

COVID 19: Data Visualization of Global Cases & Analysis of Chest X-Ray Using Python

Amruta Aney¹, Manav Gupta², Yash Daund³

¹Student, Dept. of Electronics & Telecommunication Engineering, RAIT, Maharashtra, India

²Student, Dept. of Electronics & Telecommunication Engineering, RAIT, Maharashtra, India

³Student, Dept. of Electronics & Telecommunication Engineering, RAIT, Maharashtra, India

Abstract - COVID-19, was initially widely known as Corona-virus disease. The outbreak was declared as a Public Health Emergency of International Concern on 30 January 2020 and a Pandemic on 11 March 2020 by WHO (World Health Organization) It is the catastrophe of the century and the most critical challenge that we faced since the 2nd World War. In December 2019, the first case of this infectious disease occurred in Wuhan, China. This took a toll on the livelihood of people, their jobs, the agriculture sector. It also affected the society and global economy. With this project, we intend to analyze Chest X-Rays using PyTorch and create an easy-to-use User Interface which provides people with the insights of the cases worldwide. We will use a ResNet-18 model and implement it on a COVID-19 dataset. This dataset has approximately 3200 Chest X-Ray scans, which are categorized in three classes - Normal, Pneumonia and COVID-19.

Key Words: COVID-19, Chest X-Rays, PyTorch, ResNet-18 Model, Insights of cases

1. INTRODUCTION

The world has come to halt with countless lives being taken, reason being the pandemic (Covid 19). Basically, it has brought changes on a global scale. In order to contribute towards the betterment of the world and against this deadly virus, our team has come up with this project. This work includes detection of the COVID -19 with the knowledge of the Resnet.

-18 processing and PyTorch techniques. The database used in this work is the recordings of the healthy persons, patients suffering from viral pneumonia and COVID-19. Creation of an easy-to-use User Interface which provides people with the insights of the cases worldwide. The Paper focuses on developing a tracking interface to track the rising covid cases across the world with the help of World map geo-visualization also providing the user with a convenient method to detect the disease.

We start by accessing the data of the cases by using API and after pre- processing the data we generate choropleth map, circular markers and heat maps for depiction of the distribution of Covid-19 cases in order to learn about the severity of the disease in a particular area. In addition to this, we initialize the process of analysis of X-Rays by

creating a custom dataset, preparing a data loader and then Visualizing the Data to create a trained model to classify the given data into either a normal or a viral or covid positive case.

The pre-trained model ResNet18 is used in this project, trained on ImageNet dataset. ResNet is a popular CNN architecture, it gives an easy gradient, low and more efficient training, and won the 2015 ImageNet competition.

ResNet is based on identity shortcut connection which skips one or more layers. This helps the network giving a direct path to the top layers in the network, which makes gradient updates for those layers quite easier.

1.1 Covid 19: Signs & Symptoms

1. Symptomatic: COVID-19 patients usually complain of having fever, cough and sore throat. People who previously do not have disorders related to ear, nose and throat (ENT) experience a loss of taste and smell, which are also some other major symptoms of COVID-19 that one has to be on the constant lookout for. The incubation period is the number of days that have passed from the person getting infected to the occurrence of the initial indicators. The incubation period for Corona Virus as observed in the majority of patients is between two to fourteen days, the average being 11.5 days. A huge percentage of symptomatic people have been reported of experiencing one or more than one symptoms before 12th Day of exposure to the virus.

2. Asymptomatic: Around 20% of the infected people do not develop notable symptoms or signs at any stage of the infection, they can also be considered as pre-symptomatic as it cannot be predicted whether they will stay asymptomatic throughout the viral infection or not. Asymptomatic carriers spread the disease no less than a symptomatic patient. Other infected people may develop very mild symptoms, and can equally be carriers and be responsible for spreading the Corona Virus. [1]

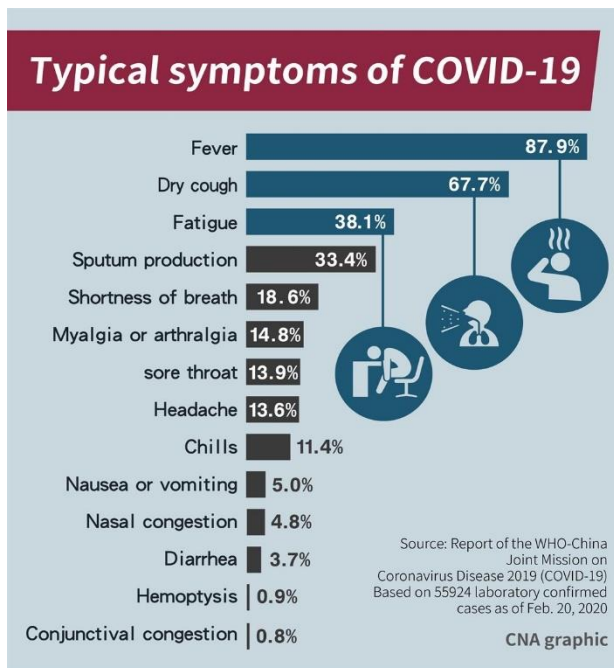


Figure 1 - Typical Symptoms of Covid-19

1.2 Virology

SARS-CoV-2 also known as coronavirus is a harsh critical respiratory syndrome. First ever cases of this anomaly were acquired from three people who suffered from pneumonia related to the clump of acute respiratory diseases, these cases were first found in Wuhan. Coronaviruses that occurs in nature has all the features similar to that of the novel SARS CoV-2 virus. SARS-CoV-2 has a lot in common to that of the original SARS-CoV. It is considered to have a zoonotic origination. It is observed that coronavirus genetically bundles with the genus Betacoronavirus, in subgenus Sarbecovirus (lineage B) together with two bat-derived strains. This has been concluded by thorough genetic analysis. [2]

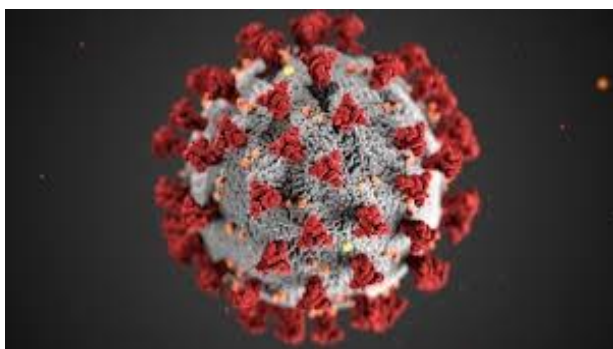


Figure 2 – Covid-19 Virus

1.3 Transmission

The spread of COVID-19 occurs mainly from person to another person for the most part which takes place by the respiratory path when after an afflicted person coughs, sneezes, sings, talks or breathes. When virus-containing particles in the form of aerosols or respiratory droplets are exhaled by an infected person and if they come in contact with a healthy person, that is if they get into the nose, eyes, or mouth of a healthy person who are in close contact with the affected person, the person gets infected. An average of 1100 infectious SARS-CoV-2 virions are considered to instigate a new infection during a case of man-to-man transmission. If an infected person does not follow social distancing norms to curb the spread, he/she can infect 406 patients in 30 days as said by the government, reported by India Today on 27th April 2021. This can be seen as a more infective version of influenza, but still less infectious than measles. It usually spreads in groups or clumps, where infections can be traced back to an geographical location or index case. Super-spreading events play a major role in transmission of the disease and causing it to affect a huge no. of people (SSEs), like family gatherings, festivities rallies, protests, funerals etc. The variants of COVID-19 variants have wrecked havoc in India ever since it recorded surges in the variant cases. B.1.1.7, a variant of COVID-19 was initially observed in the United Kingdom, was dominant in Punjab, India and Delhi, India. B.1.618, a new variant was seen in West Bengal, India. The variant B.1.617 is currently dominant in Maharashtra, India as indicated by genomic data. A rare Fungal disease - black fungus, scientifically also known as Mucormycosis has been observed to affect the people who have been infected with the Corona Virus. The current number of black-fungus cases are around 6000 in India with a very high mortality rate.

Control Methods : Some methods to curb the transmission of COVID are usage of respiratory mask, N95 masks, surgical masks or cloth masks or other coverings, observing social distancing at public place. Transmission can be reduced by not venturing outdoors and staying inside our houses with well equipped heating and ventilation systems to provide proper air circulation and thereby outdoor air usage to be increased. Using sanitizers for personal use well as keeping the surroundings hygienic and disinfected at all times. Residential, commercial and healthcare buildings are advised to carry out disinfecting procedures and also install sanitizers for people who enter the building. A warning board is put up outside building who are COVID prone zone or sealed down due to high number of cases, this prevents unnecessary visits to a building. [3]

2. Detection Of COVID-19 using RESNET-18 & PyTorch

2.1. X-Ray Database

We used the recently published datasets for COVID-19 linked to the Image Dataset for Chest X-rays, it comprises of sets of images from publications on COVID-19, gathered by Alif Rahman It contains X-Rays 1400 healthy patients, 1200 X-Rays of Pneumonic Patients, 400 X-Rays of COVID-19 patients. Only ant.-post. images are used for the prediction of COVID-19, as the medial images are necessary for the project and not the lateral ones.

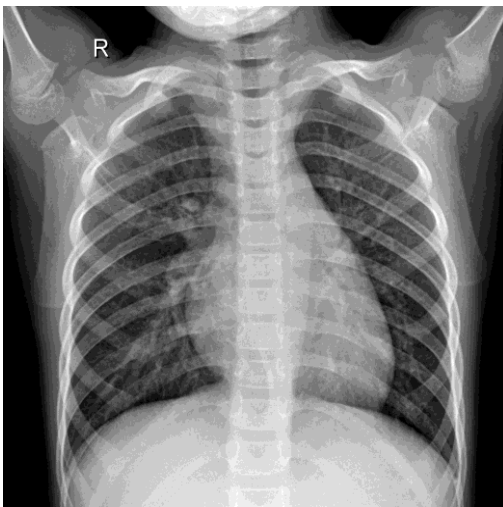


Fig 3- Healthy Patient

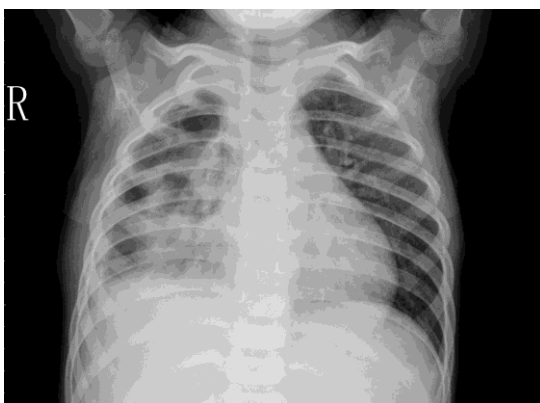


Fig 4 – Viral Pneumonia

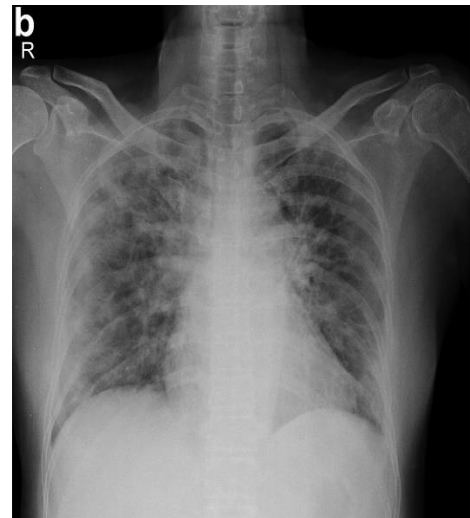


Fig 5- Covid-19 Patient

2.2 PyTorch

PyTorch is an open source machine learning library used for training and developing deep learning models based on neural network. It is primarily developed by Facebook’s AI research group. PyTorch can be used with Python as we have done the same.

2.3. ResNet -18

ResNet-18 can also perform as a 18 layers deep convolutional neural network. The pretrained version of the network can be loaded which is trained on as many as 1,000,000 images from the ImageNet database. Image classification into 1000 object categories, like keyboard, mouse, pencil, and many animals can be carried out by the pretrained network. As a result, for an accurate range of images the network has learned rich feature depiction. The picture input size of the network is 224-by-224. Residual Networks, abbreviated as ResNet can be a classic neural network used as a backbone for several computer vision tasks. Complex image analysis problems are worked out using deep learning by, network depth (stacking hundreds of layers) which is essential to extract critical features from training data and learn meaningful patterns. However, addition of neural layers often leads to computationally expensive and problematic situations due to the gradients. In this guide, problems with deep neural networks, the help ResNet provides, and therefore the things needed to use ResNet in transfer learning will all be covered and studied properly. The vanishing gradient effect on network output with regard to parameters becomes extremely small within the initial layer while training is carried out. When the gradient reached its least value it becomes further smaller. As a result, updation of weights in initial layers occurs very slowly or remain pretty much identical, resulting in a growth of error.

Architecture of ResNet-18:

It is easy to visualize how configuration of layers occurs in the architecture of ResNet-18. Firstly a convolution layer with stride 2 and 7x7 kernel size is present. Skip connection begins after the end of this convolution layer. Achieving the output by two convolution layers and 3x3 max pool layer, kernel size 3x3, 64 kernels are added together with the input received. This is called to be the first residual block.

Then the addition of the output of this residual block is carried out with the output of 2 convolution layers with 128 such filters and kernel size 3x3. This accounted for the second residual block. The third residual block consists of the output of two convolution layers with 256 such filters and filter size 3x3 and the output of the second block taken via a skip connection. The output of two convolution layers and output of third block through skip connection with same filter size of 3x3 and 512 such filters is what constitutes the fourth and final residual block.

Finally, the output of the final residual block is made to undergo average pooling and the feature map is forwarded to the entirely connected layers which is then followed by soft- max function to receive the final output.

The diagram depicts the output of each layer and input in the skip connections is manipulated according to that. [4]

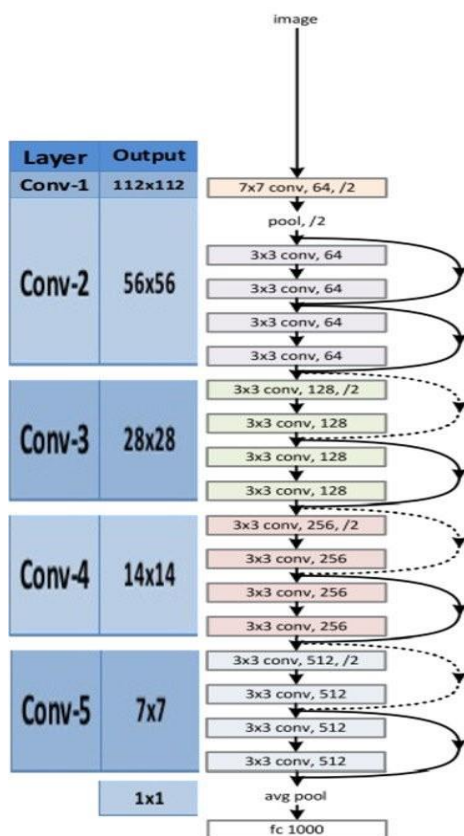


Fig 6- Architecture of ResNet-18

Applying ResNet-18 Model:

The ResNet-18 Model is applied to the set of X-rays and thus evaluated for val loss and accuracy.

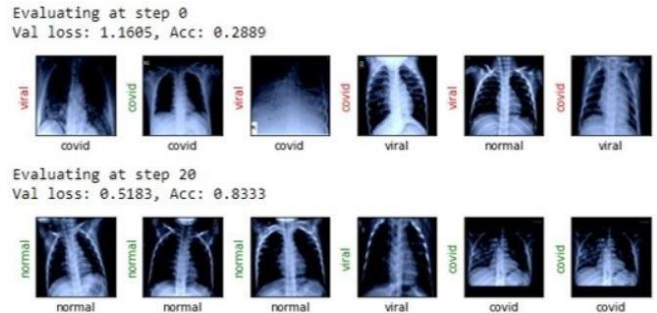


Fig 7- Evaluation of ResNet-18

Final Results:

The results of the X-ray is 96.67% accurate.



Fig 8 – Final Results

3. Python World Map Geo-visualization

3.1 Access Data via API

API is the abbreviation for application programming interface. “Interface” is the most important term of API because an API primarily communicates with the program for you. Architecture of an API. APIs consist of three parts:

- 1.User: The person who initiates a request.
- 2.Client: The computer that is sending the request to the server.
- 3.Server: The computer that is responding to the request sent by the client. The server acquires and holds data hence it is developed first. As that server is running, documentation is published by programmers, it includes the endpoints where the specific data is present. This documentation tells gives information of the structure of the data that is used. A user can then search or query the data on the server, or build a program that runs searches on the database and then transforms the gained information into a usable format.

3.2 Data Pre-Processing

Data processing technique that is employed to rework the data using an well-organized format is Data-Preprocessing. Steps required to carry out Data Preprocessing: Architecture of an API is as follows. APIs mainly comprise of three parts:

1. Data Cleaning:

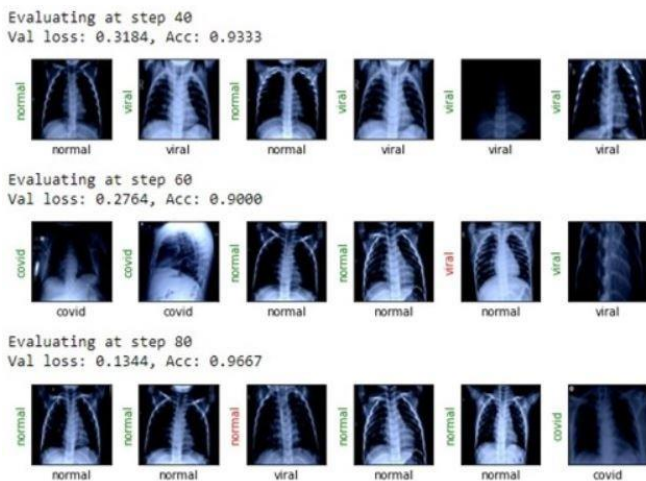
The data might consist of many missing and insignificant parts. Data cleaning is performed to get rid of the excess waste data. It consists of maneuvering missing data or noisy data etc.

a. Missing Data:

This situation arises when some data is missing in the data. It can be handled in various ways.

b. Noisy Data:

The data that can't be interpreted by machines is meaningless and hence called as noisy data. It can be generated due to data entry errors, faulty data collection, etc.



2. Data Transformation: This step takes care of reworking the info. in suitable forms appropriate for mining process. It includes Normalization, Attribute Selection, Discretization and Concept of hierarchy generation.

3. Data Reduction: Data pre-processing is a technique which may not be able to handle large amounts of data of data. While working with huge volume of knowledge, analysis became harder in such cases. In order to get rid of this, we use data reduction technique. It aims to reduce analysis costs and data storage and extend the storage efficiency. [5]

3.3 Choropleth Map

Choropleth maps are popular thematic maps used to represent statistical data through various shading patterns or symbols on predetermined geographic areas. [6] They are good at using data to easily represent variability of the desired measurement, across a region. These are maps, where areas are shaded according to a prearranged key, each shading or colour type representing a variety of values. Varying colors are used to indicate territorial

boundaries and changes. Darker colors are used to indicate greater amounts of what is being measured.

This Map illustrates the Total Confirmed Cases around the Globe.

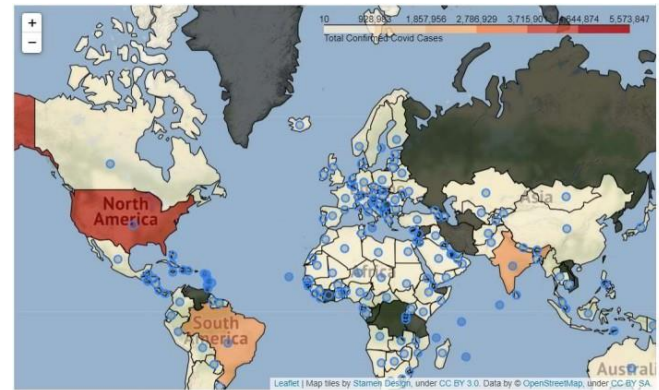


Fig 9 - Total Confirmed COVID-19 Cases (Country-Wise)

3.4 Heatmaps

A heatmap is a two-dimensional visual representation of data using colors, where the colors all represent different values. [7] Enter the data in an Excel sheet i.e the latitudes and longitudes of various countries along with their names as required in the project. Select the same dataset so as to generate a heatmap. Then we use Conditional Formatting. Conditional Formatting is used and we select Color Scales for our heat map.

The map below illustrates the Total number of deaths due to COVID-19 country-wise.

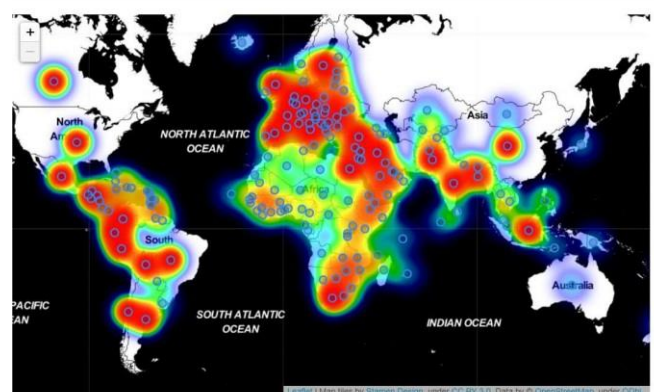


Fig 10 - Total Confirmed Deaths (Country-Wise)

3.5 State-Wise Cases

This map depicts the Total confirmed cases, Total Recovered Cases and Total Active cases of States in India, in addition to it, it also gives the statistics about cities like Mumbai, Pune etc located in Maharashtra.

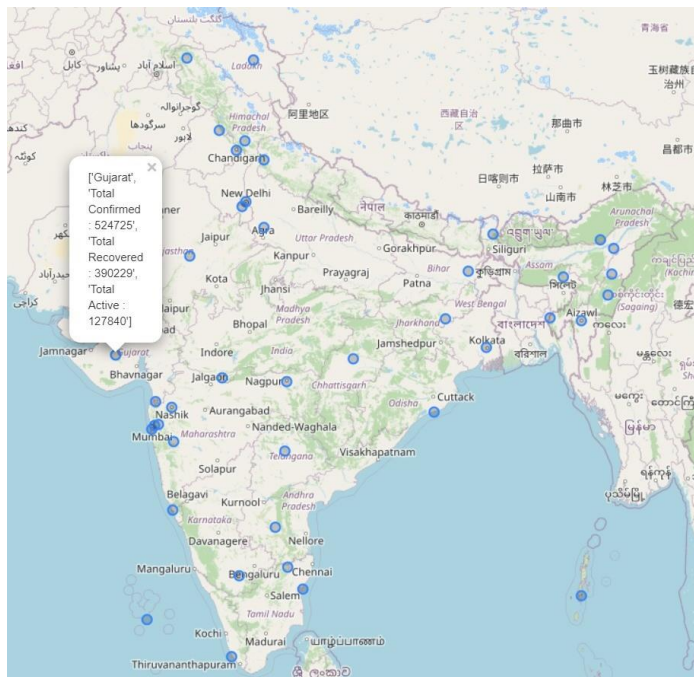


Fig. 11- Total Confirmed, Recovered and Active Cases State-Wise (India)

3.6 Vaccines Administered

This map portrays the Total no. of vaccines administered till date in the states of India.

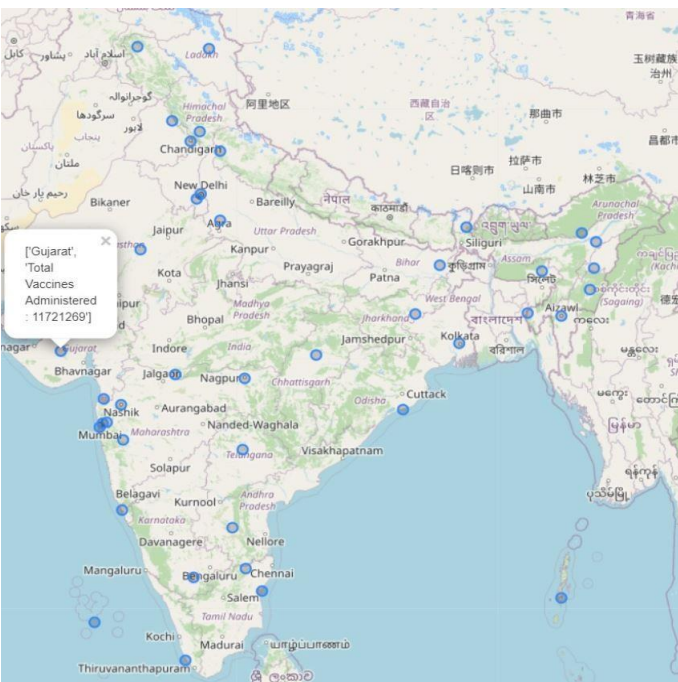


Fig. 11- Total Vaccines Administered State-Wise (India)

3.7 Graphical Representation

Another form of representation to depict the covid dataset. This representation includes three bar graphs representing three aspects of covid data. The first graph is a bar plot depicting the bar plot of active and recovered cases in different countries. The second graph represents the number of deaths in the country in descending order on a global scale. And finally the third graph represents the total recovered cases and active cases state wise in the country of India.

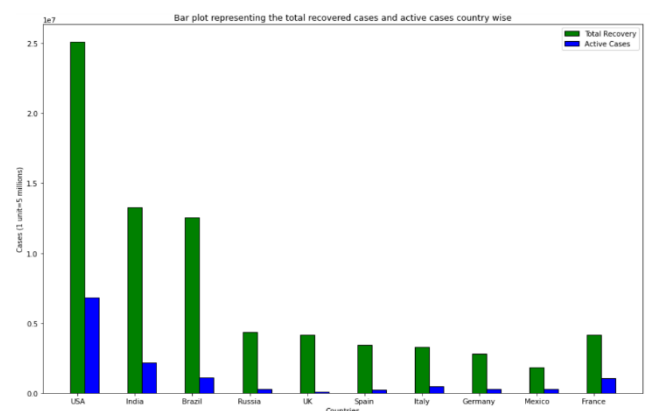


Fig. 11- Total Recovered and Active cases (Country-Wise)

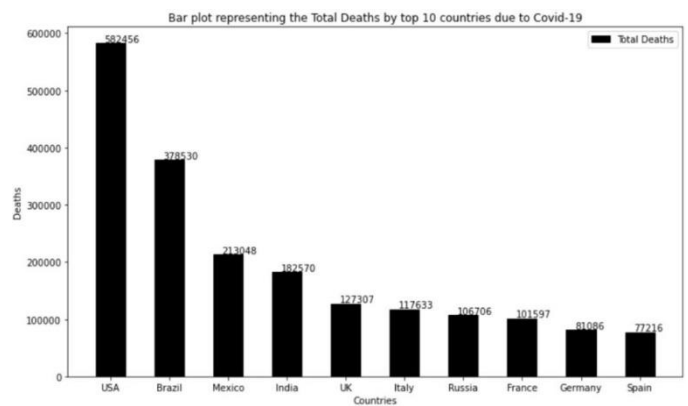


Fig. 12- 10 Countries with highest no. of fatalities

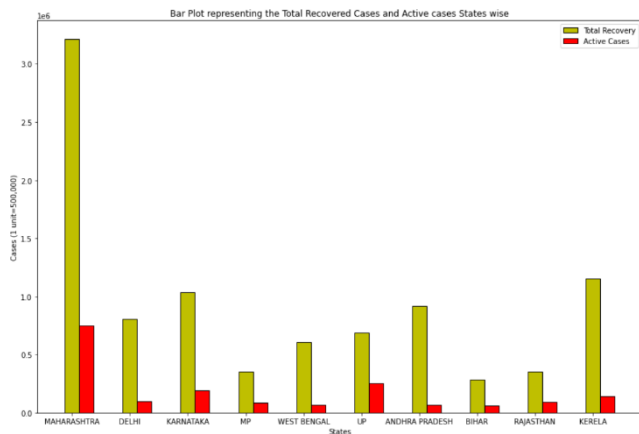


Fig. 13- Total Recovered and Active cases (Country-Wise)

[6] <https://www.internetgeography.net/choropleth-maps-in-geography/>

[7] <https://www.investopedia.com/terms/h/heatmap.asp>

3.8 Convert Jupyter Notebook to Webpage

Convert the existing Jupyter notebook document into a html file. Hosted the Website by uploading the html file on the Hosting platform.

Get Access to the Webpage by using the URL generated.

<https://covid-19-maps.000webhostapp.com>

4. CONCLUSION

The main reason of using the pre-trained ResNet18 is that, it is trained on ImageNet dataset. ResNet is one of the most popular CNN architecture, which provides easier gradient low for more efficient training. In this project we intended to provide a user – friendly access to the Insights of the current scenario of COVID-19. Using Python as the programming language we efficiently displayed the Data in the form of maps which is easy to navigate through, for the users. The Paper focuses on developing a tracking interface to track the rising covid cases and vaccines administered across the world with the help of World map geo-visualization also providing the user with a convenient method to detect the disease.

5. REFERENCES

[1] <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>

[2] <https://www.uptodate.com/contents/covid-19-epidemiology-virology-and-prevention>

[3] <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted>

[4] <https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8>

[5] <https://www.mulesoft.com/resources/api/what-is-an-api>