ARPN Journal of Engineering and Applied Sciences

DEEP LEARNING DRIVEN FRAMEWORK FOR COVID-19 DETECTION - A MOBILE/WEB APPLICATION

Syeda Dua Fatima, Muhammad Waleed Khan, Kashif Sultan and Natasha Shahid
Department of Software Engineering, Bahria University, Islamabad, Pakistan
E-Mail: kashif.buic@bahria.edu.pk

ABSTRACT

Due to the rapid spread of COVID-19, this disease has become a global threat to public health. COVID-19 detection via radiographic images such as CXR and HRCT is one of the most effective techniques. Using deep Learning models, the current study investigates COVID-19 disease on radiographic images. This project proposes a deep learning model for HRCT and CXR images to predict a patient's condition. The proposed model classifies images by incorporating VGG16 deep learning models. In addition, symptom analysis is performed to forecast COVID-19. For symptom analysis, xgboost algorithm is used. Lastly, a mobile/web application for detecting COVID-19 using deep learning models on radiographic images is developed. This work may serve as a resource for other researchers seeking to advance the development of deep learning applications for medical imaging.

Keywords: COVID-19, CNN, VGG1.

1. INTRODUCTION

COVID-19 is an infection that SARS-CoV-2 causes. The spread of COVID-19 started in Wuhan, a populated city in South China. The first confirmed case of COVID-19 was detected in November 2019. Therefore it is called COVID19. After the rapid spread of coronavirus, the World Health Organization (WHO) declared this disease a Global pandemic [1]. This disease is similar to regular flu because of the same symptoms as seasonal flu. Therefore, it was difficult to differentiate it from seasonal flu without diagnostic tests [2].

Further, it is asymptotic, with symptoms sometimes mild and sometimes severe. It is observed that mostly this disease was without symptoms. Hence, the detection of COVID-19 became a challenging task. Therefore, many engineers and researchers are trying to develop a framework that helps physicians and clinicians precisely, cheaply, and quickly detect this global pandemic.

Initially, for detection of coronavirus was performed in a virology lab. However, this process was time-consuming and expensive because the testing kits used to detect coronavirus were expensive. Moreover, the coronavirus cases were increasing exponentially; on the other hand, the testing procedure was expensive, so especially in underdeveloped countries, it was a huge challenge even to detect the coronavirus. Therefore, besides virology lab testing, the radiologist also suggested using radiographic images such as x-ray and CT to detect the virus in the lungs, as this virus affected the lungs badly. So covid-19 was also detected in the radiology lab beside virology labs. Hence, the analysis of radiographic images can be helpful in terms of time and cost [3, 4].

The main objective of diagnosing x-rays and HRCT images using deep learning is to predict whether a patient's lungs are infected with covid-19 or are expected. Therefore, we have developed a framework for detecting COVID-19; in the proposed framework, a deep learning model is used to detect COVID-19. Further, based on the proposed framework, we have also developed a mobile and web application for COVID detection. The proposed application analyzes the radiographic images and predicts the virus. Furthermore, the application also analyzes the symptoms of covid-19 and predicts the virus. Last but not least, the application also suggests nearby vaccination centers and doctors.

The rest of the paper is organized as follows: In Section II we discuss litratue reeview. In Section III. we describe overall methodology of proposed framework. In Section IV, In section IV, the simulation results are presented. Finally, we conclude our work in Section V.

2. LITERATURE REVIEW

With every passing second, the tech world is growing towards betterment and innovation. Artificial Intelligence rules the digital world now. Every little system is now AI-enabled. It allows better use of raw data and its purposeful utilization. Data Analytics and Computer Vision are the essential domains that provide better usage of one's raw data. However, it requires research. Since 2020, Covid-19 has become the talk of the town. Thus, it was a new market trend.

During the past decade, artificial intelligence, particularly deep learning algorithms, with robust data analysis and feature recognition capabilities, have become prominent and clinically recognized in the diagnosis and recognition of lung diseases, including chest X-ray quality control, lung nodule segmentation and detection, and lung cancer screening [5].

In [6], the authors provide a detailed study of neural networks for Covid-19 prediction. First, the authors collect and review the transfer learning methodology for prediction. Then, the study mentions the analysis of Covid-19 and non-Covid-19 patient samples for detection. Finally, the author used the GoogleNet model for covid detection.

In [7] authors provide a detailed study of deep neural network models for Covid-19 prediction. It involves using different models such as ResNet-101,

 ISSN 1819-6608

ARPN Journal of Engineering and Applied Sciences

Faster RCNN, Mask CNN, and ResNet-50 for model training. All these models have a specific architecture

In [7], the authors provide a detailed study of the prediction decision tree. It involves different trees like J48 and Hoeffding Tree for classification & prediction. The publish uses decision trees to create rules rather than training models. The rules help define whether the patient has mild, moderate, severe, or no Covid-19. The study shows a sequence of steps to solve the prediction problem. First, it involves the pre-processing of the dataset. Then the classifier implementation through cross-validation and the comparison of both trees. The split percentage ratio was 66% testing and 34% validation data.

Li *et al* [8] proposed a deep learning based model called COVNet that deetctts COVID 19 by analyzing chest CT images. Based on extracted features they classify non-pneumonia lung diseases and community lung diseases. Another deep learning based model is developed in [9]. They used the proposed model for classification between influenza-A viral pneumonia and COVID-19 pneumonia. In [10] based on the changes in radiographic images of lungs they predicted the covid-19 using deep learning technique. They show that from the medical image we can extract features using deep learning and use these features to classify those images for covid-19 prediction.

3. METHODOLOGY

Our work is carried out in stages such as Data Collection, Data pre-processing and deep learning approach for Covid detection, symptom analysis, and mobile application. The proposed framework is shown in Figure-1.

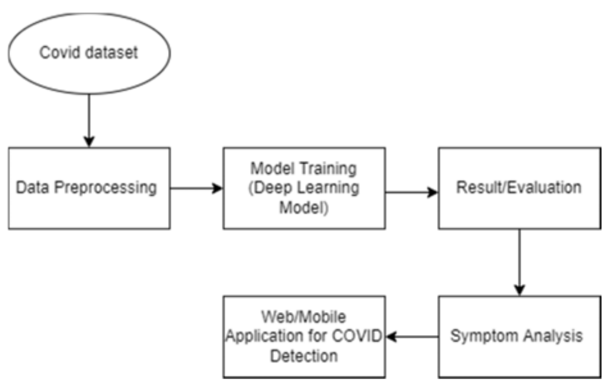


Figure-1. Methodology.

A. Dataset Collection

a) Chest X-Rays (CXR)

The dataset used for training contains 11047 images. It has two Covid and Normal. Of 1377, 360 images belong attributes to the Normal case, and 1017 belong to the Covid case. [11]. The Figure-2 shows the CXR images of infected and normal lungs.

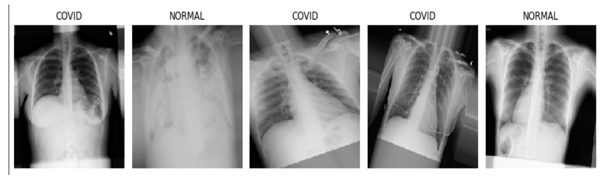


Figure-2. CXR images.

b) HRCT dataset

The dataset used for training contains 2482 CT images. It has two attributes Covid and Non-Covid. Of 2482, 1252 CT images belong to the Covid class, and 1230 belong to the non-Covid class. [12]. Figure-3 shows the HRCT images of COVID infected lungs.

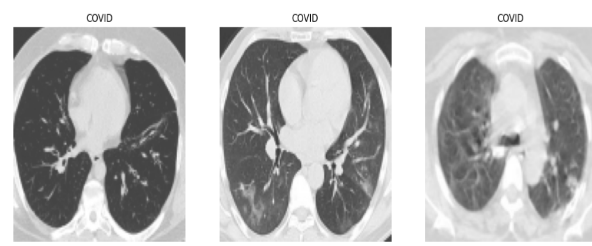


Figure-3. HRCT images.

B. Data Preprocessing

Typically, there is an increase in an intrusion in imaging and data collection procedures, which results in much noise in the dataset. We prepare the input data by removing the unwanted noise from the data using pre-processing techniques so that the dataset would be compatible with the requirements of the Deep Learning model. In this study, image datasets were standardized into resize format of (244*244*3) by changing the dimensions of the image and jpeg format. Given a typical learning rate, RGB coefficients ranging from 0 to 255 were too high for VGG16 and ResNet models to comprehend. Therefore, we focused on values between 0 and 1 by scaling the images dataset to 1/255. Standard images in the dataset were tagged as 1, and COVID-19 infected images in the dataset were tagged as 0

a) Data augmentation

We also used some data augmentation techniques to boost our training dataset artificially. It is because of the lesser number of images we have. Data augmentation will increase the data to the required number of samples. Network efficiency for a small database can also be optimized with the help of data augmentation [13].

Data augmentation procedures include flipping, rotating, shifting an image, and transforming the dataset. In this study, we are using "ImageDataGenerator" from TensorFlow Keras. These techniques apply the image augmentation techniques during the training process. Dataset was separated into two independent datasets. 80 percent of the dataset is used for training the DL model, and 20 percent of the dataset is used to test the trained DL model.

VOL. 18, NO. 2, JANUARY 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

C. Deep Learning Model

For image classification, we have used deep learning algorithms. Deep learning introduced neural networks. It refers to the system of biological neurons. CNN is one of the types of Neural Networks that is best known for its image classification ability based on statistics and predictive modeling.

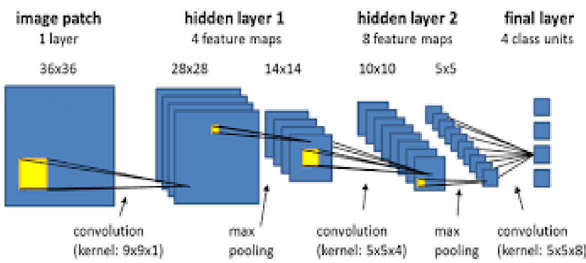


Figure-4. CNN architecture.

CNN holds a vast list of subtypes such as VGG and ResNet models. The architecture of CNN is represented in Figure-4. In this study, we have used the VGG16 training algorithm. With a consistent architecture of convolutional and max pooling layers, VGG16 is considered the best vision model [14, 15]. The Figure-5 represents VGG-16 architecture.

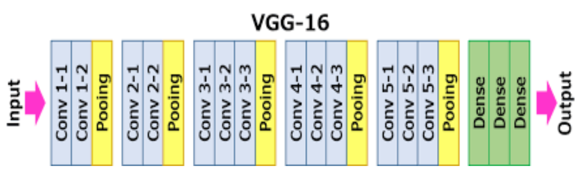


Figure-5. VGG-16 architecture.

D. Symptom Analysis

Besides, for Symptom Analysis, our system was built over an XGBoost classifier. It stands for Extreme Gradient Boosting implemented under the Gradient Boosting framework. It allows machine learning pursuits with a blend of the decision tree. [16] XGBoost is a scalable undertaking of gradient boosting that stretches the computing power for boosted tree algorithms. XGBoost allows the parallel building of trees instead of typical sequential sequences. Providing a level-wise design depletes strict rules to evaluate the quality of splits in the training set.

The model was trained using an XGBooster classifier. The dataset used for decision tree rules assignment contained a CSV file of 100 patients. The dataset possesses six attributes: SBP, DBP, Temp_C, HR, RR, SPO2, and Class. The Figure-6 represents the sample dataset used for symptom analysis.

.

	SBP	DBP	Temp_C	HR	RR	SPO2	Class
0	164	84	37.277778	87.0	18.0	97	0
1	135	89	37.000000	84.0	16.0	93	0
2	122	79	36.722222	104.0	18.0	98	0
3	137	93	36.000000	88.0	16.0	94	0
4	124	71	36.400000	64.0	17.0	98	1
5	129	78	36.500000	77.0	22.0	95	1
6	132	83	37.700000	89.0	16.0	96	1
7	111	76	36.400000	114.0	18.0	98	1
8	138	67	38.400000	97.0	16.0	96	1
9	98	57	36.777778	70.0	16.0	95	0
10	132	82	36.888889	77.0	18.0	98	0
11	57	50	36.111111	134.0	18.0	29	0
12	129	78	36.800000	62.0	18.0	94	1

Figure-6. Description of dataset.

E. Mobile/Web Application

The mobile and web applications were created using React and React Native. This cross-platform development framework uses JavaScript to create the application's user interfaces. The point of using a cross-platform framework was to make sure that the app could be used on all major platforms and React Native helps a lot with that.

In terms of processing, the backend where all of the models and computationally intensive tasks were performed was deployed on Heroku alongside Netlify. We chose REST APIs to make our applications easy to maintain and integrate. We also used information from Google Maps, WHO, and the Ministry of Health in Pakistan to find out about places to vaccinate and hospitals nearby. These data sources created an all-inclusive experience for the application users. The application framework is represented in Figure-7.

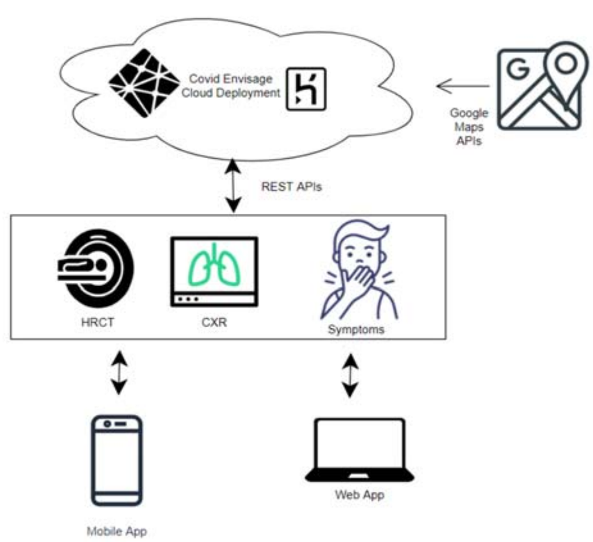


Figure-7. Mobile/Web application

VOL. 18, NO. 2, JANUARY 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

4. RESULTS

Regarding prediction, AI provides us with a confusion matrix and specific accuracy parameters to track our model performance.

A. CXR Results

a) Confusion matrix

It provides us with a summary table of predictions on any classification subject. It shows the TP (True positive), TN (True negative), FP (False positive), and FN (False negative) predicted values. It has two dimensions, i.e., Actual and Predicted, as shown below:

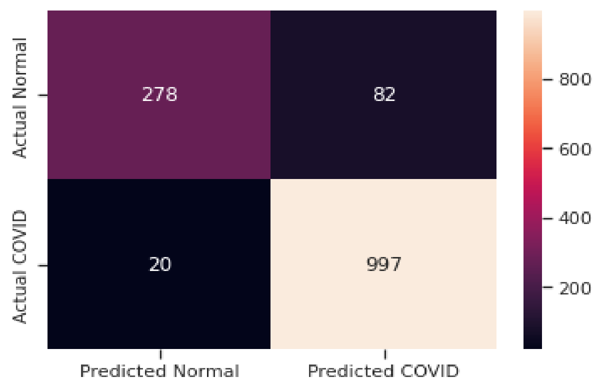


Figure-8. Confusion matrix.

The Figure-8 explains that our model made a total of 1377 predictions. Out of this, 1079 were predicted Covid-19, and 298 were Predicted Normal. However, we have 1017 Actual Covid-19 cases and 360 Actual Normal cases. So, a confusion matrix can help us easily track our model's accuracy.

b) Classification report

The classification report is imported using the sklearn library, and this report provides us with essential performance attributes such as Precision, F1 score, Recall, and support.

In Precision, we are focusing on the predicted TP and TN values. Thus, we are measuring the accuracy of positive predictions

$$\text{Precision} = TP/(TP+FP) \quad (1)$$

In Recall, we focus on correctly identifying the positive values.

$$\text{Recall} = TP/(TP+FN) \quad (2)$$

The F1 score provides a mean value of Precision and Recall. The percentage of correct optimistic prediction by our model.

$$\text{F1Score} = 2*(\text{Recall} * \text{Precision})/ (\text{Recall} + \text{Precision}) \quad (3)$$

Lastly, in support, we provide the number of actual occurrences of that specified class. In our case, a total of 360 Normal Covid-19 cases have been identified in the dataset.

c) Model accuracy chart

The graph shows the fluctuations in the accuracy of both the training and testing data. With every increment of epochs, accuracy improved.

Figure-9. VGG16 model accuracy (CXR).

d) Model loss chart

The graph shows the fluctuations in the training and testing data loss. With every iteration of optimization, the loss rate decreased.

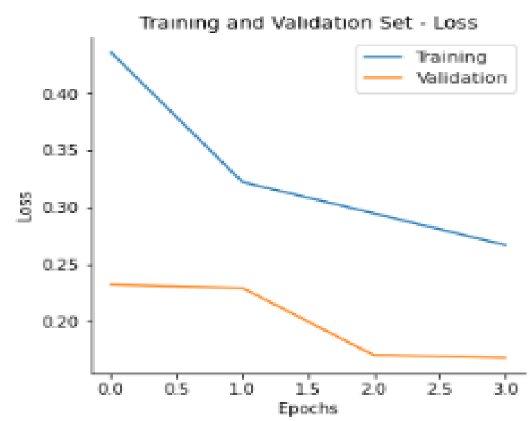


Figure-10. VGG16 model loss curve (CXR images).

B. HRCT Results

a) Confusion matrix

It provides us with a summary table of predictions on any classification subject. It shows the TP (True positive), TN (True negative), FP (False positive), and FN (False negative) predicted values. It has two

VOL. 18, NO. 2, JANUARY 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

dimensions, i.e., Actual and Predicted, as shown below. The formula estimated exponential values.

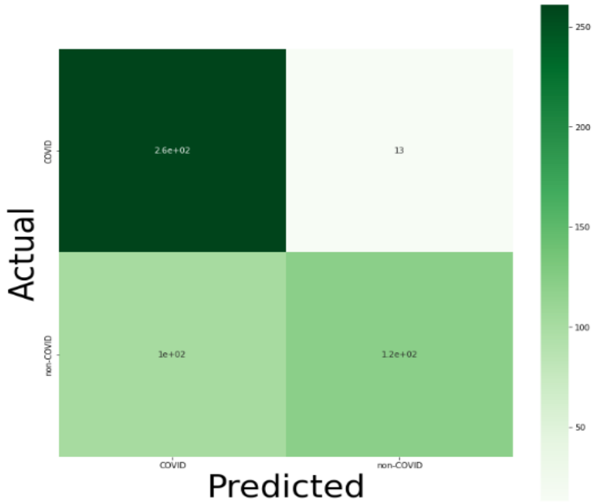
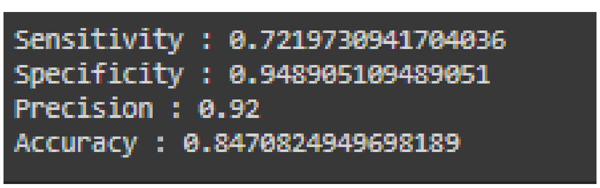


Figure-11. Confusion matrix using HRCT images

The Figure-11 explains that of total, (2.6e+02 + 3e+02) were predicted Covid-19 and (13 + 1.2e+02) were Predict non-Covid. But, in fact, we have (2.6e+02 + 13) Actual Covid-19 cases and (3e+02 + 1.2e+02) Actual Normal cases. So, a confusion matrix can help us easily track our model's accuracy.

We calculated the Recall (sensitivity), TNR (specificity), Precision, and Accuracy from the above TN, TP, FN, and FP shown in Figure-12. We applied the following formulas:

Accuracy = (TP+TN)/(TP+FP+FN+TN) (4)



```
Sensitivity : 0.7219730941704036
Specificity : 0.948905109489051
Precision : 0.92
Accuracy : 0.8470824949698189
```

Figure-12. Evaluation parameters.

b) Model accuracy chart

The Figure-13 shows the fluctuations in the accuracy of both the training and testing data. With every increment of epochs, accuracy improved.

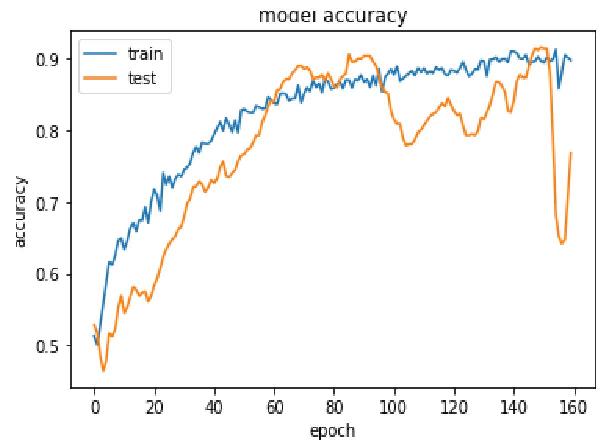


Figure-13. VGG16 accuracy curve (HRCT).

c) Model loss chart

The Figure-14 shows the fluctuations in the training and testing data loss. With every iteration of optimization, the loss rate decreased.

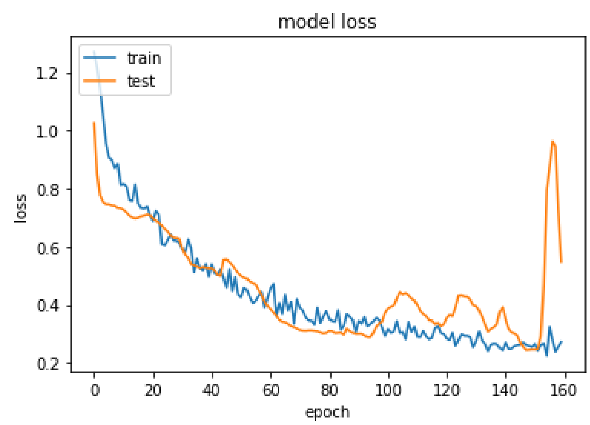


Figure-14. VGG16 model loss curve (HRCT images).

5. CONCLUSIONS

There are different existing frameworks like our project. However, all the features available in our app/framework are not present in just one existing app. Those applications have some features like our project but not all the features like our applications. So, our project is a complete package as it includes all the features in just one web and an android app to perform the Covid-19 testing and generate the reports for it. Previously some data sets and a few models have been trained on this work, and we used them for guidance. In addition, we gave it an application view which was not provided before.

REFERENCES

[1] C. Huang, et al. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395(10223): 497-506. doi: 10.1016/S0140-6736 (20)30183-5.

ARPN Journal of Engineering and Applied Sciences

[2] Shinde Gitanjali R., et al. 2020. Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Computer Science 1.4: 1-15.

[3] Kong W. and Agarwal P. P. 2020. Chest imaging appearance of COVID19 infection. Radiology: Cardiothoracic Imaging. 2(1): e200028.

[4] Lee E. Y., Ng, M. Y. and Khong P. L. 2020. COVID-19 pneumonia: what has CT taught us? The Lancet Infectious Diseases

[5] Tabik Siham, et al. 2020. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. IEEE journal of biomedical and health informatics 24.12: 3595-3605

[6] N. S. A. M. M. D. Haritha. Prediction of COVID-19 Cases Using CNN with X-rays. [Online]. Available: https://ieeexplore.ieee.org/document/9276753.

[7] A. I. Hassam Tahir. Forecasting COVID-19 via Registration Slips of Patients. [Online]. Available: https://ieeexplore.ieee.org/document/9392487.

[8] Li L., Qin L., Xu Z., Yin Y., Wang X., Kong B. and Xia J. 2020. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology.

[9] Jia G., Lam H. K. and Xu Y. 2021. Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method. Computers in biology and medicine. 134, 104425.

[10] Kamil M. Y. 2021. A deep learning framework to detect Covid-19 disease via chest X-ray and CT scan images. International Journal of Electrical & Computer Engineering (2088-8708): 11(1).

[11] NII dataset keras | Kagg https://www.kaggle.com/code/anfelsarabouachat/nii-dataset-kerasle

[12] COVID-19 CT scans | K https://www.kaggle.com/code/anfelsarabouachat/nii-dataset-kerasaggle

[13] Shorten Connor and Taghi M. Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. Journal of big data 6.1: 1-48.

[14] Abdar A. K., Sadjadi S. M., Soltanian-Zadeh H., Bashirgonbadi A. and Naghibi M. 2020, November. Automatic detection of coronavirus (COVID-19) from chest CT images using VGG16-based deep-learning. In 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME) (pp. 212-216). IEEE.

[15] Chen A., Jaegerman J., Matic D., Inayatali H., Charoenkitkarn N. and Chan J. 2020, November. Detecting Covid-19 in chest X-rays using transfer learning with VGG16. In CSBio'20: proceedings of the eleventh international conference on computational systems-biology and bioinformatics. pp. 93-96.

[16] Rochmawati N., Hidayati H. B., Yamasari Y., Yustanti W., Rakhmawati L., Tjahyaningtijas H. P. and Anistyasari Y. 2020, October. Covid symptom severity using decision tree. In 2020 Third international conference on vocational education and electrical engineering (ICVEE) (pp. 1-5). IEEE.