

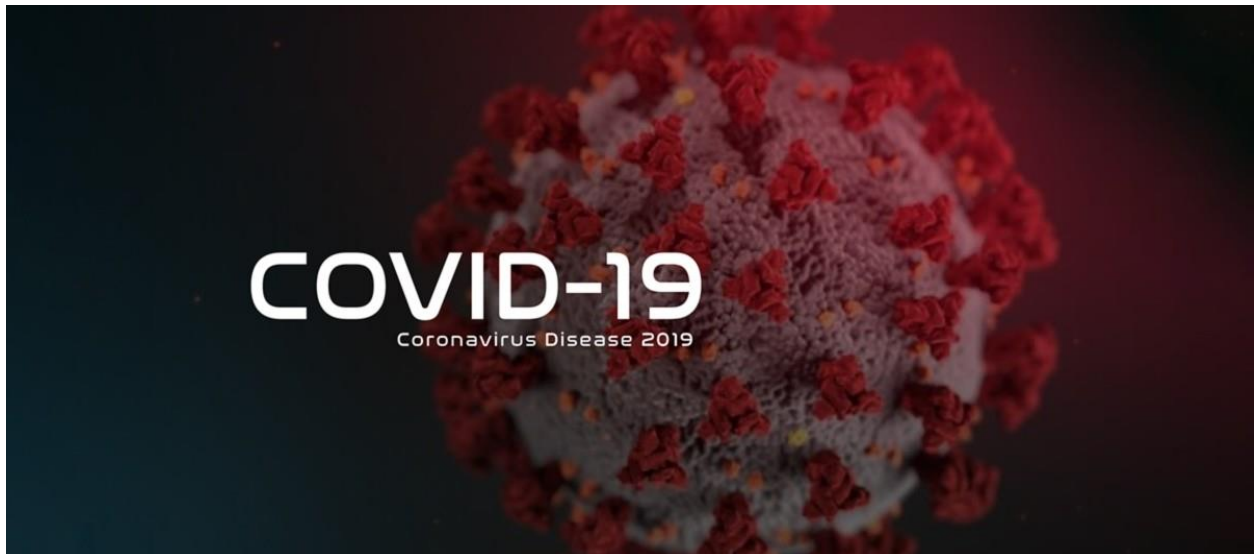
# Machine Learning Engineer Nanodegree

## Capstone Project Report

### Detecting Covid-19 in x-ray Images

Mohammed Lokma

26/8/2020



# I.Defination

## . Project Overview

These days, the world has woken up to the Covid-19 pandemic

Which is known as [corona virus](#)

The COVID-19 coronavirus is the global health crisis of our time and the biggest challenge we have faced since World War II. Since it appeared in china late last year, the virus has spread to every country except Antarctica

Not only the a health crisis, just as it is an unprecedented social and economic crisis. By emphasizing each country that touches it, this creates a devastating social state, The economic and political implications that will leave deep and long-term scars. The United Nations Development Program is the technical lead in the United Nations' social and economic recovery, along with the health response led by the World Health Organization and the Global Humanitarian Response Plan, and working under the leadership of the United Nations resident coordinators.

Let us see the covid-19's effect of economic

Every day, people lose their jobs and income, and there is no way to know when normal life will return. Small island states, which rely heavily on tourism, have empty hotels and deserted beaches. The International Labor Organization estimates that 195 million jobs can be lost.

The World Bank predicts a drop in remittances of \$ 110 billion this year, which could mean that 800 million people will be unable to meet their basic needs.

**Globally**, as of **3:37pm CEST, 20 July 2020**, there have been **24,218,739 confirmed cases** of COVID-19, including **826,603 deaths**, reported to WHO.

So I am going to try to solve this problem using some technics of Machine Learning and Deep learning

## Problem statement

Now , corona virus threatens everything . economic, people jobs , people life and our normal life .

Who And most scientists are now trying to find a vaccine or any way to reduce the spread of the pandemic

The disease is currently diagnosed by a hospital swab

According to the World Health Organization and by tracking the pathology within the human body, they discovered that the virus first attacks the respiratory system, so we can know that the person is pregnant or not through the X-ray image of the respiratory system, thus reducing the cost of detecting the disease and reducing the pressure on the detectors

## Solution

The problem here is to predict if a person is has the virus or not, depending on the x-ray scan of the chest.

## Datasets and Inputs

- I used Dr. Cohen repository ([GitHub repo.](#))
- Inside the repo you'll find example of COVID-19 cases, as well as MERS, SARS, and ARDS.
- Covid X-Ray Image Dataset for positive cases (<https://github.com/ieee8023/covid-chestxray-dataset>)
- Kaggle X-Ray Chest Images for negative cases (<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>)

## Metrics

I will use confusion matrix.

A confusion matrix is a table often used to describe the performance of a classification model .

It is used on a set of test data for which the true values are known.

It allows the visualization of the performance of an algorithm.

## II. Analysis

---

### Data Exploration And Exploratory Visualization

The data set is consist of two parts

- 1- Chest-xrays from kaggle for normal people and people who have PNEUMONIA

In this dataset we have 3 folders for training , validation and testing

Every folder has two sub folders for normal people and people who have PNEUMONIA . it has a lot of images . about 5000 photo of people .

## 2- Chest-xray for covid-19 patients .

It consists of about 1000 image foe patients . with 25 features

A sample is provided below

It also have a csv file with about 1000 record and 20 column

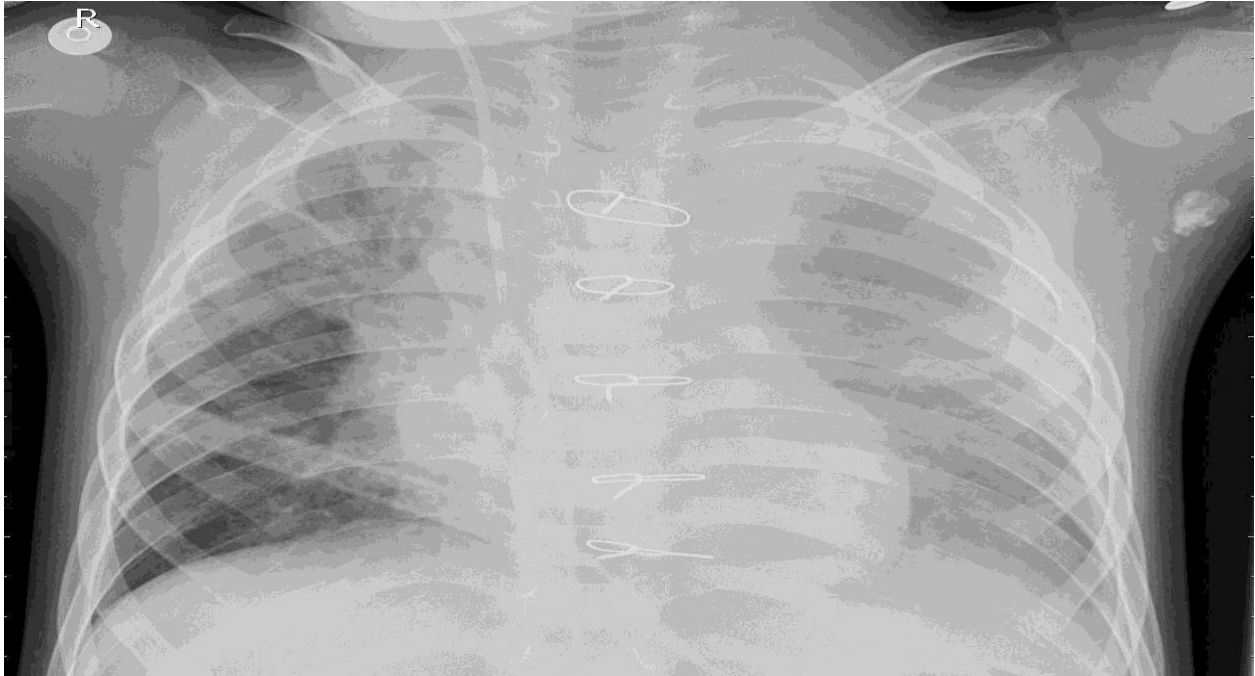
has the features

	patientid	offset	sex	age	finding	RT_PCR_positive	survival	intubated	intubation_present	went_icu	...	modality	date	location	folder	f
0	2	0.0	M	65.0	COVID-19	Y	Y	N	N	N	...	X-ray	January 22, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	a 2
1	2	3.0	M	65.0	COVID-19	Y	Y	N	N	N	...	X-ray	January 25, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	a 2
2	2	5.0	M	65.0	COVID-19	Y	Y	N	N	N	...	X-ray	January 27, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	a 2
3	2	6.0	M	65.0	COVID-19	Y	Y	N	N	N	...	X-ray	January 28, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	a 2
4	4	0.0	F	52.0	COVID-19	Y	NaN	N	N	N	...	X-ray	January 25, 2020	Changhua Christian Hospital, Changhua City, Ta...	images	r

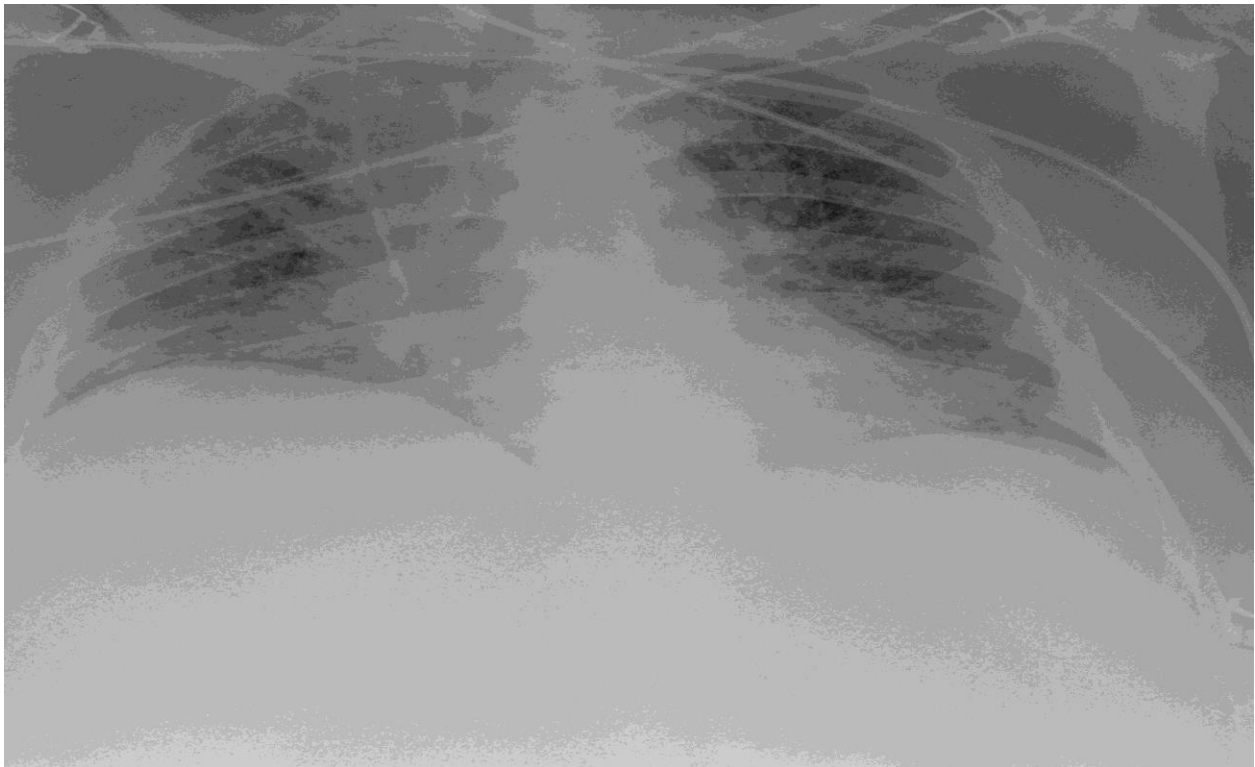
filename	doi	url	license	clinical_notes	other_notes
auntminnie-a-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	<a href="https://www.nejm.org/doi/full/10.1056/NEJMc200...">https://www.nejm.org/doi/full/10.1056/NEJMc200...</a>	NaN	On January 22, 2020, a 65-year-old man with a ...	NaN
auntminnie-b-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	<a href="https://www.nejm.org/doi/full/10.1056/NEJMc200...">https://www.nejm.org/doi/full/10.1056/NEJMc200...</a>	NaN	On January 22, 2020, a 65-year-old man with a ...	NaN
auntminnie-c-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	<a href="https://www.nejm.org/doi/full/10.1056/NEJMc200...">https://www.nejm.org/doi/full/10.1056/NEJMc200...</a>	NaN	On January 22, 2020, a 65-year-old man with a ...	NaN
auntminnie-d-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	<a href="https://www.nejm.org/doi/full/10.1056/NEJMc200...">https://www.nejm.org/doi/full/10.1056/NEJMc200...</a>	NaN	On January 22, 2020, a 65-year-old man with a ...	NaN
nejmc2001573_f1a.jpeg	10.1056/NEJMc2001573	<a href="https://www.nejm.org/doi/full/10.1056/NEJMc200...">https://www.nejm.org/doi/full/10.1056/NEJMc200...</a>	NaN	diffuse infiltrates in the bilateral lower lungs	NaN

We have known that covid-19 attack the lungs and chest .

So the challenge here is if the lung is already affected but it could be PNEUMONIA not covid-19 case



This is a PNEUMONIA case



This is a covid-19 case

## Algorithms and Techniques

The proposed solution to this problem is to use and apply some deep learning techniques that have been proved to very good and successful in the field of image classification.

The solution include a trained detector to detect covid-19 case based on chest-xray Image

We are going to include a trained detector using cnn to detect patients . CNN is the convolution Neural Network , a deep learning framework which is widely used in image classification .

CNN contains of input layers , hidden layers and output layers .

CNN uses multiple layers to extract features out of an image.

Also a the model of CNN has features hyperparameters such as number of epochs . Simply it is the no. of times the entire training set pass through the NN .

And no. hidden layers more layers lead to higher computational cost, less layers lead to underfitting.

## Benchmark

So for benchmark model , I will use the algorithms that I mentioned before . by using the dataset I have mentioned.

the target of the model is to reach multiclass loss score and accuracy above 90%.



# III. Methodology

## Data Processing

The transformed data are organized into train , test and validation directories

Below we will see examples

### Load Datasets

```
9]: TARGET_DIR = "dataset/Train/Covid"

if not os.path.exists(TARGET_DIR):
    os.mkdir(TARGET_DIR)
    print("Covid Folder has been Created")
```

Covid Folder Created

```
0]: #Copy COVID-19 images with view point PA from the downloaded directory to target directory we need there
cnt = 0
for(i,row) in df.iterrows():
    if row["finding"] == "COVID-19" and row["view"] == "PA":
        filename = row["filename"]
        image_path = os.path.join(IMAGE_PATH,filename) #IMAGE_PATH + filename
        image_copy_path = os.path.join(TARGET_DIR,filename) #TARGET_DIR + filename
        shutil.copy2(image_path,image_copy_path) #Copy image from IMAGE_PATH to TARGET_directory
        print("Moving Image",cnt)
        cnt+=1
print(cnt)
```

```
: # Sampling of Images from Kaggle dataset,As we have 180 around covid images so we will have 180 around normal images of xray

import random
KAGGLE_FILE_PATH = "dataset/chest_xray/train/NORMAL"
TARGET_NORMAL_DIR = "dataset/Train/Normal"
if not os.path.exists(TARGET_NORMAL_DIR):
    os.mkdir(TARGET_NORMAL_DIR)
    print("Normal Folder Created")

image_names = os.listdir(KAGGLE_FILE_PATH) #this contains a list of all images_names
# images_names
random.shuffle(image_names) #it will randomly shuffle names in list

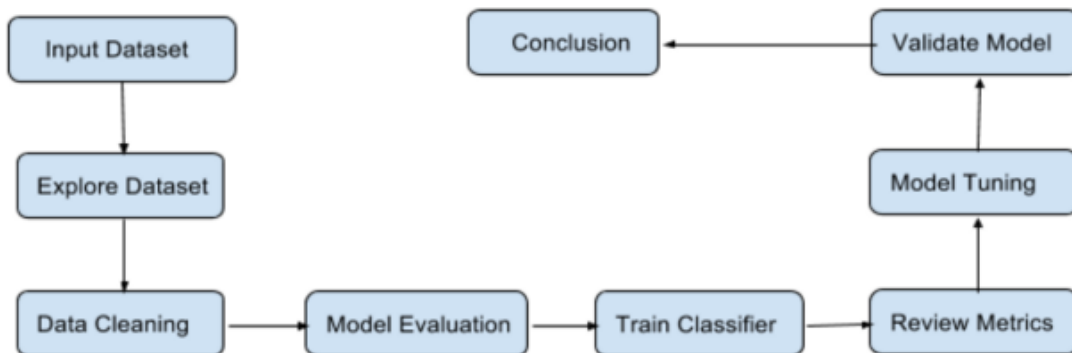
for i in range(180):
    image_name = image_names[i]
    image_path = os.path.join(KAGGLE_FILE_PATH,image_name)

    target_path = os.path.join(TARGET_NORMAL_DIR,image_name)
    shutil.copy2(image_path,target_path)
    print("Moved",i)
```

```
TRAIN_PATH = "dataset/Train"
VALIDATION_PATH = "dataset/Validation"
```

---

## Implementation



Here we are going to load the data of covid-19 first and second the data of PNEUMONIA and normal cases .

We will clean the data

. we will sample the data from kaggle to data of covid-19

Data from kaggle of chest-xrays is too much

So ex: if we have 140 covid we sample 140 of other data

After that we will train a classifier that take the training data

We will test the model , validate it and finally the prediction.

---

# Refinement

Initially the accuracy of my model is very poor .

The scratch CNN model is trained and we started with 10 epochs and the worst and the best accuracy is 52% and 92.3%.

The structure of the model is good for achieving higher accuracy.

```
Epoch 1/10
8/8 [=====] - ETA: 0s - loss: 0.6953 - accuracy: 0.5312WARNING:tensorflow:Your input ran out of data; int
errupting training. Make sure that your dataset or generator can generate at least `steps_per_epoch * epochs` batches (in this cas
e, 2 batches). You may need to use the repeat() function when building your dataset.
8/8 [=====] - 78s 10s/step - loss: 0.6953 - accuracy: 0.5312 - val_loss: 0.6930 - val_accuracy: 0.5000
Epoch 2/10
8/8 [=====] - 67s 8s/step - loss: 0.6568 - accuracy: 0.6078
Epoch 3/10
8/8 [=====] - 75s 9s/step - loss: 0.5701 - accuracy: 0.6758
Epoch 4/10
8/8 [=====] - 74s 9s/step - loss: 0.3741 - accuracy: 0.8711
Epoch 5/10
8/8 [=====] - 66s 8s/step - loss: 0.3564 - accuracy: 0.8664
Epoch 6/10
8/8 [=====] - 73s 9s/step - loss: 0.3648 - accuracy: 0.8707
Epoch 7/10
8/8 [=====] - 74s 9s/step - loss: 0.2865 - accuracy: 0.9023
Epoch 8/10
8/8 [=====] - 67s 8s/step - loss: 0.2445 - accuracy: 0.8966
Epoch 9/10
8/8 [=====] - 66s 8s/step - loss: 0.2359 - accuracy: 0.9181
Epoch 10/10
8/8 [=====] - 78s 10s/step - loss: 0.2149 - accuracy: 0.9297
```

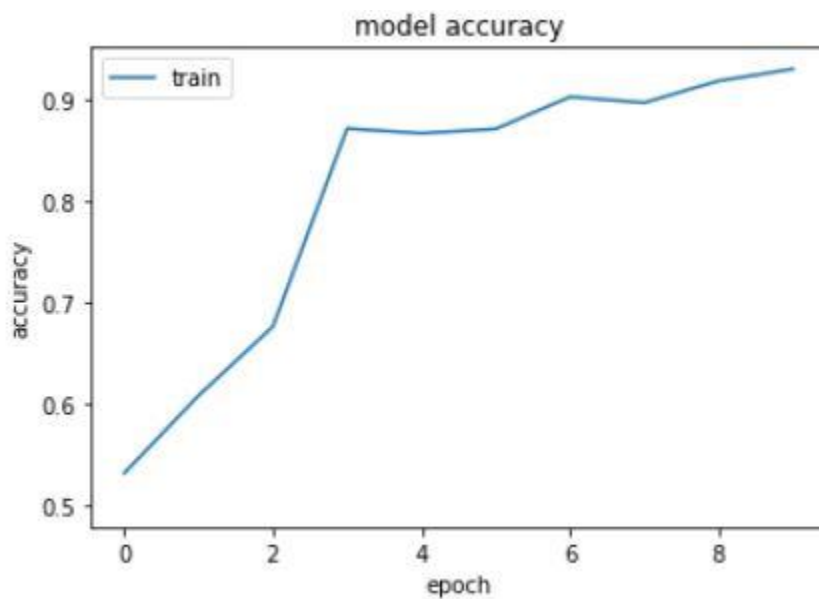
---

## IV. Results

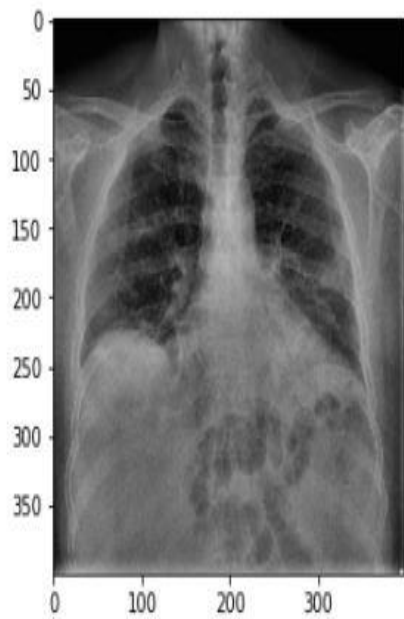
---

### Model Evaluation and Validation

I think the model is going well . we see at final we achieve 92.3% accuracy . we see that our model achieve exactly the problem we have discussed .



Prediction Of Our Model : Positive For Covid-19



Prediction Of Our Model : Negative for Covid-19



So here , we see the model detected that first person has covid-19.

The second person is a normal person.

## Test Images

```
#save the model  
model.save("Detection_Covid_19.h5")
```

## Test Images

```
model = load_model("Detection_Covid_19.h5")
```

```
train_generator.class_indices
```

```
{'Covid': 0, 'Normal': 1}
```

## Justification

Our final model achieved a classification accuracy of 92%.

The final solution performs well . we see that from the previous images

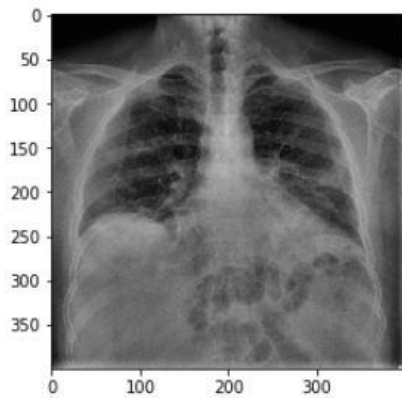
I think this is pretty good and do what we expected.

## V. Conclusion

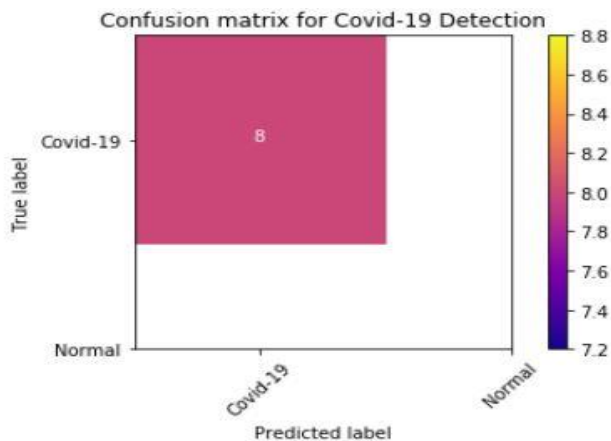
So we see that the prediction seems to be pretty accurate .

---

Prediction Of Our Model : Positive For Covid-19



So what our confusion matrix expect



## Reflection

The process used for this model can be summarized with the following:

We loaded the data , explored and analysed it .

We trained the model.

We validate it .

Finally we see the prediction of our model

The final model and solution fit our expectations for the problem and I think by some improvements we can use it in a general setting .

## Improvement

I know I can improve the model little more .

I can detect if the person does not have covid-19 and complain from anything in his chest what is exactly the problem .

I would train my model with more than 30 epochs to improve it .

I could contact with a real doctor and see what we can do to improve this.

So finally , this is my project. I hope it is doing well and pretty good .

I hope I can use this in real life and help people .

Thanks For your interest of reading this.