Deliverable 1 Description
Katty Lu

1. **Dataset:**
   https://github.com/ieee8023/covid-chestxray-dataset
   https://www.kaggle.com/raddar/tuberculosis-chest-xrays-shenzhen

   The first dataset contains patients' age, sex, offset, their x-ray or Ct images at different phases, and the type of pneumonia (Covid-19 or other diseases) they get.
   The second dataset contains 2 types of chest x-ray images: the normal ones and the ones with tuberculosis. I will only be using the normal ones from this dataset.

2. **Methodology**
   a. **Data Preprocessing**
      I want to use x-ray images from patients with Covid-19 and normal x-ray images to do a classification. I will find a way to eliminate all the images that are not X-ray, such as CT scan and MRI scan images, by going through the metadata.

   b. **Machine learning model**
      My plan is to predict whether a person has Covid or not from his/her/their chest X-ray image. I will probably use CNN image classification, but I am not sure for the moment, since I haven't learnt that yet. Hopefully it gets clearer with the course moving forward.

   c. **Evaluation Metric**
      I might use confusion matrix, accuracy or precision are potential evaluation metrics. I will decide which one to use when I dig deeper into the datasets and learn more about these evaluation metrics.

3. **Application**
   For the moment, my idea is to create a webapp that classifies chest x-ray image to detect whether the owner of the x-ray image has Covid or not. My idea is still vague for the moment, but I will update this part hopefully soon.