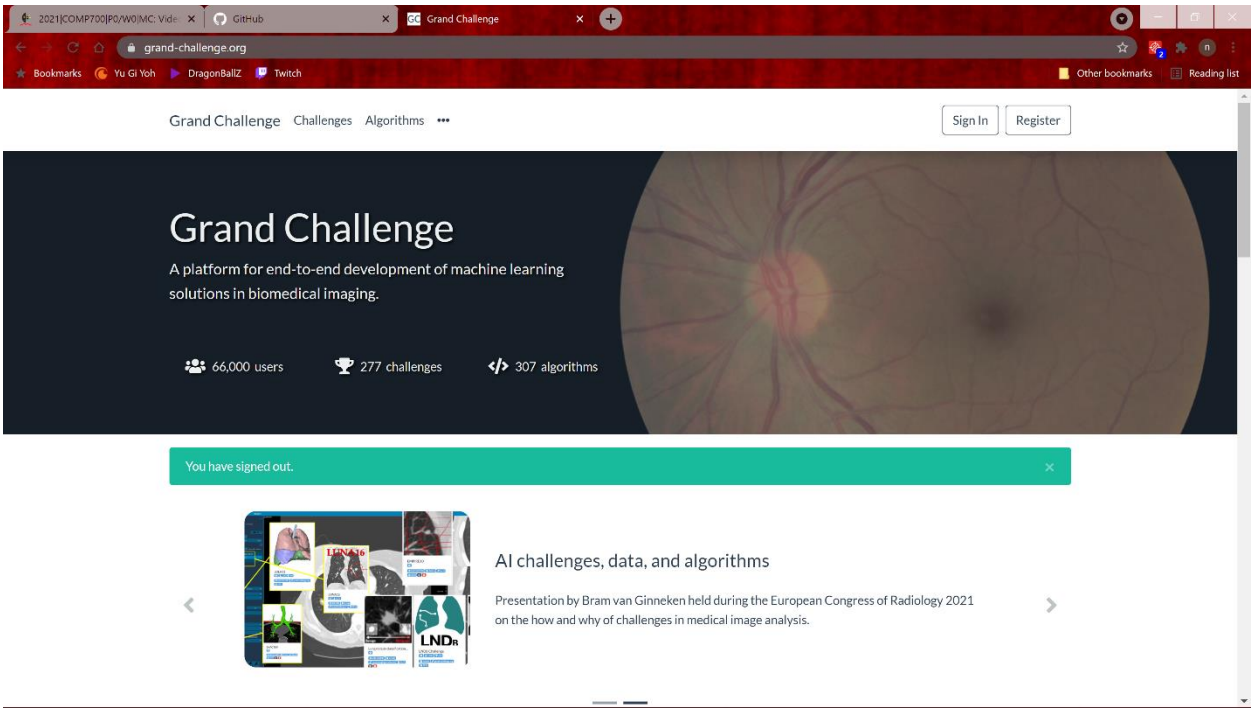
How To Run The programs

Clarifications

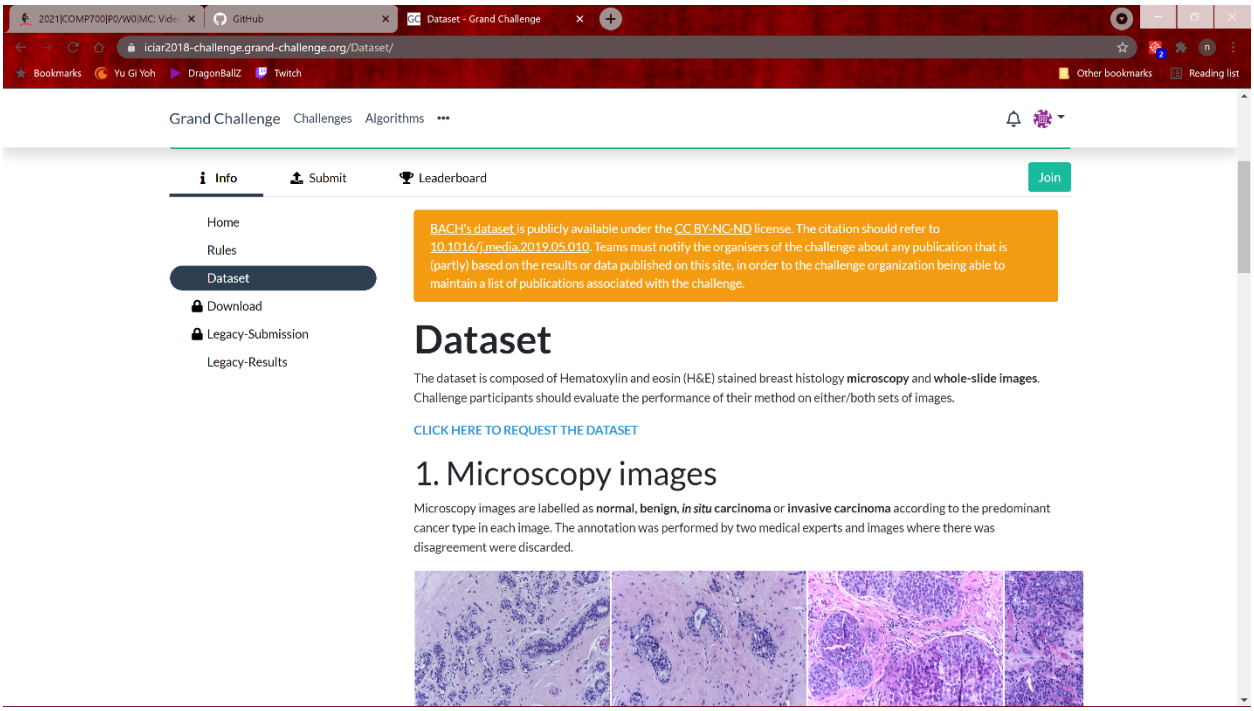
- The project assumes you will be using Pycharm and Google colab
- The project assumes you will upload relevant data to your google drive. Make sure to mount the google drive in runtimes for each notebook.
- The program assumes you will run 1 notebook at a time and disconnect the other when not in use.
- Pycharm code must be run before proceeding to the notebooks.
- Warning – google colab will terminate the session if it is left to idle for too long. Make sure to frequently check and move mouse at 10 minute intervals. Alternatively visit the site https://stackoverflow.com/questions/57113226/how-to-prevent-google-colab-from-disconnecting for information on how to stop colab from disconnecting.

OBTAINING THE DATASET

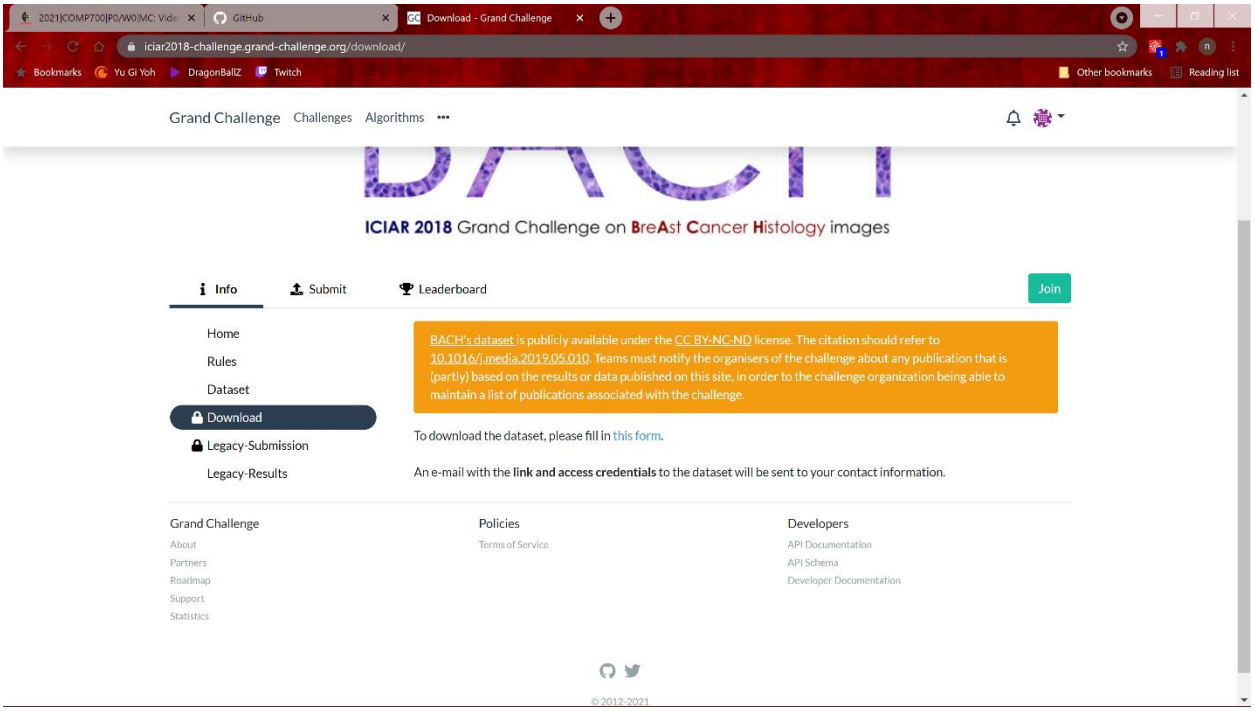
The ICIAR 2018 dataset can be requested from the following site - https://iciar2018-challenge.grand-challenge.org/Dataset/

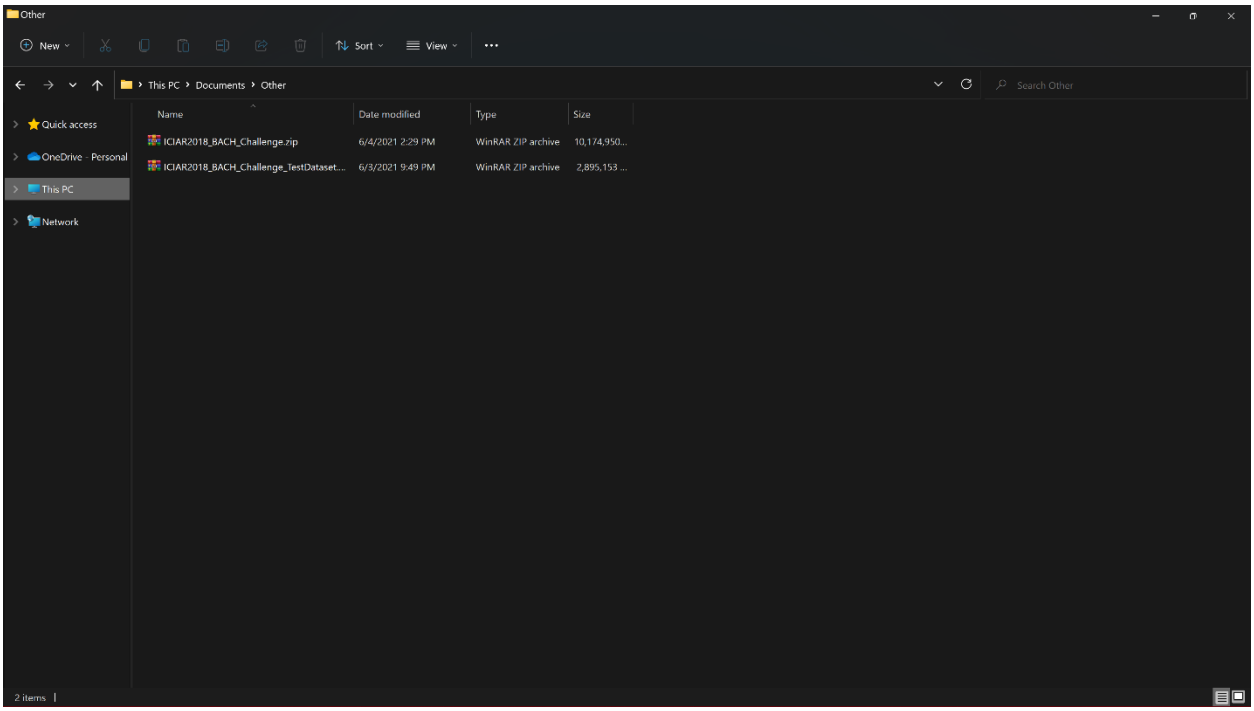


You will need to sign up before requesting the dataset



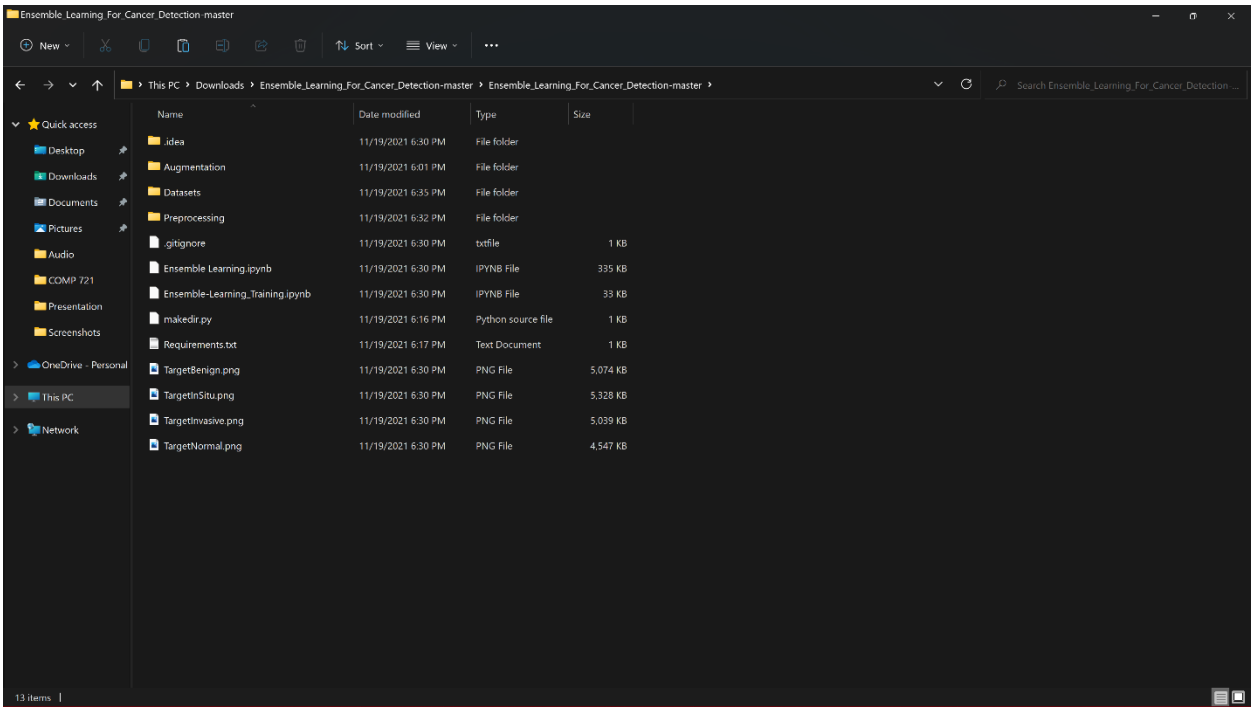
Request the dataset and follow the prompts until you are in possession of the dataset. Once you have

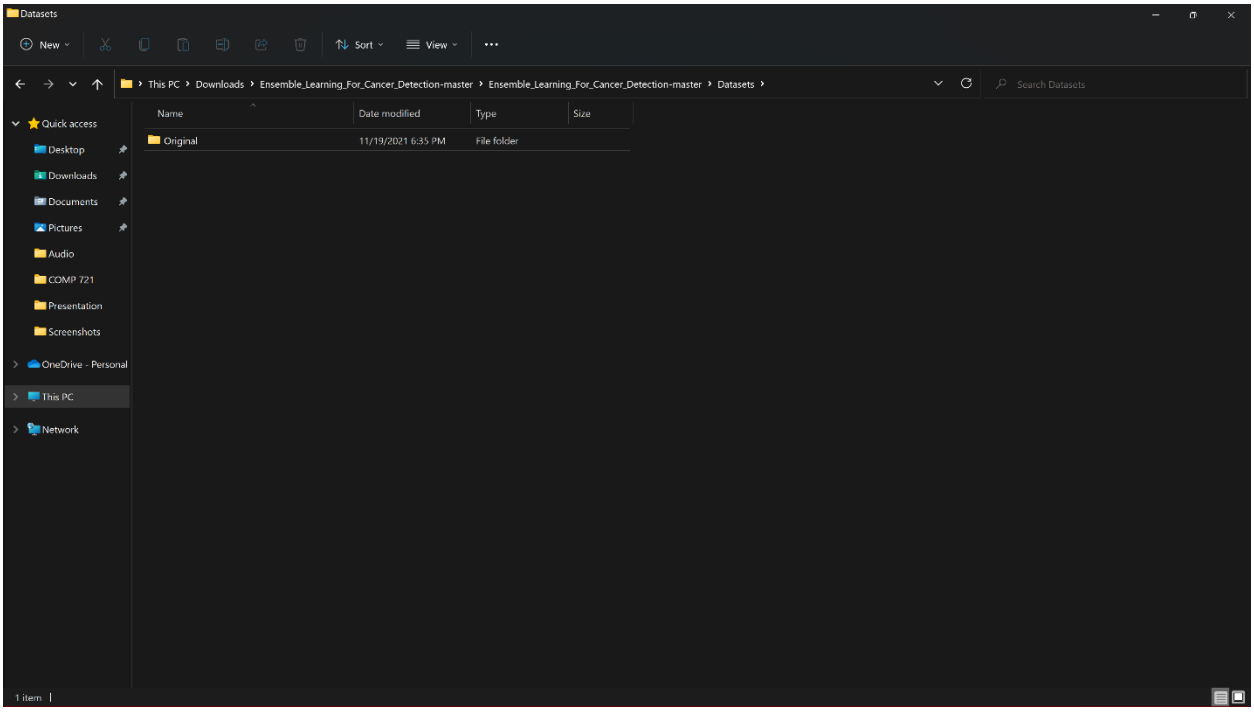




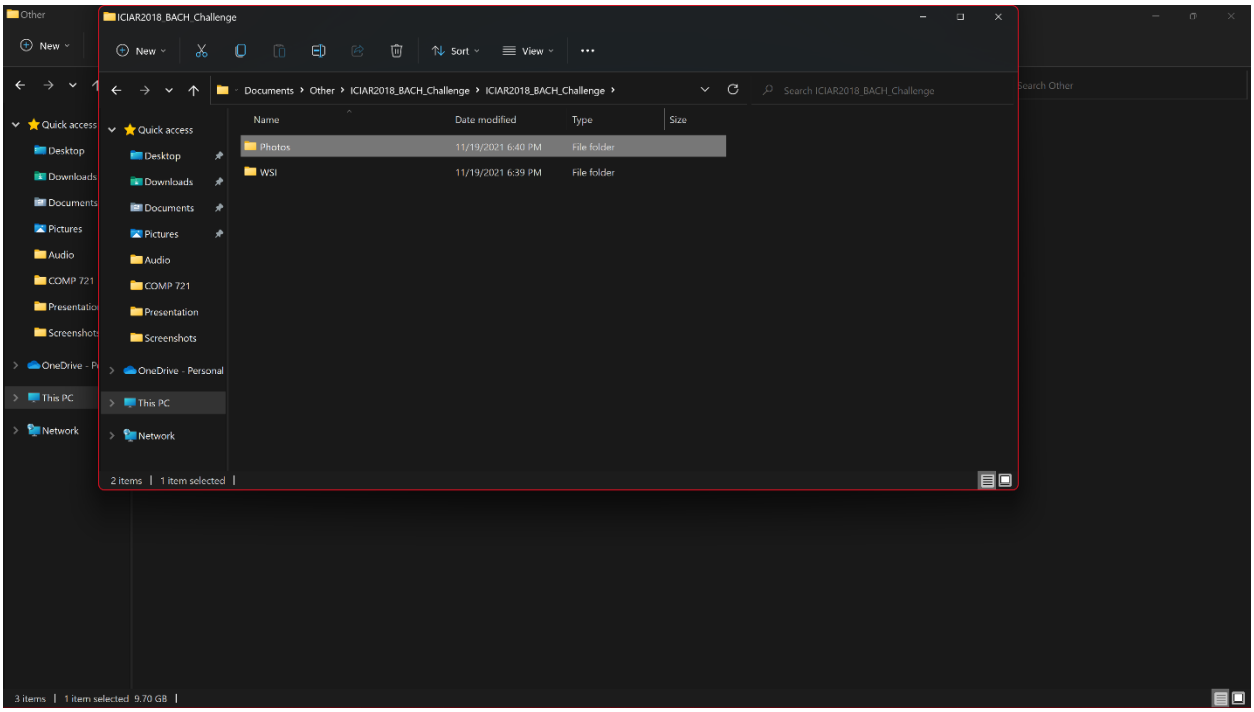
Once you are in possession of the dataset unzip the file names ICIAR2018_BACH_Challenge.zip . the other file may be disregarded.

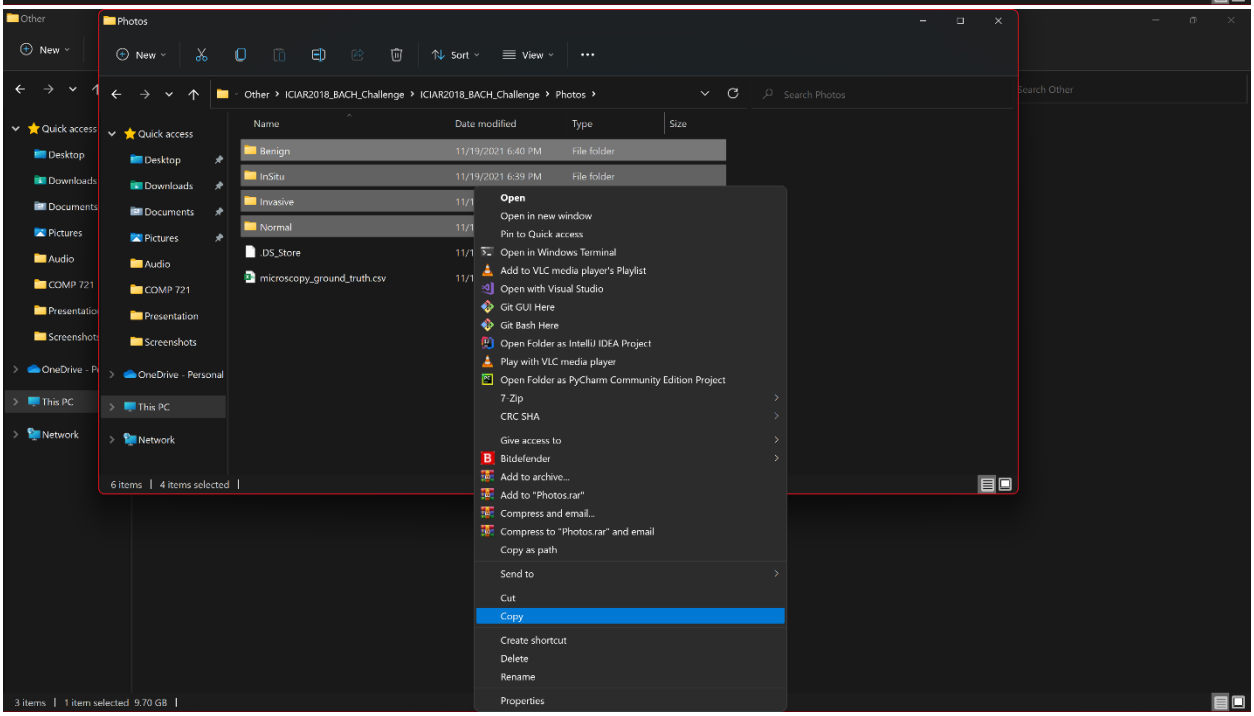
Once you have obtained the code from github unzip it and create a folder called Datasets. Create another folder called Original inside of it

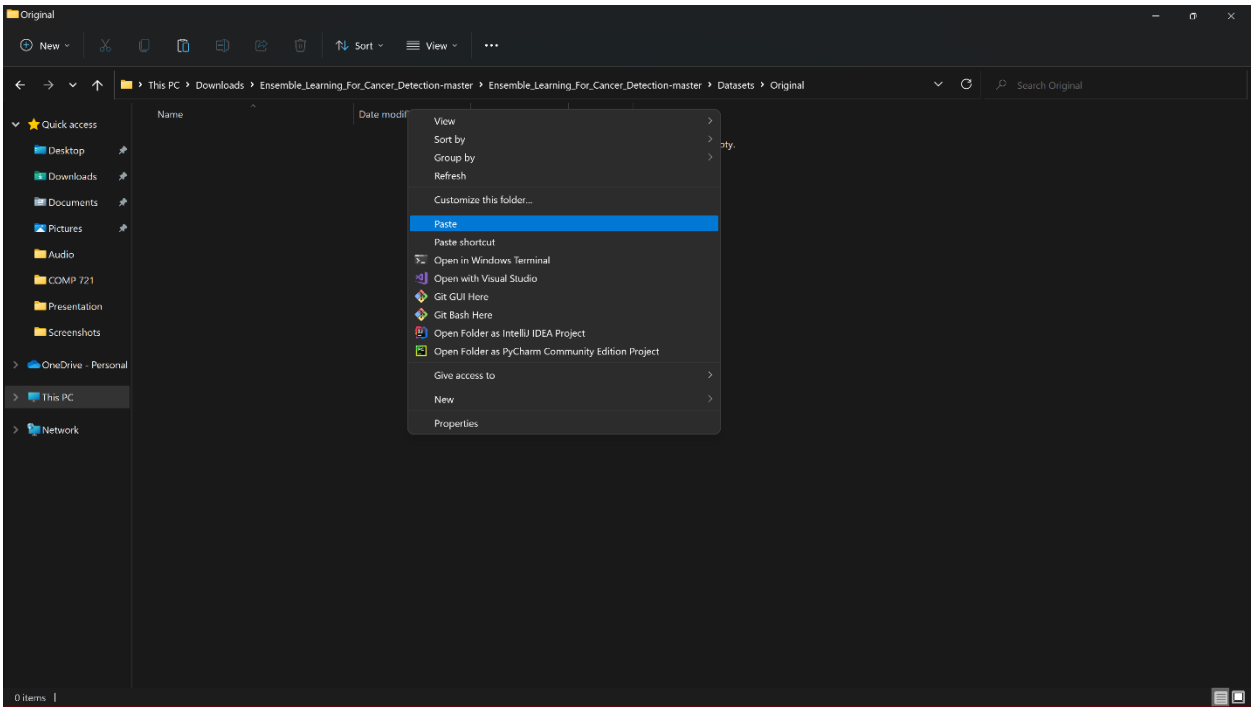

Ensemble_Learning_For_Cancer_Detection-master
New
Sort
View
This PC › Downloads › Ensemble_Learning_For_Cancer_Detection-master › Ensemble_Learning_For_Cancer_Detection-master ›
Search Ensemble_Learning_For_Cancer_Detection-...
Quick access
Desktop
Downloads
Documents
Pictures
Audio
COMP 721
Presentation
Screenshots
OneDrive - Personal
This PC
Network
Name
Date modified
Type
Size
.idea 11/19/2021 6:30 PM File folder
Augmentation 11/19/2021 6:01 PM File folder
Datasets 11/19/2021 6:35 PM File folder
Preprocessing 11/19/2021 6:32 PM File folder
.gitignore 11/19/2021 6:30 PM txtfile 1 KB
Ensemble Learning.ipynb 11/19/2021 6:30 PM IPYNB File 335 KB
Ensemble-Learning_Training.ipynb 11/19/2021 6:30 PM IPYNB File 33 KB
makedir.py 11/19/2021 6:16 PM Python source file 1 KB
Requirements.txt 11/19/2021 6:17 PM Text Document 1 KB
TargetBenign.png 11/19/2021 6:30 PM PNG File 5,074 KB
TargetInSitu.png 11/19/2021 6:30 PM PNG File 5,328 KB
TargetInvasive.png 11/19/2021 6:30 PM PNG File 5,039 KB
TargetNormal.png 11/19/2021 6:30 PM PNG File 4,547 KB
13 items

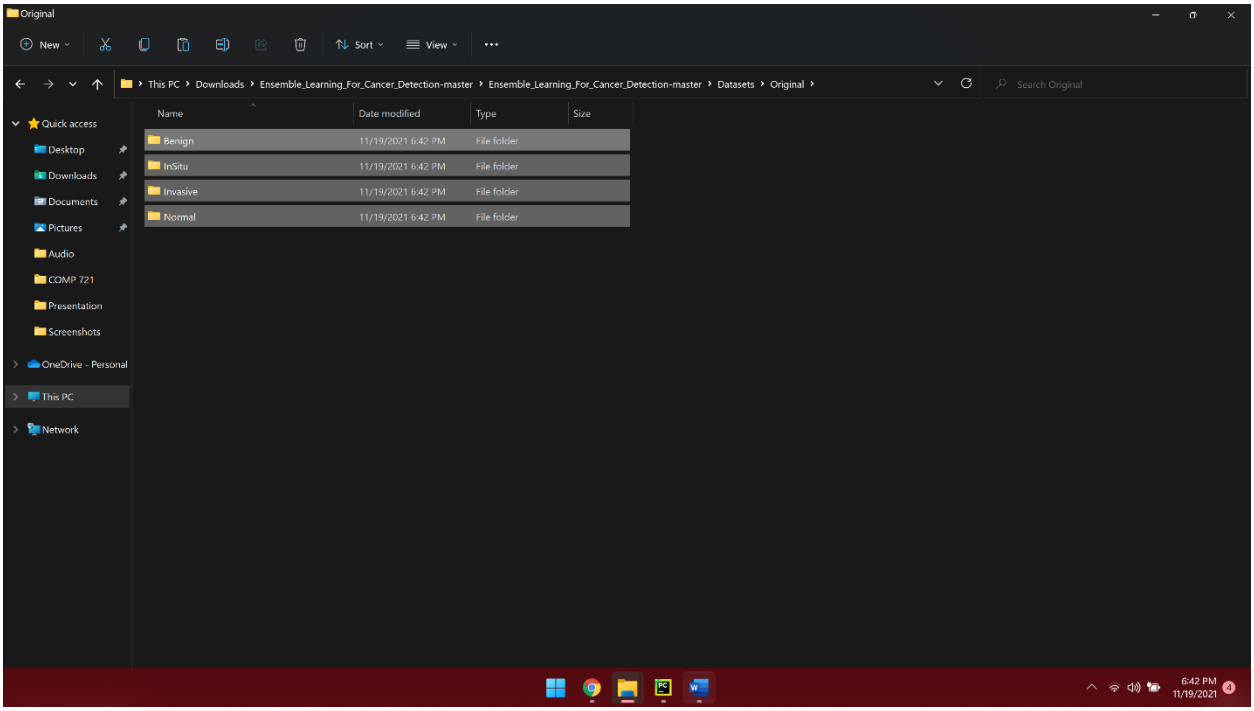

Datasets
New
Sort
View
This PC › Downloads › Ensemble_Learning_For_Cancer_Detection-master › Ensemble_Learning_For_Cancer_Detection-master › Datasets ›
Search Datasets
Quick access
Desktop
Downloads
Documents
Pictures
Audio
COMP 721
Presentation
Screenshots
OneDrive - Personal
This PC
Network
Name
Date modified
Type
Size
Original 11/19/2021 6:35 PM File folder
1 item

Open the unzipped ICIAR2018_BACH_Challenge and navigate until you see the following folders. Copy the folders named Benign, InSitu, Invasive and Normal and paste them in the folder Original that you created.



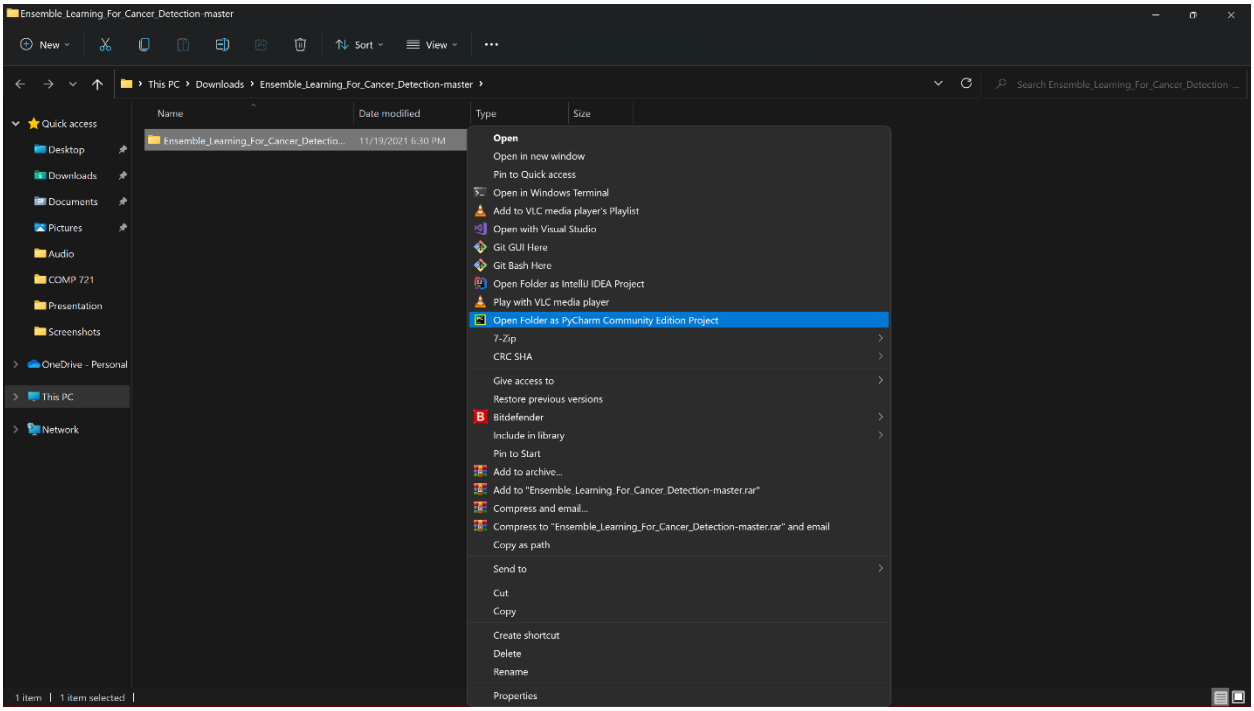



Original
New
Sort
View
This PC › Downloads › Ensemble_Learning_For_Cancer_Detection-master › Ensemble_Learning_For_Cancer_Detection-master › Datasets › Original
Search Original
Quick access
Desktop
Downloads
Documents
Pictures
Audio
COMP 721
Presentation
Screenshots
OneDrive - Personal
This PC
Network
Name
View
Sort by
Group by
Refresh
Customize this folder...
Paste
Paste shortcut
Open in Windows Terminal
Open with Visual Studio
Git GUI Here
Git Bash Here
Open Folder as IntelliJ IDEA Project
Open Folder as PyCharm Community Edition Project
Give access to
New
Properties
0 items

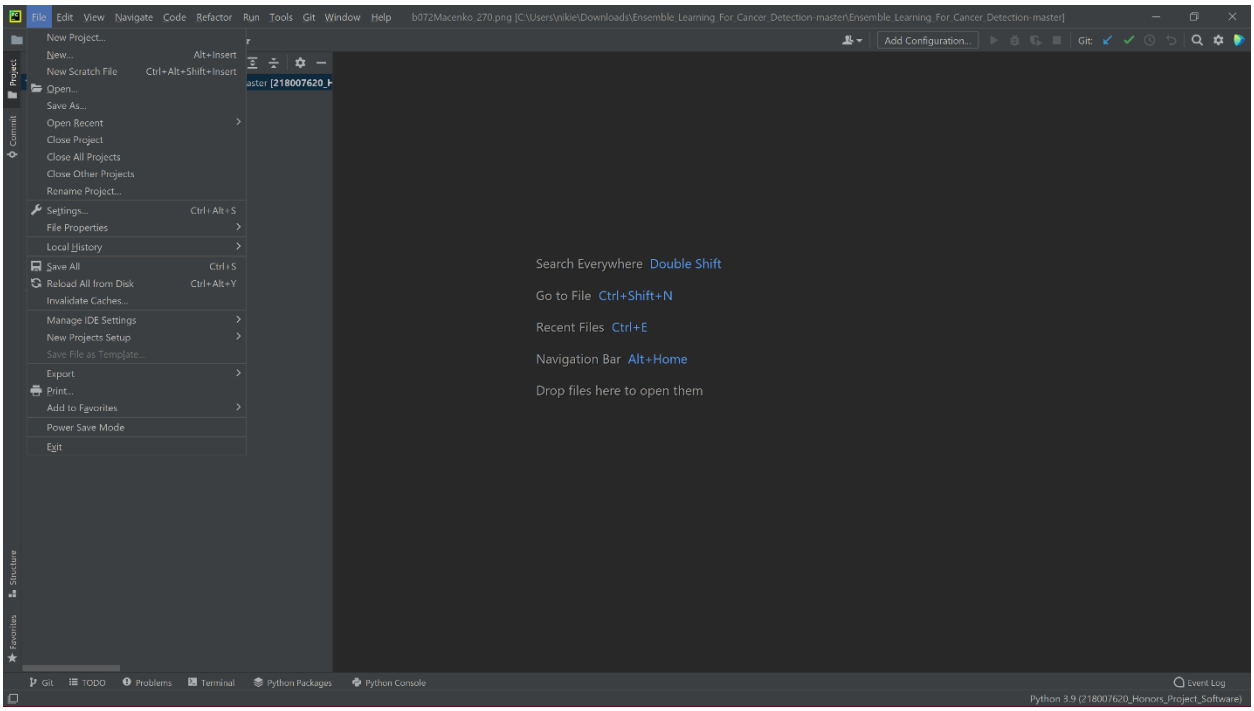

Original
New
Sort
View
This PC › Downloads › Ensemble_Learning_For_Cancer_Detection-master › Ensemble_Learning_For_Cancer_Detection-master › Datasets › Original ›
Search Original
Quick access
Desktop
Downloads
Documents
Pictures
Audio
COMP 721
Presentation
Screenshots
OneDrive - Personal
This PC
Network
Name
Date modified
Type
Size
Benign 11/19/2021 6:42 PM File folder
InSitu 11/19/2021 6:42 PM File folder
Invasive 11/19/2021 6:42 PM File folder
Normal 11/19/2021 6:42 PM File folder
6:42 PM
11/19/2021

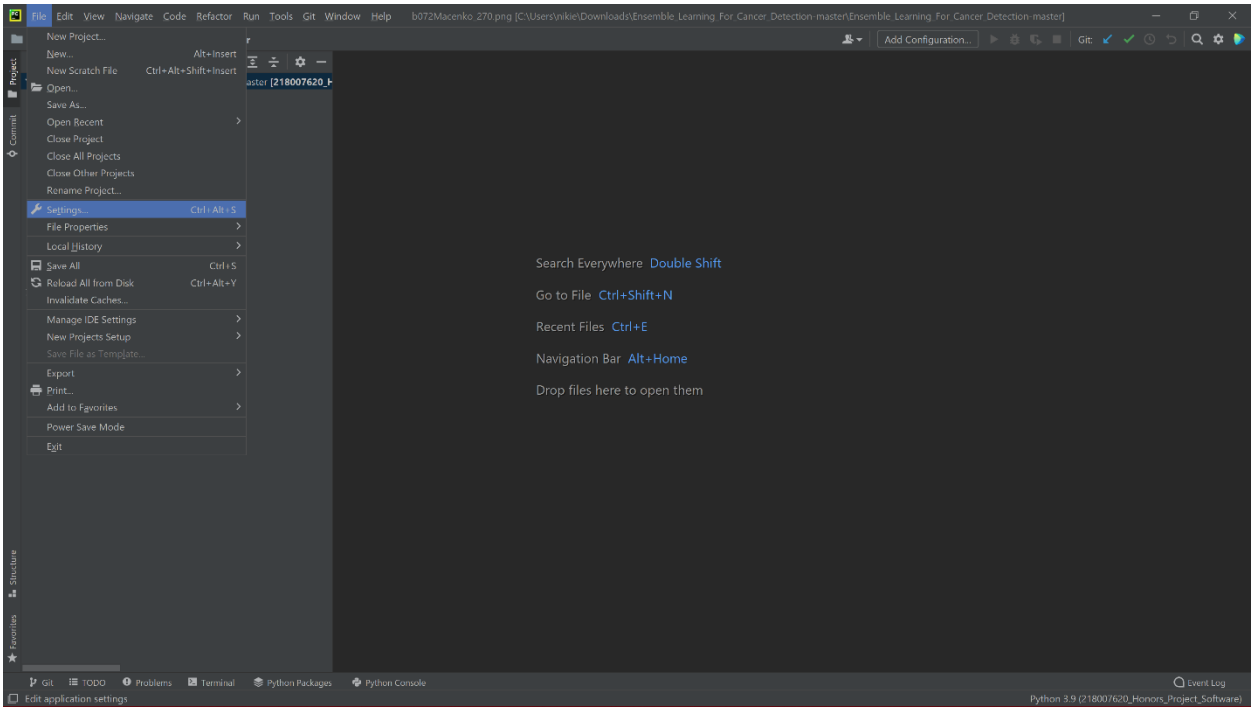
SETTING UP VRITUAL ENVIROMENT ON PYCHARM

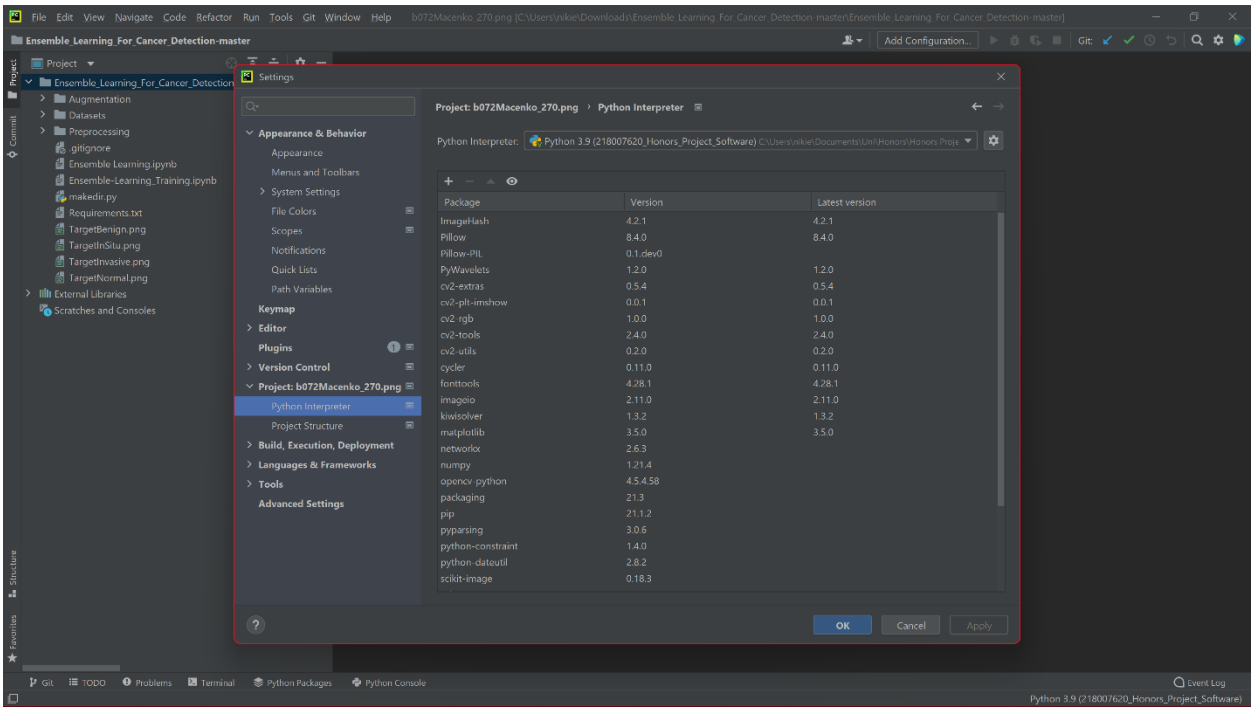
Once you have obtained the code from Github unzip the file and open up the folder in pycharm

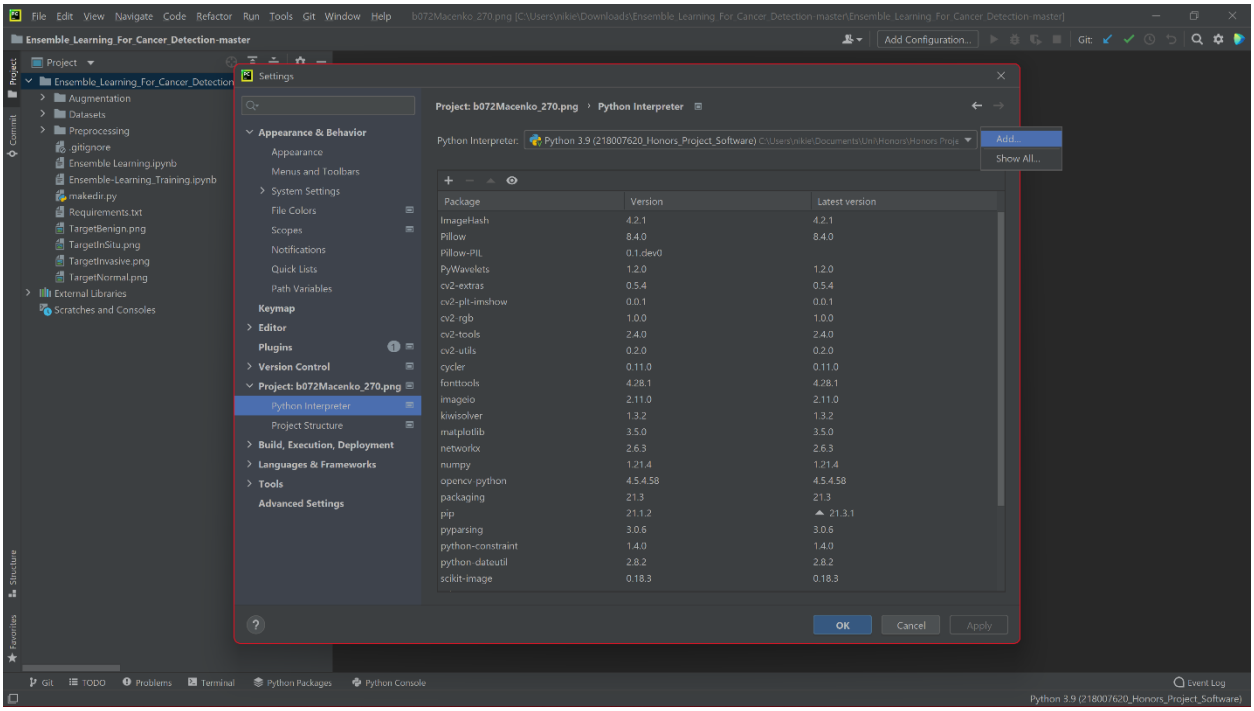


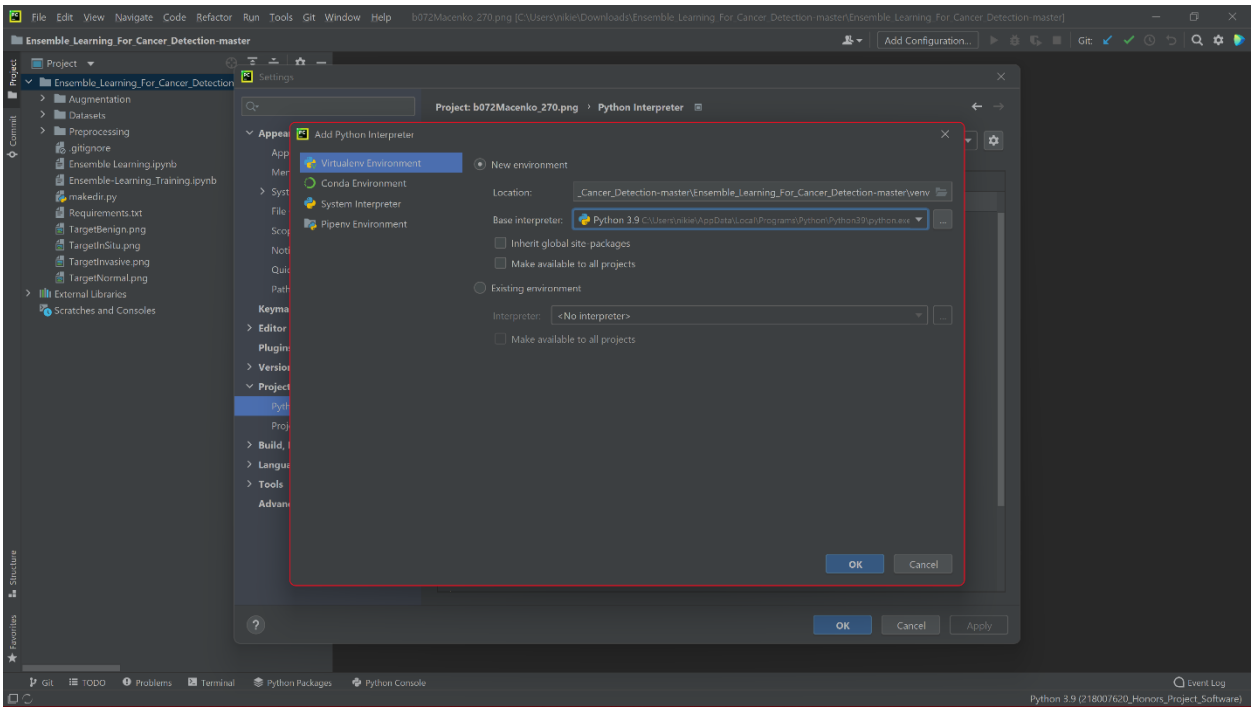
Once opened click File -> Settings -> Python Interpreter -> Add to add a new virtual environment


File Edit View Navigate Code Refactor Run Tools Git Window Help
b072Macenko_270.png [C:\Users\nikie\Downloads\Ensemble_Learning_For_Cancer_Detection-master\Ensemble_Learning_For_Cancer_Detection-master]
New Project...
New... Alt+Insert
New Scratch File Ctrl+Alt+Shift+Insert
Open...
Save As...
Open Recent
Close Project
Close All Projects
Close Other Projects
Rename Project...
Settings... Ctrl+Alt+S
File Properties
Local History
Save All Ctrl+S
Reload All from Disk Ctrl+Alt+Y
Invalidate Caches...
Manage IDE Settings
New Projects Setup
Save File as Template...
Export
Print...
Add to Favorites
Power Save Mode
Exit
Add Configuration...
Git:
Search Everywhere Double Shift
Go to File Ctrl+Shift+N
Recent Files Ctrl+E
Navigation Bar Alt+Home
Drop files here to open them
Git TODO Problems Terminal Python Packages Python Console Event Log
Python 3.9 (218007620_Honors_Project_Software)

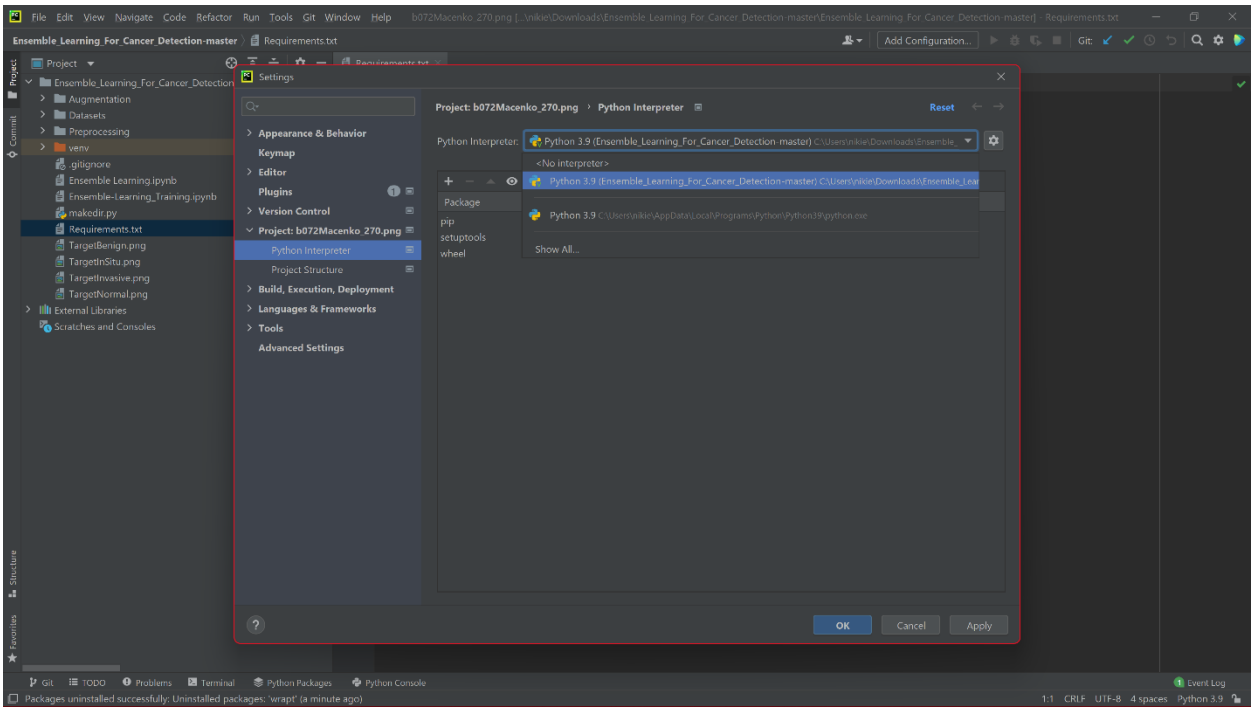

File Edit View Navigate Code Refactor Run Tools Git Window Help
b072Macenko_270.png [C:\Users\nikie\Downloads\Ensemble_Learning_For_Cancer_Detection-master\Ensemble_Learning_For_Cancer_Detection-master]
New Project...
New... Alt+Insert
New Scratch File Ctrl+Alt+Shift+Insert
Open...
Save As...
Open Recent
Close Project
Close All Projects
Close Other Projects
Rename Project...
Settings... Ctrl+Alt+S
File Properties
Local History
Save All Ctrl+S
Reload All from Disk Ctrl+Alt+Y
Invalidate Caches...
Manage IDE Settings
New Projects Setup
Save File as Template...
Export
Print...
Add to Favorites
Power Save Mode
Exit
Add Configuration...
Git:
Search Everywhere Double Shift
Go to File Ctrl+Shift+N
Recent Files Ctrl+E
Navigation Bar Alt+Home
Drop files here to open them
Git TODO Problems Terminal Python Packages Python Console Event Log
Edit application settings
Python 3.9 (218007620_Honors_Project_Software)

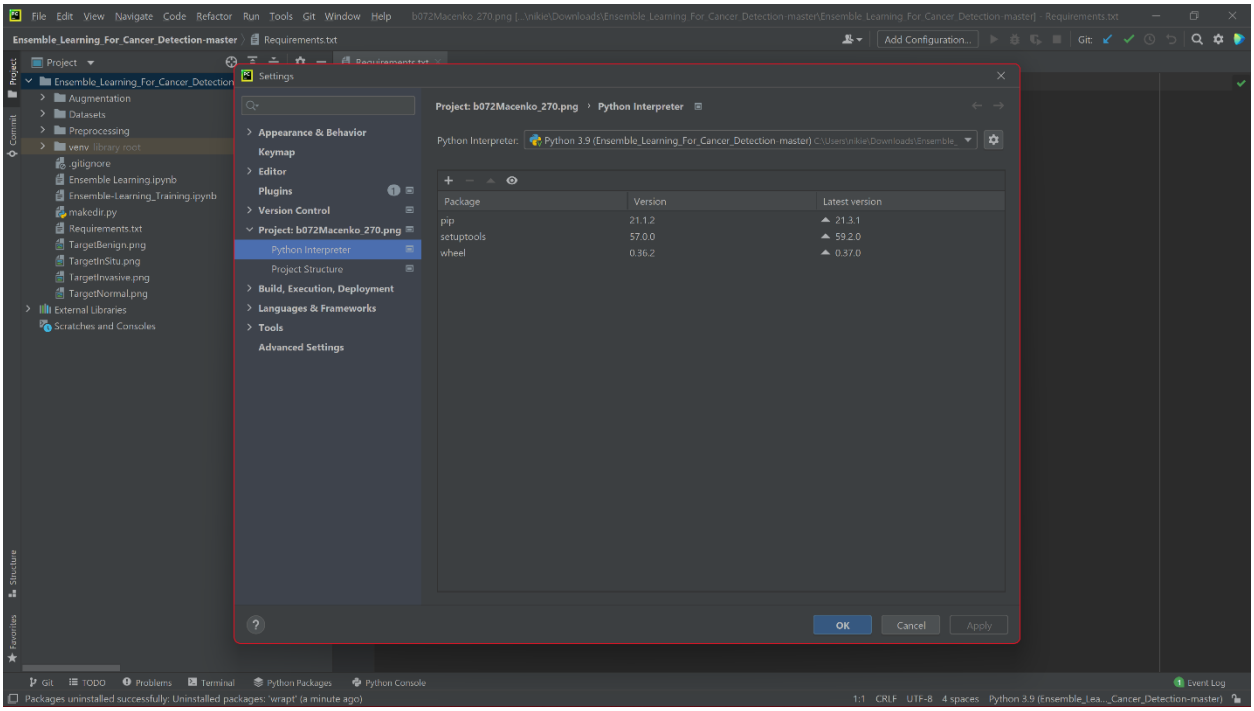

File Edit View Navigate Code Refactor Run Tools Git Window Help
Ensemble_Learning_For_Cancer_Detection-master
Add Configuration...
Settings
Project: b072Macenko_270.png › Python Interpreter
Python Interpreter: Python 3.9 (218007620_Honors_Project_Software)
Appearance & Behavior
Appearance
Menus and Toolbars
System Settings
File Colors
Scopes
Notifications
Quick Lists
Path Variables
Keymap
Editor
Plugins
Version Control
Project: b072Macenko_270.png
Python Interpreter
Project Structure
Build, Execution, Deployment
Languages & Frameworks
Tools
Advanced Settings
Package Version Latest version
ImageHash 4.2.1 4.2.1
Pillow 8.4.0 8.4.0
Pillow-PIL 0.1.dev0
PyWavelets 1.2.0 1.2.0
cv2-extras 0.5.4 0.5.4
cv2-plt-imshow 0.0.1 0.0.1
cv2_rgb 1.0.0 1.0.0
cv2-tools 2.4.0 2.4.0
cv2-utils 0.2.0 0.2.0
cycler 0.11.0 0.11.0
fonttools 4.28.1 4.28.1
imageio 2.11.0 2.11.0
kiwisolver 1.3.2 1.3.2
matplotlib 3.5.0 3.5.0
networkx 2.6.3
numpy 1.21.4
opencv-python 4.5.4.58
packaging 21.3
pip 21.1.2
pyparsing 3.0.6
python-constraint 1.4.0
python-dateutil 2.8.2
scikit-image 0.18.3
OK
Cancel
Apply
Git
TODO
Problems
Terminal
Python Packages
Python Console
Event Log
Python 3.9 (218007620_Honors_Project_Software)


File Edit View Navigate Code Refactor Run Tools Git Window Help
Ensemble_Learning_For_Cancer_Detection-master
Add Configuration...
Settings
Project: b072Macenko_270.png › Python Interpreter
Python Interpreter: Python 3.9 (218007620_Honors_Project_Software)
Add...
Show All...
Appearance & Behavior
Appearance
Menus and Toolbars
System Settings
File Colors
Scopes
Notifications
Quick Lists
Path Variables
Keymap
Editor
Plugins
Version Control
Project: b072Macenko_270.png
Python Interpreter
Project Structure
Build, Execution, Deployment
Languages & Frameworks
Tools
Advanced Settings
Package Version Latest version
ImageHash 4.2.1 4.2.1
Pillow 8.4.0 8.4.0
Pillow-PIL 0.1.dev0
PyWavelets 1.2.0 1.2.0
cv2-extras 0.5.4 0.5.4
cv2-plt-imshow 0.0.1 0.0.1
cv2_rgb 1.0.0 1.0.0
cv2-tools 2.4.0 2.4.0
cv2-utils 0.2.0 0.2.0
cycler 0.11.0 0.11.0
fonttools 4.28.1 4.28.1
imageio 2.11.0 2.11.0
kiwisolver 1.3.2 1.3.2
matplotlib 3.5.0 3.5.0
networkx 2.6.3 2.6.3
numpy 1.21.4 1.21.4
opencv-python 4.5.4.58 4.5.4.58
packaging 21.3 21.3
pip 21.1.2 21.3.1
pyparsing 3.0.6 3.0.6
python-constraint 1.4.0 1.4.0
python-dateutil 2.8.2 2.8.2
scikit-image 0.18.3 0.18.3
OK
Cancel
Apply
Git
TODO
Problems
Terminal
Python Packages
Python Console
Event Log
Python 3.9 (218007620_Honors_Project_Software)



Choose your new python environment and click ok



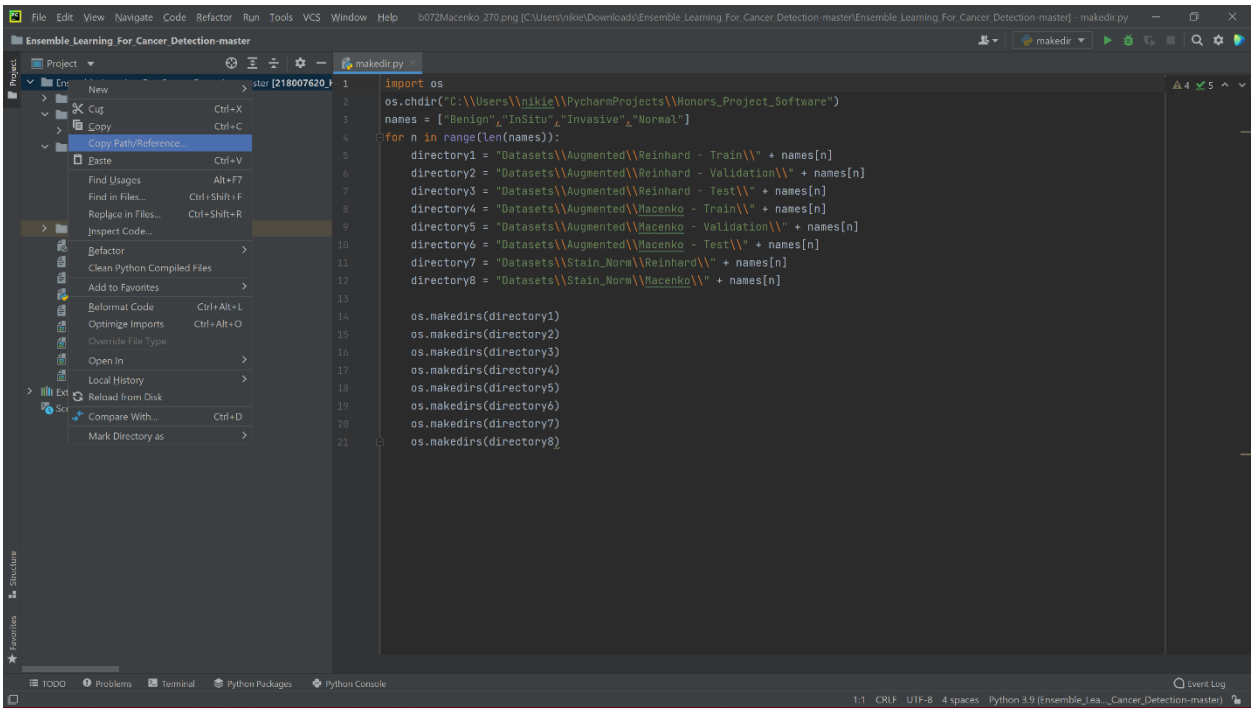


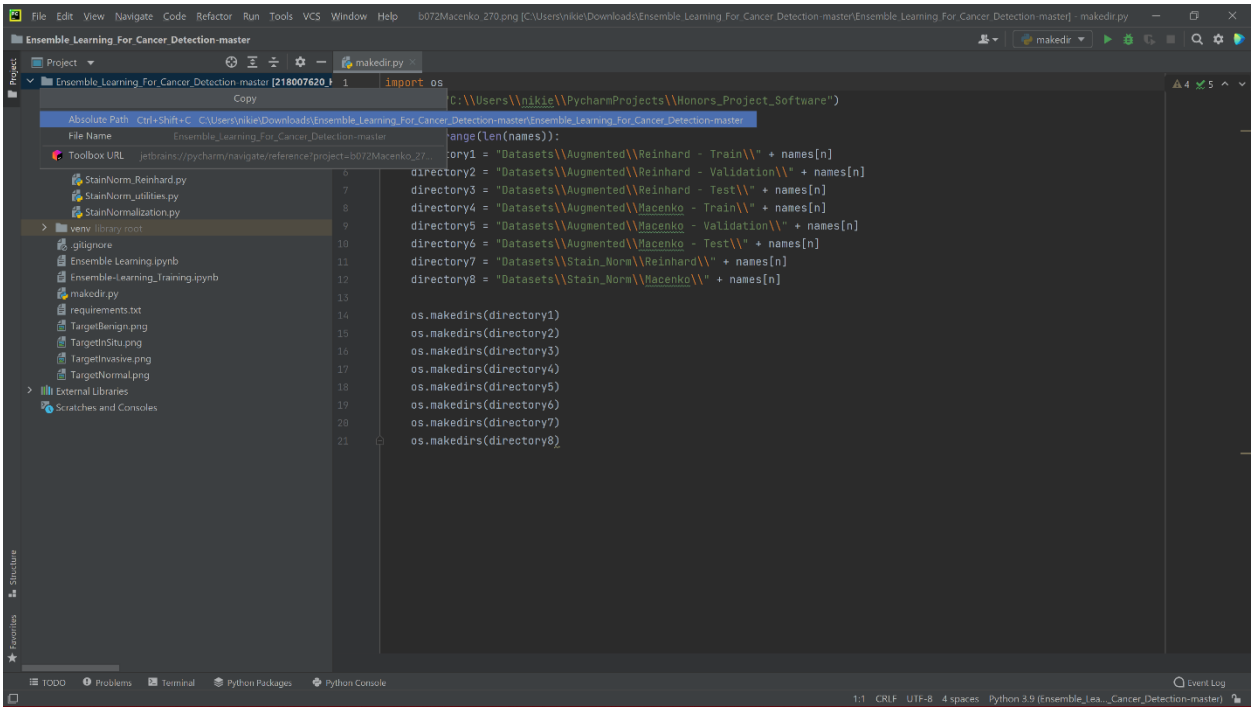
In the terminal run the following command -> pip install -r requirements.txt

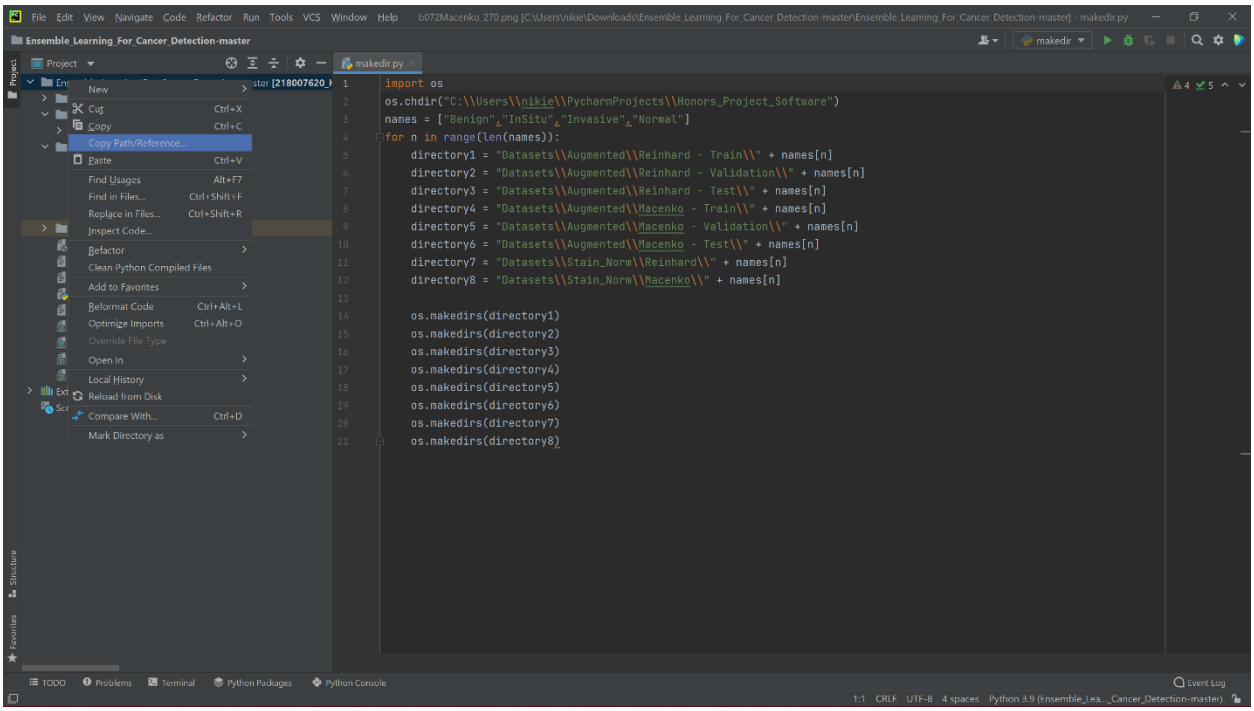
Your Environment is now configured to run the programs.

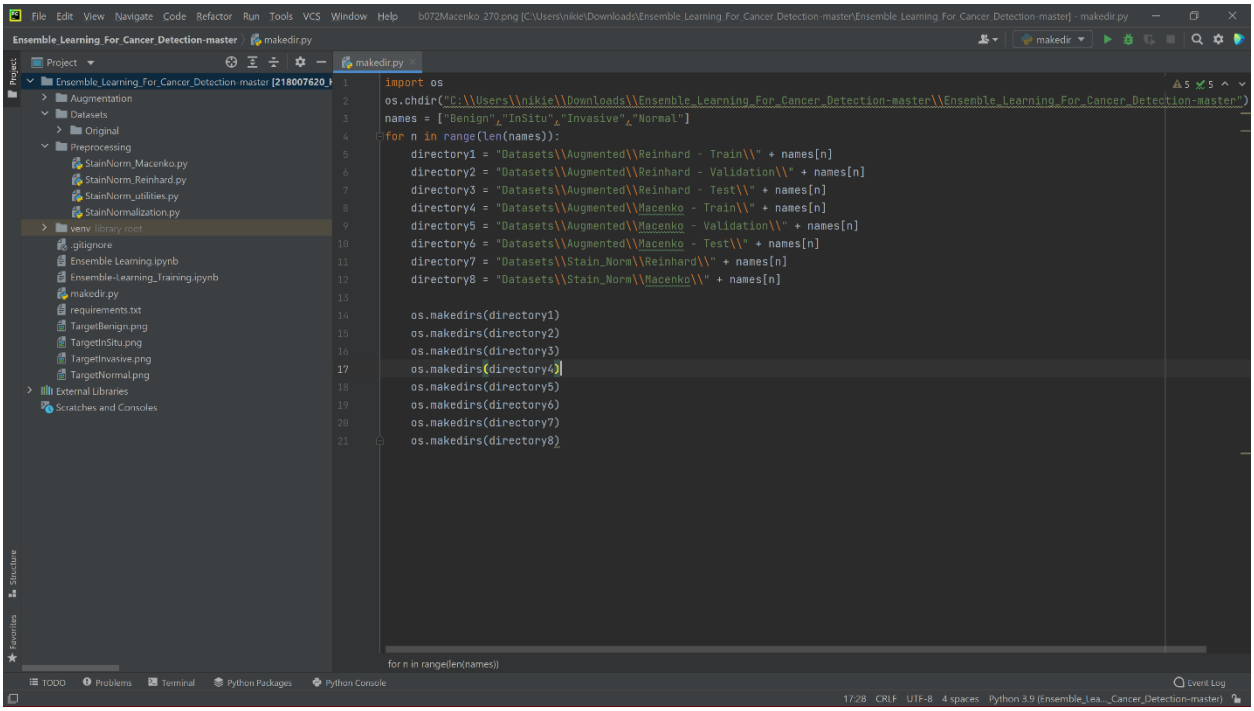
SETP 1 PREPROCESSING – STAIN NORMALIZATION AND AUGMENTATION.

Change the line os.chdir to the current directory of your program. Following screenshot is a shortcut





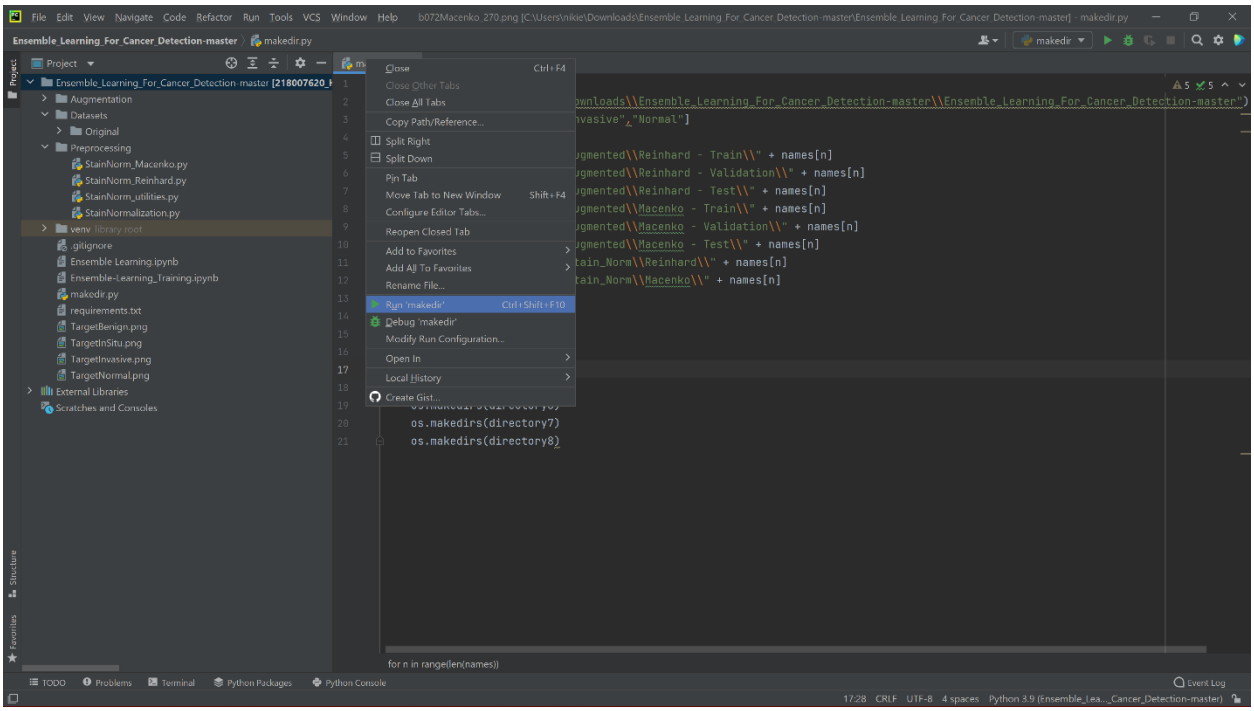




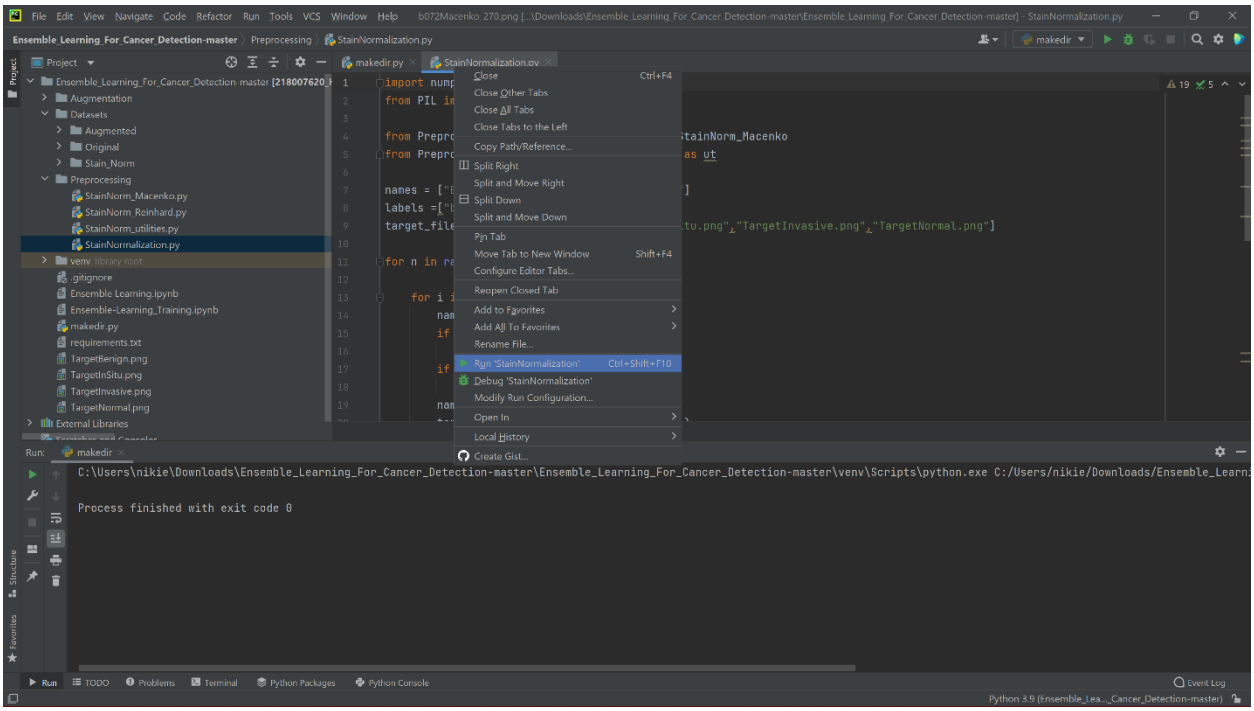
For safety when using directories always use “\\”

Once you’ve made these changes

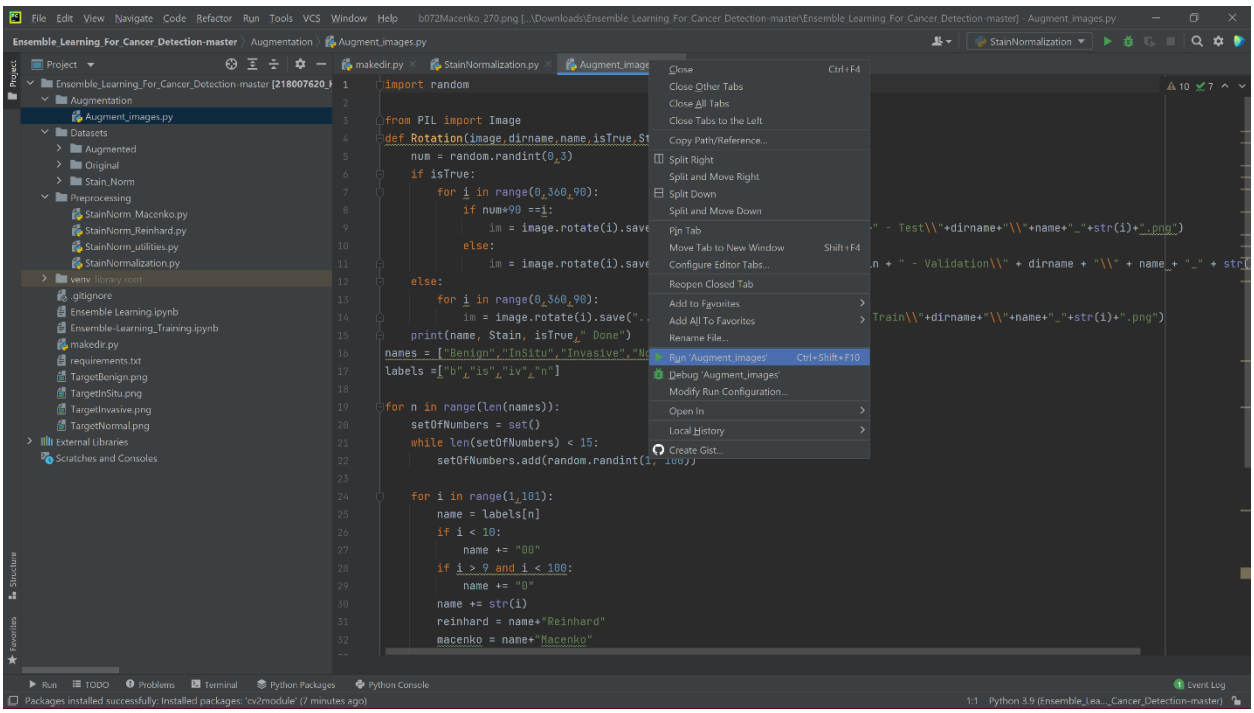
Run the file makedir.py first to create all necessary directories for reading and writing



Once all necessary directories are created you can process to run the StainNormalization.py file



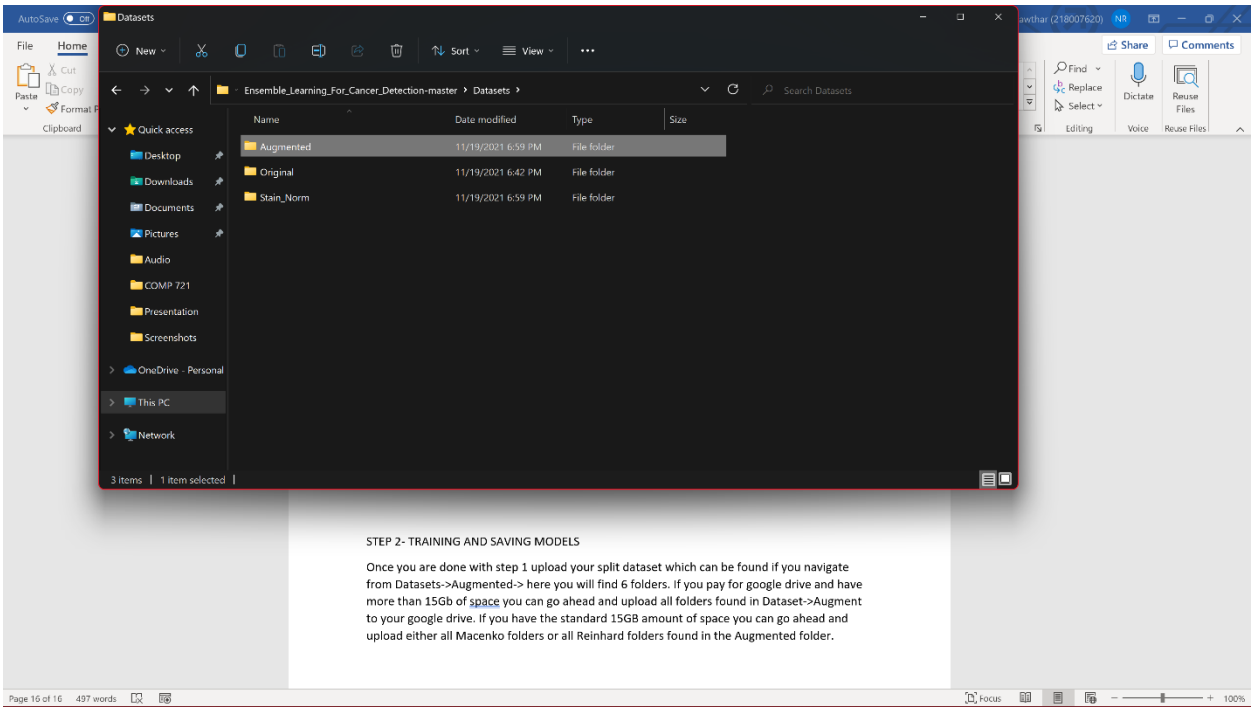
Once The StainNormalization.py file has finished running. You can proceed to run the Augment_images.py file

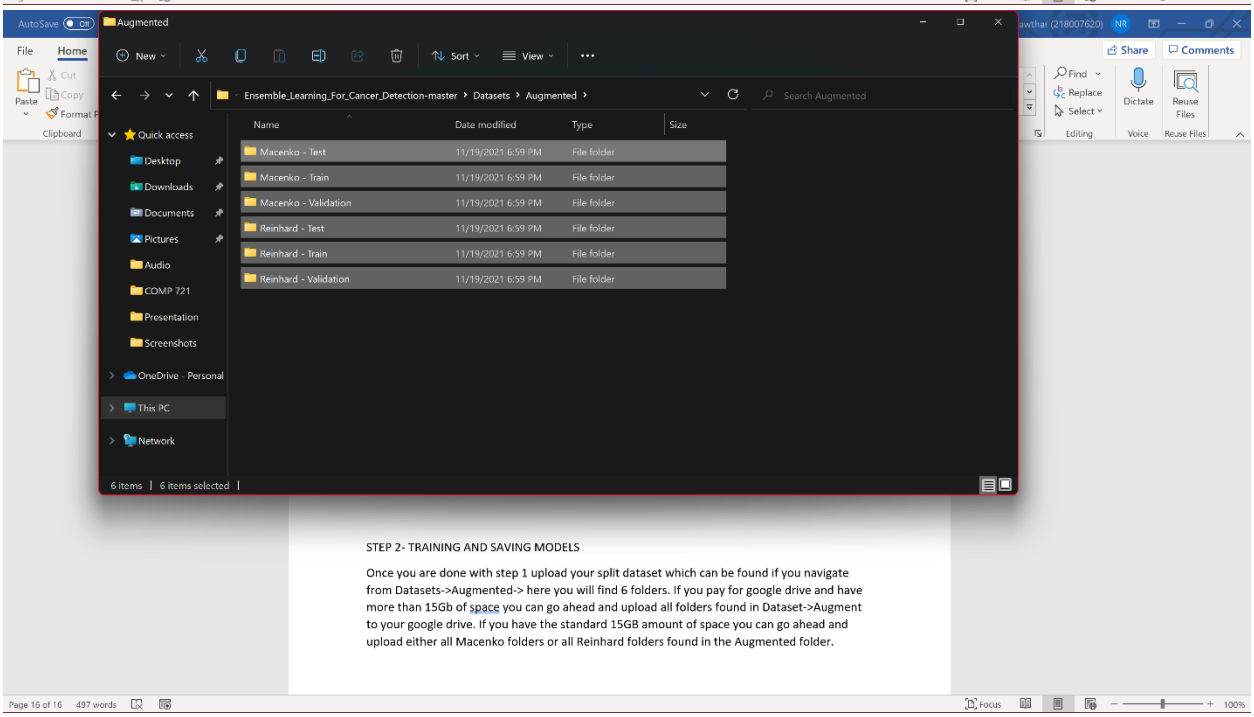


You are now all set to begin training on Google Colab the next few steps will show you how to train a model and run the ensemble learning model.

STEP 2- TRAINING AND SAVING MODELS

Once you are done with step 1 upload your split dataset which can be found if you navigate from Datasets->Augmented-> here you will find 6 folders. If you pay for google drive and have more than 15Gb of space you can go ahead and upload all folders found in Dataset->Augment to your google drive. If you have the standard 15GB amount of space you can go ahead and upload either all Macenko folders or all Reinhard folders found in the Augmented folder.





Upload Ensemble-Learning_Training.ipynb and Ensemble Learning.ipynb to your google drive and open them in colab

*if you are using the Macenko dataset please update all names in the file i.e if something is named Reinhard change it to Macenko. This would've been automated but because you may change locations and naming structure this wass the only possible way

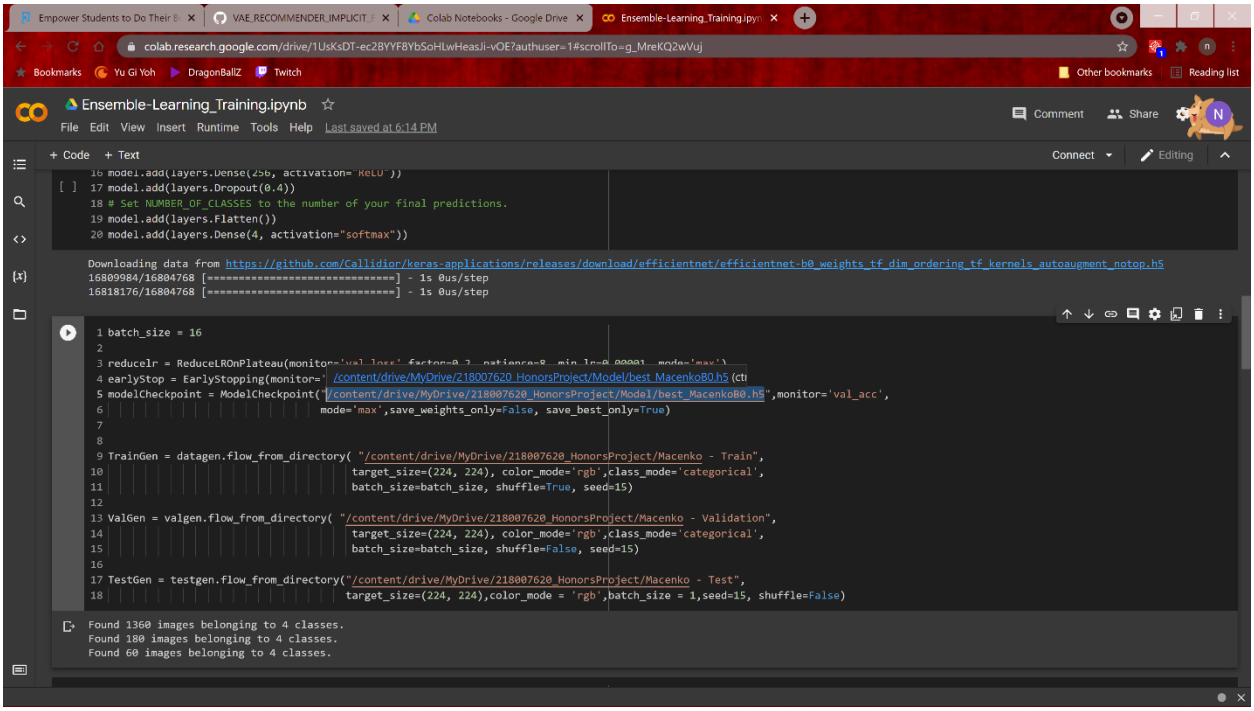
Once you have opened Ensemble-Learning_Training.ipynb and uploaded the datasets to Google Colab you can make a few changes before running. The following directories need to be changed to where you stored the dataset in your google drive.

Step A - The first highlighted text deals with where you would like to save your model and the name to associate it with. The default directory for google drive is /content/drive/MyDrive/. After /MyDrive/you may either same the model as a .h5 file or enter the location along with a name for your saved model

*Please save models using following guidelines best_"Macenko/Reinhard"B"0-7 depending on the architecture you chose".h5

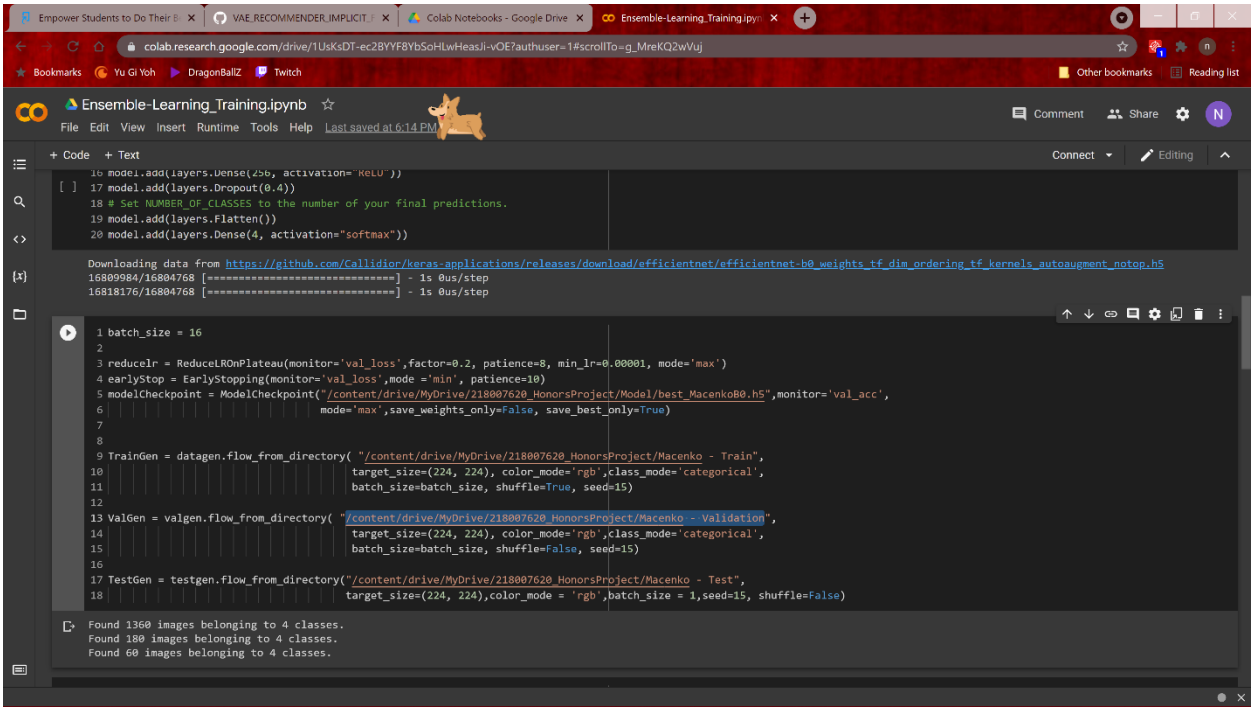
Example if youre using the Macenko dataset and using efficientnet architecture B5

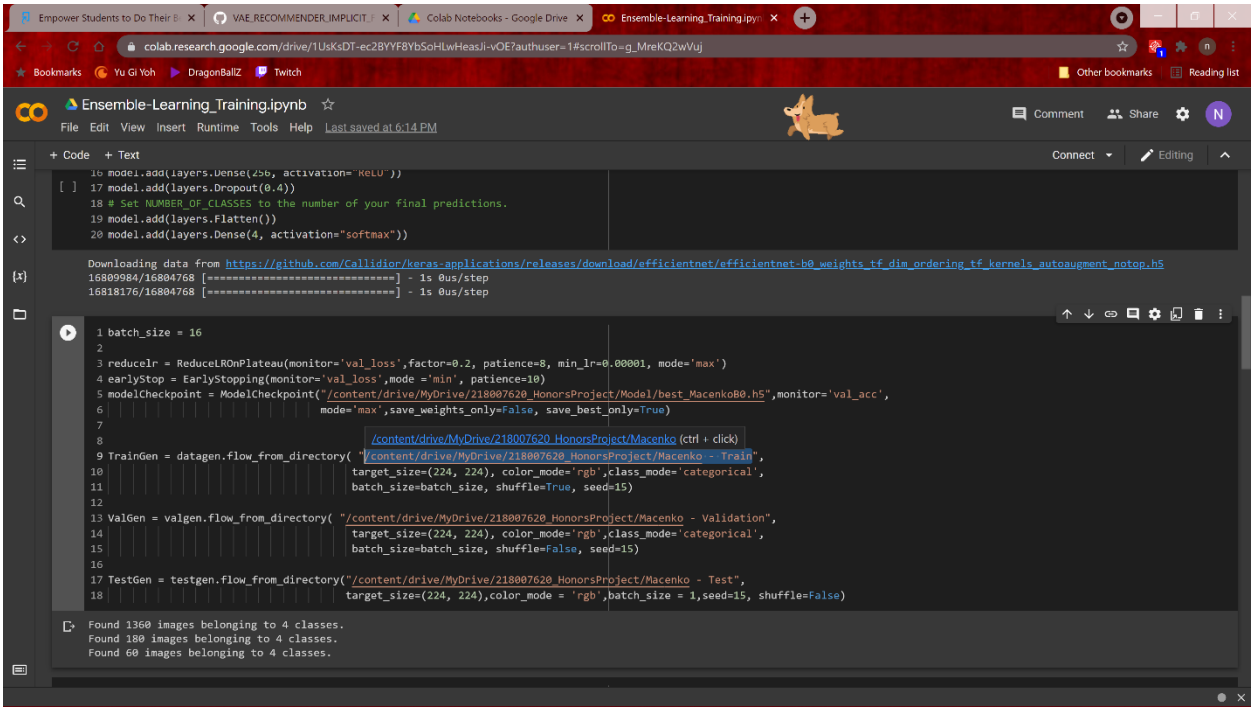
Best_MacenkoB5.h5

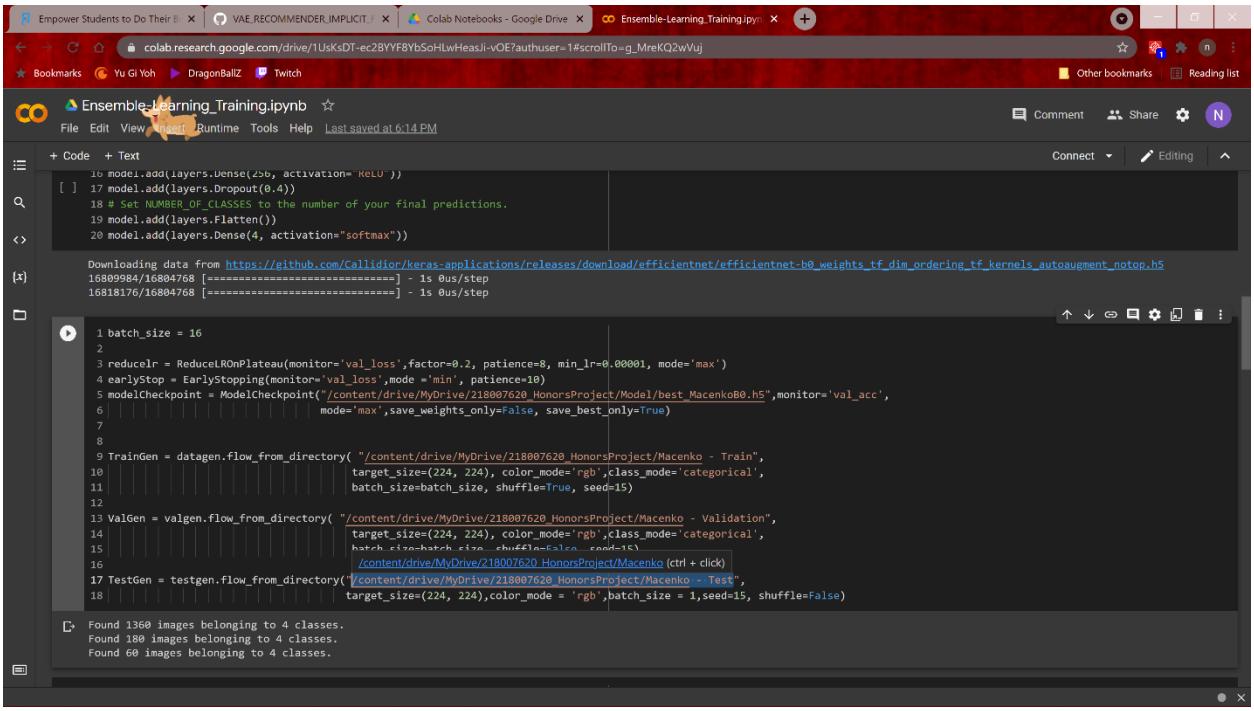


Step B - The second, third and fourth highlighted texts deal with the location of the training, validation and testing data you can go ahead and change these to the location you have them stored in google drive.

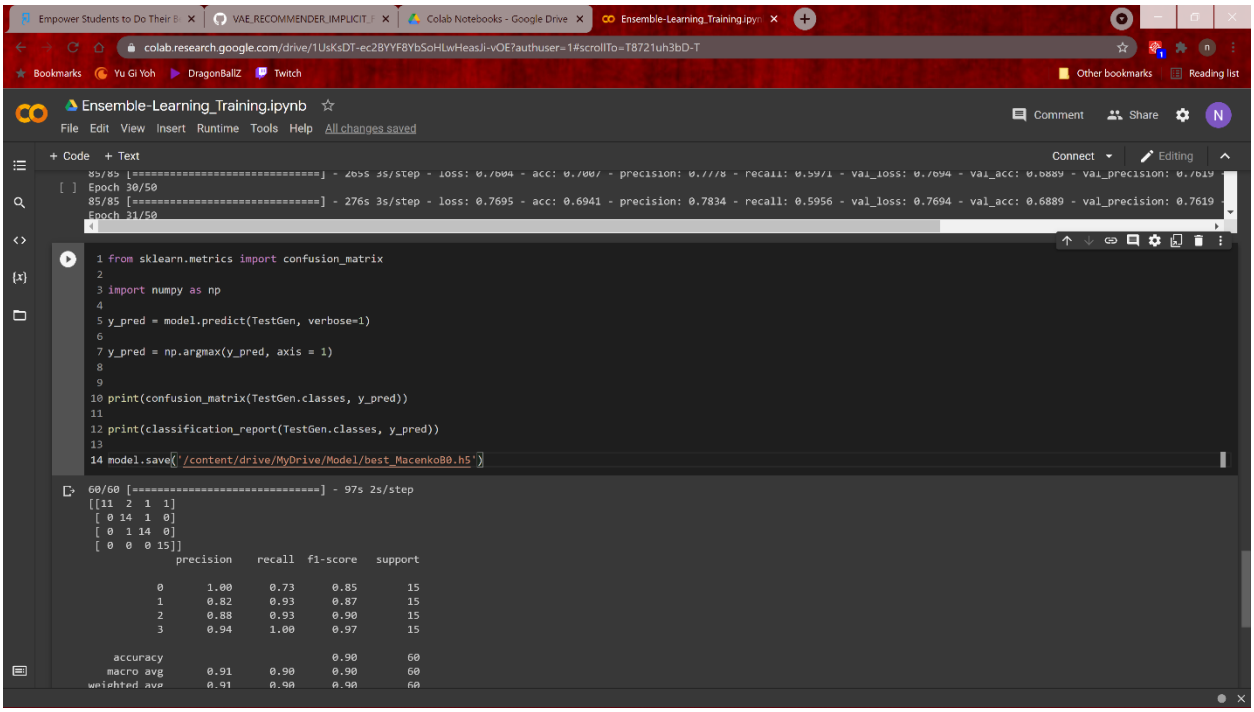
*note /content/drive/MyDrive/ is the root and needs to be added when accessing files







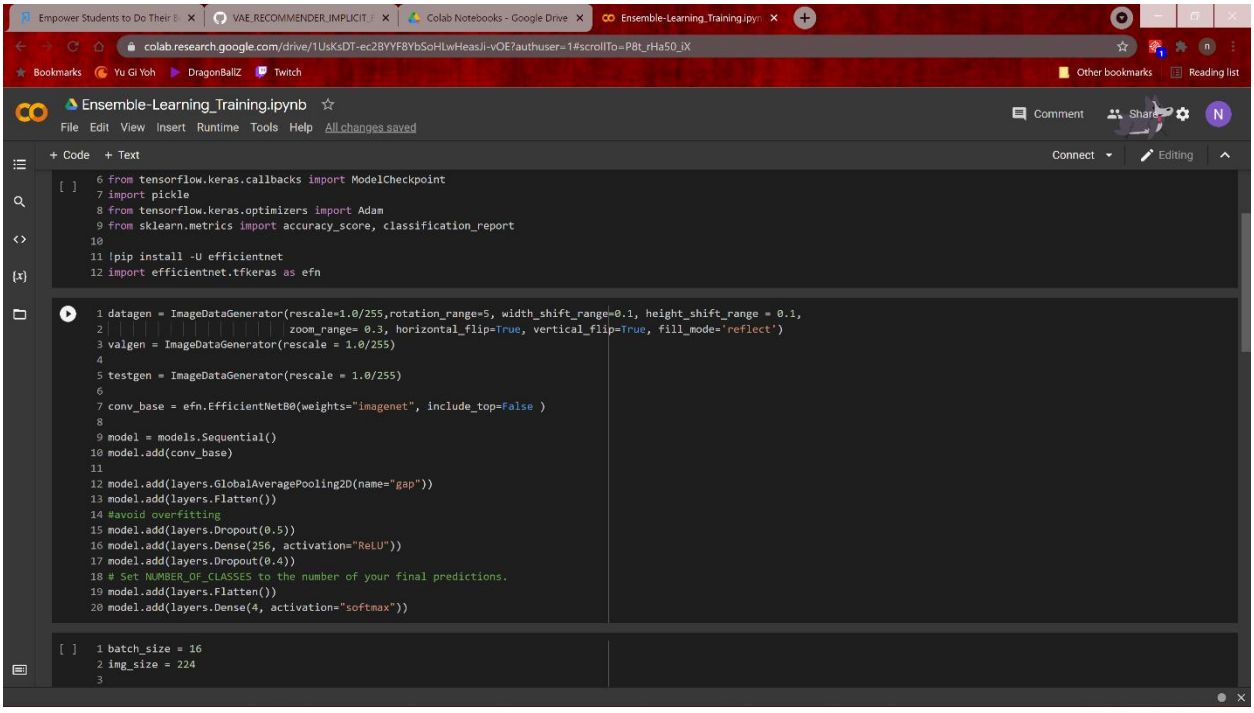
The last adjustment that needs to be made is again where you are saving the model in Step A- go ahead and copy that directory and past it here.



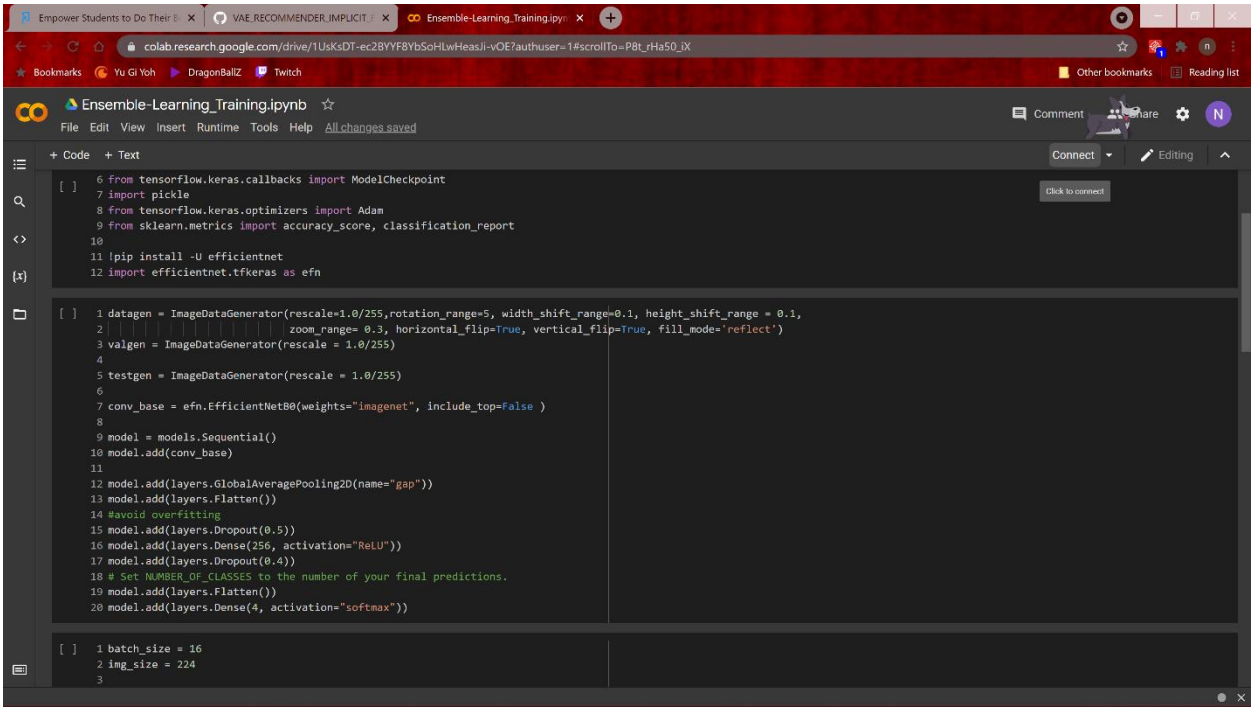
*Note there are different EfficientNets architectures. Each with a different img_size required. Depending on the model you choose change the img_size variable in block 4 to the corresponding size

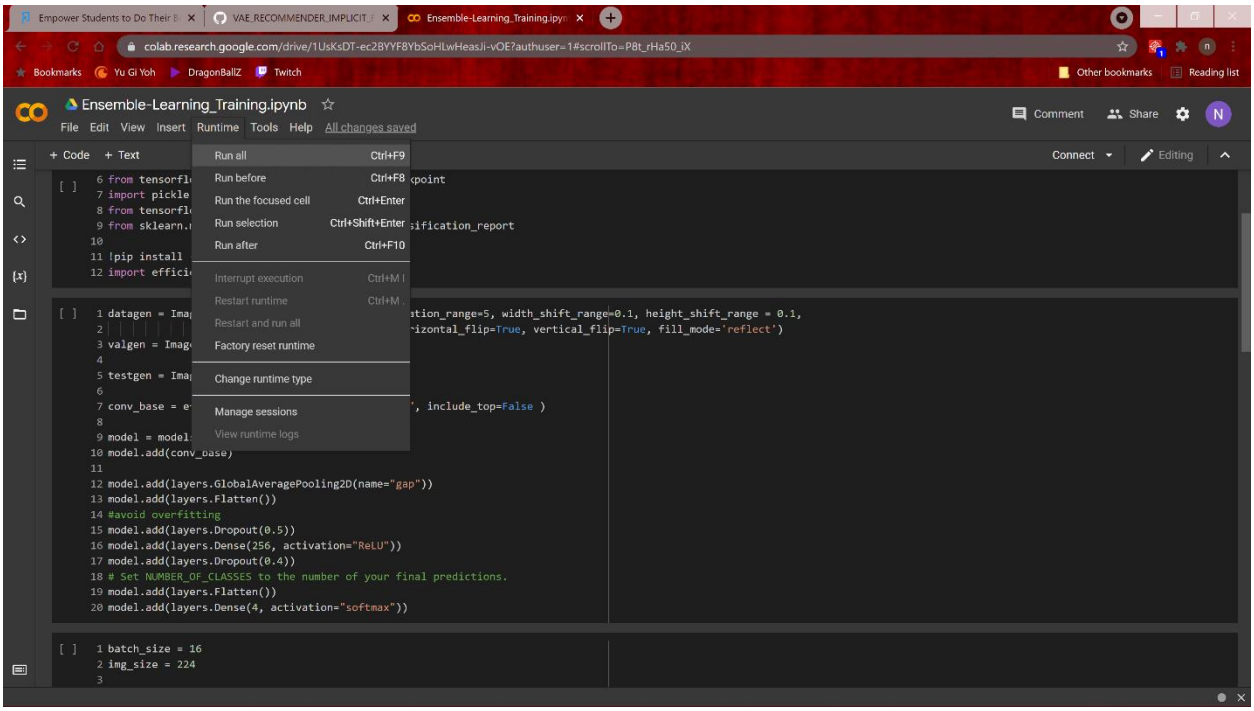
B0	224
B1	240
B2	260
B3	300
B4	380
B5	456
B6	528
B7	600

The final change is the selection of what model you want to train. On line 7 of code block 3 the number after B can be changed so you can choose different architectures to train. The range is 0-7 and remember to update the img_size to the corresponding image size needed for each architecture.



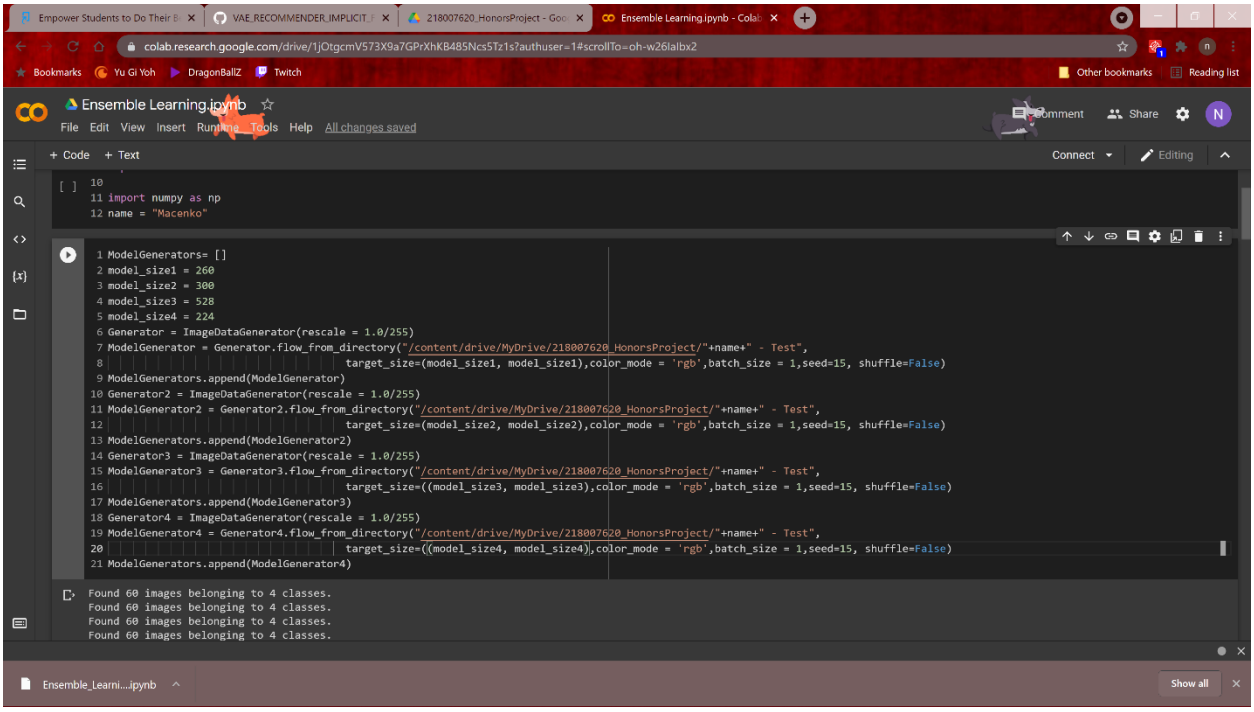
Once all these changes are made connect to a run time and click run all





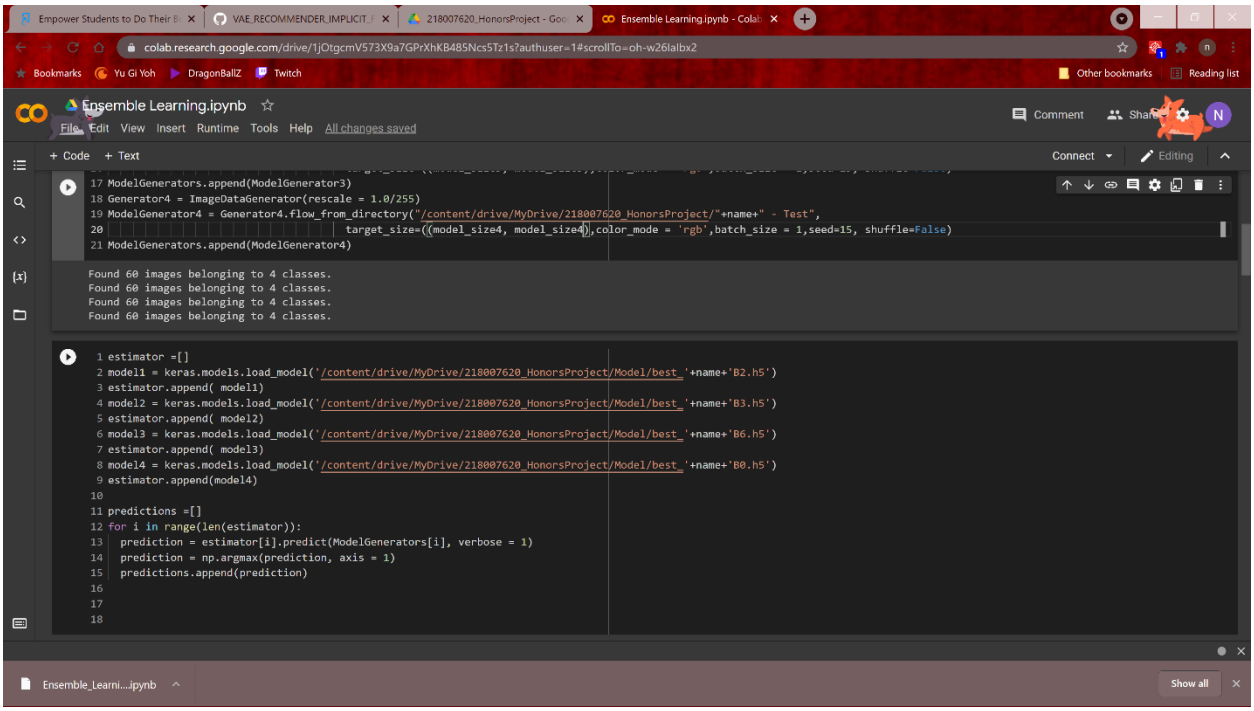
Once you have done training a minimum of 4 different models you may proceed to use the Ensemble Learning.ipynb file

A few changes need to be made here as well



Firstly for each ModelGenerator please pass through the location you have placed the test set in on google drive. Please try to remember the model you plan on using and update the model_size variable to the corresponding efficientnets architecture used. If you plan on using the B5 model for ModelGenerator2 then change the size for model generator 2 to the appropriate size.

The final change you need to make is the location you stored your models. Once again this is dependent on the Model Generator sizes you updated. The program will perform poorly if it does not get the required target size variable. For each model in the below code update the location of each of your chosen 4 models. Once you have updated this you may click connect and run all just like you did in the previous notebook



*Please note the models take about 5-10 hours to train. It may require tedious amounts of time. Disconnections may result in re running the code. Please do not run if you do not have a UPS (due to loadshedding) as it may be frustrating to run and wait all over again.

Hope this helps and feel free to contact me via email – nikielramawthar@gmail.com if you are unclear about certain details pertaining to this.