**COVID-19 Image Classification**

**UVA School of Data Science DS 6050 Deep Learning**

**Professor Stephen Baek**

**Paul Hicks, Jay Hombal, Sud Luthra, Matt Dakolios**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Project Proposal**

*Motivation*

Our aim for this project is to build a deep learning model that can correctly identify chest x-rays into three classes: COVID-19, Viral Pneumonia, or Normal. We were initially interested in a project related to public health and believe that this project aligns with the spirit of the Machine Learning for Virginia Project’s mission of using real world situations with applicability to the state of Virginia in order to prepare us for machine learning in the real world. Additionally, we feel that COVID-19 in many ways has placed a significant role in our graduate experience in the UVA School of Data Science; this project seems befitting of our growth as aspiring data scientists in an online program in the age of telecommunications and a global pandemic.

*Dataset*

<https://www.kaggle.com/pranavraikokte/covid19-image-dataset>

We seek to utilize a dataset from Kaggle that includes a test set of 26 COVID-19 images, 20 Viral Pneumonia images, and 20 normal images. The provided training set includes 111 images of COVID-19, and 70 each of Viral Pneumonia and normal. After meeting with Dr. Baek, who advised that this project would be challenging without more data to train a deep learning network, we have considered using a dataset such as the one below in order to supplement our training process.

<https://www.kaggle.com/nih-chest-xrays/data>

*Related Work - Literature Review*

<https://pubmed.ncbi.nlm.nih.gov/33037291/>

The study above, found from the National Library of Medicine, details a deep learning neural network application that uses CT scans to predict COVID-19 in patients. The authors attempted the task using 16 different convolutional neural networks and found that DenseNet-201, the deepest in their study, produced the best results in terms of accurately classifying COVID-19 pateinted from non COVID-19 patients from CT scans. They also looked at the effects of using data augmentation, and found that accuracy decreased when using data augmentation. This study, and others we may find to be similar, could prove to be valuable in terms of establishing best practices, norms and borrowing ideas for model building, tuning and visualization.

*Intended Experiments*

At this point in the process, we are still learning about the best practices and methods for building and tuning a neural network that would be effective with image classification of this kind. Certainly deciding how many hidden layers and nodes to include will be central to our tuning process, as well as recently learned regularization strategies. We anticipate attempting data augmentation, and we will look forward to comparing performance and accuracy using this strategy. Assignment 2 in Module 6 will likely prove to be invaluable to us in this pursuit, as we will then learn and hone our skills with an image system.