**Programming Final Exam**

**Introduction to Deep Learning**

20250612

You need to submit the final trained model along with the test results file (.csv) by **2025/6/24 23:59** to ecourse2. The test results file (.csv) should be submitted to Kaggle at the same time. The two .csv file submissions (one on ecourse2 and one on Kaggle) should be the same one. This part of the final exam covers 50% of your final grade.

#### Go to the following dataset and click on the “Starred” in the upper right corner.

#### <https://github.com/ICCC-Platform/Air-Pollution-Image-Dataset-From-India-and-Nepal>

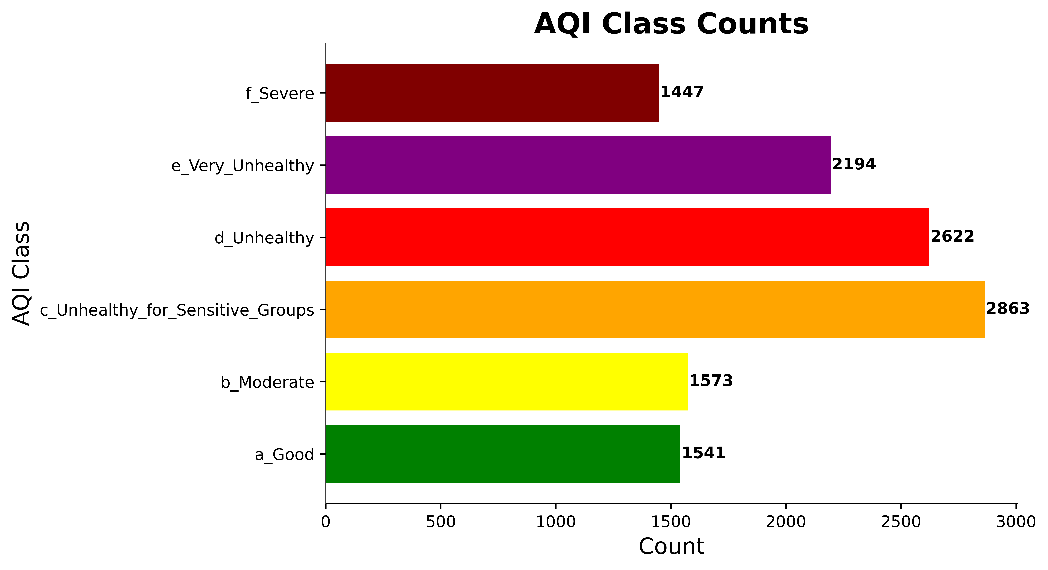
#### Download the dataset of air pollution images from GitHub/Kaggle from the above site.

Contents of dataset: 12,240 images of size 224\*224 for air pollution detection, collected from cities of India and Nepal. There are two folders: Combined\_Dataset and Country\_wise\_Dataset. Use the All\_img folder in Combined\_Dataset for training your model, and use Combined\_Dataset/IND\_Nep\_AQI\_Dataset.csv file to retrieve the details for each image.

Files

* **Readme.docx**: Readme file
* **Combined\_Dataset/All\_img: All images**
* **Combined\_Dataset/IND\_Nep\_AQI\_Dataset.csv**: Use this CSV file to read the information about all the images.
* **Dataset\_for\_AQI\_Classification:** In this folder, you can get three CSV files with details of the images that are to be used as training set, validation set, and testing set while training your model. The CSV files are as follows:
  + train\_data.csv for training set
  + val\_data.csv for validation set
  + testing\_data.csv for testing set

**Recommended:** Use above file to train the model, also you need to pass the image folder path (Combined\_Dataset/All\_img) for training the image classification model.



**Task**

* **Mandatory tasks:**
* You need to create a deep neural network model that can classify each image of the dataset into one of the above 6 classes with high accuracy.
* You need to show us the final (training and testing) accuracies in terms of average ROC AUC scores for the 6 classes, with graphs.
* You can also show the F1-score, overall MAE, overall RMSE and any other metrics you think reasonable.
* You need to show the 6x6 confusion matrix for the 6 classes in training, validation and testing, that is, 3 matrices.
* **Bonus Task:** if your model can also predict the AQI and the PM2.5 for each of the test images, then we can give you bonus depending on your prediction accuracy.
* **Additional Bonus:** if your model can also predict all the other air quality metrics including AQI, PM2.5, PM10, O3, CO, SO2, NO2, we can give you more bonus!

**Combined\_Dataset/IND\_Nep\_AQI\_Dataset.csv (Attributes)**

There are 14 attribute labels for each image as follows.

Columns:  
# Location

# Filename

# Year

# Month

# Day

# Hour

# AQI

# PM2.5

# PM10

# O3

# CO

# SO2

# NO2

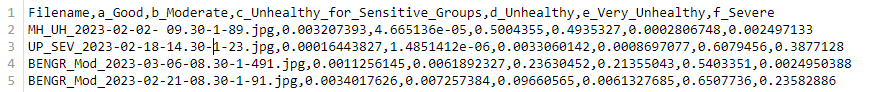
# AQI\_Class   
(f\_Severe, e\_Very\_Unhealthy, d\_Unhealthy, c\_Unhealthy\_for\_Sensitive\_Groups, b\_Moderate, a\_Good)

* **Please use these columns for the bonus part:**
  + **To predict the AQI and the PM2.5 for each of the test images, please use:**
    - # Filename
    - # PM2.5
  + **To predict all the other air quality metrics including AQI, PM2.5, PM10, O3, CO, SO2, NO2 please use these columns:**
    - # Filename
    - # AQI
    - # PM2.5
    - # PM10
    - # O3
    - # CO
    - # SO2
    - # NO2
* **File format to be submitted on Kaggle:**

Students have to do prediction only on test\_data.csv file and submit the csv file to the competition on the Kaggle.

**CSV file should contain all these columns:**

* + **Filename,**
  + **a\_Good,**
  + **b\_Moderate,**
  + **c\_Unhealthy\_for\_Sensitive\_Groups,**
  + **d\_Unhealthy,**
  + **e\_Very\_Unhealthy,**
  + **f\_Severe**

**and their values as shown in the screenshot below:**

Corresponding values

Columns

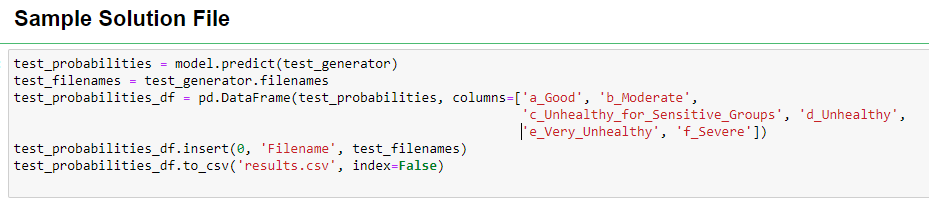
**Kaggle competition :**

You need to submit the test results file (.csv) to Kaggle. Kaggle will automatically rank your score. Below is the competition link and description.

[https://www.kaggle.com/t/b860260f92f04a07ba36976c3c5146b3](%20https:/www.kaggle.com/t/b860260f92f04a07ba36976c3c5146b3)

**1. Save Test Results in a CSV File**

Students can obtain this file after training their models using "train\_data.csv" for training and "val\_data.csv" for validation.

Students can use the following code to test their data and save the file in CSV format.

**2. Upload Results to Kaggle**

Go to the Kaggle competition page and log in to your Kaggle account.

Find the "Submit Predictions" button and submit your test results file (.csv).

**3. Ranking**

You can view the ranking on the leaderboard. Additionally, you need to submit your account name, final trained model, and the test results file (.csv) to ecourse2.