**COT 6930: Advanced Data Science Capstone**

**Project Option #1**

**Deep Neural Network for COVID-19 image classification**

The project #1 intends to apply deep neural network to classify X-Ray images into four different categories. A dataset with four groups of X-Ray images will be provided. Your project will carry out research study on the images, and report your results and findings.

A picture containing X-ray film

Description automatically generated A picture containing X-ray film

Description automatically generated  A picture containing X-ray film, person

Description automatically generated

COVID-19 Lung-Opacity Viral Pneumonia Normal

**Data source and tasks**

1. Please download the COVID-19 X-Ray images (700+MB zip file) from the following URL

<http://www.cse.fau.edu/~xqzhu/courses/cot6930/dataset/covidImages.zip>

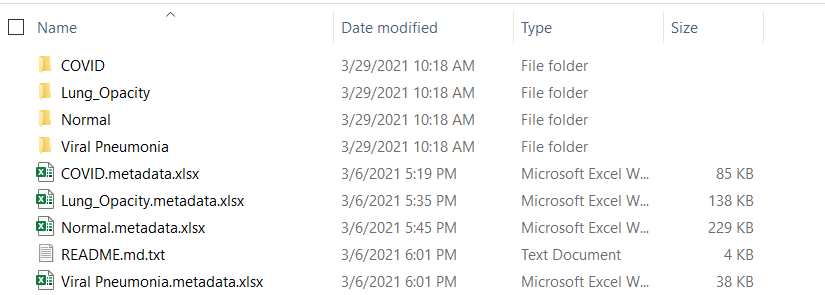
The original images were downloaded from the Kaggle site (<https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>). You can also download the images from Kaggle.

Kaggle also provides useful resources to explain certain category of images (such as lung opacity)

<https://www.kaggle.com/zahaviguy/what-are-lung-opacities>

Please refer to these resources to obtain domain knowledge, and understand the difference between four groups of images.

1. Unzip the downloaded (zip) file. There are four sub-folders, four excel files, and a readm.md.txt file. The readme file shows original source of the data, and detailed information about each group of images. Each excel file shows the size, format, and the URL source of the image. Each subfolder includes images in each group.



1. Please follow “[CNN Image Classification [pdf]](https://canvas.fau.edu/courses/50073/modules/items/1248564)” notebook to create your own CNN image classifiers to classify images into four categories.
2. The implemented project should include following major tasks (required tasks):
   1. The project must include a cross-validation (either 5-fold or 10-fold). The reports must include both training and test results.
   2. Please vary the network structure, such as the number of convolution layers, the filter sizes, dropout layers (and different dropout degrees), and report the best model performance.
   3. Please report the filters and their feature maps from at least one convolutional layer. Select one sample images from each class, and observe the filters and the feature maps.
3. Additional suggested tasks (not required, but suggested):
   1. Because deep neural network takes tremendous amount of time for training, and with limited resource, you may not be able to train a good classifier. Therefore, you can leverage pretrained network (such as VGG16 or VGG19: <https://keras.io/api/applications/vgg/#vgg16-function>) or Resnet (https://keras.io/api/applications/) to obtain features, and then add classification models.
   2. You can compare your model vs. pretrained network for COVID-19 image classification.
   3. You can also add changes to the network (when fine tuning the pretrained network), such as add drop out layer, to compare the algorithm performance.

**Project Report Instruction**

1. You will need to follow the project report instructions to prepare your report [40 pts]. The project report instructions will be released soon.
2. You also need to prepare a poster to report your research motivation, designs, and findings [20 pts]. The term project report poster template will be published on Canvas soon.