COMP4471 Project Proposal

LI Tianle, SUN Yushi

tliax@connect.ust.hk, ysunbp@connect.ust.hk

Nowadays, COVID-19 is spreading rapidly all around the world and causing jitters among the population. Fast and accurate diagnostic methods are urgently needed to combat the disease. Thereby, we want to exploit the cutting-edge technology learned from the course to accelerate the diagnosis of pneumonia from chest X-ray images to save time for septic patients and enhance disease control.

The key reason why we opt for chest X-ray images instead of CT scan is that CXRs are cheaper and more accessible for people in rural and isolated areas. Thus, this project will enhance the COVID-19 diagnosis for all people regardless of their social class and financial status.

The dataset we will use is from Kaggle: Dataset of validated Chest X-Ray images for pneumonia disease classification. It consists of two general classes: the normal chest X-Ray images and the ones with pneumonia which contain two subclasses. To outperform the existing classification models for the datasets, we want to further distinguish the images with symptoms as bacteria pneumonia and virus pneumonia. We also found a small dataset of COVID-19 CXRs scrapped from publicly available research which could be utilized to provide more pragmatic and rewarding research results if we can solve the challenge ultimately.

Due to the limited amount of data we can collect, we are likely to implement data augmentation in the first step to enlarge the training set, and try transferable learning with diverse eminent architectures such as an Inception, ResNet and so on to obtain the baseline for the validation accuracy. Then, we may attempt to build our own original model to improve the accuracy with the goal of extracting the features with lower quality in CXRs images. To realize the target, we will do some research towards the appropriate architecture for highly skewed data and subclass classification with low quality featured images. (Try omni-supervised learning with data distillation)

To evaluate our results, we may plot some informative graphs of the relationship between crucial parameters and test accuracy and visualize the checkpoints' staged results with tensor board to make pinpoint evaluation about our model.