

Project Deliverable 1

1. Dataset

The project I choose is to detect and classify Pneumonia from chest X-ray images (<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>). The reason why I choose this project is China is struggling to contain the outbreaks of coronavirus which may cause respiratory illnesses such as pneumonia. By using this dataset, we can classify normal and Pneumonia chest, which can facilitate rapid referrals for patients.

2. Methodology

i. Data processing

The dataset is a set of images of normal and Pneumonia chests. It is not classified in many categories. I can just simply create a dataframe that has columns of index, image and label. Label 0 indicates normal and Label 1 indicates Pneumonia.

ii. Machine learning model

It is a classification problem, so I will use classification model to process my data. We haven't gone through classification algorithm yet. I will think about it as the course progresses. The main problem is how to use few samples to build a model of relative high accuracy.

iii. Final conceptualization

I will integrate my model in a landing-page webapp. Users basically can upload an image of X-ray chest to predict whether the chest is infected with Pneumonia. I will add more functions if time is enough.