Covid Prediction and Covid Detection via X-Ray Images

Vibhor Chauhan
Department of Computer Science & Engineering
Dr Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India.
1vibhorchauhan99@gmail.com

Prashant Johri
School of Computing Sciences and Engineering, Galgotia University, Gr Noida, India
Galgotia University, Gr Noida, India.
7prashant.johri@galgotiauniversity.edu.in

Prashant Singh
Department of Information Technology
Dr Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India.
5prashant.ert@gmail.com

Shivam Gupta
Department of Computer Science & Engineering
Dr Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India.
4shivgupta152@gmail.com

Amrutha P
Department of Computer Science & Engineering
Dr Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India.
2amrutha1999@gmail.com

ShivamGupta
Department of Computer Science & Engineering
Dr Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India.
3shivamgupta3734@gmail.com

Dr Akhilesh Das Gupta
Institute of Technology & Management, New Delhi, India.
6dhyanendra.jain@gmail.com

2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 978-1-6654-3789-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICACITE53722.2022.9823691

***Abstract* – The situation of the novel coronavirus is deteriorating day by day. There are more than 373 million cases recorded across the globe to date. The first incident of the disease was reported in China's Wuhan province in the month ofDecember 2019. The virus is called COVID-19 which is an abbreviation for Corona Virus Disease 2019 or 2019-nCoV. The covid virus belongs to the same virus family as SARS, which stands for the severe acute respiratory syndrome. To date, approx. 5.5 million people have died due to the virus. The worst affected countries are the USA, India, Brazil, France, and Turkey. Most countries face the Second Wave of the coronavirus, which is more dangerous than the previous wave. India is currently passing through a rough phase, registering more than 50,000 positive cases daily for the past few weeks. Around 1000 people are dying every day as per the official data issued by government bodies. With a population of 1.4 billion people, it is challenging to break the chains of the spread of the disease. The country is facing serious medical shortages which include Oxygen support, medicines, ICUs, and Ventilators. A sudden surge of the cases created a panic situation across the country. It is nearly impossible to calculate the actual cases and the number of deaths. The conventional Rapid Antigen tests and RT-PCR tests (i.e., real-time polymerase chain reaction) are not completely efficient and quick. The Rapid Antigen tests have just 50 -60% accuracy, and the RT-PCR test takes 24 to 48 hours to declare whether a person is Covid positive or negative. Time plays an important role during covid. The early the infection can be detected, the early that person can start his/her medications, consult a doctor, and isolate himself/herself to prevent further spreading of the virus. So, in this study, X-Ray scans are used to ascertain whether a person is covid positive or not in few seconds. A Machine Learning model is also included in the study to forecast the number of instances in the next few days.**

Keywords: Coronavirus, WHO, COVID-19, SARS, RT-PCR Test, Machine Learning, Rapid Antigen Test, Pneumonia, X-Ray images,Deep Learning, Convolutional Neural Network, Sklearn, Regression, Support Vector Machine (SVM), Django.

I. INTRODUCTION

In December 2019, the first coronavirus case was detected in the Chinese region of Wuhan. Various clusters of cases of pneumonia of unknown cause were found in a local area. In the year 2020, in March, a major outbreak was seen throughout the world as the people migrate and carried the virus with them across the globe. The covid-19 virus is a coronavirus, which is a type of virus that causes illnesses ranging from simple colds to lethal diseases such as MERS, which is an abbreviation for Middle East Respiratory Syndrome and SARS, which is an abbreviation for Severe Acute Respiratory Syndrome. The WHO i.e., World Health Organization has declared this novel coronavirus disease as a "pandemic" on 11th March 2020.

Initially, the people were diagnosed with respiratory issues like shortness of breath, breathing difficulties. Some of the people who got infected did not even show any symptoms. The remaining people who develop symptoms have mild to moderate symptoms initially, like the seasonal flu or a normal cold. The symptoms shown by the infected people are broadly divided into three groups. Exhaustion, Fever, and dry coughare the most frequent symptoms reported. Aches and pains, sore throat, headaches, deficit of taste and smell, and diarrhoea are some of the less common symptoms. Breathing difficulties,lack of speech or movement, and chest painare all serious indications [1].

India has crossed the mark of 23 million cases since the beginning of the pandemic and has a 1% death percentage. More than 494,000 people have lost their lives. With a huge population of over 1.35 billion people, it was important for a country like India to take tough decisions in the early stages. The governments applied various lockdowns and somehow controlled the disease to spread. The second wave hit the country critically. With the mutations, the virus became stronger and deadly. The reputed US Centres for Disease Control and Prevention (CDC) has designated it "airborne", meaning it may spread through the air. So, the chances of spread are even higher now. Now with the daily cases of more than 700,000 (worldwide), it is difficult to test, supply and medicate all the people.

The capacity to diagnose and isolate infected individuals early enough and place them under special care to break the chain is critical in fighting and preventing COVID-19. When a new pathogen or a virus enters a body of a person whose immunity is low, it can cause negative effects. As soon as the immunity builds up in a sample population, the relationship with the pathogen changes. With limited testing kits, testing every patient with a respiratory sickness or other respiratory disorders using traditional RT-PCR techniques is nearly difficult. The tests are also time-consuming, have little sensitivity, and are inefficient. While test results are awaiting, detecting suspected COVID-positive infections on Chest X-Rays will assist frontline workers and officials in quarantining high-risk patients. There isn't any wastage of transportation time or transportation cost for the samples because X-Ray devices are now available in most healthcare systems, and most current X-ray machines have already been digitized. In this work, we recommend that a chest X-ray can be used to diagnose those who are most likely infected and to make a straightforward selection of patients for future RT-PCR testing[2]. This approach can also be followed in an inpatient situation, where current systems are unsure whether to hold a patient either in a general ward with other patients or isolate them in special COVID-19 areas. It will also aid in determining patients with a high risk of COVID and a false negative RT-PCR result who more likely would need to be tested again [3].

A Machine Learning model will be included in the project to forecast the number of instances in the future, allowing for improved planning and attentiveness. The datasets will be taken from trusted sources to maximize the accuracy and efficiency of the model [4].

An infectious disease involves quick-spreading threat, which endangers the health of huge numbers of people, and lead to weakening the health care infrastructure of the country. Therefore, we aim to design a model for COVID-19 detection and covid cases predictor.

II. Data, Methods And Discussion

A. *COVID DETECTION USING CHEST X-RAY IMAGES*

The COVID-19 infection can be detected using a deep learning framework based upon chest X-ray or computed tomography (CT) images[2]. COVID-19 can be detected more quickly using image tests, which can help curb the virus's transmission. The imaging modalities of computed tomography (CT) and chest X-ray (CXR) are helpful in diagnosing COVID-19 disease[5].

Convolutional neural networks (CNN) for image recognition and classification have achieved high accuracy thanks to the large number of large-scale picture datasets accessible.

The signs of pneumonia disease and chest X-ray examinations are linked to the identification of COVID-19 infection. The initial imaging technique used in the diagnosis of COVID-19 disease, determining whether a person is positive or negative, is the chest X-ray.

III. Materials And Methods

Dataset: We used a dataset that is publicly available on GitHub and Kaggle for this investigation.The chest X-ray scans of Covid-19 positive patience's are taken from the Kaggle [6] and GitHub[7]. These repositories consisting of various X-raysandcomputed tomography (CT) images of around 255 COVID-19 Positive patients. In addition to that, 255 X-ray/ computed tomography (CT) images of 255 Normal Patients are taken from the Kaggle [6].

This data set is then split into two parts i.e., training and testing dataset. Testing datasetconsists of 10% of the original dataset, the rest 90% of the dataset is utilized as the training data. All the images were reshaped to 224x224 pixels in the dataset [8].

Classes / Dataset	Training		Testing	
	Covid-19	Normal	Covid-19	Normal
Dataset-1	230		25	
Dataset-2		230		25

Fig.1. Images divided as pr class of each dataset

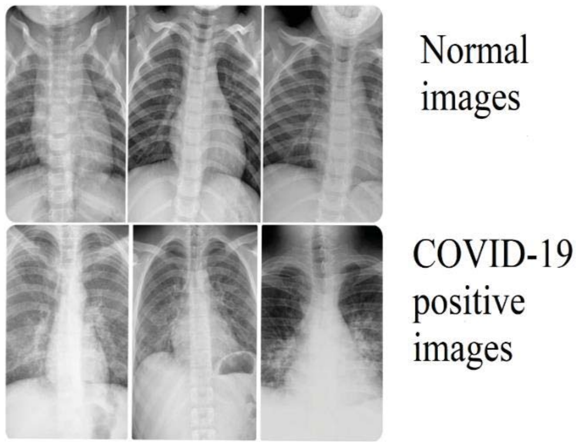

Fig.2. Pictorial representation of chest X-ray images of patients with normal(healthy) and COVID-19.

B. *The Formulation of Model*

The Dataset with the patient's information is taken from the Kaggle and GitHub repository is raw data and refined as per the need. To work on deep learning technologies, we require a huge number of datasets. But in the field of medical-related issues it is not possible to have a dataset (images in our case) are available for every possible case.

For a better analysis of the available dataset, a deep learning model is constructed using CNN (Convolutional Neural network) [9],[10],[11]. In the given model we have implemented ReLU activation so that during the back-propagation process, not all the neuron's weight and biases get updated. We used the Sigmoid function after ReLU in the last 2 layers. This function now converts the input into a value between 0.0 and 1.0, yielding a binary result of 0 which represents Covid-19 positive and 1 which represents either Covid-19 negative or Normal[4][14].

The given deep learning model is a newly constructed model using various convolutional layers followed by the max-pooling layers and dropout layers that is used to randomly drop a few parameters to avoid the case of overfitting of the Dataset. After several sequential layers, two Dense layers are added, one with the ReLU activation and another followed by the Sigmoid activation function.

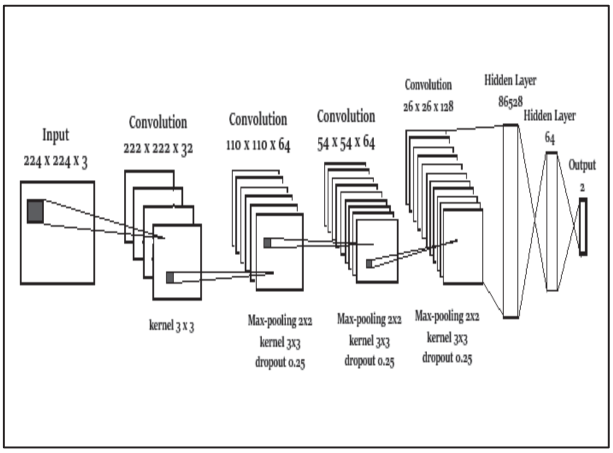

Fig.3. Model image

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
conv2d_1 (Conv2D)	(None, 220, 220, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 110, 110, 64)	0
dropout (Dropout)	(None, 110, 110, 64)	0
conv2d_2 (Conv2D)	(None, 108, 108, 64)	36928
max_pooling2d_1 (MaxPooling2	(None, 54, 54, 64)	0
dropout_1 (Dropout)	(None, 54, 54, 64)	0
conv2d_3 (Conv2D)	(None, 52, 52, 128)	73856
max_pooling2d_2 (MaxPooling2	(None, 26, 26, 128)	0
dropout_2 (Dropout)	(None, 26, 26, 128)	0
flatten (Flatten)	(None, 86528)	0
dense (Dense)	(None, 64)	5537856
dropout_3 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65

Total params: 5,668,097
Trainable params: 5,668,097
Non-trainable params: 0

Fig.4. Description of the Model

C. *The loss function used: Binary cross-entropy*

To train our model, we employed the Binary cross-entropy loss. It's utilized to determine the best parameter value for the model. The goal is to reduce the loss function with each epoch [12],[13]. For training the model, we utilize the learning rate of 0.001 of the Adam optimizers.

Loss=−sizeoutput1i=1∑sizeoutputyi·logy^i +(1−yi).log(1−y^i)

D. *Proposed algorithm*

The way we utilized to implement the proposed model is detailed further below. [9],[10]: -

Step 1: Image pre-processing i.e., picture= X,

Using the Keraspreprocessing data generator for preprocessing the image:

1) Reshape image (X) to (224,224)
2) Rescaling image(X) by 1. /255
3) Horizontal Flip = True
4) Zoom Range = 0.2
5) Shear range = 0.2

Step 2: Apply the image to one of the inputs of the Constructed model.

Step 3: Obtain the result of the supplied model's previous convolution layer.

Step 4: Reducesize of the dimension.

Step 5: Apply a dense layer.

No. of Units = 64 for the model.

$$Z = W * A + b$$

where, W represents the Weights and b represents the bias.

Activation function = ReLU

$$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$$

Step 6: Put the dense Layer for inference

Z = W ∗ A + b

where, W represents the Weights and b represents bias.

Activation function = Sigmoid

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

```
8/8 [==============================] - 6s 695ms/step - loss: 0.4031 - accuracy: 0.8658
Epoch 7/20
8/8 [==============================] - 6s 775ms/step - loss: 0.3793 - accuracy: 0.8438
Epoch 8/20
8/8 [==============================] - 6s 689ms/step - loss: 0.4066 - accuracy: 0.8398
Epoch 9/20
8/8 [==============================] - 6s 758ms/step - loss: 0.2871 - accuracy: 0.9023
Epoch 10/20
8/8 [==============================] - 6s 787ms/step - loss: 0.2952 - accuracy: 0.8828
Epoch 11/20
8/8 [==============================] - 6s 759ms/step - loss: 0.2098 - accuracy: 0.9414
Epoch 12/20
8/8 [==============================] - 6s 767ms/step - loss: 0.1788 - accuracy: 0.9453
Epoch 13/20
8/8 [==============================] - 6s 762ms/step - loss: 0.2204 - accuracy: 0.9297
Epoch 14/20
8/8 [==============================] - 6s 700ms/step - loss: 0.1882 - accuracy: 0.9264
Epoch 15/20
8/8 [==============================] - 6s 704ms/step - loss: 0.1703 - accuracy: 0.9351
Epoch 16/20
8/8 [==============================] - 6s 781ms/step - loss: 0.1428 - accuracy: 0.9648
Epoch 17/20
8/8 [==============================] - 6s 746ms/step - loss: 0.1931 - accuracy: 0.9336
Epoch 18/20
8/8 [==============================] - 5s 683ms/step - loss: 0.1678 - accuracy: 0.9524
Epoch 19/20
8/8 [==============================] - 6s 764ms/step - loss: 0.1629 - accuracy: 0.9453
Epoch 20/20
8/8 [==============================] - 6s 760ms/step - loss: 0.1581 - accuracy: 0.9437
```

Fig.5. The model's loss and accuracy during training.

E. *Experimental Setup*

Python, an object-oriented programming language, was used to train the suggested deep transfer learning models.All the tests were done on a Google Collaboratory (Colab) Linux server running Windows 10 Home for free and using a central processor unit (CPU) and a Tesla K80 graphics processing unit with an online cloud service (GPU).In this model, pictures from a normal chest X-ray/CT are compared to X-Ray/CT scans from the affected person.The model's accuracy is fairly good, but owing to a lack of data, we trained it with 510 photos and recommend that it be trained and tested on the updated Dataset in the future.

In this experiment, we used the GPU to train the model for 20 epochs with a batch size of 32(Graphical Processing Unit) on the Goggle Collaboratory. Figure 5 depicts the trained model's final loss and accuracy.

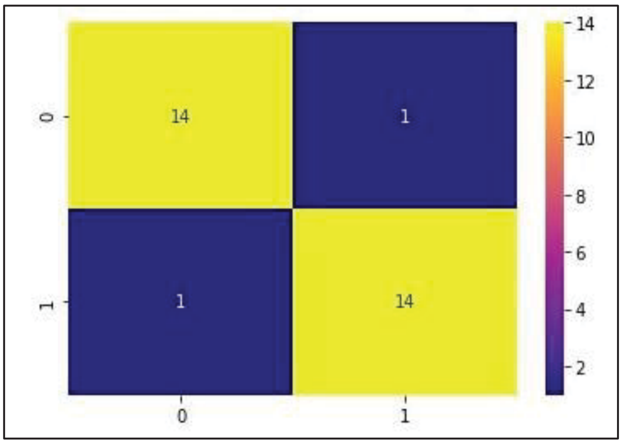

Fig.6. Confusion Matrix

IV. COVID PREDICTOR, DATA, METHODS AND DISCUSSION

A. Estimation of the Number of Novel Coronavirus Cases in 2019 (COVID-19)

The number of covid-19 instances is predicted using a date-time prediction model. The sklearn library, polynomial regression, and SVM are used to train the dataset (Support Vector Machine). The mean squared error and mean absolute error are also used to calculate the error and determine the loss[15][16].

V. MATERIALS AND METHODS

Dataset: In this study, we have taken the dataset, which is publicly available on GitHub, Kaggle and covid19india website. The data set has the data of world cases that are categorised according to Continent, Country, and States. Dataset has the cases according to the date on which they are reported. Data set contains various parameters like longitude, latitude, number of cases reported, deaths, recovered, etc.

Dataset has noise and various NAN or blank values that are discarded to make the dataset smooth. All the unnecessary columns are dropped from the table so to increase the accuracy of the model.

	Date	State	TotalSamples	Negative	Positive
0	2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
1	2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2	2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3	2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4	2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0
...	...	...	...	...	...
14093	2021-06-03	West Bengal	12645747.0	NaN	NaN
14094	2021-06-04	West Bengal	12716953.0	NaN	NaN
14095	2021-06-05	West Bengal	12789625.0	NaN	NaN
14096	2021-06-06	West Bengal	12859678.0	NaN	NaN
14097	2021-06-07	West Bengal	12919787.0	NaN	NaN

[14098 rows x 5 columns]

Fig.7. Table of Datasets

Some figures of the daily trend of the various states in India.

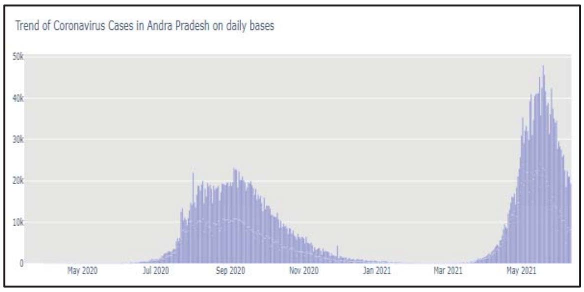

Fig.8. Daily Trends of cases in various India states.

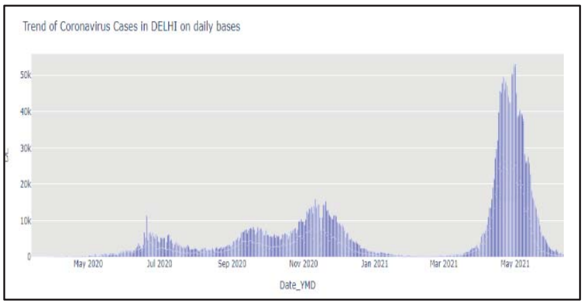

Fig.9. Daily Trends of cases in variousIndian states.

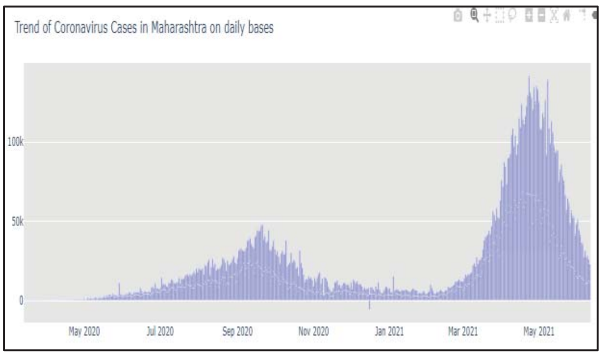

Fig.10. Daily Trends of cases in various Indian states.

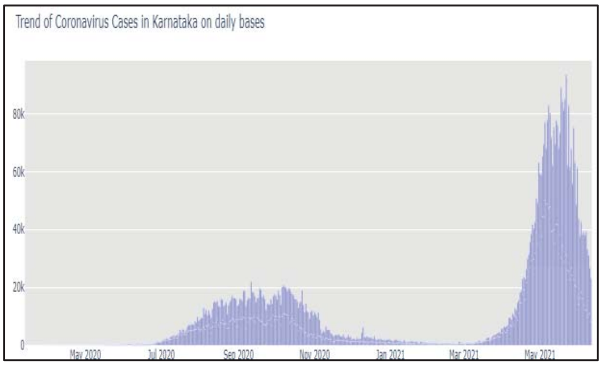

Fig.11. Figure 11. Daily Trends of cases in various Indian state.

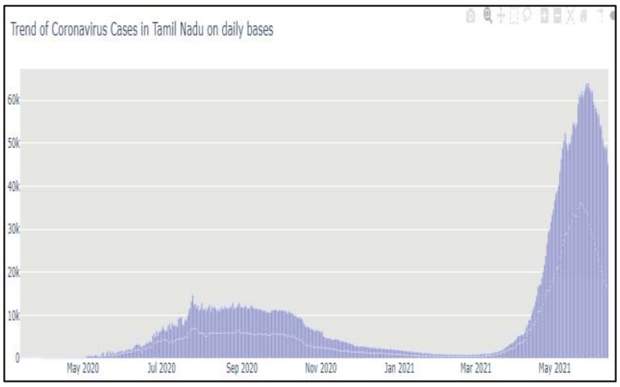

Fig.12. Figure 12. Daily Trends of cases in various Indian states.

A. Mean squared error and mean absolute error were the Loss functions employed.

To train our model, we employed the mean squared error and mean absolute error. It's utilized to determine the best parameter value for the model. Our goal is to reduce the loss function with each epoch[13],[14].

B. Mean Squared Error (MSE)

Let's say we want to estimate the value of an unobserved random variable XX based on the fact thatY=yY=y. Our estimate x^x^ is a function of yy in general:

$$x\wedge = g(y).x\wedge = g(y).$$

The equation determines the error in our estimation,

$$X\sim = X - x\wedge = X - g(y).X\sim = X - x\wedge = X - g(y).$$

And often the value of mean squared error (MSE) is determined by the equation.

$$E[(X - x\wedge)2|Y = y] \; = \; E[(X - g(y))2|Y = y].$$

C. Mean Absolute Error (MAE)

It's defined as the difference in errors between matched observations with the same trend. Comparisons of predicted against observed, starting time versus subsequent time, and one method of measurement versus another way of measurement are all examples of Y versus X. The Mean Absolute Error (MAE) is determined as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}.$$

Analysis: In Figure 13, the blue line depicts the total number of actual cases in Delhi, while the red points depict the total number of cases expected in the next seven days.

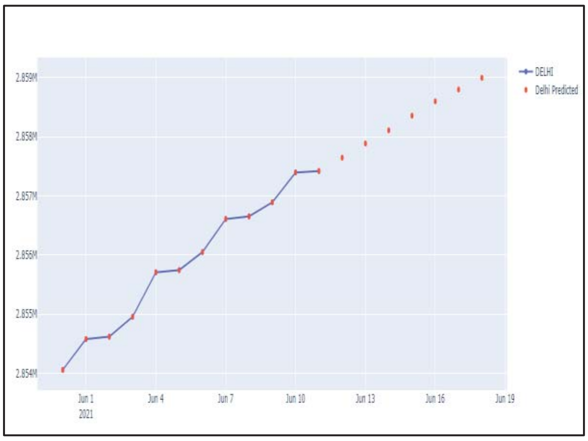

Fig.13. Actual cases versus predicted cases

VI. CONCLUSION

In this research, we have presented the use of Machine Learning and Artificial Intelligence for the effective prediction detection and classification of COVID 19. The proposed methodology is trained on publicly available datasets on Kaggle and GitHub.

The algorithm used in the research to "detect whether the person in Covid positive or not", is CNN (Convolutional Neural Network). The accuracy comes out to be around 93%.

The algorithm used in the research to predict the possible number of cases in upcoming days, is sklearn library, polynomial Regression, SVM (Support Vector Machine). The accuracy comes out to be 95%.

This brings down the time for testing the infection significantly. To make the model more effective and to makea clinically effective prediction of COVID 19, training with more datasets will be truly useful. In the future, we will more closely look for various enhancements associated with the model. One of the main achievements of Machine learning is to predict and help human beings. Using the model, many serious illnesses can be cured, and deaths can be prevented on time which is important for humanity.

A fully functional website relates to the COVID predictor and COVID detector. Django, a python-based web development platform, was utilized to create the website.The website will provide all the basic and necessary information related to the COVID 19 pandemic, a dedicated webpage where the predictions can be seen and a webpage where a user can upload his scans and get the result in a matter of seconds.

VII. REFERENCES

[1] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China". Lancet 395(10223):497–506 2020.

[2] Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song, "Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a

pictorial review," European Radiology, vol. 30, no. 8, pp. 4381–4389, 2020.

[3] N. Chen et al., "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," The Lancet, vol. 395, no. 10223, pp. 507-513, 2020.

[4] D. Wang et al., "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," JAMA, vol. 323, no. 11, pp. 1061-1069, 2020.

[5] S. Salehi, A. Abedi, S. Balakrishnan, and A. Mohd, "Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients," American Journal of Roentgenology, pp. 1-7, 2020.

[6] www.kaggle.com/praveengovi/coronahack-chest-xraydataset (Kaggle data)

[7] www.github.com/ieee8023/covid-chestxray-dataset (GitHub)

[8] Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, and Rodrigues JJ,"Identifying pneumonia in chest X-rays: A deep learning approach. Measurement", 145:511-518, 2019.

[9] Das NN, Kumar N, Kaur M, Kumar V, and Singh D.,"Automated Deep Transfer Learning Based Approach for Detection of COVID-19 Infection in Chest X-rays". IRBM, https://doi.org/10.1016/j.irbm.2020.07.001, 2020.

[10] Alqudah AM, Qazan S, Alqudah A,"Automated Systems for Detection of COVID-19 using chest X-ray images and lightweight convolutional neural networks", 2020.

[11] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim and U. Rajendra Acharya, "Automated detection of COVID 19 cases using deep neural neworks with X-ray images", 2020.

[12] MohdZulfaezal Che Azemin, Radhiana Hassan, MohdIzzuinTamrin and Mohd Adli Md Ali.,"Covid 19 Deep learning prediction model using publicly available Radiologist", 2020.

[13] Rachna Jain, Meenu Gupta, Soham Taneja & Soham Taneja & D. Jude Hemanth.,"Deep learning-based detection and analysis of COVID-19 on chest X-ray images", 2020.

[14] Ali Narin, Ceren Kaya, Ziynet Pamuk., "Automatic detection of Coronavirus Disease (COVID 19) using X-ray images and deep convolutional neural networks", 2020.

[15] AlisonJenkins, Vinika Gupta, Mary Lenoir., "General Regression Neural Networks, Radial Basis Function Neural Networks, Support Vector Machines, and Feedforward Neural Networks", 2019.

[16] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion and G. Varoquaux., "Machine learning for neuroimaging with scikit-learn", 2014