VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Macche, Santibastwada Road, Belagavi-590018, Karnataka

MINI PROJECT REPORT : COVID-19 DATA ANALYSIS USING X-RAYS VIA CNN WITH OUTBREAK PREDICTION AND DATA VISUALISATION

Submitted in partial fulfilment of the requirement for the degree of

Bachelor of Engineering
in
Electronics & Communications Engineering

Submitted by

1DS19EC128: SHRISHTI SINGH
1DS19EC132: SOURAV SHARMA
1DS19EC143: TANISH RAI
1DS19EC150: VAIBHAV KHANNA

SEMESTER: 6TH
SECTION/BATCH: C / C01

SECTION COORDINATOR :- PROF. VIDYASHREE K.N.

MINI PROJECT GUIDE :- PROF. NIRMALA Y.N.

Department of Electronics & Communication Engineering
Dayananda Sagar College of Engineering
(An Autonomous College affiliated to VTU Belgaum & accredited by NBA/NAAC)
Shavige Malleshwara Hills, Kumaraswamy Layout,
Banashankari, Bengaluru-560078, Karnataka

Certificate

Certified that the mini-project work **(19EC6DCMPR)** entitled "Covid-19 Data Analysis Using X-Rays Via CNN With Outbreak Prediction And Data Visualisation" carried out by **Shrishti Singh** (USN-1DS19EC128), **Sourav Sharma** (USN-1DS19EC132), **Tanish Rai** (USN-1DS19EC143), **Vaibhav Khanna** (USN-1DS19EC150) are bonafide students of the ECE Dept. of Dayananda Sagar College of Engineering, Bangalore, Karnataka, India in partial fulfilment for the award of Bachelor of Engineering in Electronics & Communication Engineering of the Visvesvaraya Technological University, Belagavi, Karnataka for the VI Semester course during the academic year 2021-22. It is certified that all corrections / suggestions indicated for the mini-project work have been incorporated in the mini-report submitted to the ECE department. This Mini-Project report has been approved as it satisfies the academic requirement in respect of mini-project work prescribed for the said degree.

Mini-Project Guide Sign : _____

Name : **Prof. Nirmala YN**

Mini-Project Section Coordinator Sign :
_____

Name : **Prof. Vidyashree KN**

Mini-Project Convener & Chief Coordinator Sign : _____

Name : **Dr. Shashi Raj K.**

Dr. T.C. Manjunath : _____

HOD, ECE, DSCE.

External Mini-Project Viva-Voce (SEE)

Name of the mini-project examiners (int & ext) with date :

1 : Signature : _____

2 : Signature : _____

Declaration

Certified that the mini-project work entitled, “Covid-19 Data Analysis Using X-Rays Via CNN With Outbreak Prediction And Data Visualisation” with the course code **19EC6DCMPR** (2 Credits, 100 Marks, CIE & SEE 50 marks each) is a bonafide work that was carried out by ourselves in partial fulfilment for the award of degree of Bachelor of Engineering in Electronics & Communication Engg. of the Visvesvaraya Technological University, Belagavi, Karnataka during the academic year 2021-22 for the VI Semester Autonomous Course. We, the students of the mini-project group/batch no. **C01** does hereby declare that the entire mini-project has been done on our own & we have not copied or duplicated any other’s work. The results embedded in this mini-project report have not been submitted elsewhere for the award of any type of degree.

Student Name-1 : SHRISHTI SINGH
USN : 1DS19EC128

Sign : _____

Student Name-2 : SOURAV SHARMA
USN : 1DS19EC132

Sign : _____

Student Name-3 : TANISH RAI
USN : 1DS19EC143

Sign : _____

Student Name-4 : VAIBHAV KHANNA
USN : 1DS19EC150

Sign : _____

Date : 02 /07/2022

Place : Bengaluru -560078

Acknowledgement

In this section, a brief acknowledgment has to be presented as a thanks giving to all the people who have directly or indirectly helped in the successful completion of the undergraduate mini-project work.

Thanksgiving hierarchy to the members in this order :

1. Chairman – Dr. Hemachandra Sagar
2. Vice Chairman – Dr. Premachandra Sagar
3. Secretary – Galiswamy
4. Joint Secretary – Tintisha Sagar
5. Principal – Dr. C.P.S. Prakash
6. HOD – Dr. T.C. Manjunath
7. Mini Project Guide – Prof. Nirmala YN
8. Mini-Project Section-wise Coordinators–
   Kumar P. (A) ; Kavitha S. Guddad (B) ; Vidyashree K.N. (C) ; Shobha A.S. (D)
9. Mini-Project Convener & Chief Coordinator – Dr. Shashi Raj K.
10. Dept. staff members – Teaching & Non-teaching
11. Family members & relatives
12. Friends
13. Almighty

Table of Contents:

LIST OF FIGURES:

INTRODUCTION

The COVID-19 pandemic, also known as the coronavirus pandemic, is a global pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The novel virus was first identified from an outbreak in Wuhan, China, in December 2019. Attempts to contain it there failed, allowing the virus to spread worldwide. The World Health Organisation (WHO) declared a Public Health Emergency of International Concern on 30 January 2020 and a pandemic on 11 March 2020. As of 29 June 2022, the pandemic had caused more than 545 million cases and 6.33 million confirmed deaths, making it one of the deadliest in history.

COVID-19 symptoms range from undetectable to deadly, but most commonly include fever, dry cough, and fatigue. Severe illness is more likely in elderly patients and those with certain underlying medical conditions. COVID-19 transmits when people breathe in air contaminated by droplets and small airborne particles containing the virus. The risk of breathing these in is highest when people are in close proximity, but they can be inhaled over longer distances, particularly indoors. Transmission can also occur if contaminated fluids reach the eyes, nose or mouth, and, rarely, via contaminated surfaces. Infected persons are typically contagious for 10 days, and can spread the virus even if they do not develop symptoms. Mutations have produced many strains (variants) with varying degrees of infectivity and virulence.

The pandemic COVID-19 is a global challenge which has infected and killed people worldwide. Some people do not show any symptoms while some have fever, cough, sore throat, general weakness and fatigue and muscular pain and in most cases, severe pneumonia, acute respiratory distress syndrome, sepsis and septic shock all leading to death. It has adversely affected the economy and social integrity of countries. There is rising concern about the mental health challenges of the general population (children, adults, or elderly), along with health workers and families of infected people. This study aims to determine the effect of COVID-19 on mental health of people in India. It also focuses on the stigma and discriminating factors in our society and ways to cope with such conditions. A structured survey was conducted with 250 participants of different age groups. Our analysis focuses on the factors affecting mental health of any person, changes in behaviour and daily routine due to stress, anxiety or fear of transmission of virus in their family and friends, some are worried for their lifestyle and career. There is a need to understand that a pandemic is affecting everyone, either physically or mentally. There must be increase in the study of the aspects of mental health during the pandemic and methods to cope with issues like discrimination for better mental health during pandemic period.

ABSTRACT

Most detection methods of coronavirus disease 2019 (COVID-19) use classic image classification models, which have problems of low recognition accuracy and inaccurate capture of modal features when detecting chest X-rays of COVID-19. This study proposes a COVID-19 detection method based on image modal feature fusion. This method first performs small-sample enhancement processing on chest X-rays, such as rotation, translation, and random transformation. Five classic pre-training models are used when extracting modal features. A global average pooling layer reduces training parameters and prevents overfitting. The model is trained and fine-tuned, the machine learning evaluation standard is used to evaluate the model, and the receiver operating characteristic (ROC) curve is drawn. Experiments show that compared with the classic model, the classification method in this study can more effectively detect COVID-19 image modal information, and it achieves the expected effect of accurately detecting cases. Deep learning techniques are widely used to design robust classification models in several areas such as medical diagnosis tasks in which it achieves good performance. In recent years Pneumonia causes 15% of the total number of deaths in children under the age of 5. It can be caused by viruses, bacteria or fungi, which led the researchers to focus their studies on identifying pneumonia based on Chest X-ray images, using deep learning techniques. In this paper, we propose a CNN model (Convolutional Neural Network) for the classification of Chest X-ray images. The proposed method is based on a non-complex CNN and without the use of transfer learning. Our proposed model uses fewer parameters and thus reduces the training time. In this study, and because of the low availability of data, we used the data augmentation method to eliminate overfitting and to improve the accuracy of the validation and classification of the model. The obtained results prove the efficiency of the proposed architecture compared with the state of the art methods.

Lack of Coordination and Direction

The WHO and government guidelines were changing continuously given the disease is new and previous knowledge is little. Consequently, doctors remained uncertain about the line of treatment. These uncertainties created additional mental stress for medical professionals. The participants reported that patients were unaware of any safety protocols. COVID-19-positive patients often come to medical facilities to receive standard medical consultation, which puts COVID-negative patients as well as the medical workers at-risk. In several cases, doctors and nurses got infected because patients did not reveal that they were COVID-19-infected. A high-level coordination failure was prevalent in the healthcare administrations.

Moreover, healthcare workers were dissatisfied about some discriminatory initiatives taken up by the authority. In contrast, healthcare professionals did double or triple shifts, which was frustrating. Besides, they did not have any training regarding how to function

correctly in a virus outbreak. It was also perceived that the authority involved more administrators and fewer specialists to tackle this pandemic.

LITERATURE SURVEY :

The list of papers along with the publications, author and their publish dates have been mentioned below:

1. Research on Classification of COVID-19 Chest X-Ray Image Modal Feature Fusion Based on Deep Learning. Author : Daniel Cafoulla, Yanzhong Zhao, Qianchuan Zhao.

   Publish Year : 25 August, 2021.

2. Impacts of COVID-19 Pandemic on Geopolitics, Health, Economics, Education and Sociocultural Events. Authors : Hamrouni AM , Sharif RS, Sharif SI, Hassanein MM , Abdullkarem AR**.**

   Publish Year : 5 May, 2022.

3. Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network. Authors : Sayantani Ghosh, Rohit Dutta, Sachin Agarwal.

   Publish Year: 19 March, 2021.

Deep learning techniques are widely used to design robust classification models in several areas such as medical diagnosis tasks in which it achieves good performance.
In recent years Pneumonia causes 15% of the total number of deaths in children under the age of 5. It can be caused by viruses, bacteria or fungi, which led the researchers to focus their studies on identifying pneumonia based on Chest X-ray images, using deep learning techniques. Here, we propose a CNN model (Convolutional Neural Network) for the classification of Chest X-ray images. The proposed method is based on a non-complex CNN and without the use of transfer learning. Our proposed model uses fewer parameters and thus reduces the training time. In this study, and because of the low availability of data, we used the data augmentation method to eliminate overfitting and to improve the accuracy of the validation and classification of the model. The obtained results prove the efficiency of the proposed architecture compared with the state of the art methods. The detection of COVID-19 in this article requires several stages, as shown in Figure. The

original X-ray image is preprocessed, including size adjustment, rotation, position translation, cross-cutting transformation, scaling, and flip processing. The dataset is then divided into training and validation (test) sets. The preprocessed data are used to extract the modal feature information of the X-ray images through pretraining models by transfer learning, and this is input to the fully connected (FC) layer and trained after fusion. The first two layers of the FC layer contain 512 hidden units, followed by the ReLU activation function, and the last layer contains a hidden unit, followed by the sigmoid activation function, which is used to detect COVID-19. The performance of the system is evaluated by indices such as accuracy, recall rate, precision, and F1-score.The number of active cases of COVID-19 and deaths is markedly escalating. The pandemic had affected almost every aspect of our lives including geopolitics, health, economics, education, and sociocultural events. However, besides the negative impacts of the pandemic, there are some positive impacts as well, such as improving our awareness of the daily hygienic practices, emphasising digital inequality, and increasing global collaboration in combating the crisis by intensifying scientific research to establish a promising vaccine.

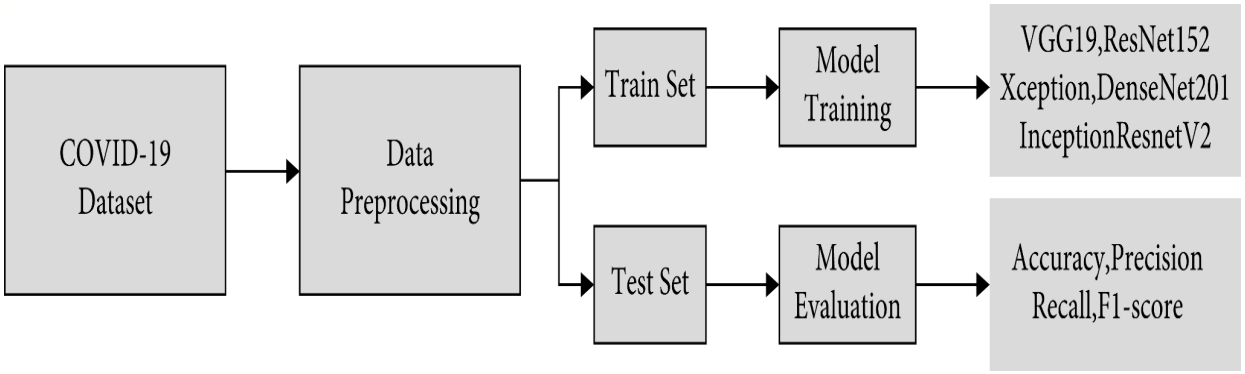

Figure 1

PROBLEM STATEMENT

- People with suspected COVID-19 need to know quickly whether they are infected, so they can receive appropriate treatment, self-isolate, and inform close contacts but there are various shortcomings.
- Blood tests are costly (not affordable by all sections of the society).
- Blood test results can take several minutes to several weeks, depending on the test. It is common to wait a day or two for most results to come back. A person should talk to their doctor or lab about what to expect regarding their test results.
- Rapid Spread Of the Virus poses a serious threat to even the life of the doctors and the

medical team. The disease is spread by droplet transmission. As of April 2020, the total number of infected individuals stands at around 3 million, with ~200,000 deaths and more than 1 million recoveries globally. The virus thus has a fatality rate of around 2% and an R_0 of 3 based on current data. However, a more recent report from the CDC, Atlanta, USA, claims that the R_0 could be as high as 5.7. It has also been observed from data available from China and India that individuals likely to be infected by the virus from both these countries belong to the age groups of 20–50 years.

- Extent of The Spread In the Body Can be detected using Deep Learning Models And CNN within minutes.
- If we can analyse the present situation due to this havoc-causing Pandemic,
- We can use this meaningful information for predicting similar probable occurring future events also.

METHODOLOGY

Chest imaging is commonly used in medicine, and it plays an important role in the detection of COVID-19. Through the diagnosis of chest imaging, medical staff can more accurately grasp the imaging modal characteristics of COVID-19 cases, such as multiple small patchy shadows and interstitial changes in the early stage, which are obvious outside the lungs. It then develops into multiple ground glass and infiltration shadows in both lungs. In severe cases, lung consolidation and pleural effusion are rare. It has important guiding value for accurately judging the condition and its development, formulating treatment plans, and evaluating prognosis.
Using the following methods for the detection of COVID-19 positive or negative patient, we can use the following algorithms :-

TRANSFER LEARNING

Transfer learning improves learning by transferring knowledge from related tasks that have been learned, i.e., transferring learned and trained parameters to a new model to help with its training. The architecture of deep learning models is complex and data dependent, requiring much data to train them. Much COVID-19 data are published online, but the number of samples is small, making it difficult to train a deep learning model from start to finish. Transfer learning can facilitate the training of such a small sample dataset to achieve the research purpose.

Apostolopoulos adopted transfer learning to detect the performance of different models in a small sample of pneumonia image datasets. Rafi used chest X-ray images to identify patients with COVID-19, using transfer learning methods to train DenseNet121 and ResNet152 series

models. Taresh et al. discussed the effectiveness of artificial intelligence in the rapid and reliable detection of COVID-19 based on chest X-ray images and applied transfer learning technology to detect COVID-19 from chest radiographs. Majeed et al. compared 12 transfer learning CNNs in the detection of COVID-19 from chest X-rays. The COVID-19 samples collected for our experiment were limited. To obtain better experimental results, different CNN models trained on ImageNet, a database of approximately 14 million images, were used to train the COVID-19 dataset.

CONVOLUTIONAL NEURAL NETWORKS

Xception is an improvement of Inception V3, replacing its convolution operation with depthwise separable convolution, which divides traditional convolution into the steps of depthwise and pointwise convolution.

The InceptionResNetV2 model is a CNN with top accuracy on the ILSVRC image classification benchmark. It is based on Google's Inception V3 model and draws on the ideas of ResNet, a 152-layer neural network successfully trained by using the ResNet Unit. The error rate on Top5 is 3.57%. It has fewer parameters than VGGNet, and the effect is outstanding. It introduces the idea of residual learning, which effectively solves the problem of network degradation.

The VGG family is used in face recognition and image classification, where VGG19 has better performance. VGG19 has 19 hidden layers, consisting of 16 convolutional layers and three fully connected layers. The input is set to 224 × 224 RGB images. The RGB average of all images is calculated on the training set image, and the image is passed as input and enters the VGG19 convolutional network.

DenseNet builds a connection relationship between layers, makes full use of features, and further alleviates the problem of gradient disappearance. The use of a bottleneck layer, translation layer, and smaller growth rate makes the network narrower, reduces the parameters, effectively suppresses overfitting, and reduces calculation.

FUSION MODEL

The experimental architecture uses pretrained Xception, ResNet152, DenseNet201, VGG19, and InceptionResNetV2 CNN to extract the feature information of COVID-19 X-ray lung images. Each network has three FC layers, where the last layer is FC for classification. In this experiment, the FC layer behind each network is replaced by the global average pooling layer, which can effectively reduce the training parameters. The dataset contains chest X-ray images of COVID-19 cases and healthy people. We set the label of COVID-19 images to 1 and images of healthy people to 0, for training and evaluating the model. The DarkCovidNet model is trained to detect two classes: COVID-19 and *No-Findings*

categories. The performance of the proposed model is evaluated using the 5-fold cross-validation procedure for both the binary and triple classification problem. Eighty percent of X-ray images are used for training and 20% for validation.

BLOCK DIAGRAM

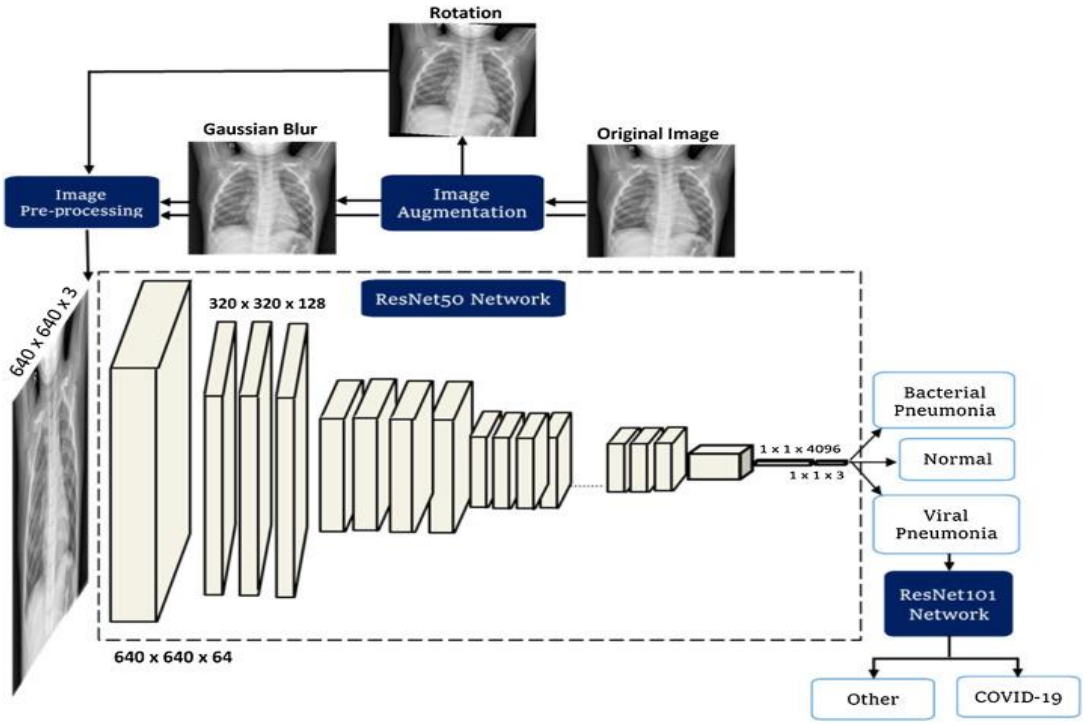

Figure 2

WORKING PRINCIPLE

Image classification is the process of segmenting images into different categories based on their features. A feature could be the edges in an image, the pixel intensity, the change in pixel values, and many more. An image consists of the smallest indivisible segments called pixels and every pixel has a strength often known as the pixel intensity. Whenever we study a digital image, it usually comes with three colour channels, i.e. the Red-Green-Blue channels, popularly known as the “RGB” values.

Now if we take multiple such images and try to label them as different individuals we can do it by analysing the pixel values and looking for patterns in them. However, the challenge here is that since the background, the colour scale, the clothing, etc. vary from image to image, it is hard to find patterns by analysing the pixel values alone. Hence we

might require a more advanced technique that can detect these edges or find the underlying pattern of different features in the face using which these images can be labelled or classified. This is where a more advanced technique like CNN comes into the picture.

CNN(Convolutional Neural Networks):

CNN or the convolutional neural network (CNN) is a class of **deep learning neural networks**. In short, think of CNN as a machine learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. CNN works by extracting features from the images. Any CNN consists of the following:

1. The input layer which is a grayscale image
2. The Output layer which is a binary or multi-class labels
3. Hidden layers consisting of convolution layers, ReLU (rectified linear unit) layers, the pooling layers, and a fully connected Neural Network.

It is very important to understand that ANN or Artificial Neural Networks, made up of multiple neurons, is not capable of extracting features from the image. This is where a combination of convolution and pooling layers comes into the picture. Similarly, the convolution and pooling layers can't perform classification hence we need a fully connected Neural Network.

ALGORITHM PERFORMING IMPLEMENTATION OF CNN :-

- Neural network with one convolutional layer and four filters.
- Initialises the weights of the convolutional layer to be the weights of the 4 defined filters.
- Assuming there are 4 grayscale filters; declare the CNN layer here.
- Size of the kernel equals size of the filter, usually the Kernels are smaller in size
- Calculation of the output of a convolutional layer pre- and post-activation and then return both the layers

- Instantiate the model and set the weights
- Print out the layer in the network
- We create the visualisation layer, call the class object, and display the output of the Convolution of four kernels.
- Convert the image into an input tensor
- print(type(gray_img_tensor))
- print(gray_img_tensor)
- Get the convolutional layer (pre and post activation)
- Visualise the output of a convolutional layer

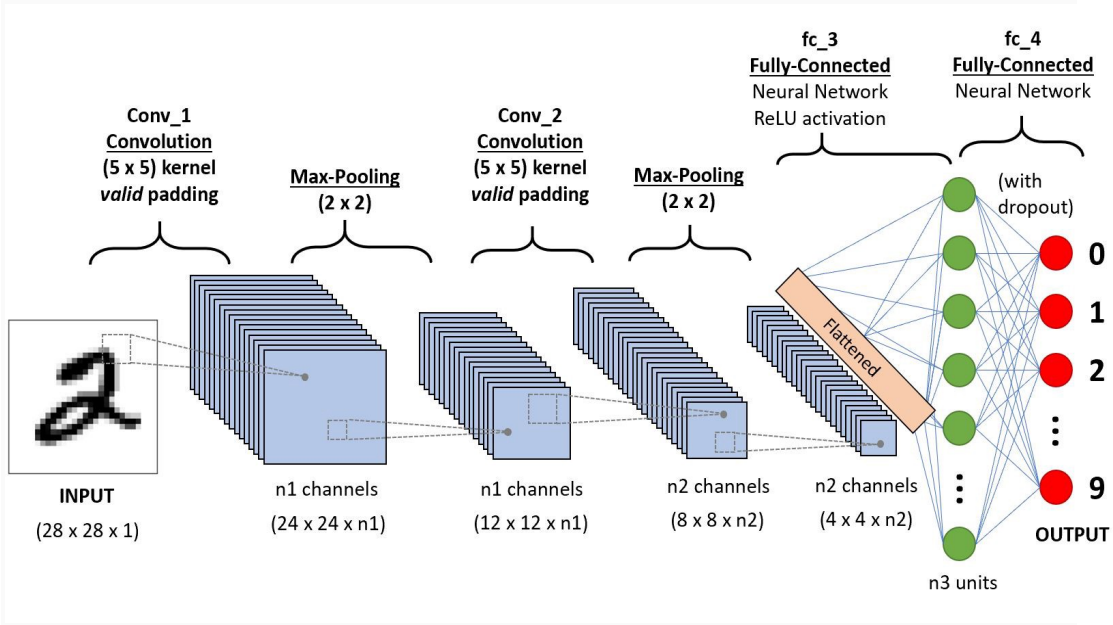

Figure 3

CONVOLUTIONAL LAYERS

In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM).The output is termed as the Feature map which gives us information about the image such as the corners and edges. Convolution leverages three important ideas that motivated computer vision researchers: sparse interaction, parameter sharing, and equivariant representation. Trivial neural network layers use matrix multiplication by a matrix of parameters describing the interaction between the input and output unit. This means that every output unit interacts with every input unit. However, convolution neural networks have

sparse interaction. This is achieved by making the kernel smaller than the input e.g., an image can have millions or thousands of pixels, but while processing it using a kernel we can detect meaningful information that is tens or hundreds of pixels. This means that we need to store fewer parameters that not only reduces the memory requirement of the model but also improves the statistical efficiency of the model.

POOLING LAYERS

The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. The pooling layer applies a non-linear down-sampling on the **convolved feature** often referred to as the **activation maps**. This is mainly to reduce the computational complexity required to process the huge volume of data linked to an image. Pooling is not compulsory and is often avoided. Usually, there are two types of pooling, **Max Pooling,** that returns the maximum value from the portion of the image covered by the Pooling Kernel and the **Average Pooling** that averages the values covered by a Pooling Kernel. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon the method used, there are several types of Pooling operations.In Max Pooling, the largest element is taken from the feature map.This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually.There are several pooling functions such as the average of the rectangular neighbourhood, L2 norm of the rectangular neighbourhood, and a weighted average based on the distance from the central pixel.

FULLY CONNECTED LAYERS

In this, the input image from the previous layers are flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place. Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.The FC layer helps to map the representation between the input and the output.

DROPOUT

When all the features are connected to the FC layer, it can cause overfitting in the

training dataset. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model's performance when used on new data.To overcome this problem, a dropout layer is utilised wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model.

ACTIVATION FUNCTIONS

They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. It decides which information of the model should fire in the forward direction and which ones should not at the end of the network.It adds non-linearity to the network.There are several commonly used activation functions such as the ReLU, Softmax, tanH and the Sigmoid functions. Each of these functions have a specific usage. For a binary classification CNN model, sigmoid and softmax functions are preferred. Since convolution is a linear operation and images are far from linear, non-linearity layers are often placed directly after the convolutional layer to introduce non-linearity to the activation map.

There are several types of non-linear operations, the popular ones being:

1. Sigmoid

The sigmoid nonlinearity has the mathematical form $\sigma(\kappa) = 1/(1+e^{-}\kappa)$. It takes a real-valued number and "squashes" it into a range between 0 and 1.However, a very undesirable property of sigmoid is that when the activation is at either tail, the gradient becomes almost zero. If the local gradient becomes very small, then in backpropagation it will effectively "kill" the gradient.

2. Tanh

Tanh squashes a real-valued number to the range [-1, 1]. Like sigmoid, the activation saturates, but — unlike the sigmoid neurons — its output is zero centred.

3. ReLU

The Rectified Linear Unit (ReLU) has become very popular in the last few years. It computes the function $f(\kappa)=\max(0,\kappa)$. In other words, the activation is simply threshold at zero.

EXPLORATORY DATA ANALYSIS ON COVID-19 DATASET :

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypotheses. Visual exploratory data analysis (V-EDA) offers a user-friendly data visualisation model to evaluate the impact of the pandemic. It allows one to observe visual patterns of trends. We will be making use of multiple Python Libraries to visualise the data using Bar-Plots, Scatter-Plots, Counter-Plots, Pie-Charts, etc. EDA is primarily used **to see what data can reveal beyond the formal modelling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them**. It can also help determine if the statistical techniques you are considering for data analysis are appropriate. The four types of EDA are univariate non-graphical, multivariate non-graphical, univariate graphical, and multivariate graphical.

TOOLS REQUIRED FOR EXPLORATORY DATA ANALYSIS:

Some of the most common tools used to create an EDA are:

1. R: An open-source programming language and free software environment for statistical computing and graphics supported by the R foundation for statistical computing. The R language is widely used among statisticians in developing statistical observations and data analysis.
2. Python: An interpreted, object-oriented programming language with dynamic semantics. Its high level, built-in data structures, combined with dynamic binding, make it very attractive for rapid application development, also as to be used as a scripting or glue language to attach existing components together. Python and EDA are often used together to spot missing values in the data set, which is vital so you'll decide the way to handle missing values for machine learning.

Apart from these functions described above, EDA can also:

Perform k-means clustering: Perform k-means clustering: it's an unsupervised learning algorithm where the info points are assigned to clusters, also referred to as k-groups, k-means clustering is usually utilised in market segmentation, image compression, and pattern recognition

EDA is often used in predictive models like linear regression, where it's wont to predict outcomes.

It is also utilised in univariate, bivariate, and multivariate visualisation for summary statistics, establishing relationships between each variable, and understanding how different fields within the data interact with one another.

STEPS UNDER EXPLORATORY DATA ANALYSIS IN PYTHON

1. Description of data using functions such as **describe() , info().**
2. Handling the missing data using the functions **drop() , null() , fill()** =>
    a. Data can either be missing during data extraction or collection due to several reasons.
    b. Missing values need to be handled carefully because they reduce the quality of any of our performance matrices.
    c. It can also lead to wrong prediction or classification.
3. Handling outliers => An outlier is something which is separate or different from the crowd. Outliers can be a result of a mistake during data collection or it can be just an indication of variance in your data. Some of the methods for detecting and handling outliers => **ScatterPlots, Boxplots, IQR, etc.**
4. Understanding relationships and new insights related to the data through plots.

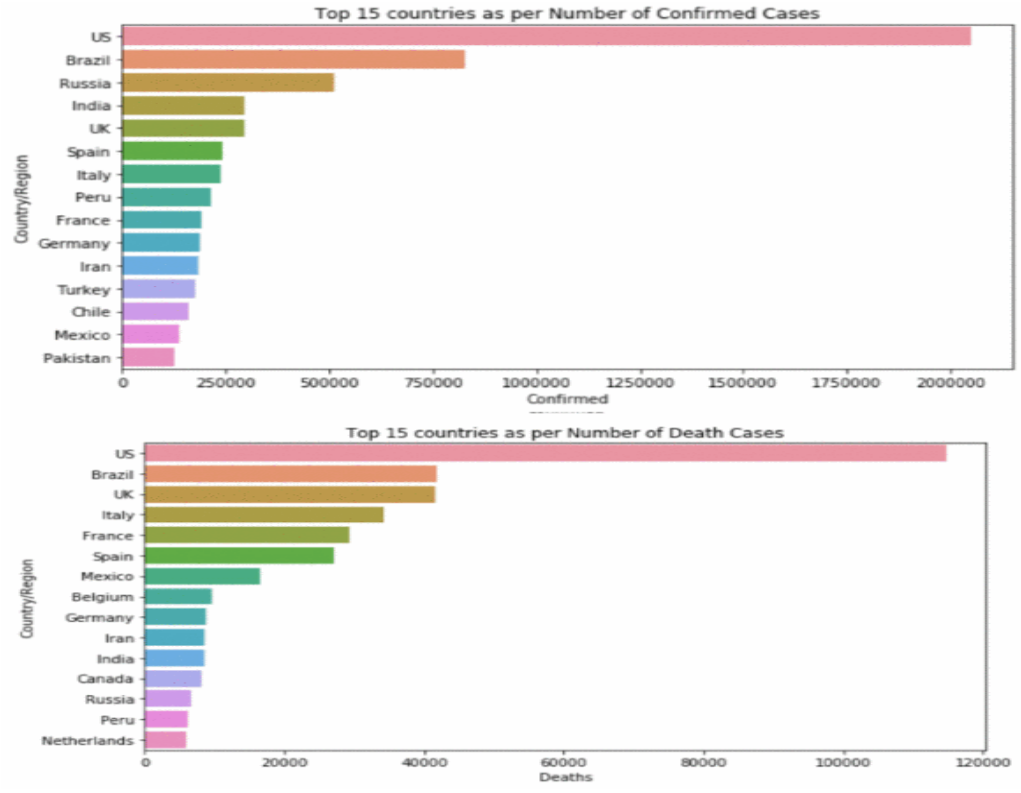

Figure 4

HARDWARE / SOFTWARE TOOLS BRIEFING FOR THE PROJECT :

HTML

The HyperText Markup Language or HTML is the standard markup language for

documents designed to be displayed in a web browser.

CSS

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML or XML.

JAVASCRIPT

JavaScript, often abbreviated JS, is a programming language that is one of the core technologies of the World Wide Web, alongside HTML and CSS.

As of 2022, 98% of websites use JavaScript on the client side for web page behaviour, often incorporating third-party libraries. All major web browsers have a dedicated JavaScript engine to execute the code on users' devices.

HEROKU

Heroku is a cloud platform as a service (PaaS) supporting several programming languages. One of the first cloud platforms, Heroku has been in development since June 2007, when it supported only the Ruby programming language, but now supports Java, Node.js, Scala, Python, Closure, etc.

GOOGLE COLAB

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

LIBRARIES USED:

NumPy

NumPy basically provides n-dimensional array objects. NumPy also provides mathematical functions which can be used in many calculations. Command to install: *pip*

install numpy.

SciPy

SciPy is a collection of scientific computing functions. It provides advanced linear algebra routines, mathematical function optimization, signal processing, special mathematical functions, and statistical distributions.

Command to install: *pip install scipy*

MATPLOTLIB

Matplotlib is a scientific plotting library usually required to visualise data. Importantly, visualisation is required to analyse the data. You can plot histograms, scatter graphs, lines etc. Matplotlib is a library in Python and it is a numerical – mathematical extension for NumPy library. Pyplot is a state-based interface to a Matplotlib module which provides a MATLAB-like interface. There are various plots which can be used in Pyplot are Line Plot, Contour, Histogram, Scatter, 3D Plot, etc.

Command to install: *pip install matplotlib*

BOOTSTRAP

Bootstrap is the most popular HTML, CSS, and JavaScript framework for developing responsive, mobile-first websites. Bootstrap is completely free to download and use! Bootstrap is a potent front-end framework used **to create modern websites and web apps**. It's open-source and free to use, yet features numerous HTML and CSS templates for UI interface elements such as buttons and forms. Bootstrap also supports JavaScript extensions.

VISUAL STUDIO CODE

Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and

embedded Git. Users can change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.

GEOPLUGIN API

geoPlugin provides a free geolocation API in multiple different programming languages in a single API call. There is no software installation required, no API key and whether your programming language of choice be Javascript, PHP, XML, JSON, ASP, or CSV,

geoPlugin has a way to simply and efficiently geo-localise your visitors.

PANDAS

Pandas is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named Numpy, which provides support for multi-dimensional arrays. As one of the most popular data wrangling packages, Pandas works well with many other data science modules inside the Python ecosystem, and is typically included in every Python distribution, from those that come with your operating system to commercial vendor distributions like ActiveState's Active Python.

KERAS

Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code. The code is hosted on GitHub, and community support forums include the GitHub issues page, and a Slack channel.

TENSORFLOW

TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. TensorFlow provides a collection of workflows

to develop and train models using Python or JavaScript, and to easily deploy in the cloud, on-prem, in the browser, or on-device no matter what language you use. The tf. data API enables you to build complex input pipelines from simple, reusable pieces.

RESULTS :

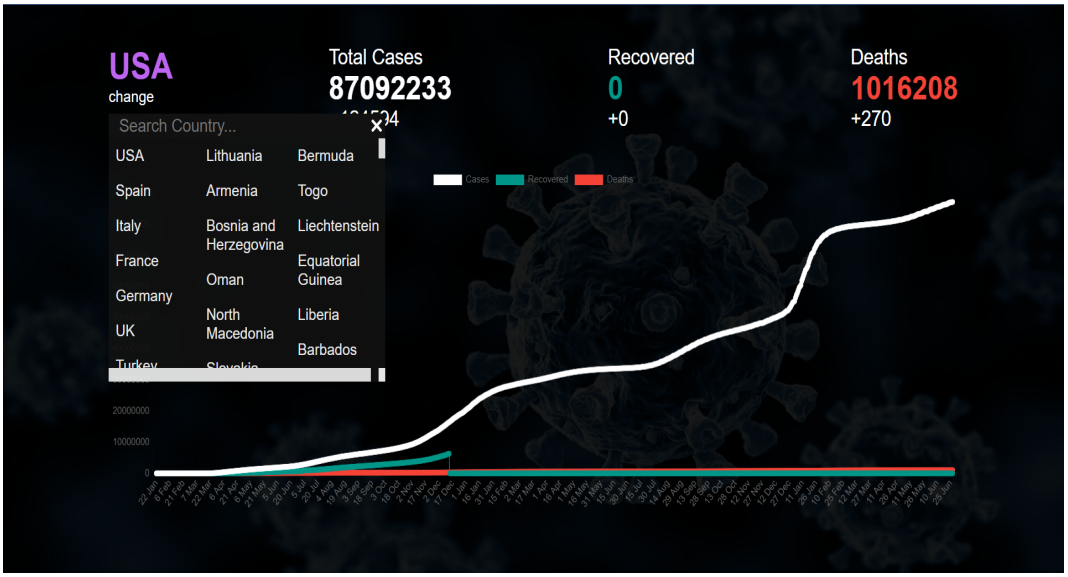

Figure 5

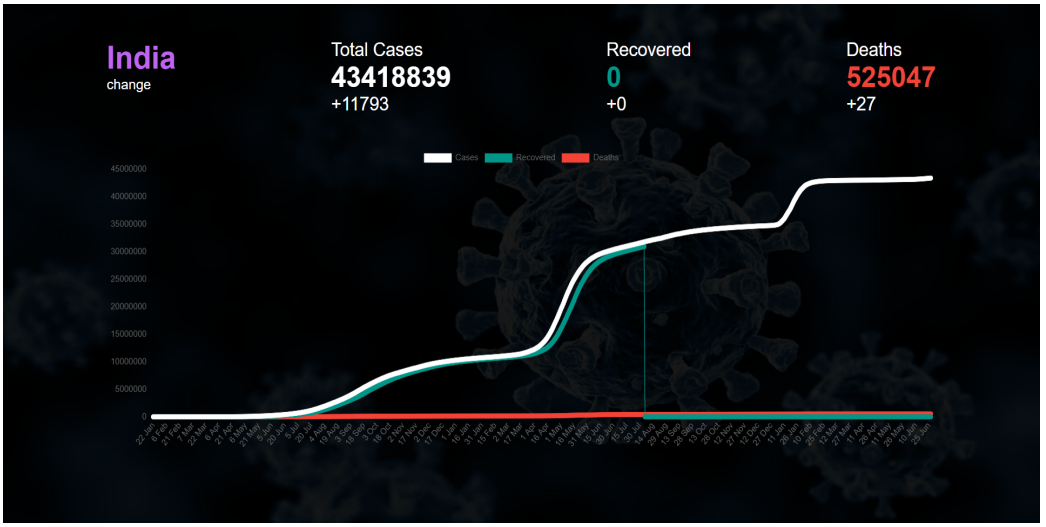

Figure 6

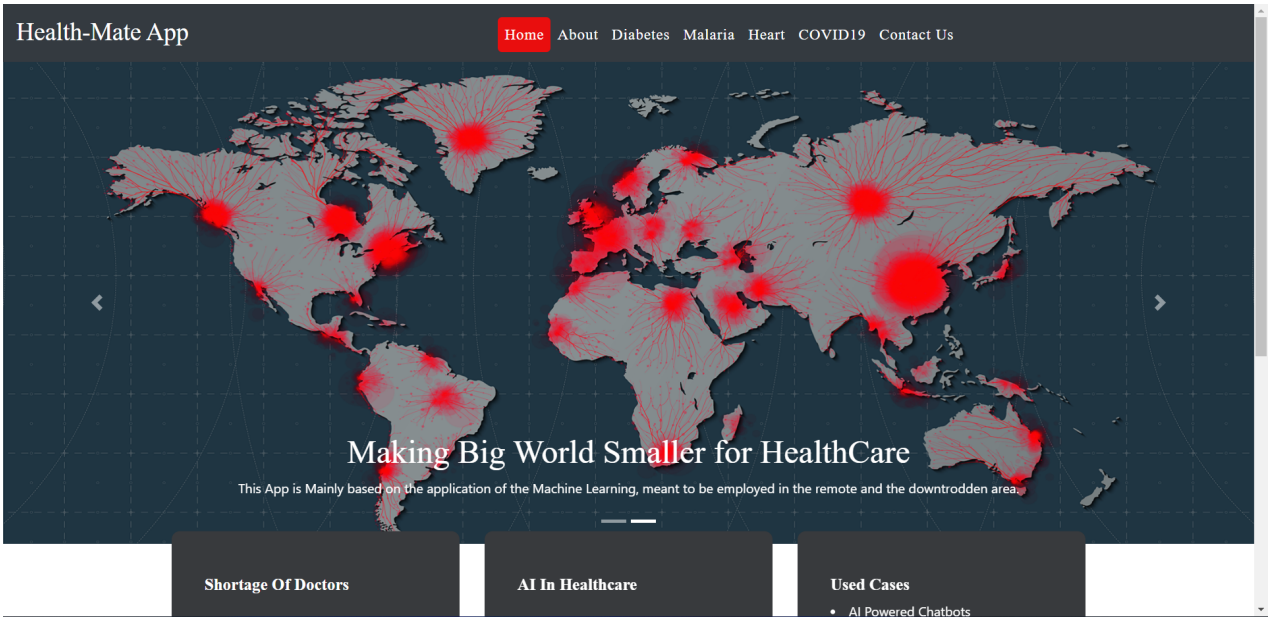

Figure 7

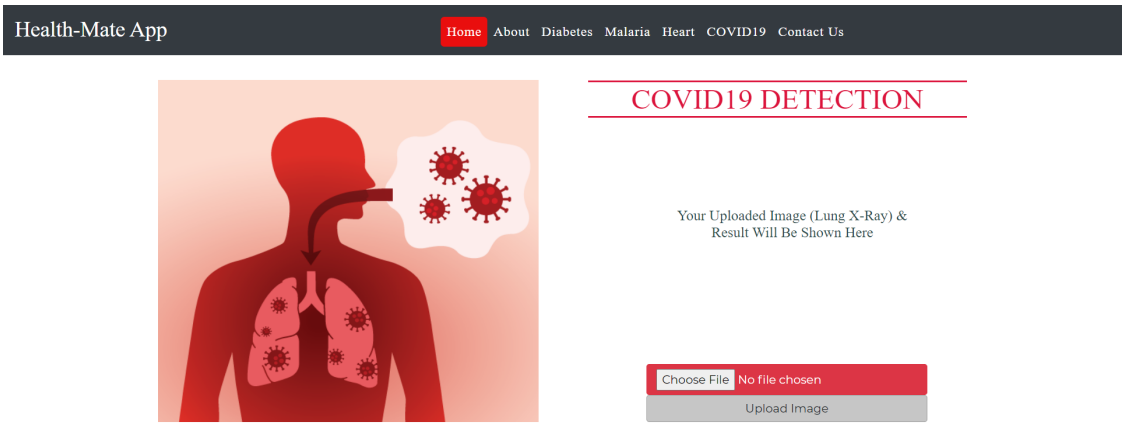

Figure 8

Screenshots of working project modules have been attached for reference and the project is deployed and made live.

APPLICATIONS AND ADVANTAGES:

An effective screening process is required to eliminate images that are not useful present in the database. Since COVID-19 is a new epidemic, fewer X-ray images are available for developing the automated diagnostic system. The main advantages of the model are as follows:-

- The model classified chest X-ray images without using a feature extraction technique.
- It is an effective approach that can assist experts for diagnosis.
- The heat maps produced by the model are evaluated by an expert radiologist. The model focuses on localising effective regions on chest X-ray images.
- The proposed model can be used for the diagnosis of COVID-19 using X-ray radiographs. X-ray radiographs are preferred because they are readily accessible for disease diagnosis.
- They are widely used in health centres worldwide during the pandemic. The model has the ability to diagnose COVID-19 within seconds.
- CT is a costly process and not readily accessible as they are usually only located in larger health centres. In addition, when CT is compared to X-ray, the amount of radiation received by the patient is more. Hence, it is recommended to use a deep learning model with X-ray imagery, as it is more accessible with lower radiation dose as compared to CT.
- Patients diagnosed as COVID positive by the model can be directed to advanced centres for confirmation, followed by treatment without delay. In addition, patients who have been diagnosed negatively by the model can be prevented from undergoing PCR tests and occupying health centres unnecessarily.

LIMITATIONS:

- The performance of the developed model has to be assessed by expert radiologists and only then, it would be ready to be tested with a larger database.
- This system can be used in remote places in countries affected by COVID-19 to overcome a shortage of radiologists.
- Also, such models can be used to diagnose other chest-related diseases including tuberculosis and pneumonia, provided the database size should be

large enough to train the model better.

- A limitation of the study is the use of a limited number of COVID-19 X-ray images. We intend to make our model more robust and accurate by using more such images from our local hospitals.

CONCLUSIONS:

- In this project, we have tried to implement a deep learning based model to detect and classify COVID-19 cases from X-ray images.
- Our model is fully automated with an end-to-end structure without the need for manual feature extraction.
- Our developed system is able to perform binary and multi-class tasks with an accuracy of 98.33% on validation dataset and upto 98.96% accuracy on training dataset in detecting **COVID-19 +ive** and **COVID-19 -ive** patients.
- Performed Interactive Visualisations using Plotly and Folium and Chloropleth Maps along with animations.
- Analysed the World-Data Trends along with dynamically updating and changing Visualisations.
- Performed TreeMap Analysis on the World COVID-19 Data.
- Performed The Comparison between the Other Past/Previously Occurred Pandemics and COVID-19.