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 coronavirus COVID-19 pandemic is the defining global health crisis of our time and the greatest challenge we have faced since World War Two. Since its emergence in Asia late last year, the virus has spread to every state except Antarctica
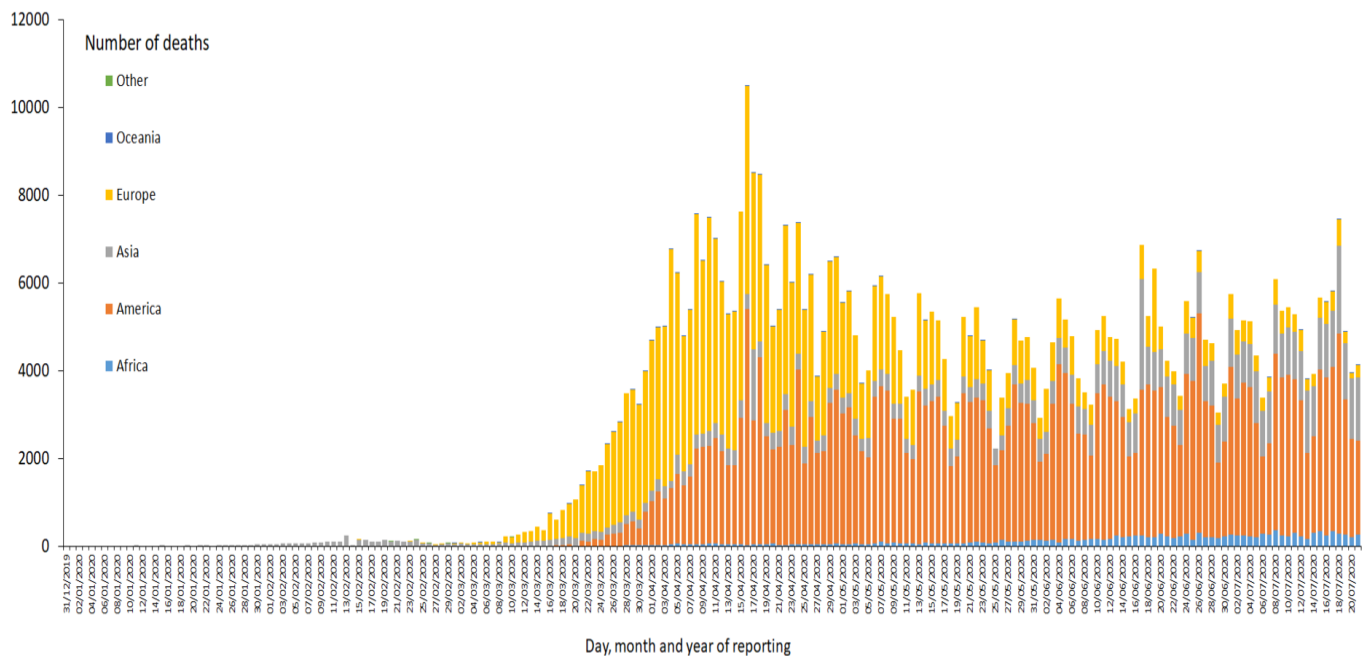
But the pandemic is much more than a health crisis, it's also an unprecedent socio-economic crisis. Stressing every one of the countries it touches, it has the potential to create devastating social, economic and political effects that will leave deep and longstanding scars. UNDP is the technical lead in the UN's socio-economic recovery, alongside the health response, led by WHO, and the Global Humanitarian Response Plan, and working under the leadership of the UN Resident Coordinators.

Statics

Every day, people are losing jobs and income, with no way of knowing when normality will return. Small island nations, heavily dependent on tourism, have empty hotels and deserted beaches. The International Labour Organization estimates that 195 million jobs could be lost.

The World Bank projects a US$110 billion decline in remittances this year, which could mean 800 million people will not be able 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.

Number of deaths
- Other
- Oceania
- Europe
- Asia
- America
- Africa

Day, month and year of reporting

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
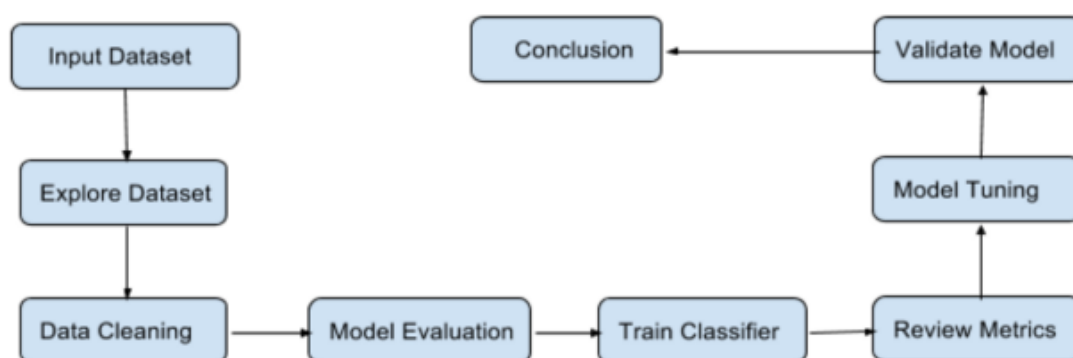
Solution

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

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 is occurring 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 .Table below shows the confusion matrix for a two class classifier .

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

Classification accuracy is defined 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 defined as the number of true positives over the number of true positives plus the number of false positives

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

Recall is defined as the number of true positives over the number of true positives plus the number of false negatives .

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

Sensitivity refers to the rate of correctly classified positive and is equal to true positives divided by the sum of TP and FN . Sensitivity may be referred 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