Detecting and Predicting COVID-19 Impact on Patients through Chest X-Rays and Machine Learning

Gavin O'Sullivan, Ben Temkin

September 25, 2020

Introduction

While COVID-19 can often be asymptomatic in individuals that contract it, there is a large percentage of cases that require medical intervention or hospitalization. In these cases, it is critical to determine what level of care a patient requires and when they may need it. However, due to the lack of historical data from the new virus it could be difficult to form an accurate diagnosis for a given patient. Using a dataset of chest X-rays from patients with COVID or other pneumonias, our goal is to create a prognosis tool doctors could use as a supplemental source when working with patients that have or are suspected of having the coronavirus. While the current COVID machine learning studies have been focused on diagnosis, our project seeks to further this line of research by not only diagnosing COVID-19 cases but also determining their severity.

Problem Definition and Algorithm

One of the most important components of combating COVID-19 is speed. However, with limited medical resources hospitals can easily find themselves over capacity and short on supplies. The severity of each COVID-19 case changes over time, and as such assessing risk early is an important factor in preventing medical facilities from being overwhelmed. Our project looks to address the problem of medical supply and apparati deficiency by enabling doctors to distribute aid more efficiently. The formal inputs and outputs of our planned model are as follows:

Inputs
- Chest X-ray image(s) from patient
- Optional metadata about the individual to aid in classification and prediction

Outputs
- Diagnosis of patient (COVID-19, Pneumonia, Healthy)
- Prognosis of illness to predict likelihood of survival and level of care the patient may need (probability of requiring intubation)

Dataset

Our main dataset for this project is a collection of chest x-ray images from patients that have or are suspected of having COVID-19 or other pneumonias. It currently consists of 470 patients and 931 corresponding images for these patients (multiple chest views for patients) that have been collected from different hospitals and public sources with more data being added regularly. The data is labelled in a database csv file that contains metadata about each of the patients as well as the outcome of their illness (survival, intubation required, etc.). In addition, we have a Kaggle dataset with thousands of chest x-ray images that are either healthy or have a form of pneumonia. This dataset is also labelled and we aim to use it to help train the diagnostic piece of our program to determine the type of illness a patient may have based on their chest x-ray. We do not expect to need special hardware for this project, as all members of our group have modern laptops that we believe will be capable of training the image classification algorithms.

Related Work

In this[1] machine learning study, researchers used a classification approach to distinguish between COVID-19 cases and non-COVID-19 cases using chest x-ray images. It was created with the intention to be used as a diagnostic tool to quickly identify what illness the patient has. Their method "used a dimensionality reduction approach to generate a model with an optimized set of synthetic features that can distinguish COVID-19 images with an accuracy of 94% from non-COVID-19 cases."[1]. Our project aims to expand on this study and build a prognostic tool in addition to the diagnosis that can be used to predict what level of care a given COVID-19 patient may need based on a chest x-ray image and potential metadata about the individual.

Another[2] study is using AI to analyse images of COVID patients to allow doctors to diagnose and understand COVID-19. Mainly for reaffirmation purposes, the project aims to "give physicians an edge and allow them to act with more confidence while they wait for the analysis of a radiologist by having a digital second opinion confirm their assessment of a patient's condition"[2]. The secondary goal of the project is to create a repository of COVID-19 data, both images and analysis, for further studies. Our project plans utilize the images and metadata compiled to extrapolate further information about an individual's case rather than simply determining the cause of a patient's symptoms.

Bibliography

[1]   COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7273278/

[2]   COVID-19 image data collection. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[3]   Chest X-Ray Images (Pneumonia). [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia