Detecting COVID-19
using X Ray images
Project Report

10-May-2020

Team Members: Nihal Nihalani, Shawn Gonsalves, Rachel Culver

Version History

Name	Date	Reason For Changes	Version
Change of dataset	03/22/20	CT scans are more proper to detect covid19	1.0
Change of learning rate from 3 to 6	03/25/20	To decrease Validation loss	1.1
Learning rate change to 3	04/02/20	To decrease Validation loss	1.2
Change of epochs	04/10/20	From 25 to 100 epochs to decrease loss and increase accuracy	1.3
Change of dataset back to x ray	04/13/20	Not a large dataset was available for CT scans	2.0
Change of architecture VGG16 to VGG19	04/18/20	Professors suggestion	2.1
Architecture change to VGG16	04/25/20	Better suited architecture over VGG19	2.2
Change epochs from 100 to 600	05/01/20	Increase validation and decrease loss	2.3
Adding tensorflow.js and website	05/10/20	Requirement to demonstrate our model	2.4

Table of Contents

1. Deliverables

Code Repo (URL): https://github.com/shawngonsalves/COVID-19-detection
Presentation Slides (URL): Detecting COVID-19

2. Executive Summary

In the development of our model, we managed to grab the dataset containing Chest X ray images of patients suffering from CoronaVirus as well as X-ray of normal patients.The system was then trained for a number of epochs and generated a model.

This model is capable of identifying if a person suffers from the virus or not provided a Chest X-ray as an input image.

The most interesting part of our model is that we get a sensitivity of 1.0000 when we feed the data for 600 epochs.

3. Project Overview

3.1 Problem Statement

COVID-19 is spreading at an unprecedented rate all around the globe. Health care systems are experiencing a tremendous amount of pressure and health professionals work on exhausting shifts to deal with the demand. As patients flood clinics and hospitals for a checkup to see if they have the virus, it becomes a burden for the doctors to check all the incoming patients since it's impossible to know without a test. While carrying out the test using Xray scan images, a drawback is that the analysis requires an expert and takes significant time which is precious when people are sick around the world.

3.2 Scope

This deep learning model takes X-Ray images of a patient as input and predicts if the person suffers from the virus based on the scan of the persons' Chest.
This will significantly reduce the time taken in analyzing the scan images and thereby increasing the efficiency of detecting the virus.

3.3 Project Team

Name of the project: Detection Of COVID-19

Name of the Team member	**Responsibility**	**Contribution %**
Nihal Nihalani	Training of models. Conversion to Tensorflow.js and adding the model to website.	33.3%
Shawn Gonsalves	Refining the model and data. Decreasing loss and playing with epochs and batch size to find optimal one.	33.3%
Rachel Culver	Gathering dataset and Processing of dataset Converting from radiograph to JPEG format. UI for website	33.3%

4. Machine Learning Aspects

4.1 Dataset

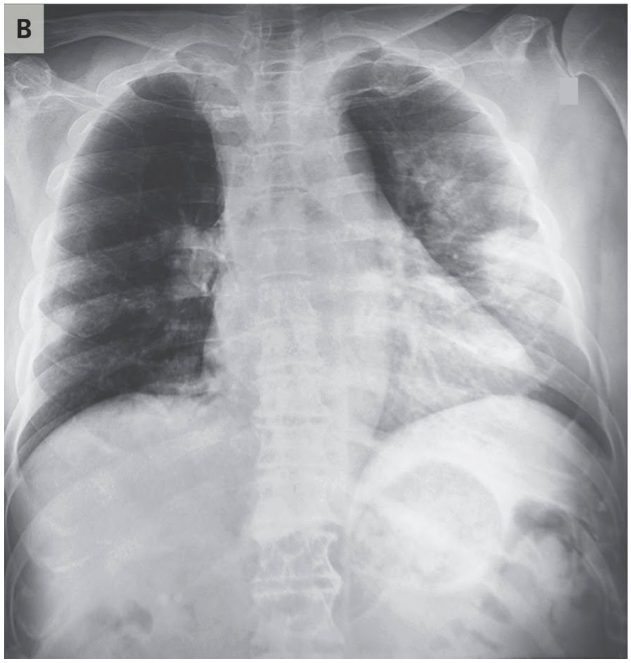
The dataset used in this project was the x-ray images of both healthy and COVID-19 infected patient Chest.
Dataset for this project was acquired from Kaggle.

The upstream dataset currently lists the following upstream data sources:

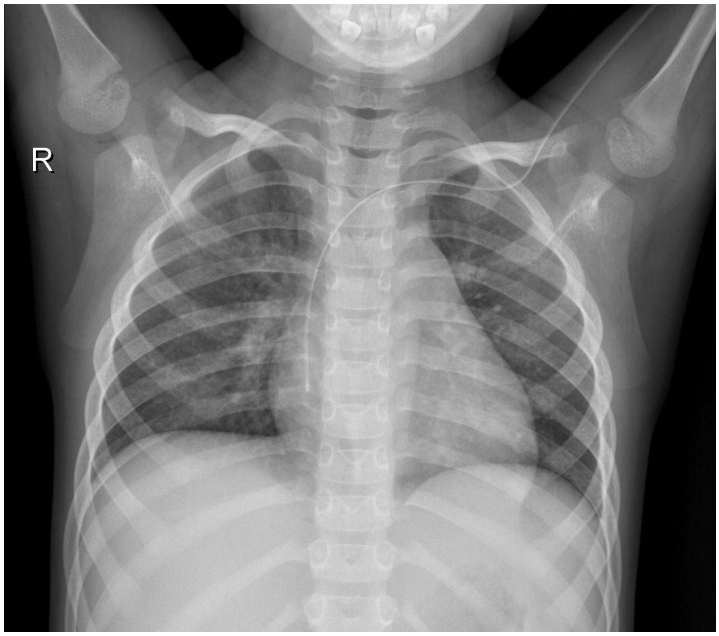
- World Health Organization (WHO): https://www.who.int/
- DXY.cn. Pneumonia. 2020. http://3g.dxy.cn/newh5/view/pneumonia
- BNO News: https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/
- National Health Commission of the People's Republic of China (NHC): http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml
- China CDC (CCDC): http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm
- Hong Kong Department of Health: https://www.chp.gov.hk/en/features/102465.html
- Macau Government: https://www.ssm.gov.mo/portal/
- Taiwan CDC: https://sites.google.com/cdc.gov.tw/2019ncov/taiwan?authuser=0
- US CDC: https://www.cdc.gov/coronavirus/2019-ncov/index.html
- Government of Canada: https://www.canada.ca/en/public-health/services/diseases/coronavirus.html
- Australia Government Department of Health: https://www.health.gov.au/news/coronavirus-update-at-a-glance
- European Centre for Disease Prevention and Control (ECDC): https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
- Ministry of Health Singapore (MOH): https://www.moh.gov.sg/covid-19
- Italy Ministry of Health: http://www.salute.gov.it/nuovocoronavirus

Data and labels were converted to NumPy arrays, one-hot encoding was done to the labels and the data was partitioned into using 80% for training and 20% for testing.

The image shown below is an example of a patient with positive Covid-19



The image shown below is an example of a healthy individual.



4.2 Model Creation and Training

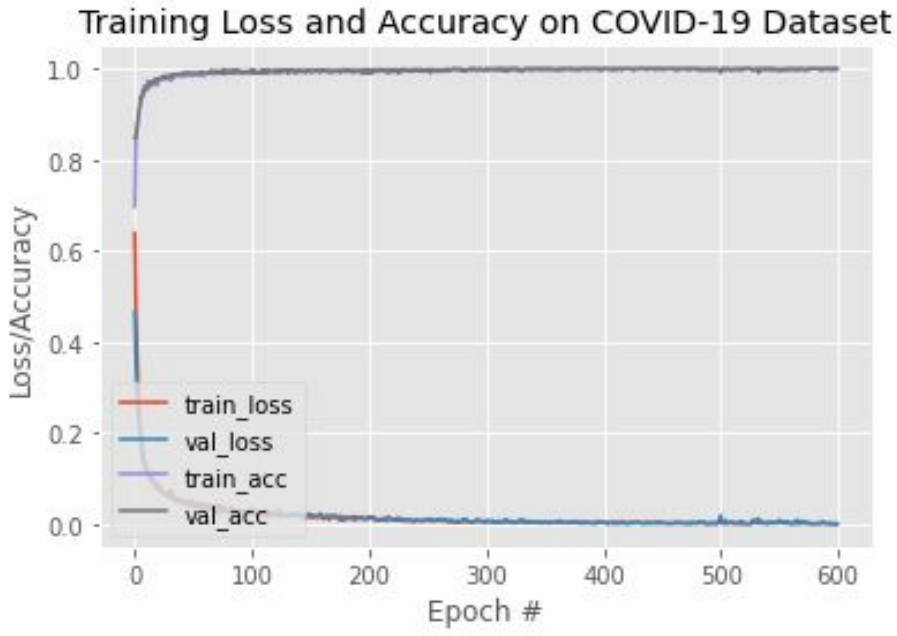
The generated model can classify if a patient has COVID-19 or not based on the images of his/her Chest X-Ray. We construct the head model that gets placed on top of the base model. For the head model we first start by considering it as output of the base model. Then we averagePool it with a size of 4*4 which down samples it to an average value in that square. Next, we flatten the head model and create a dense layer and use relu as the activation function. Lastly, we have added a Dropout layer which prevents our model from overfitting.

For the training process, we have set the learning rate of 3 for 600 epochs with the Batch size of 100 samples per training pass.

4.3 Inference

We are using Tensorflow js to deploy our machine learning model. Now, to test the model we use the same website and choose an X Ray image from the file storage and test that image on the model.

4.4 Evaluation of the Model



	precision-	recall	f1-score	support
covidct	1.00	1.00	0.99	49
normalct	1.00	1.00	1.00	273
accuracy			1.00	322
macro avg	1.00	1.00	1.00	322
weighted avg	1.00	1.00	1.00	403

acc: 1.0000
sensitivity: 1.0000
specificity: 1.0000

5. Software

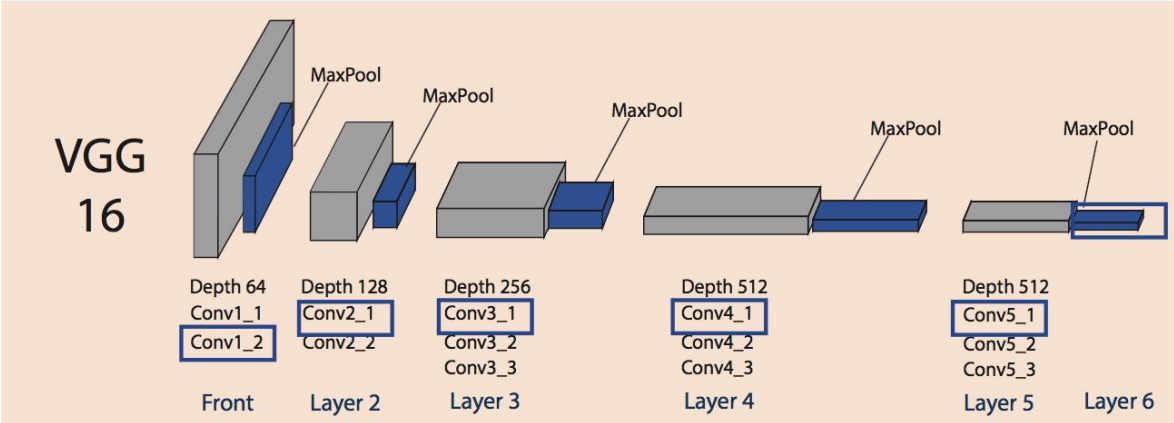
5.1 System as a Software

We are using Keras layers. Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow. We are using this library as we are implementing our model using Convolutional Neural networks and Keras is capable of supporting both convolutional networks and recurrent networks.

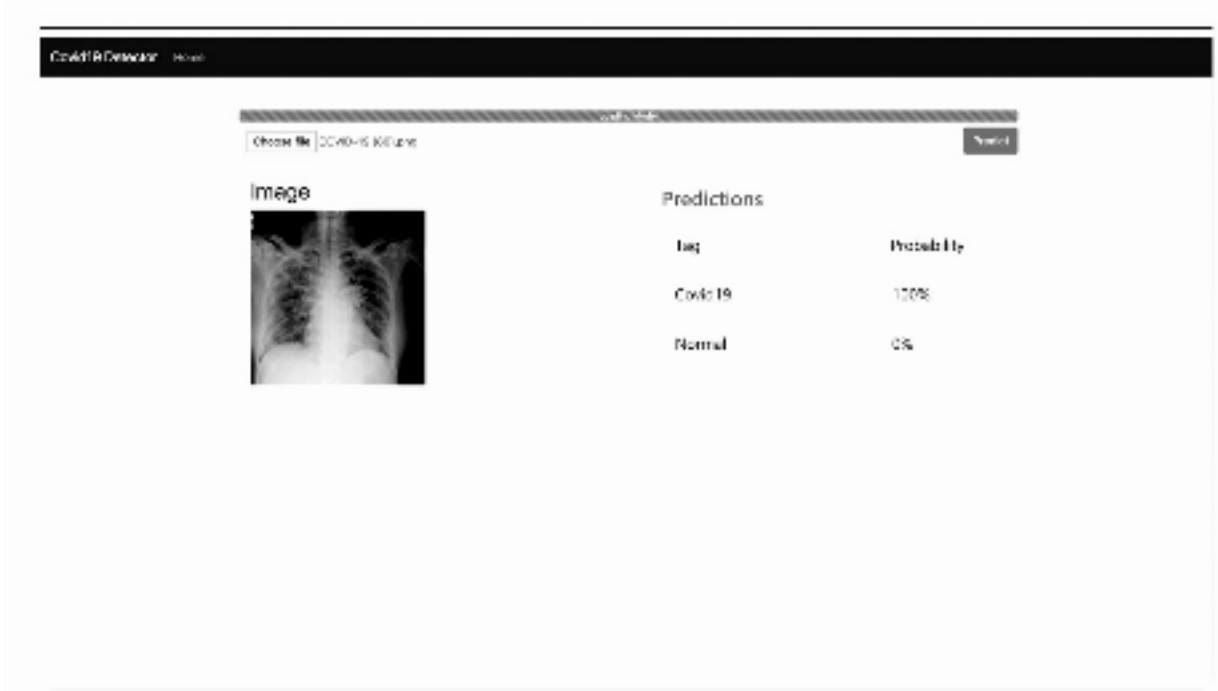
The X-Ray images of Chest are classified using VGG16 architecture (VGG16 is Convolution Neural Networks used to train large datasets of images) with a learning rate of 3 for 600 epochs with the Batch size of 30 samples per training pass. Convolution is a technique that allows us to extract visual features from an image in small chunks. Each neuron in a convolution layer is responsible for a small cluster of neurons in the preceding layer.

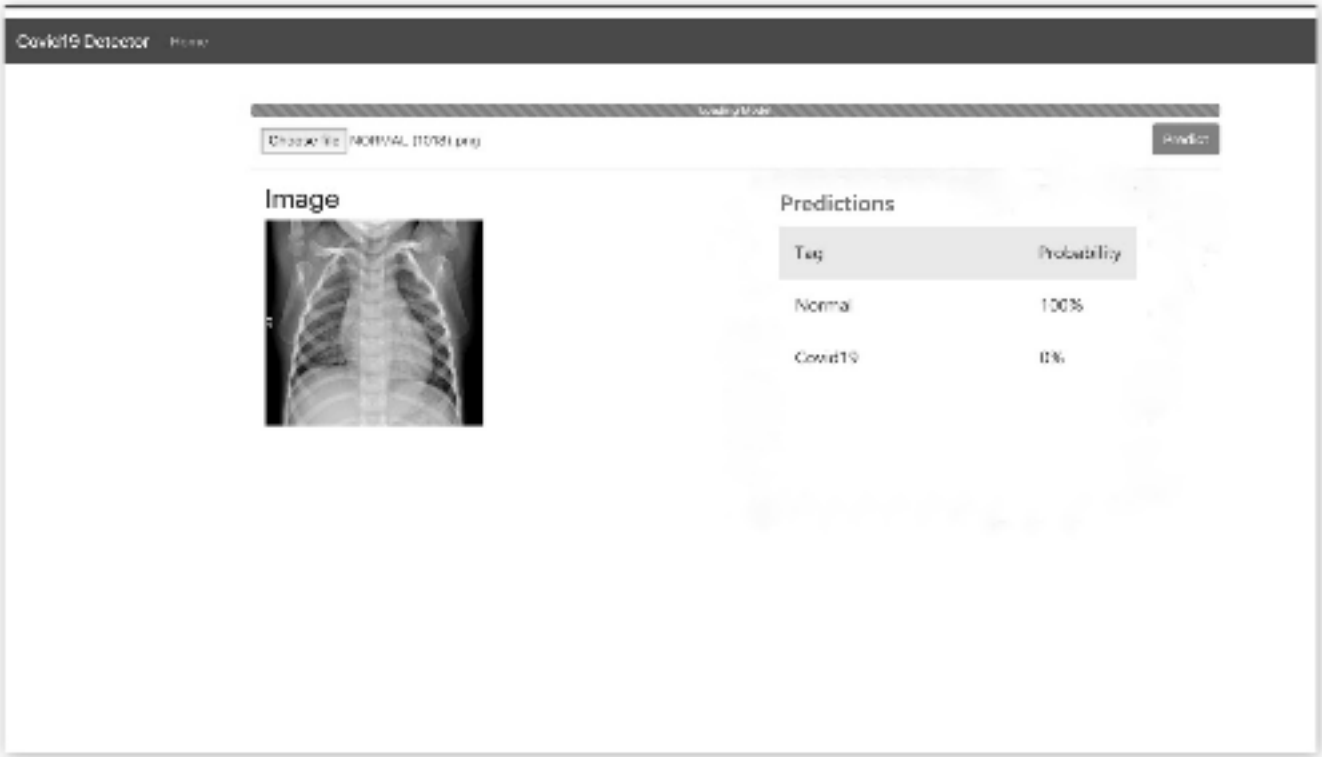
5.2 System Architecture

In the development of our model we used Convolutional Neural Network Architecture called VGG-16 using Keras, Tensorflow, and Deep Learning. VGG-16 is a convolutional neural network architecture, it's named VGG-16 comes from the fact that it has 16 layers. Its layers consist of Convolutional layers, Max Pooling layers, Activation layers, Fully connected layers.



5.3 User Interface Overview





5.4 Use Cases

- The model can be used by Doctors to accelerate their process of detecting if a patient is suffering from COVID-19 or not.
- Radiologists and healthcare workers are at the risk of contracting the virus; this model can be used even without them and in turn preserving them.
- This would 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.
- Also, these tools can provide quantitative scores to consider and use in studies. Image data linked with clinically relevant attributes in a public dataset that is designed for ML will enable parallel development of these tools and rapid local validation of models.
- Furthermore, this data can be used for completely different tasks.

6. References

- https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7187882/
- https://www.medrxiv.org/content/10.1101/2020.05.04.20090423v1.full.pdf