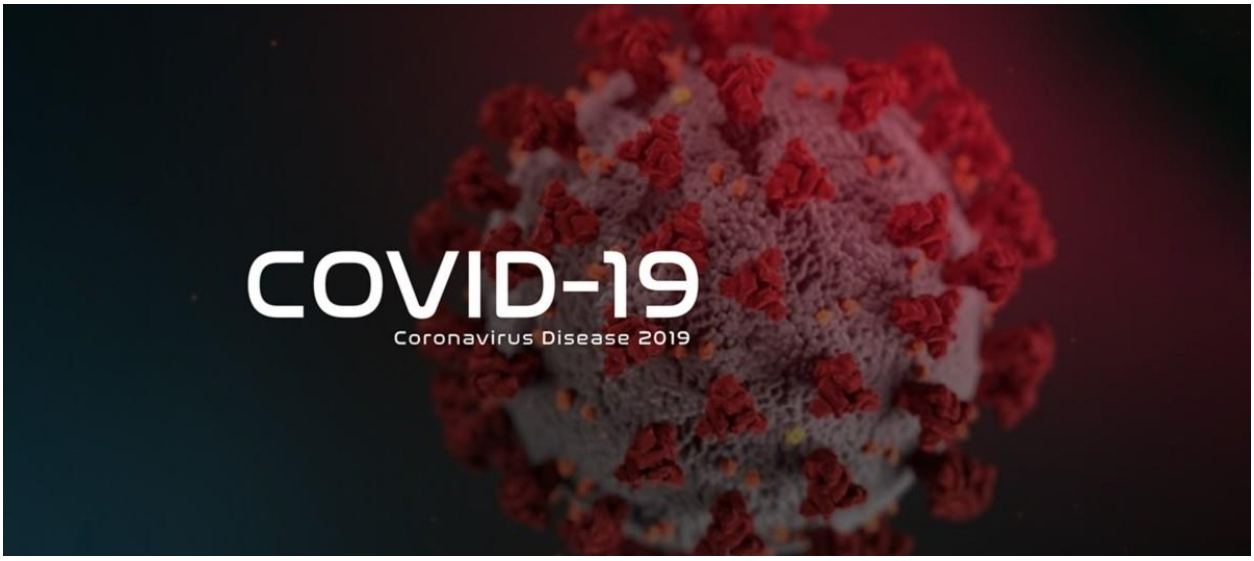
Machine Learning Engineer Nanodegree

Capstone Proposal

Detecting Covid-19 in x-ray Images

Mohammed Lokma

22/7/2020



Domain Background

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 stressing every country that touches it, it has the potential to create a devastating social, 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.

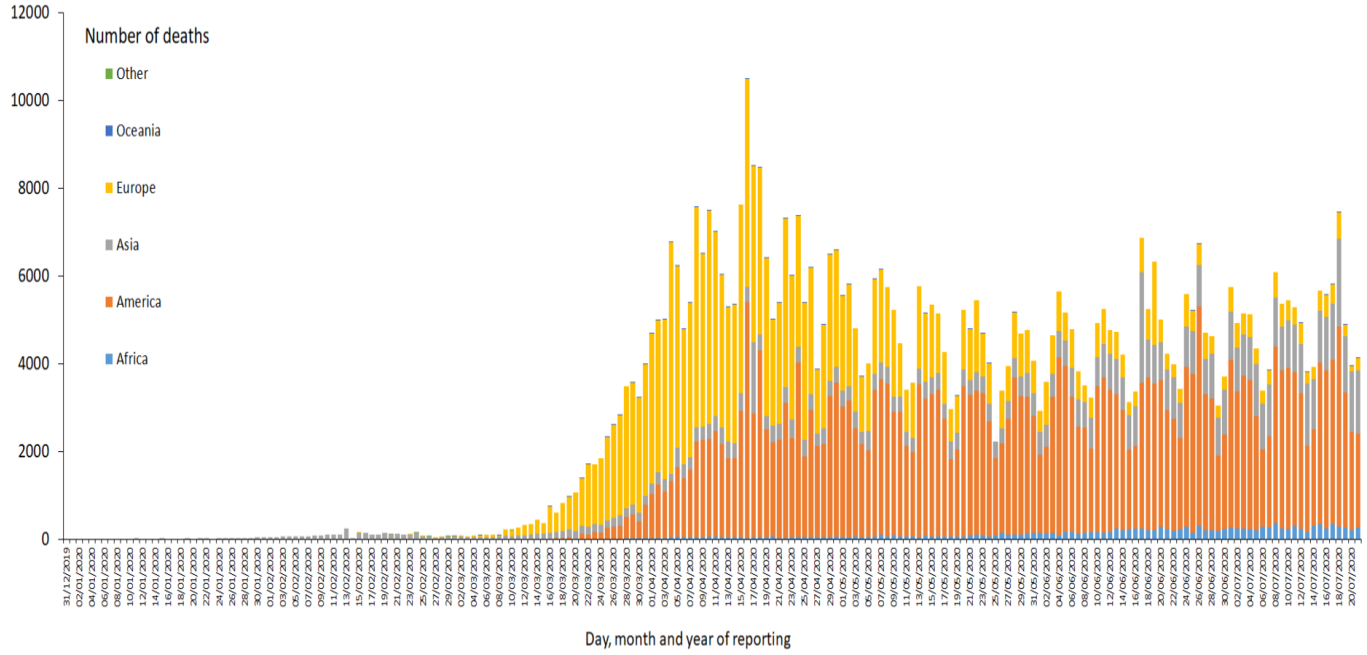
Statics

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 **14,538,094 confirmed cases** of COVID-19, including **607,358 deaths,** reported to WHO.



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.

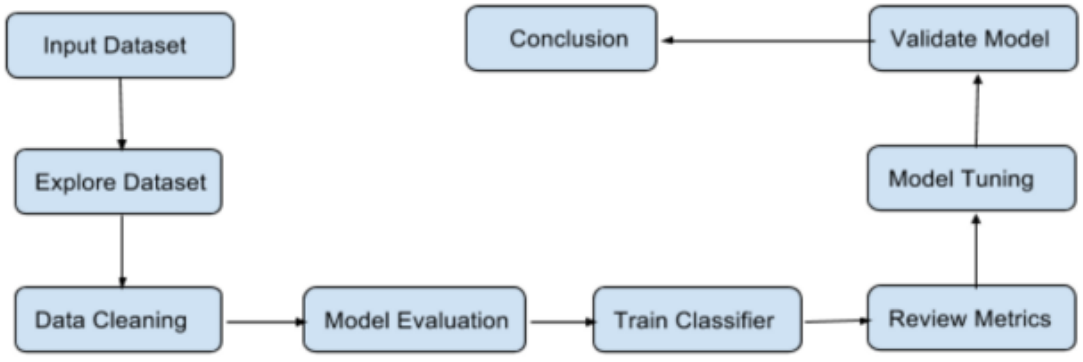
Datasets and Inputs

- 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)

Implementation



Benchmark Model

I will use the Convolutional Neural Networks (CNN) model . I will create it using transfer learning which has more accuracy rather than the model which is being created from scratch

This is a binary classification model

Evaluation metrics

For this multi class classification ,Multi class log will be used to evaluate the model .

Confusion metrics (What is this ?)

How is the classification being evaluated ?

The performance of each classification model is evaluated using three statistical measures ; classification accuracy ,sensitivity and specificity . It is using true positives ,true negative ,false positive and false negative . The percentage of correct and incorrect classification is the difference between the actual and predicted values of variables. True positives is the number of correct predictions that an instance is true , or in the words ; it occurs when the positive prediction of the classifier coincided with a positive prediction of target attribute . True negative is presenting a number of correct predictions that an instance is false ; it occurs when both the classifier , and the target attribute suggests the absence of positive prediction. The false positive is the number of incorrect predictions that an instance is true . Finally , False Negative is the number of incorrect predictions that an instance is false .The next table shows us the confusion matrix for a two class classifier .

	Predicted No	Predicted Yes
Actual no	TN	FN
Actual yes	FP	TP

Classification accuracy is known as the ratio of the number of correctly classified cases and is equal to the sum of TP and TN divided by the total number of cases (TN + FN + TP + FP).

$$\text{Accuracy} = \frac{TP + TN}{TN + FN + TP + FP}$$

Precision is known as the number of true positives over the number of true positives in addition to the number of false positives

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall is known as the number of true positives over the number of true positives in addition to the number of false negatives .

$$\text{Recall} = \frac{TP}{TP + FN}$$

The sensitivity indicates the correctly rated positive rate and is equal to the true positives divided by the sum of TP and FN. sensitivity may be referred to as a true positive rate.

$$\text{Sensitivity} = \frac{TP}{FN + TP}$$

Specificity refers to the rate of correctly classified negative and is equal to the ratio of TN to the sum of TN and FP

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Project Design

Step 1 : import the necessary libraries

Step 2 : pre-process the data , validate it and create the train , validation and test dataset

Step 3 : Build CNN model

Step 4 : train the model

Step 5 : Deploy the model

Step 6 : test it

References

- https://www.undp.org/content/undp/en/home/coronavirus.html
- https://covid19.who.int/?gclid=Cj0KCQjwpNr4BRDYARIsAADIx9zwAu9ZUMiqtVJ_liv_5VOw3du1f0K_Vg9jjUycN8DASK3SoHvLPScaAqHFEALw_wcB
- https://www.xlstat.com/en/covid
- https://medium.com/@madhavvyas_10553/bank-marketing-campaign-prediction-project-99fe1c3029a4