

# Sprint 4- Status Check In 2

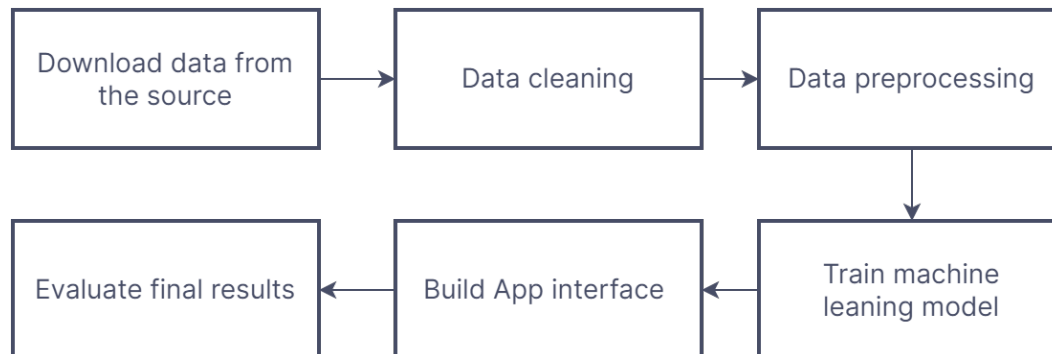
Chang Liu  
chl769@gatech.edu

## 1 PROJECT INTRODUCTION

Over the last two years, Covid19 has quickly spread worldwide, resulting in the Covid-19 pandemic which causes thousands and thousands of infections and death.

In this project, the aim is to leverage machine learning for Covid 19 classification and display the result in an App/Web page. The App/Web page will generate results based on patients' chest X-rays in real-time.

## 2 ACCCOMPLISHMENTS



*Figure 1*—Project Task

Based on the previously proposed timeline, I have been working on training the machine learning models in the past weeks.

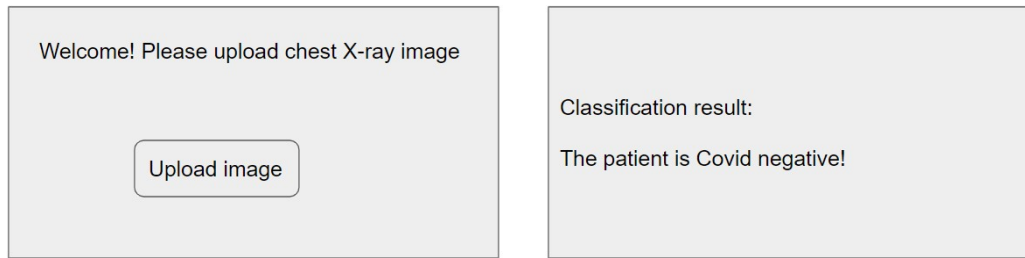
Data cleaning and preprocessing: 10/17 - 10/30

Machine learning model training 11/1 - 11/21

### 2.1 Download dataset

-After some search, I have downloaded the covid chest x-ray data set from the Kaggle challenge. It is the largest covid 19 dataset which consist of 3616 covid images and 10701 normal chest x-ray images.

Covid chest X-ray data has been downloaded from this source:



*Figure 2*—Machine learning interface (demo)

<https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>

## 2.2 Visualize the data

I have loaded the images and visualized them in a jupyter notebook. It looks like patients with covid have a more 'foggy' lung image in x-ray.



*Figure 3*—Normal image



*Figure 4*—Covid image

## 2.3 Data preprocessing

After some literature search, I decide to use 2 different methods for image pre-processing.

1. Just the baseline approach apart from size reduction.
2. Use histogram equalization to see if we can somehow increase the range of the pixel intensity and potentially a higher contrast.

## **2.4 Machine learning training**

Based on the discussion with my mentor, I have established several benchmarks for model evaluations. I used these 3 benchmarks to evaluate my model.

1. Accuracy for classification is greater than 0.5.
2. AUC for classification is greater than 0.5.
3. Also, because we don't want to miss any covid patients, I think specificity greater than 0.50 can be another benchmark.

I started with a subset of the large dataset for training. and tried to fine-tune the model to select some best models. I split the data into training and testing set and tried some different models on the training set.

## **3 CHALLENGES**

1. Currently, I don't have any major roadblocks. I am working on improving the machine model's performance.
2. Building the application might be difficult since I don't have much web development experience. I have been looking into some tutorials on FAST API and hopefully, I will be able to implement it after I train the model.

## **4 SPRINT PLANS**

In the next week, I am planning play around with different models and hyper-parameters and finalize the machine-learning model for covid 19 classification. Then I will try to build a web/app for displaying the results. To better prepare myself for that, I will look into some online materials for web/app developing.