MAIS 202
PROJECT DELIVERABLE 2

Simaan Arshad

COVID19 X-RAY DETECTION

1. The initial project I had proposed is the Covid-19 Xray detection, which uses a Kaggle dataset available on https://www.kaggle.com/bachrr/covid-chest-xray. The dataset has images which will be used for training the model further in the course of time.
2. The dataset I'm working has been stated above, as of now there are no changes to the dataset.

Here is a list of each metadata field, with explanations:

- Patient Id (internal identifier, just for this dataset)
- offset (number of days since the start of symptoms or hospitalization for each image, this is very important to have when there are multiple images for the same patient to track progression while being imaged. If a report says "after a few days" let's assume 5 days.)
- sex (M, F, or blank)
- age (age of the patient in years)
- finding (which pneumonia)
- survival (did they survive? Y or N)
- view (for example, PA, AP, or L for X-rays and Axial or Coronal for CT scans)
- modality (CT, X-ray, or something else)
- date (date the image was acquired)
- location (hospital name, city, state, country) importance from right to left.
- filename
- doi (DOI of the research article
- url (URL of the paper or website where the image came from)
- license
- clinical notes (about the radiograph in particular, not just the patient)
- other notes (e.g. credit)

It has about 350 X-Ray images which will be useful to preprocess the data. I will be changing the dimensions and resolutions of the images before using it for training. Because some of the X-Rays might be of different dimension which can be an issue when the model gives out predictions for different dimensions for images.

3. The model I choose earlier was KNN but after a discussion with the TPM, I've decided to use CNN for my implementation. Clearly because it's most commonly applied to analyzing visual imagery. Which is what the dataset I'm dealing with is, X-Ray images. It will help in processing the connectivity pattern between neurons.
4. My next step would be get the preliminary results ready for the development of the model, and then compiling all the data. After which I will implement the CNN!