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

CS 254 Project Report

Ben Temkin, Gavin O'Sullivan

October 26, 2020

---

1. 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.

Our initial research plan to build a binary classifier with black box deep learning and convolutional neural networks to distinguish between COVID-19 positive and negative cases based on chest x-ray data. We plan to determine which of these models can perform better with the current chest x-ray data we have through the use of the accuracy performance metric as our classes are evenly split. Based on these results, we will expand and hone the more successful model for the remainder of the project in an attempt to add successful multi-class classification and achieve better overall classification results.

2. Problem Definition and Algorithm

2.1 Task Definition

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

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

2.2 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. Currently, after removing the CT scans and unlabeled images, we are working with a set of 1017 images; 504 Covid and 513 non-COVID. However, we do plan to increase the number of images in our dataset as they become available or needed.

From the metadata that corresponds to each image we then relabel the image. For this first step of our project we focused on diagnosis, i.e. classifying images as either COVID or non-COVID. Moving forward however, we will also be taking into account prognosis, and will work on multiclass classification, classifying the images as either non-COVID or by the severity of the COVID case.

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.

2.3 Algorithm Definition

Currently we are using all of the deep learning algorithms shown below as a black box for classification, and we plan to go into further detail on the exact techniques at a later date. The data being classified are the images detailed in section 2.2 reduced to a standard size and converted to grayscale. Initially we tried using a multi-layered neural network to classify the images however given the level of detail needed the results, as seen in section 3.2, were less than satisfactory. Moving on, we then tried training a convolutional neural network as a classifier which performed considerably better and more consistently than the initial neural network. As such we are planning to use the convolution neural network moving forward. As mentioned we also plan to train a neural network to segment out the lungs from the images in the dataset to enhance the current system, and to extend the model from binary to multiclass classification of COVID cases.

3. Experimental Evaluation

3.1 Methodology

Due to the nature of our dataset being binary and almost a perfectly even split of classes (504 COVID CXR's and 513 non-COVID), we believe accuracy is a metric that can accurately measure the performance of our algorithms. Our experiment will be testing the performance of a normal deep learning approach vs a convolutional neural network to determine which is a better suited method for our data. Additionally, we will be using K-fold cross-validation on both models to get a better estimate on how each will perform on real-world data. The training and testing data used are chest x-rays from different publications and hospitals, which gives us a realistic dataset that can be used as a representation of the entire population of chest x-ray data.

For all algorithms, we trained on 10 epochs and collected the training and validation accuracy for each epoch, as well as the final test accuracy of the trained model. Graphs were created (Fig. 1-4 below) to visualize the performance of our models over the course of training. Using these, we will be measuring the training vs validation accuracy for each to determine how well the model was fit as well as comparing the final test accuracies to evaluate how well each can perform with real-world data.

3.2 Results

Below are the results of our different classifiers for comparison using accuracy as our performance metric. The accuracy ranking of the different classifiers is as follows:

Multi-layer NN < Multi-layer NN with CV < CNN < CNN with CV

Both the Multi-layer NN and CNN models were trained on the same 75%/25% split on training/testing data, while each of the cross-validation models were trained with a k = 10 K-Fold method (90%/10% train/test split). For the CV models, the model with the best test performance was selected and plotted out of the 10 that were trained.

```
Test Accuracy: 0.6941
```

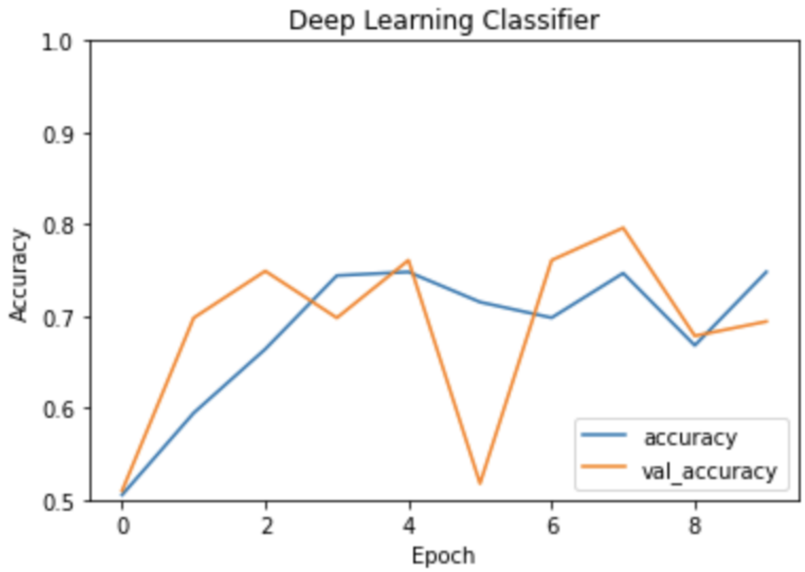

Fig 1. Multi-layer neural network classifier without cross validation

```
Test Accuracy: 0.7745
```

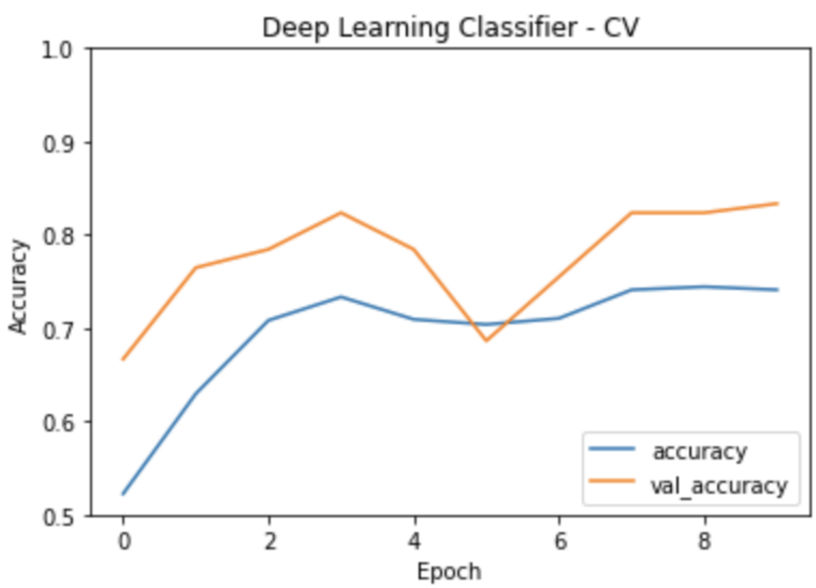

Fig 2. Multi-layer neural network classifier using cross validation

`Test Accuracy: 0.8039`

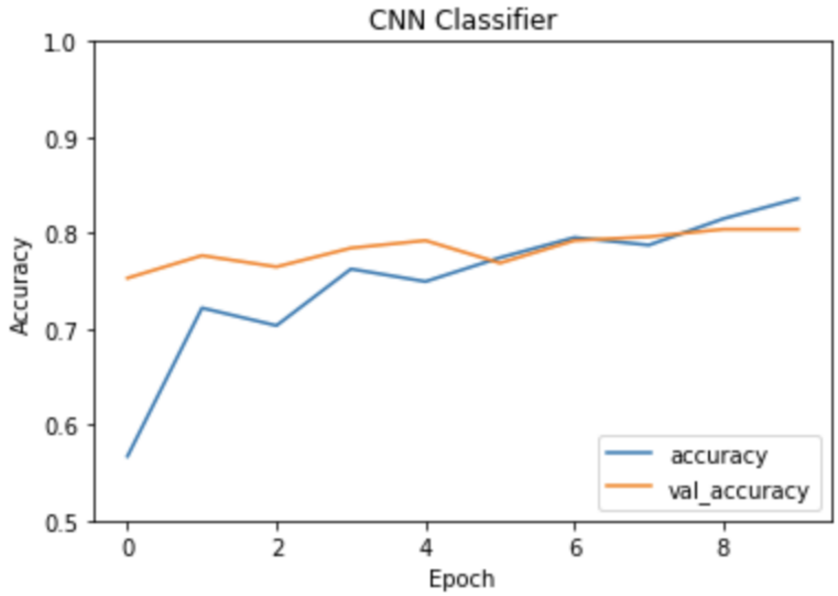

Fig 3. Convolutional neural network without cross validation

`Test Accuracy: 0.8218`

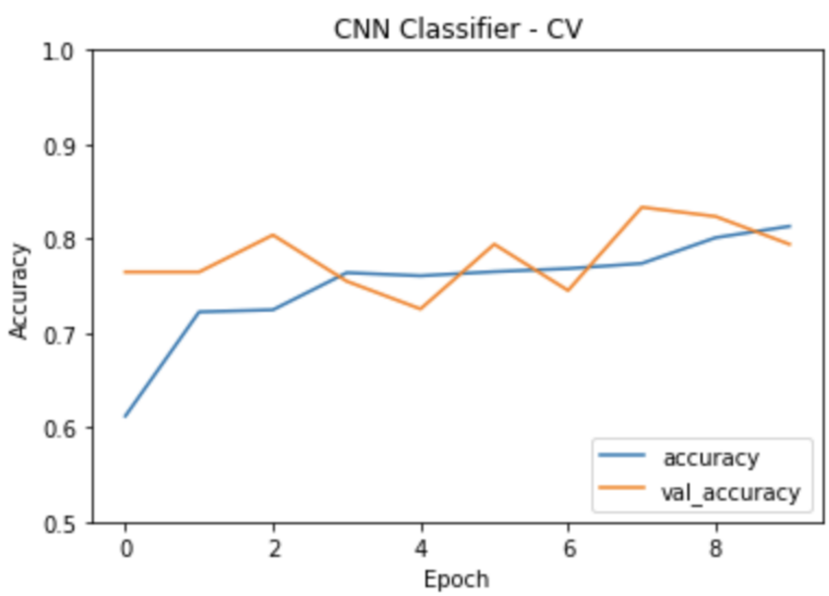

Fig 4. Convolutional neural network using cross validation

3.3 Discussion
After training our classifier in the four ways seen above, in Figures 1-4, the results are straightforward. The convolutional neural network performed was more accurate at classifying the images compared to the regular multi-layer neural network in every test. However, while ~80% accuracy is not terrible for an initial model as we want to transition into multiclass classification, which will hurt the balance of our model, we are aiming for >95% for this binary classification through the use of lung segmentation to remove noise from our data. We surmise that the problem holding back the accuracy of our model is the amount of noise in the images. The only part of the x rays that are useful for classification are the lungs, but most, if not all, of the images include the skeleton of the patient's chest and arms. It is likely that this is causing many of our images to appear to be more similar than they should to the neural network. Once we have achieved this improvement on the binary classification, our target is to achieve at least an ~80% accuracy on the multi-class predictors we plan to develop as outlined in section 5.

4. 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.

A similar study[4] to ours classifies COVID x rays after extracting the images of the lungs by using a neural network. The study claims that removing the excess data from the images "potentially prunes away possible bias sources, like for example the presence of medical devices(typically correlated to sick patients), various text which might be [embedded] in the scan etc."[4]. As our project seeks to use multiclass classification to also give a prognosis for the COVID patients we are also planning to segment out the lungs of the patients in order to reduce the noise in the images. Currently we are planning on following the steps outlined in this[5] notebook to train our own model for lung segmentation of our dataset.

5. Next Steps

Our main goal for the final project is to expand on our binary classifier into multiple classes to add helpful prognosis data to our target values. In addition to COVID/Non-COVID, for each sample we will have survival, intubation, and admission to the ICU - resulting in a collation of classes where class 0 = Non-COVID and class > 0 is COVID and some unique combination of the classes above. This will result in a greater skewing of our data, and as such moving forward we will be evaluating our models with performance metrics other than accuracy such as ROC and precision/recall curves.

Additionally, the current CXR's we have used are regularized to 128x128 images before being passed to the algorithm. Since the majority of COVID/Non-COVID identification comes from the state of the lungs in the x-ray, this introduces a considerable amount of noise to the input data, which we believe is negatively affecting our results. To resolve this, we plan to segment out the lungs through neural networks in a similar fashion as [4] to remove a large amount of noise in our data. This will also help the computational time for running our algorithms, as a large percent of the pixels in each of our images will be ignored.

Due to the size of our team being only two programmers, we have used and plan to continue to use pair programming practices for the development of our machine learning algorithms. This has allowed us to participate in the creation of each of the algorithms we have developed, regardless of who was writing the code. However, as we learn more about neural networks, when we move to write our own, assuming it is a feasible option, we will assign programming responsibilities as evenly as possible.

6. Code and Dataset

After cloning the repository from github, the notebook should run properly just by running each cell in order.

Link to github containing code and dataset: https://github.com/gdosulli/CS254_COVID

7. Conclusion

Overall the work done so far is a good first step towards the project goals. We have evaluated that a convolutional neural network is more effective at classifying our data set than a multilayered neural network. Additionally, we have implemented cross validation to increase the efficacy of our classifier, which is at an acceptable level for an initial model (>80% test accuracy). We identified both a potential problem, in an excess of noise in our dataset, and the corresponding solution to segment out the important information from our images. This change should greatly increase the ability of our model to correctly classify our dataset. Finally, and most importantly, we have created a solid base for transitioning from binary to multiclass classification which is the end goal of the project.

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

[4] Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data. [Online]. Available: https://arxiv.org/pdf/2004.05405.pdf

[5] Lung segmentation from Chest X-Ray dataset. [Online]. Available: https://www.kaggle.com/nikhilpandey360/lung-segmentation-from-chest-x-ray-dataset

[6] Basic classification: Classify images of clothing. [Online]. Available: https://www.tensorflow.org/tutorials/keras/classification

[7] Convolutional Neural Network (CNN) tutorial. [Online]. Available: https://www.tensorflow.org/tutorials/images/cnn