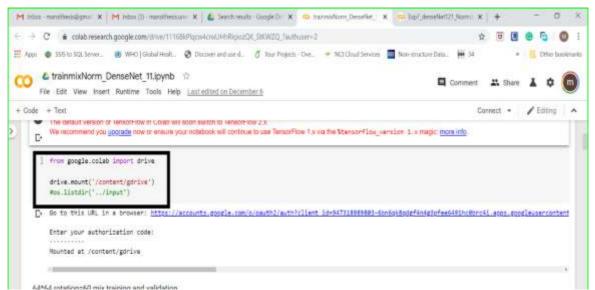
Figure 2:DataSet from Kaggle

II Neo	<b>Q</b> 110	55 SGL	Severa 👘	WHO ( Chasel Healt	ð Disim	e ind a	es it. d' Northands Ove. 🔫 i	NCI Coard Iamètes 🛛 🔤 Norr Husson's Data	H 14			Otter	
۵	Drive			Q Search Driv				<i></i>		Ø	۲	Ш	6
+	New			My Drive -								m	0
-	My Dra-						C remember, Secolde, Tauré	Encoded(2)_2008_1 ayes	(D. health, Jre	etherites C	-	-94-	
		-	New folder				Angest and many transmission and the many transmission of the many transmission of the	Anne and a start band a star	ALC: NO.	1.1.1.1			
8	Shated	-					And in a real to be the second term	And have been some stated back to have been	And the second second	Concernant	in the second		
0	Recett	G	Upload files		_		<ul> <li>Solar Solar (1994) Solar Sola</li></ul>	And large inflations (both the Analysis and the second state of the Analysis and the second state of the Analysis and the second state of the second state of the Analysis and the second state of the second state of the Analysis and the second state of the second state of the second state of the Analysis and the second state of the second state of the second state of the Analysis and the second state of the second state of the second state of the Analysis and the second state of the			to beauty them		
tit.	Starre	B	Upload folder	E)	140	<b>1</b> -1	👄 trainmishterm, Densafied	10 Detsafie(121_200X_1 py.,	- trainfu			D8.	
B	Treate		Gongle Docs	2	- 22		You offer just how	This willing by this paint woold	you withd t	r the pai	t marth		
			Gongle Sheet	а			0.200314	100000000		1.111.0			
C.	Inclug	=	Google Silder	2 19			Owner	Last modified by	ne.	Pike-ski	LF.		
-			More				THE	Phile, 25, 23119					
=	Stornet			Color	-		- The	Nov.22, 2019	-				
	HUY STO		used										
	Jacob and a			Colub Notebo	oka		(THE	PHON 223, 2017/0					
				40X Densethe		a land	1 ma	Face 28, 2019		219.62			

Figure 3: Upload data on Drive

2. Mount Google drive in the colaboratory notebook:

Figure 4 shows the steps for mounting drive in the python code. Click on the url and select the gmail account for the colaboratory and enter the authorization code as shown in figure 5.



**Figure 4: Mounting of Google Drive** 



### Sign in

Please copy this code, switch to your application and paste it there:

Ē

4/uAFkc0hLcOszubPo241EWnPqencCwub4xdD33kz1UIbAMyNqBcsZXs

0	from google.colab import drive
	<pre>drive.mount('/content/gdrive') wos.listdir('/input')</pre>
***	Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth/client id=947318989883-6bn6pi8odgf4n4g3pfee6491hc8brc4i.acos.googleusercontent
	Enter your authorization code:
-	

**Figure 5: Enter Authorization code** 

3. Access the image data using below command in the notebook:

Figure 6 shows the code for accessing images from folder present on google drive.

```
[ ] train_gen=ImageDataGenerator(rotation_range=60,
                                  width_shift_range=0.1,
                                  height_shift_range=0.1,
                                   rescale = 1./255,
                                  #rescale = 0.8,
                                  shear_range = 0.2,
                                  zoom_range = 0.5,
                                  horizontal_flip = True,
                                  #featurewise_std_normalization=True,
                                  fill mode="nearest")
     train = train_gen.flow_from_directory("/content/gdrive/My Drive/breast_cancer/fold1/train_mix_Norm/
                                                   Idss mode- categorical ,
                                                  target_size=(64,64),
                                                  color_mode="rgb",
                                                  shuffle=True,
                                                  batch size=32)
```

Figure 6: Accessing images in the python notebook

4. Import Libraries for image normalization and augmentation:

Libraries required for Histogram normalisation are imported in python notebook as shown in figure7. ImageDataGenerator is used from Keras library to augment the images. Images are saved in the same directory for further use. Figure 7 shows the libraries required for the upsampling method.

$\mathbf{co}$	Histogran_normalization_processing.ipynb  File Edit View Insert Runtime Tools Help <u>All changes saved</u>
+ Code	+ Text
C	<pre>import os import cv2 import numpy as np from PIL import Image import pandas as pd import matplotlib.pyplot as plt from PIL import ImageOps from numpy import expand_dims from matplotlib import pyplot as plt</pre>
	from IPython.display import display, Math, Latex import matplotlib.pyplot as plt
	import matplotlib.pyplot as plt Figure 7: Libraries for Histogram Normalization
0	<pre>import matplotlib.pyplot as plt Figure 7: Libraries for Histogram Normalization     upsampling.ipynb</pre>
00	import matplotlib.pyplot as plt Figure 7: Libraries for Histogram Normalization
<b>CO</b> + Code	<pre>import matplotlib.pyplot as plt Figure 7: Libraries for Histogram Normalization     upsampling.ipynb</pre>
<b>CO</b> + Code	<pre>import matplotlib.pyplot as plt Figure 7: Libraries for Histogram Normalization Upsampling.ipynb ☆ File Edit View Insert Runtime Tools Help <u>All changes saved</u></pre>
CO + Cod	<pre>import matplotlib.pyplot as plt Figure 7: Libraries for Histogram Normalization Upsampling.ipynb ☆ File Edit View Insert Runtime Tools Help <u>All changes saved</u> e + Text import os </pre>

**Figure 8: Image Augmentation** 

# **3** Deep Learning and Transfer Learning Model Execution

This section will explain which libraries need to be imported for the execution of CNN and DenseNet-121

1. CNN Execution:

For CNN execution keras and tensorflow libraries are used from python. Figure 9 shows all the required libraries for image augmentation, model definitions, model compiling and model execution. Also, libraries are required for statistical analysis and graph plotting. The random seed is set every time so that every time different output is generated for different experiments.

- All layers in the CNN model are imported from keras.layers libraries.
- The model is built using tensorow as tf library.
- The model layers are plotted using tensorow.keras.utils library

€ →	• mansithesis@gmail × M Inbox (3) - mansithesis.unel × Colab Notebooks - Googie × CO Copy_Exp7_denseNet_Nor × G Success co C
S 7	Colab.research.google.com/ drive/ rer2steb/synyphints/requiv (korrer24/authuser=2#scrollito=ro4ojDAuAop_
0	SSIS to SQL Server (例) WHO   Global Healt ③ Discover and use d ⑤ Your Projects - Ove ≪ NCI Cloud Services III Non-structure Copy_Exp7_denseNet_Normdata.ipynb 台
+ Code	+ Text
> O	<pre>import os import keras from keras.models import Model from keras.layers import Conv2D, MaxPooling2D, Dense, Input, Activation, Dropout, GlobalAveragePooling2D, \ BatchNormalization, concatenate, AveragePooling2D from keras.optimizers import Adam from keras.optimizers import Adam from keras.layers.core import Activation, Flatten, Dropout, Dense,Reshape from keras.import backend as K from keras.preprocessing.image import ImageDataGenerator from keras.utils import np_utils from sklearn.metrics import confusion_matrix from keras.preprocessing imort image from keras.preprocessing import image from keras.preprocessing import image_to_array from sklearn.preprocessing import MultiLabelBinarizer import matplotlib.pyplot as plt import tensorflow as tf from pathlb import Path import cv2 from mlxtend.plotting import plot_confusion_matrix tf.set_random_seed(1238) from sklearn.metrics import fl_score</pre>

**Figure 9: Libraries for CNN execution** 

- 2. DenseNet-121 Execution:
  - For DenseNet-121 execution keras and tensorflow libraries are used from python. Figure 10 shows all the required libraries for image augmentation, model definitions, model compiling and model execution. Also, libraries are required for statistical analysis and graph plotting.
  - DenseNet121 library is imported from keras package.
  - Plotting of testing and training plots: matplotlib.pyplot as plt, seaborn as sns libraries are used for plotting training and validation graphs.

### 🔥 🖕 Copy of DenseNet121\_400X\_1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
[6] import os
    from keras.layers.normalization import BatchNormalization
    from keras.layers.convolutional import Conv2D, MaxPooling2D
    from keras.layers.core import Activation, Flatten, Dropout, Dense, Reshape
    from keras import backend as K
    from keras.preprocessing.image import ImageDataGenerator
    from keras.optimizers import Adam
    from keras.utils import np utils
    from keras.preprocessing import image
    from keras.preprocessing.image import img_to_array
    from sklearn.preprocessing import MultiLabelBinarizer
    import matplotlib.pyplot as plt
    import numpy as np
    import argparse
    import tensorflow as tf
    import cv2
    from mlxtend.plotting import plot_confusion_matrix
    from keras.applications import densenet
    from keras.models import Sequential, Model, load model
    import seaborn as sn
    from sklearn.metrics import confusion_matrix, classification_report
    tf.set random seed(1234)
    from keras import layers
    from keras.applications import DenseNet121
```

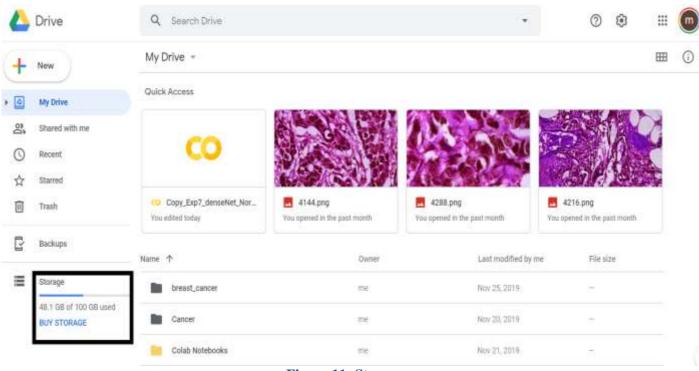
Figure 10: Libraries for execution of DenseNet-121

### **4** Settings done for accelerating Computation time

This section will explain about how the drive storage is extended and GPU setting is done from the google colaboratory.

1. Drive Storage:

Extra 100 GB drive storage is subscribed to store normalized and augmented images. Extra storage was required for storing different python files with the executed models. Figure 11 shows the utilization of google drive for this project.



#### Figure 11: Storage

#### 2. GPU:

For boosting the execution speed for both the models GPU accelerator hardware is selected from runtime option as shown in figure 12 and figure 13.

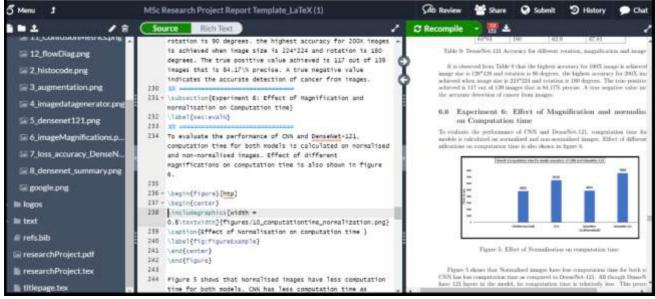
	NetNewData_5(with_weight).	- C00	ment # Share 🛦 🗘 🌀
Code + Test		1479. 1478	Connect • 🖌 Editing
<ul> <li>import os import seaborn a import pandes as import keras</li> </ul>	B		
from keras.model from keras.layer Batchkormalizati from keras.optim from keras.model	s on ir	Input, Activation, Dropout, GlobalAveragePooling2D, \	
from keras.layer from keras.layer from keras.layer from keras.layer	<ul> <li>Reset all runtimes</li> </ul>	ition .	
from keras.layer from keras impor from keras.prepr from keras.optim from keras.utils	t Manage sessions	ppout, Dense,Reshape	
from sklearn.met from sklearn.met	rics import classification_repor rics import confusion_matrix ocessing import image		

Figure 12: Runtime selection of hardware

Runtime type		
Python 3	<b>—</b>	
Hardware accelerator		
GPU	- O	
<u>+1-</u>		
Omit code cell o	tout when saving this notebook	
Omit code cell o	utput when saving this notebook	
Omit code cell o	utput when saving this notebook	

### 5. Other Software used

For visualization of evaluation results Microsoft Excel is used. Research report is written using Overleaf tool. Figure 14 shows the utilization of Overleaf for this project.



**Figure 14: Overleaf Project** 

# References

https://keras.io/preprocessing/image/

https://medium.com/hugoferreiras-blog/confusionmatrixandothermetricsinmachinelearning-894688cb1c0a

https://keras.io/getting-started/sequential-model-guide/

https://matplotlib.org/3.1.1/tutorials/introductory/pyplot.html