Predictive Analysis of Corona Virus Using CNNs

Prateek Nayak[1], Shruti Jain[2], Shruti Katiyar[3], Shubham Agarwal[4,] Harshita Bhardwaj[5]
[1]*Faculty at Computer Science and Engineering Department, Raj Kumar Goel Institute of Technology, AKTU*
[2,3,4,5]*Student at Computer Science and Engineering Department, Raj Kumar Goel Institute of Technology, AKTU*

Abstract **- Corona Virus has plunged the world into a "crisis like no other". Covid-19 has threatened the entire world. For the health services providers, it became a challenge to check whether a person is affected by virus in the fast way. But, an early diagnosis of COVID-19 patients is indispensable to prevent the spread of the disease to others. The objective of this briefing is to develop an efficient decision support system to predict the possibility of Covid-19 virus in a patient using the techniques of Machine Learning. Machine Learning is used to discover patterns in the data, detect and analyze trends and then make predictions with the help of algorithms. It provides methods, techniques and tools that can help in solving diagnostic problems in a variety of medical domains. It offers a principled approach for developing sophisticated, automatic, and objective algorithms for biomedical data. The purpose of this study is to achieve an accurate model to predict whether the particular case is corona positive or negative through the utilization of CNN algorithm of machine learning using the Chest X-ray data of the patients.**

Index Terms **- Machine Learning, Convolutional Neural Networks (CNN), Corona Virus, Covid-19.**

I.INTRODUCTION

The COVID-19 pandemic has led to a dramatic loss of human life. As of 31 December 2020, COVID-19 had infected over 82 million people and killed more than 1.8 million worldwide. But preliminary estimates suggest the total number of global “excess deaths” directly and indirectly attributable to COVID-19 in 2020 amount to at least 3 million, 1.2 million higher than the official figures reported by countries to WHO.[6]

According to experts, to prevent the spread of the Covid-19, early detection of the disease is crucial. The most important question is why fast detection is necessary and how it can be done. 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. Currently, a formal diagnosis of COVID-19 requires a laboratory test (RT-PCR) of nose and throat samples. RT-PCR requires specialist equipment and takes at least 24 hours to produce a result. It is not completely accurate and may require a second RT-PCR or a different test to confirm diagnosis. COVID-19 is a respiratory disease. Clinicians may use chest imaging to diagnose people who have COVID-19 symptoms, while awaiting RT-PCR results or when RT-PCR results are negative, and the person has COVID-19 symptoms. [7]

The objective of this study is to analyze the Chest X-Ray images and build an effective machine learning model using CNN algorithm for early detection of Corona Virus in a patient.

II. MACHINE LEARNING

Machine Learning is the one of the most emerging area in the field of technology. Machine Learning is an area of Artificial Intelligence (AI), which provides machines the ability to learn automatically from their own and improve from their experiences without being explicitly programmed. It includes learning the theory automatically from the data, through a process of inference, model fitting, or learning from examples. It automates extraction of useful information from a body of data by building good probabilistic models and ideally suited for areas with lots of data in the absence of a general theory.[9] A formal definition of machine learning is given by Mitchel: A computer program is said to learn from experience E with respect to some class of tasks T and performance

measure P, if its performance at tasks in T, as measured by P, improves with experience E. [8] There are three types of Machine Learning Algorithms: Supervised, Unsupervised and Reinforcement. [9]

III. ALGORITHM USED – CNN

A convolutional neural network (CNN) is a specific type of artificial neural network that uses perceptrons, a machine learning unit algorithm, for supervised learning, to analyze data. CNNs apply to image processing, natural language processing and other kinds of cognitive tasks. A convolutional neural network is also known as a ConvNet. Using this technique, a (CNN) architecture is created and trained to process the chest X-ray image of the patient to test whether a person is covid positive or not.[12]

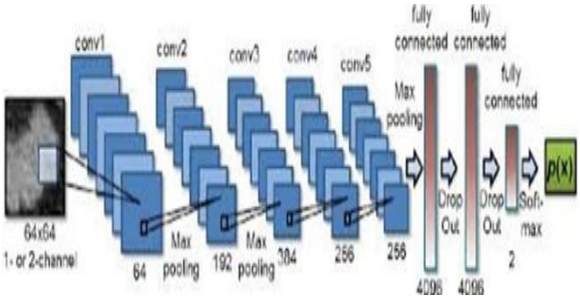


Fig. 1 Structure of CNNs

Different Layers in the structure of CNN are:

A. Input Layer

The input layer is responsible for holding the raw input of the image with width 200, height 200 according to our dataset. [2]

B. Convolutional Layer

This layer calculates the output volume by calculating the dot product between all the filters and image patch. It is the layer consists of weights which are needed to be trained. [2]

C. Activation Function Layer

This layer applies the element wise activation function to the result of convolution layer. It is the function which decides whether to fire a neuron or not. Some examples of activation functions include:

- RELU: max (0, x) and leaky RELU.
- Sigmoid:1/(1+e^-x)
- Tanh: (1-e^-2x)/ (1+ e^-2x) [2]

D. Pool Layer

The main function of Pool Layer is to reduce the size of volume which makes calculations fast and reducing the memory and also prevents from over fitting. It is used to reduce the number of parameters. This layer is periodically inserted in the convolution layer. Two common types of pooling layers are max pooling and average pooling. [2]

E. Fully Connected Layer

Fully Connected Layer: This layer is regular neural network layer which takes input from the preceding layer and calculates the class scores and gives the 1-D array of size equal to the number of classes as output. [2]

IV. METHODOLOGY USED

In this project, Convolutional neural network (CNNs) has been used to classify the Covid-19 positive and negative cases using X-Ray Chest dataset. The input image size is 200*200*1 in gray scale.

First layer has 256 filters, so the kernel size= (3,3,256). Then, Max pooling has been applied with stride=2 to reduce the size of the layer, thus making the output layer of 100*100*256 with the “relu” activation function.

Second layer has 64 filters, so the kernel size= (3,3,64). Then, Max pooling has been applied with stride=2 to reduce the size of the layer, thus making the output layer of 50*50*64, with the “relu” activation function.

Our third layer has 16 filters, so the kernel size= (3,3,16). Then, Max pooling has been applied with stride=2 to reduce the size of the layer, thus making the output layer of 25*25*16 with the “relu” activation function.

After feature engineering has been done, the next step includes flattening of the output layer and forming the column vector 7 to make the first dense layer(D1) having input size of 10000 and output size of 64. The second dense layer(D2) will take the output generated by D1 as input and form the output layer with 2 neurons associated with sigmoid activation function.

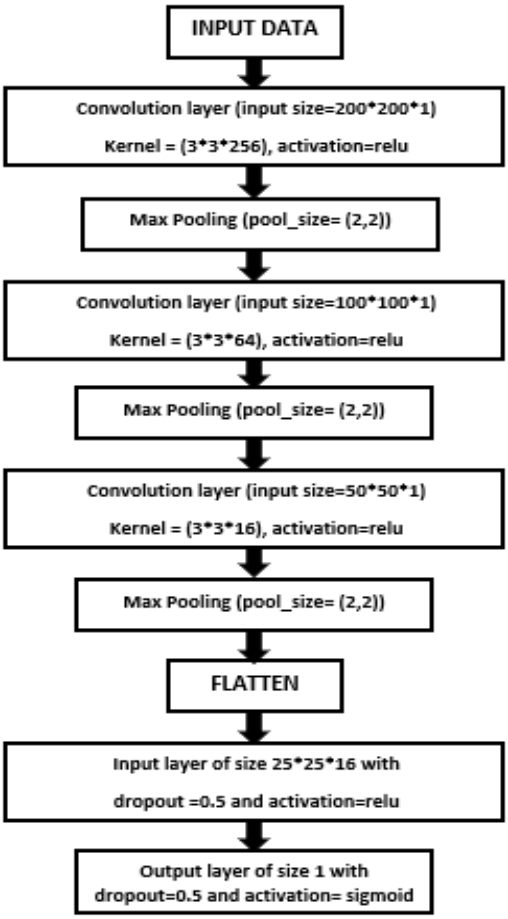


Fig. 2 Flowchart of the Model

V. IMPLEMENTATION

The implementation phase has various steps of machine learning and the flow of implementation is shown in following figure.

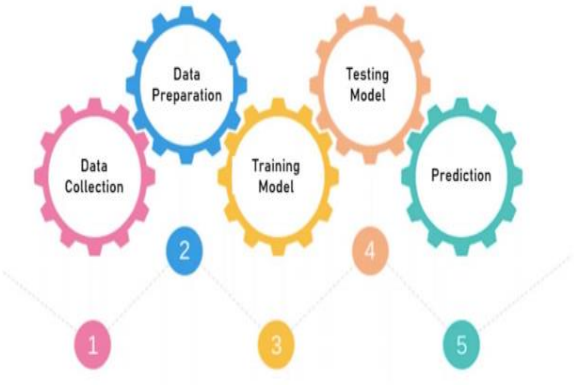
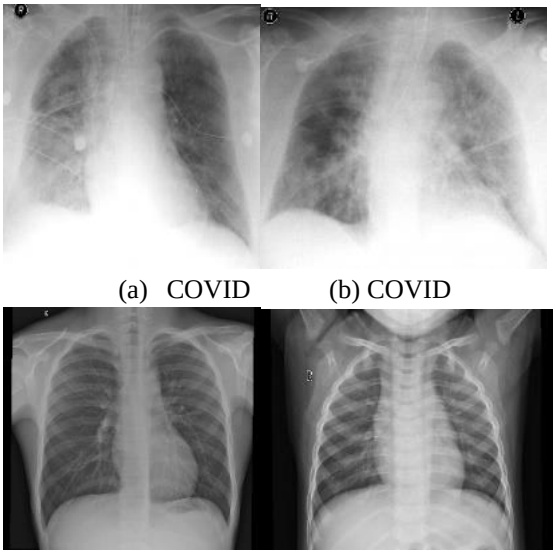


Fig. 3 Phases of Implementaion

A. Data Collection

Data Collection is the first and most crucial step in the implementation of machine learning project. Better the dataset, better the results. The dataset used in this project has been taken from the Kaggle Community.



(a) COVID (b) COVID

(c) NORMAL (d) NORMAL

Fig. 4 Snapshot of the Dataset

B. Data Preparation

Data preparation refers to cleaning the data and adding appropriate missing data. This step is also called Data Pre- processing [2]. Data preparation is the process of readying data for the training, testing, and implementation of an algorithms [11]. As this paper deals with medical data, much changes have not been made in order to preserve the patient's data.

As the dataset contain the images having different height and width, so we resize the dimensions of all the images to 200*200*3, to fit the images in our model having 200*200 neurons as input layer

C. Training Model

Training is basically the process of giving the machine capability to make further predictions after learning from the training dataset [2]. The datasets are trained using Convolutional neural network with input layer of shape (200,200,3), three Conv layers, three pool layers having the 'relu' activation function on both the pool layer and the output layer having with the softmax function.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 200, 200, 256)     2560
_________________________________________________________________
activation (Activation)      (None, 200, 200, 256)     0
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 100, 100, 256)     0
_________________________________________________________________
batch_normalization (BatchNo (None, 100, 100, 256)     400
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 100, 100, 64)      147520
_________________________________________________________________
activation_1 (Activation)    (None, 100, 100, 64)      0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 50, 50, 64)        0
_________________________________________________________________
batch_normalization_1 (Batch (None, 50, 50, 64)        200
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 50, 50, 16)        9232
_________________________________________________________________
activation_2 (Activation)    (None, 50, 50, 16)        0
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 25, 25, 16)        0
_________________________________________________________________
batch_normalization_2 (Batch (None, 25, 25, 16)        100
_________________________________________________________________
flatten (Flatten)            (None, 10000)             0
_________________________________________________________________
dropout (Dropout)            (None, 10000)             0
_________________________________________________________________
dense (Dense)                (None, 64)                640064
_________________________________________________________________
activation_3 (Activation)    (None, 64)                0
_________________________________________________________________
dropout_1 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
_________________________________________________________________
activation_4 (Activation)    (None, 1)                 0
=================================================================
Total params: 800,141
Trainable params: 799,791
Non-trainable params: 350
_________________________________________________________________
```

Fig 5. CNN Model

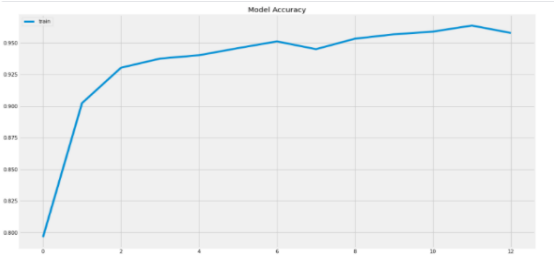


Fig. 6 Model Accuracy Graph

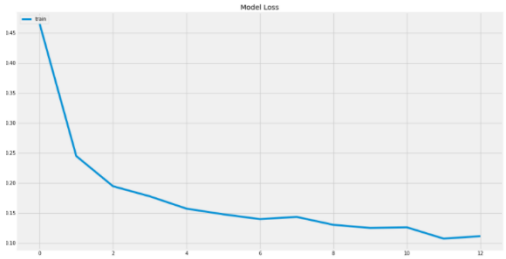


Fig. 7 Model Loss Graph

D. Testing Model

Testing of a model is done to check the performance of the algorithms in term of accuracy, precision etc. In testing whether the prediction is correct or not is checked using already predefined dataset. We tested the dataset on the validation dataset our original dataset.[2]

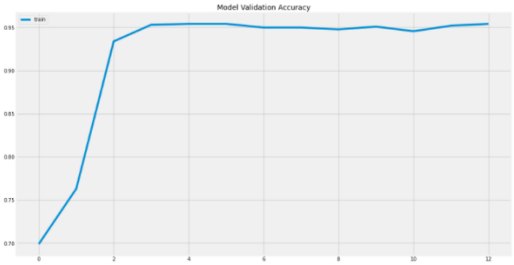


Fig. 8 Model Validation Accuracy Graph

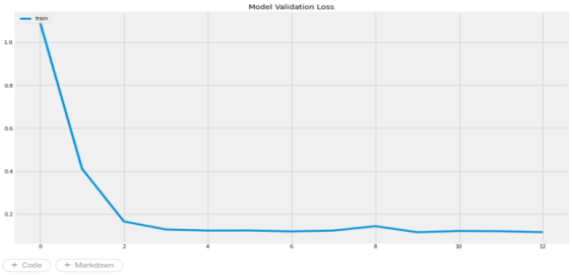


Fig. 9 Model Validation Loss Graph

E. Prediction

Prediction refers to the output of a model after it has been trained on a predefined dataset and applied to new data when forecasting the likelihood of an outcome. The model will generate values for an unknown variable for each record in the new data, allowing the model to identify what that value will most likely be [2]. Our model has the accuracy of 96.24%.

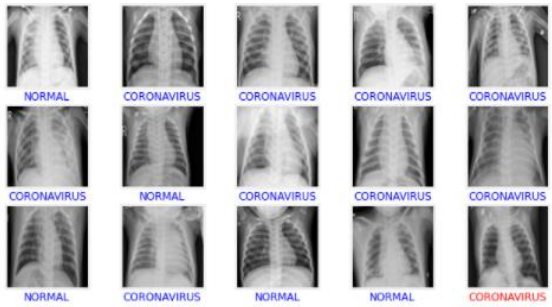


Fig. 10 Predicted Results

```
37/37 [==============================] - 1s 38ms/step - loss: 0.1032 - acc: 0.9625
[0.10320743173360825, 0.9624573588371277]
```

Fig. 11 Model Accuracy

VI. CONCLUSION

Machine learning has emerged as a field critical for providing tools and methodologies for analyzing the data generated by the biomedical sciences. This review is an extensive analysis that shows that the proposed model can classify the chest CT images as covid-positive or negative at a good accuracy rate. Fusion of disparate multimodal and multi-scale biomedical data continues to be a challenge. Further improvements in data can be made like having more features and least null values.

All the ideas are just the beginning, the applications of machine learning are beyond the boundaries of imagination, which are practically limitless. A potential future development of the presented work is

to apply ML models to other data with different features, concerning the survival prognosis of the patients and early detection of the disease and it can also be developed in web-based application with additional service.

REFERENCES

[1] A Comparative Study of Machine Learning Classifiers for Medical Diagnosis, Bhavnath Thakur, Harshit Rohela, Kanishk Gupta, Chhaya Sharma. "A Comparative Study of Machine Learning Classifiers for Medical Diagnosis", Volume 8, Issue IV, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 1748-1752, ISSN: 2321-9653.

[2] Prateek Nayak, Sarthak Garg, Rajat Shukla International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653 Volume 8 Issue VIII Aug 2020.

[3] E. Irmak, "A Novel Deep Convolutional Neural Network Model for COVID-19 Disease Detection," 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299286.

[4] Maior CBS, Santana JMM, Lins ID, Moura MJC (2021) Convolutional neural network model based on radiological images to support COVID-19 diagnosis: Evaluating database biases. PLoS ONE 16(3): e0247839.https://doi.org/10.1371/journal.pone.0247839

[5] Wang, L., Lin, Z.Q. & Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *SciRep* **10,** 19549 (2020). https://doi.org/10.1038/s41598-020-76550-z

[6] [World Health Organization] https://www.who.int/news-room/spotlight/the-impact-of-covid-19-on-global-health-goals

[7] [World Health Organization] https://www.cochrane.org/CD013639/INFECTN_how-accurate-chest-imaging-diagnosing-covid-19

[8] T. Mitchell, Machine learning, 0-07-042807-7, McGraw Hill (1997), p. 2.

[9] Ayodele, Taiwo. (2010). Machine Learning Overview. 10.5772/9374

[10] [Kaggle][Dataset] https://www.kaggle.com/khoongweihao/covid19-xray-dataset-train-test-sets

[11] [Data - Preparation] [Machine Learning] https://lionbridge.ai/articles/data-preparation-for-machine-learning-the-ultimate-resource-guide/

[12] https://www.techopedia.com/definition/32731/convolutional-neural-network-cnn