Coronavirus Detection using Deep Pre-trained Model on Chest X-ray Imaging

Aqsa Noreen*, Moazzam Jawaid, Sanam Narejo, Shahnawaz Talpur, Bisharat Ayub

Department of Computer Systems Engineering, MUET, Jamshoro, Pakistan
*Corresponding author: noreenaqsa9@gmail.com

Abstract

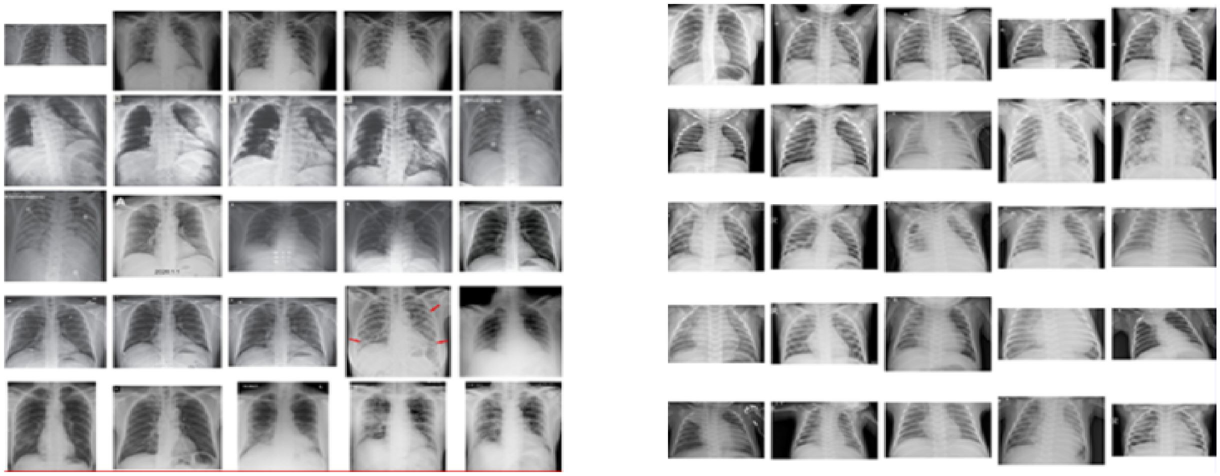
The constant increase in the number of patients suffering from the pandemic Covid-19 has overpowered the medical management system over the world. This pandemic has affected every aspect of human life all over the world. However, the limited number of testing kits is making it more challenging to deal with the global pandemic. It is near to impossible for people to maintain social distancing all the time, and lengthy testing procedure is one of the major cause of the rapid spread of the epidemic Covid-19. The polymerase chain reaction test that is done in real-time (RT-PCR) is time-consuming for each patient with gasping disease to be tested. However, the use of a chest x-ray machine tool can accelerate testing time. X-ray machines are now accessible in almost every medical management healthcare system, with instant results. The proposed research aims to detect the pandemic. The dataset was collected from an open-source repository that consists of the chest x-ray images of patients who were Covid-19 positive, and covid-19 negative which means not infected by the pandemic coronavirus but suffering from normal pneumonia. The approach of this paper uses a deep convolutional neural network model i-e VGG16, to begin the process of classifying images into positive or negative. In the proposed research VGG16, the Deep Con-volutional Network model is used, this is a pre-trained model, previously trained on the ImageNet dataset that com-prises 14M images that belong to 1K different classes. This model is fine-tuned on 1202 images. The preprocessing of collected data is done by converting it into an RGB channel and resizing it to match the requirement of our model i-e VGG16. Data augmentation is performed on the preprocessed dataset. Our proposed model gives 99.1%, 100%, and 98.59% accuracies, sensitivity specificity respectively for the detection of Covid-19 on patients' cxr images.

Keywords—Covid-19; Deep Learning; Pre-trained model; CNN and X-ray Images

1 Introduction

Coronaviruses (CoVs) are enclosed viruses with immensely large single-stranded RNA genomes ranging from 26 to 32 kilobases in length.[1] Middle East Respiratory Syndrome (MERS) Severe Acute Respiratory Syndrome (SARS) viruses are major causes of coronavirus infection. The infection produced by SARS-CoV-2 is COVID-19, a new coronavirus (Zhang, 2020). In Wuhan capital of China's Hubel province, the first COVID-19 cases were regis-tered, in December 2019. COVID-19 has been declared a pandemic by the World Health Organization (WHO) [2]. Till January 2022, 363 million cases have been documented worldwide, with 5.63 million deaths. These illnesses produce respiratory issues that can be addressed without the use of specific medication or equipment. Despite this, underlying medical conditions such as diabetes, cancer, cardiovascular disease, and lung disease might exacerbate the illness. Test kits took too much cost so many countries cannot get more test kits and have limited kits to test people. For these normal test kits, every clinic needs a technical person that can arrange and control all the systems in every clinic. Rapid diagnostic tests for Covid-19 are also a very big issue the world is facing. However, Immunoassays have known as ”antigen tests” look for a particular viral antigen, which can be a sign of active viral infection. Cur-rently, saliva, nose swab, or nasopharyngeal specimens deposited directly into the assay's extraction buffer or rea-gent may be used for antigen testing. Although much faster than antigen tests, the other common coronavirus test is less precise. It is less sensitive and can require confirmation tests (more false negative results). Due to covid, there are many global

ISSN: 2523-0379 (Online), ISSN: 1605-8607 (Print)
DOI: https://doi.org/10.52584/QRJ.2001.08



(a) Covid-19 +ve chest x-ray dataset

(b) Covid-19 -ve chest x-ray dataset

Fig. 1: Covid-19 +ve -ve chest x-ray dataset

issues, a few of them being hunger, unemployment, and vaccine. Therefore, we have trained our model using an open-source dataset, there are just a few substantial datasets of COVID-19 X-Ray images that are open to the public. The model proposed in this approach is trained to detect covid-19 in x-ray images using a Convolutional Neural Network. As a result, we can detect whether are provided X-ray image is covid positive or covid negative. Chest x-ray pictures have been utilized for the prediction of Covid-19 patients. Famous pre-trained models e.g Res-net50, Inception ResNetV2, and InceptionV3 have been prepared tried on chest x-ray pictures. Resnet50 per-formed very well among the 3 used models and gave 98% accuracy.[3] Researchers in paper [4], indicated a model which can be considered as s promising to describe analyze Covid-19 diseases is ResNet-101. This model does not force significant expense can be utilized as an adjuvant strategy during CT imaging in radiology offices.

In the paper [5], a novel technique is introduced to identify Covid-19. This technique comprises pre-processing and includes extraction with Lingering-Exemplar-Local-Binary-Pattern (ResExLBP), including determination with interactive Relief (IRF), grouping stages. In this preprocessing stage, the info X-Ray picture is changed into a gray-scale picture and is resized (512*512) measured pictures. ResExLBP isolates input pictures into (128*128) estimat-ed models and LBP removes include from the input picture and models of it. The created highlights are connected, what's more, separating ones of them are chosen by utilizing IRF. The choice included by IRF are used as the con-tribution of the classifiers by utilizing Leave-One-Out-Cross-approval (LOOCV), 10-crease Cross-Validation (CV), and holdout approval. Researchers in paper [7,8] worked on the ML classifier that was fed with features obtained from the existing Res-Net-50 CNN model. The accuracy of this deep hybrid learning-based system was reported to be 95%. Features from fine-tuned pre-trained ResNet-152 have also been retrieved and fed into machine learning classifiers. COVID-19 identification using Random-Forest and XGBoost classifiers, on the other hand, had an accuracy of 97.3 percent and 97.7%, respectively. Zang et al. in paper [10,] ResNet network model, is developed COVID-19 may be diagnosed from X-ray imaging. In their experiment, they used a collection of 1078 images including COVID-19 images and healthy patient images. According to their findings, the sensitivity, specificity, and AUC values attained were 96.0%, 70.7%, and 95%, respectively.

In the paper[11], the authors used a publicly available COVID-19 chest X-ray dataset of 70 lung X-ray pictures of patients, researcher designed a model for covid-19 prediction i-e Bayesian Convolutional Neural Network (BCNN). They developed a Bayesian-based CNN model and achieved a 92.9 percent accuracy score in their experiments. For this problem i-e lack of kits for the Covid-19 test and the delayed result, we have proposed a convolution neural network(CNN) based model named "Covid-19 Detection Using Deep Learning on Chest X-Ray Images", In which we will be using VGG16 pre-trained model for training of data that we have taken from Kaggle and Github, then binary-level classification will be done to detect either patient's chest is covid positive or negative.

2 Material and Methods

2.1 Data Gathering Preprocessing

Dataset: The dataset we analyzed to train and test our proposed approach is referred to as the detection of Covid-19 Using Deep Learning Pre-Trained Model on CXR Images, the proposed dataset is taken data from an open repository which is comprised of thousands (Chest Radiograph/Chest X-Ray) images of 13,870 patient suffering from covid-19, we have used 1202 images in total, covid-19 cxr image dataset for covid +ve patients from Github [6] cxr image dataset for covid -ve patients from Kaggle[7]as shown in Fig. 1(a) and Fig. (b).

The decision to use these two datasets to create Covid-19 was influenced because they are both freely available for researchers and the general public. More specifically, the data was sorted in the following way from each of the data repositories to build the covid-19 dataset by combining and modifying both data sources. The preprocessing of the collected dataset is done by converting its channel into RGB and resizing it to match the requirement of our model i-e VGG16. Data augmentation is performed on the preprocessed dataset to improve performance to make the model more robust as as shown in Fig. 2.

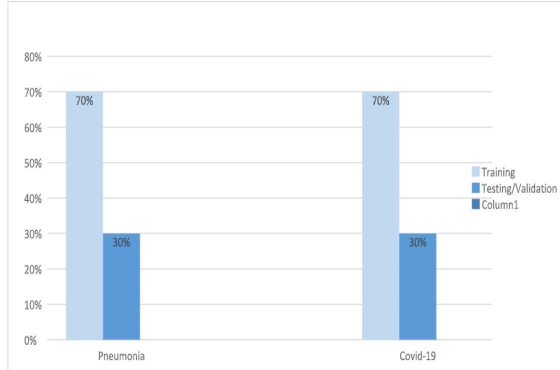


Fig. 2: shows the distribution of the images for both categories.

2.2 VGG-16 Model / Pre-trained CNN model

The importance and originality of this study are that it facilitates the medical doctors who are under continuous stress due to the increase in the number of covid patients, patient's x-ray is used to determine either its covid +ve or -ve. Convolutional Neural Networks paired with Computer Vision are capable of executing difficult operations such as picture classification, astronomical issue solving, and self-driving car development. The Convolutional Neural Network (CNN) is a combination of Convolutional Layers and Neural Networks. Any Neural Network used for image processing is made up of the layers listed below. Convolutional Layer, Pooling Layer, and Dense Layer are the layers that make up the input layer. Convolution is a filter that is applied to a picture to extract features from it. We'll use different convolutions to extract different features from the image, such as edges and highlighted patterns.

Following is the application of convolutions, there is a notion known as pooling. Pooling is a technique for reducing the size of an image. Pooling can be divided into two categories:

1. Max Pooling: It's nothing more than picking the highest value from a matrix of a certain size. This method is useful for extracting important features or features that are highlighted in the image. A feature that is highlight-ed is a section of an image with high pixel values.

2. Average Pooling: It takes the average of every single pixel value in the pooling layer's matrix. The majority of the time, maximum pooling is employed since it outperforms average pooling. The first convolutional layer, when defining a neural network, needs the shape of the image that is passed as input. The image will be delivered to the dense layer after passing from all convolutional and pooling layers. Because the output of the convolutional layer is multi-dimensional, we can't pass it directly to the dense layer because the dense layer requires input in a one-dimensional shape.

3 Experimental Setup

3.1 Tools

Python is used as the coding language. It's a high-level language of programming. It is interpreted object-oriented. It has high-level data structures Development-(RAD), thanks to its dynamic binding and dynamic binding. Python language is easy to learn straightforward and hence it decreases the amount of time spent learning it. Python package that comes preconfigured with a variety of useful modules for data analysis. The reason behind using anaconda is that it combines many tools needed for data science, data analytics, and machine learning into one single installation, making it more suitable for rapid development. Anaconda also used environments to distinguish between different libraries and versions. Anaconda's package management lets you install libraries.
Therefore, in between the convolutional and dense layers, we'll use the Flatten() method. The Flatten() technique reduces a multi-dimensional matrix to a

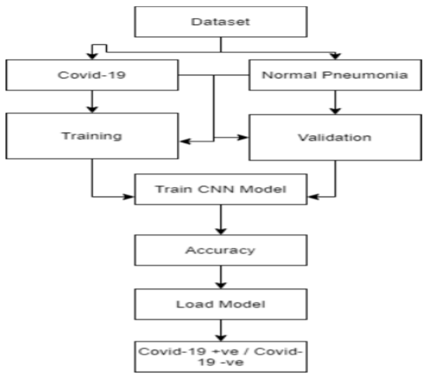


Fig. 3: shows the distribution of the images for both categories.

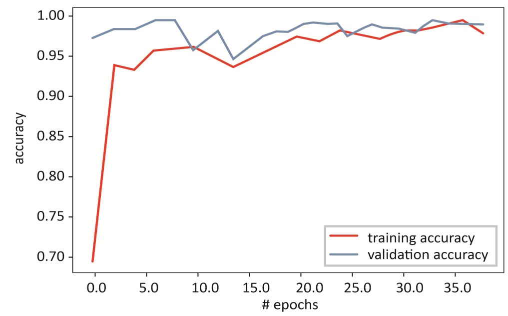


Fig. 4: Training validation accuracy.

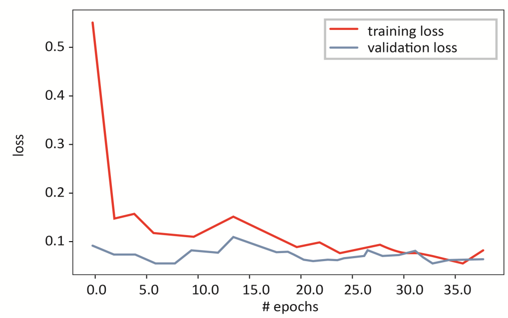


Fig. 5: Training validation loss.

single dimension. The activation function of a neural network is non-linear. The below Fig.3 represents the flow of the methodology of the project.

3.2 Performance Measure

Machine learning performances are evaluated by using a confusion matrix. It has information about real and predicted classification by classifiers. The actual performance of the model on a collection of test data for which the true values are known can be determined using the confusion matrix.

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)} \quad (1)$$

As expressed in Equation 1, precision computes the accuracy i-e how much our model is accurate out of predicted positives and how many are positive.

$$Recall = \frac{True\ Positive(TP)}{true\ Positive + false\ Negative(FN)} \quad (2)$$

Equation 2 is used to calculate recall which means how many are positive by our model by marking it as true positive.

$$F1Score = \frac{2(Precision\ +\ Recall)}{Precision\ +\ Recall} \quad (3)$$

F1score is determined using equation 3, f1score can be defined as when you want a balance between Recall and Precision. It is a function of precision recall.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Equation 4 is used to calculate Accuracy for evaluating classification models and the fraction of predictions of our model got right. Whereas,

True Positive (TP)=Number of samples classified as true while they were true.

True Negative (TN)=Number of samples classified as false while they were false.

False Positive (FN)=Number of samples classified as true while they were false.

False Negative (FP)= True Negative (TN)=Number of samples classified as false while they were true.

4 Result and Discussion

In this study, a total of 1202 chest X-ray images were investigated to detect covid-19 using a deep convolutional neural network model namely VGG16, the dataset used was taken from publicly accessible repositories [6][7]. Alt-hough VGG16 is the pre-trained model i-e it is already trained on the ImageNet Dataset but to achieve good accuracy, we need to fine-tune it on the dataset that we have chosen to work with. 7% of the dataset is used for training purposes 30% of the dataset was used for testing purposes. After that data augmentation was performed. The model can detect covid-19 from patient chest X-ray images, this can be used for rapid computer-aided screening, this can help

TABLE 1: Overall Model Performance

VGG16		
Accuracy	**Sensitivity**	**Specificity**
99.10%	100%	98.59%

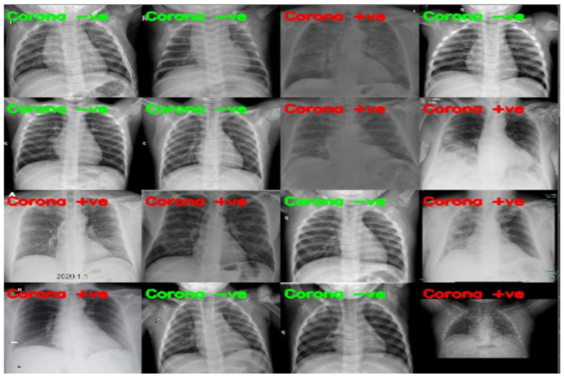


Fig. 6: result of a new x-ray image

practitioners improve screening while also increasing trust. Table 1 shows the overall model performance.

Fig. 4 and 5 shows the accuracy loss curves of deep learning training which demonstrates that our model is not overfitting despite the limited Covid-19 Chest X-Ray training data used in our model.

The Table 2 shows the classification report of the confusion matrix giving the Precision, Recall, and F1 score for our VGGNet Model. Results of the proposed model can be seen in the above table, 99% accuracy has been obtained from the used dataset based on X-Ray images. And the model obtained 100%, and 98% sensitivity specificity respectively. We loaded our model and made predictions on Chest X-Ray images to detect whether the patient is either covid positive or negative.

Fig. 6 shows that the patient is corona +ve whereas Fig. 7 shows that the patient is corona -ve. It's great to

Fig. 7: result of a new X-ray image.

TABLE 2: Classification report of the confusion matrix giving the Precision, Recall

VGGNet16	Precision	Recall	F1 Score
Covid	98%	100%	99%
Normal	100%	99%	99%

be able to diagnose Covid-19 with this much accuracy, but our true negative percentage is troubling. Medical partitioners would not like to identify a patient as Covid-19 +ve and confine them with other positive pa-tients, only to infect another person who had never been exposed to the virus.

Sensitivity specificity are exceedingly difficult to achieve in medical approaches, particularly in the treatment of contagious infections that spread quickly, such as Covid-19. We must all be mindful that the predictive models that we are working on, will have real-world implications, and that a missed diagnosis will cost lives. As a result, our findings are solely for academic purposes.

5 CONCLUSION

The present research aimed to detect covid-19 using patients' chest X-ray images. The epidemic covid 19 has be-come one of the most dangerous diseases of all time, which is constantly affecting human life in every aspect. The pandemic was announced in late December 2019 first in China. As this virus is continuously spreading around the globe and the continuously growing number of patients is not a good sign. Early predictions are important to pre-vent the spread of this virus to other people. The coronavirus leads to pneumonia which is an infection that inflames your lungs. One of the methods by which these inflames can be observed is by using an X-ray of the Chest. The proposed work is to develop a method that can help doctors instantly diagnose the virus to avoid overwhelming the healthcare system. This study introduced pneumonia Chest X-ray detection based on VGGNet16 and a fined tuned deep transfer learning using CXR images obtained from covid-19 patients and normal patients, this dataset is open source and available to the general public. This Covid-19 detector system predicts the infectious viral disease corona virus which has already infected most people around the world. From this model, early diagnosis of this disease can be predicted with a trained CNN model using Chest Xray. This system showed that the VGGNET model has achieved 99.10% accuracy and is considered promising performance as a good model to detect Covid-19 positive patients. The model does not impose any significant expense and can be used in radiology departments for early diagnosis. Through this, the test rate will be increased help to control this pandemic and more lives can be saved.

References

[1] Su, Shuo, Gary Wong, Weifeng Shi, Jun Liu, Alexander CK Lai, Jiyong Zhou, Wenjun Liu, Yuhai Bi, and George F. Gao. "Epidemiology, genetic recombination, and pathogenesis of coronaviruses." Trends in microbiology 24, no. 6 (2016): 490-502.

[2] Arias-Garzón, Daniel, Jesús Alejandro Alzate-Grisales, Simon Orozco-Arias, Harold Brayan Arteaga-Arteaga, Mario Alejandro Bravo-Ortiz, Alejandro Mora-Rubio, Jose Manuel Saborit-Torres et al. "COVID-19 detection in X-ray images using convolutional neural networks." Machine Learning with Applications 6 (2021): 100138.

[3] Narin, Ali, Ceren Kaya, and Ziynet Pamuk. "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks." Pattern Analysis and Applications 24, no. 3 (2021): 1207-1220.

[4] Ardakani, Ali Abbasian, Alireza Rajabzadeh Kanafi, U. Rajendra Acharya, Nazanin Khadem, and Afshin Mohammadi. "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks." Computers in biology and medicine 121 (2020): 103795.

[5] Tuncer, Turker, Sengul Dogan, and Fatih Ozyurt. "An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image." Chemometrics and Intelligent Laboratory Systems 203 (2020): 104054.

[6] Covid positive image dataset: https://github.com/ieee8023/covid-chestxray-dataset/tree/master/images.

[7] Covid negative image dataset: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[8] Khan, Saddam Hussain, Muhammad Haroon Yousaf, Fiza Murtaza, and Sergio Velastin. "PASSENGER DETECTION AND COUNTING FOR PUBLIC TRANSPORT SYSTEM." NED University Journal of Research 17, no. 2 (2020).

[9] Khan, Saddam Hussain, Anabia Sohail, Asifullah Khan, Mehdi Hassan, Yeon Soo Lee, Jamshed Alam, Abdul Basit, and Saima Zubair. "COVID-19 detection in chest X-ray images using deep boosted hybrid learning." Computers in Biology and Medicine 137 (2021): 104816.

[10] Zhang, Jianpeng, Yutong Xie, Yi Li, Chunhua Shen, and Yong Xia. "Covid-19 screening on chest x-ray images using deep learning based anomaly detection." arXiv preprint arXiv:2003.12338 27 (2020).

[11] Ghoshal, Biraja, and Allan Tucker. "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection." arXiv preprint arXiv:2003.10769 (2020).

[12] Zaki, Muhammad Ahmed, Sanam Narejo, Sammer Zai, Urooba Zaki, Zarqa Altaf, and Naseer u Din. "Detection of nCoV-19 from hybrid dataset of CXR images using deep convolutional neural network." International Journal of Advanced Computer Science and Applications 11, no. 12 (2020).

[13] Yang, Li, Shasha Liu, Jinyan Liu, Zhixin Zhang, Xiaochun Wan, Bo Huang, Youhai Chen, and Yi Zhang. "COVID-19: immunopathogenesis and Immunotherapeutics." Signal transduction and targeted therapy 5, no. 1 (2020): 1-8.

[14] Andreadakis, Z., Arun Kumar, Raúl Gómez Román, Stig Tollefsen, Melanie Saville, and Stephen Mayhew. "The COVID-19 vaccine development landscape." Nat Rev Drug Discov 19, no. 5 (2020): 305-306.

[15] Ciotti, Marco, Massimo Ciccozzi, Alessandro Terrinoni, Wen-Can Jiang, Cheng-Bin Wang, and Sergio Bernardini. "The COVID-19 pandemic." Critical reviews in clinical laboratory sciences 57, no. 6 (2020): 365-388.

[16] Tahamtan, Alireza, and Abdollah Ardebili. "Real-time RT-PCR in COVID-19 detection: issues affecting the results." Expert review of molecular diagnostics 20, no. 5 (2020): 453-454.

[17] Giri, Basant, Shishir Pandey, Retina Shrestha, Krisha Pokharel, Frances S. Ligler, and Bhanu B. Neupane. "Review of analytical performance of COVID-19 detection methods." Analytical and bioanalytical chemistry 413, no. 1 (2021): 35-48.

[18] Salman, Fatima M., Samy S. Abu-Naser, Eman Alajrami, Bassem S. Abu-Nasser, and Belal AM Alashqar. "Covid-19 detection using artificial intelligence." (2020).

[19] Ji, Tianxing, Zhenwei Liu, GuoQiang Wang, Xuguang Guo, Changchun Lai, Haoyu Chen, Shiwen Huang et al. "Detection of COVID-19: A review of the current literature and future perspectives." Biosensors and Bioelectronics 166 (2020): 112455.

[20] Vaid, Shashank, Reza Kalantar, and Mohit Bhandari. "Deep learning COVID-19 detection bias: accuracy through artificial intelligence." International Orthopaedics 44, no. 8 (2020): 1539-1542.

[21] Alazab, Moutaz, Albara Awajan, Abdelwadood Mesleh, Ajith Abraham, Vansh Jatana, and Salah Alhyari. "COVID-19 prediction and detection using deep learning." International Journal of Computer Information Systems and Industrial Management Applications 12, no. June (2020): 168-181.