CS 542 Class Challenge: Image Classification of COVID-19 X-rays

Total Points: 100

In this class challenge, we will classify X-ray images. The data we will use has been collected by Adrian Xu, combining the Kaggle Chest X-ray dataset with the COVID-19 Chest X-ray dataset collected by Dr. Joseph Paul Cohen of the University of Montreal. The data can be downloaded **here**. When you extract the data you will have two folders: ***two*** that will be used for a binary classification task (Task1), and ***all*** that will be used for multi-class classification (Task2). An ipython notebook template is provided for each task.

- **[30 points]** ***Task1*** Train a deep neural network model to classify normal *vs.* COVID-19 X-rays using the data in the folder ***two.*** Starting from a pre-trained model typically helps performance on a new task, e.g. starting with weights obtained by training on ImageNet. After training is complete, visualize features of training data by reducing their dimensionality to 2 using t-SNE. If your extracted features are good, data points representing a specific class should appear within a compact cluster.

- **[30 points]** ***Task2*** Train a deep neural network model to classify an X-ray image into one of the following classes: normal, COVID-19, Pneumonia-Bacterial, and Pneumonia-Viral, using the folder ***all.*** Explore at least two different model architectures for this task, eg. AlexNet *vs.* VGG16. After training is complete, visualize features of training data by reducing their dimensionality to 2 using t-SNE. If your extracted features are good, data points representing a specific class should appear within a compact cluster.

- **[10 points]** ***Challenge*** How well your best model performs with respect to the class.

- **[30 points]** ***Report***
    - **[5 points]** Describe the architectures used in detail: layers, layer dimensions, dropout layers, etc. for both tasks. List the optimizer, loss function, parameters, and any regularization used in both tasks
    - **[10 points]** Comparison of the performance of different architectures for the second task and relating this to the architecture and parameter settings used
    - **[10 points]** Plot and comment on the accuracy and the loss for both tasks
    - **[5 points]** Plot and comment on the t-SNE visualizations
    - **[Bonus: 5 points]** Run the training on a GPU on the SCC cluster and include a CPU *vs.* GPU training time comparison by taking snapshots from your terminal

Submission:

Please complete the class challenge and submit a pdf file containing: task 1 code, task 2 code, and the report on GradeScope. The deadline for this class challenge is: Apr 29,2021.