

Deliverable 2 Description

Katty Lu

1. Problem statement

Predict whether a person has Covid-19 or not from the person's chest X-ray image, using CNN. My datasets include X-ray images of patients with Covid-19 and X-ray images of normal/healthy people.

2. Data Preprocessing

The two datasets that I will be using are:

1. <https://github.com/ieee8023/covid-chestxray-dataset>

The above dataset contains 584 covid-19 cases, however a small amount of them are CT images which are to be eliminated. I will only use the X-ray images.

2. <https://www.kaggle.com/raddar/tuberculosis-chest-xrays-shenzhen>

The above dataset contains 326 normal chest X-ray images.

For both datasets, I will slice them into 3 parts: training (60%), validation (20%) and testing (20%). The percentage in brackets is the proportion of data that I plan to use in each section.

3. Machine learning model

I will be using the convolutional neural network (CNN) model. For the moment, I am still trying to figure out how CNN works. Hopefully I can make some progress soon.