Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

# COVID-19 Prediction using X-Ray Images

G. Aparna Dr. S. Gowri R. Bharathi Dr.Vimali J S Dr. Jabez J Dr. Ajitha P
*Department of IT, Sathyabama Institute of Science and Technology*
aparnagopinath2000@gmail.com
gowriamritha2003@gmail.com



***Abstract*—Coronavirus disease (COVID-19) is a pandemic caused by the coronavirus SARS-CoV-2 that was not previously seen in humans. COVID-19 is spreading rapidly throughout the world. COVID-19 can be detected by a lung infection of the patients. The standard method for detecting COVID-19 is the Reverse transcription-polymerase chain reaction (RT-PCR) test. But the availability of RT-PCR tests is in short supply. As a result of this, the early detection of the disease is difficult. The easily obtainable modes like X-rays are often used for detecting infections in the lungs. It is confirmed that X-ray scans can be widely used for efficient COVID-19 diagnosis. But a physical diagnosis of X-rays of an outsized number of patients is a long-term process. A deep learning-based diagnosis process can help radiologists in detecting COVID-19 from X-ray scans. Pre-trained CNNs are commonly used in detecting diseases from datasets. This paper proposes a CNN model with a parallelization strategy that extracts the features in the X-ray images by applying filters parallelly through the images. Our proposed method aims to attain higher accuracy and a less loss rate with precision. To do so, the accuracy and loss rates of three types of CNN - VGG-16, MobileNet, and CNN are compared with the parallelization technique. Since, VGG-16 and MobileNet are pre-trained models; those two models are directly imported from Keras. Moreover, this paper utilizes two datasets consisting of COVID X-ray images and Non-COVID X-ray images for the prediction of COVID-19 using Convolution Neural Network [CNN].***

***Keywords*—COVID-19, Tensorflow Keras, NumPy, Deep Learning, X-Ray Image, Parallelization, Convolutional Neural Network (CNN), Corona Virus, RT-PCR, React Web App.**

## I. INTRODUCTION

On New Year's Eve 2019, when China informed the world about a bunch of pneumonia cases of unknown origin in Wuhan City, Hubei Province, the coronavirus epidemic came to light. The disease rapidly spread to more Chinese provinces and eventually all over the world. It has now been declared as a pandemic by the World Health Organization. The virus was named SARS-CoV-2, hence the disease is known as COVID-19. The outbreak spread quickly all around the world. Aged individuals and others with existing medical conditions, such as diabetes, cancer, and respiratory illness are most likely to experience severe illnesses. They may require immediate medical attention and a ventilator system to make breathing easier if the patient's condition is critical.

The Reverse transcription-polymerase Chain Reaction (RT-PCR) test is the standard method for the diagnosis of COVID-19. The facilities for RT-PCR tests, however, are in short supply, making early detection of the disease difficult. It is proven that for effective diagnosis of COVID-19, X-ray images can be used. The abnormalities in the X-ray images can be detected easily.

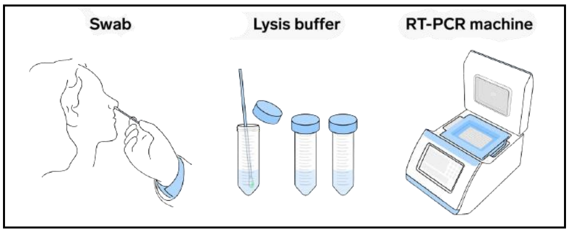
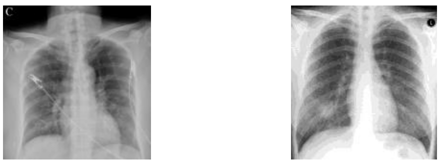



Fig: 1 – RT-PCR Test

As the patients in the Intensive Care Unit (ICU) are extremely contagious, X-rays are better, safer, and more affordable than CT scans and with reduced radiation exposure. Rapid and precise testing for COVID-19 is observed due to the accessibility of open-source deep-learning-based approaches. Radiology images are visual indicators of anomalies and a key screening tool. This method of diagnosing COVID-19 using X-ray images is often used in hospitals nowadays.

To detect COVID-19 patients cost-effectively and to get the results in less than 10 minutes, X-rays can be used, said Dr. Vishal Rao, who is a member of the consultative group for COVID and a surgical expert. Dr. Rao who started the project said that an artificial intelligence system can be used for analyzing the X-ray, in which a computerized method can scan for the abnormal area that will show the relevant results.



PA View AP View

Fig: 2 – Types of views in X-ray image

The proposed research work has used two datasets. One is the Covid Chest X-ray dataset from GitHub with a collection of COVID-19 positive X-ray scans. Another one is the COVID-19 Radiography Database dataset which consists of Covid and normal X-ray images separately. The projection of X-ray scans have PA and AP view. The standard view used for chest X-ray is PA view. So, PA is utilized in this research work to view chest X-rays in this work.

Any abnormality in lungs can be easily identified through X-ray images. When there is an abnormality, there will be an increase in lungs density which will cause whiteness in the X-ray image as in fig 3. This is called as ground-glass pattern. Radiologists can identify the disease with this ground-glass pattern.

Fig: 3 – Ground-Glass Opacity

This paper applies a deep learning-based CNN model to predict COVID-19 with a higher accuracy rate. Convolutional Neural Networks are popularly known for automatic feature extraction in images. For this reason, CNNs are widely used in medical fields for detecting diseases from images. Here, our proposed CNN will extract this ground-glass pattern and will be able to detect it. Since this paper is related to medical field, the main goal of this proposed work is to achieve higher accuracy with less loss rate. Here, the two pre-trained models – VGG-16 and MobileNet are used for comparison.

## II. LITERATURE SURVEY

In [9], according to The New Indian Express' report, Dr. Rao, a member of the COVID consultant group said that for diagnosing COVID-19 patients immediately and cost-efficiently X-rays can be used for better results. With this as an idea, this paper has proposed a CNN model which could predict COVID-19 with a higher accuracy rate.

In [2], the authors developed a deep learning model to predict COVID-19 using patients' CT images. Using deep learning, they have trained a neural network to diagnose COVID-19 using the patients' CT images. They have attained an accuracy of 76% using their model.

The authors of [5] proposed the parallelization concept in their research paper. From this, the parallel convolution neural network concept is selected for this proposed work.

In [1] the proposed structure is divided into three stages. To begin, features are extracted from CT scans using AlexNet, a Convolutional Neural Network (CNN). Second, a proposed features selection algorithm based on Stochastic Fractal Search (SFS), Guided Whale Optimization Algorithm (Guided WOA), is used, followed by feature balancing. Finally, a proposed voting classifier, Guided WOA based on Particle Swarm Optimization (PSO), combines the predictions of various classifiers to determine the most voted class.

The existing artificial intelligence-based models for COVID-19 detection and the challenges these systems face is addressed in [13]. Several deep learning architectures, such as ResNet, Inception, Googlenet, and others, are used to detect COVID-19. A thorough analysis of these approaches can be found here.

Qiu H [12] compiled a large data set of X-ray and CT scan images from various sources and used deep learning and transfer learning algorithms to create a simple and efficient COVID-19 detection system.

The importance of medical image in controlling and treating COVID-19 disease is demonstrated in a succinct review by Dong et al. [14]. One of the most important points raised is that an automated tool can be a useful complement to the laboratory-based real-time polymerase chain reaction (RT-PCR) test.

Harsh Sharma in [11] deployed four types of models to explain the efficiency of adding dropout layers. These dropout layers play a major role in preventing overfitting.

To characterize the various manifestations of COVID-19 and community-acquired pneumonia (CAP) infections, a dual sampling attention network [16] was proposed. A lung mask was used to eliminate the image context of non-lung areas in chest CT to focus on the lungs.

Sammy V. Militante in [4] proposed a VGG-16 model in their research to predict COVID-19 and pneumonia on X-ray scan images using the Convolutional Neural Network method. With this model, they have obtained 95% accuracy.

Abdel-Basset et al. used the x-ray segmentation technique to extract similar small regions that could be COVID-19 features in [18]. They used an improved marine predator's algorithm to create a hybrid COVID-19 detection model (IMPA).

In the work done by Wu et al. [3], they proposed a model of weakly supervised CNN which detects and classifies COVID-19 with the help of CT scans and it has achieved 96.2% accuracy.

## III. METHODOLOGY

The modules and methodologies used in this paper are as follows:

1. Data Collection
2. Data Preprocessing
3. Training the CNN model
4. Web app implementation

### *A. Data Collection*

The first and foremost step in deep learning is data collection. Two datasets are utilized to apply the proposed model. One of which is the Covid Chest X-Ray dataset from GitHub consisting of covid positive X-Ray scan images.

The second dataset is the COVID-19 Radiography Database dataset which consists of Covid and normal X-ray images. For the first dataset, with the help of the metadata file, we split the dataset into two different folders as Covid-19 Positive and Covid-19 Negative. In the second dataset, the images are already categorized into covid and normal. So, we move them directly into the positive and negative folders.

*B. Data Preprocessing*

After splitting datasets into positive and negative categories, we convert them into grayscale images and resize them all to the same size. The dataset consists of images with various sizes. Since the images are of different sizes, we resize the images into 100x100 pixels. Compared to other formats, absolute feature extraction can be done in grayscale images. So we convert RGB to grayscale using the following equation:

Grayscale = ((Wr x R) + (Wg x G) + (Wb x B)) where,

W indicates the weights of red, green, and blue colors respectively. The weights of red, green, and blue are 0.30, 0.59, and 0.11 respectively which sums to a total of 1. So the new equation will be:

Grayscale = ((0.30 x R) + (0.59 x G) + (0.11 x B))

The above is the formula for converting images into grayscale. After modifying them, we store the modified data into a NumPy array as .npy. NPY is a python programming library that stores all the information required to reconstruct an array on any computer. We use the NumPy array to reduce the storage size and we use them for training.

*C. Training the CNN model*

The training phase is where the deep learning algorithms are used. For the proposed model, i.e., the CNN model, after data preprocessing, we train the dataset with Convolutional Neural Network. We use the NumPy array data from data preprocessing to train the model.

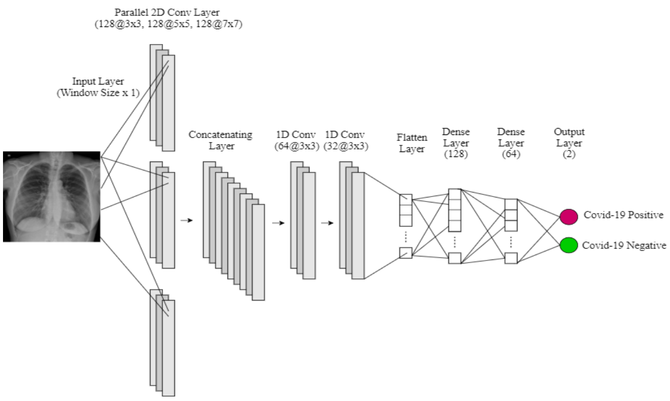



Fig 4: Architecture Diagram of Proposed Work

We train a sequential model with the parallel convolutional layer. We apply a parallel 2D convolutional layer to the input layer with 128 filters to the kernels of sizes 3, 5, and 7. We concatenate all the conv2D layers into a single layer with padding as same to get the convolution output size same as the input then reduce spatial dimensions using max pooling. Then we apply a convolutional layer with 64 and 32 filters to kernel size 3x3. For the activation parameter, we have used Relu to get either 0 or 1, depending on whether the input is negative or not. Then reduce spatial dimensions using max pooling.

We flatten the output of the convolutional layers into a single-dimensional array. This is connected to the 128 and 64 neurons dense layer, also called a fully-connected layer. To prevent the overfitting of the training data, we've added dropout layers with a dropout rate of 0.5. We've grouped the dataset into 2 batches by giving batch_size=2 since the amount of data is large and requires more storage space. We've given a total of 20 epochs for the training process.

Model: "sequential"

| Layer (type) | Output Shape | Param # |
|---|---|---|
| model (Functional) | (None, 100, 100, 384) | 11008 |
| conv2d_3 (Conv2D) | (None, 98, 98, 64) | 221248 |
| activation (Activation) | (None, 98, 98, 64) | 0 |
| max_pooling2d (MaxPooling2D) | (None, 49, 49, 64) | 0 |
| conv2d_4 (Conv2D) | (None, 47, 47, 32) | 18464 |
| activation_1 (Activation) | (None, 47, 47, 32) | 0 |
| max_pooling2d_1 (MaxPooling2 | (None, 23, 23, 32) | 0 |
| flatten (Flatten) | (None, 16928) | 0 |
| dropout (Dropout) | (None, 16928) | 0 |
| dense (Dense) | (None, 128) | 2166912 |
| dropout_1 (Dropout) | (None, 128) | 0 |
| dense_1 (Dense) | (None, 64) | 8256 |
| dropout_2 (Dropout) | (None, 64) | 0 |
| dense_2 (Dense) | (None, 2) | 130 |

Total params: 2,426,018
Trainable params: 2,426,018
Non-trainable params: 0

Fig 6: Model Summary

The model summary given above shows the layers used in our proposed model and their types, namely convolution layer, flatten layer, dense layer, activation and pooling layer.

The output layer will be of two neurons for the positive and negative results of COVID-19. So once the model is fully trained, the best model with the least loss and high accuracy is saved by setting save_best_only=True. The best model will be saved with an extension of '.h5'. This saved model will be loaded into the web app to predict Covid-19.

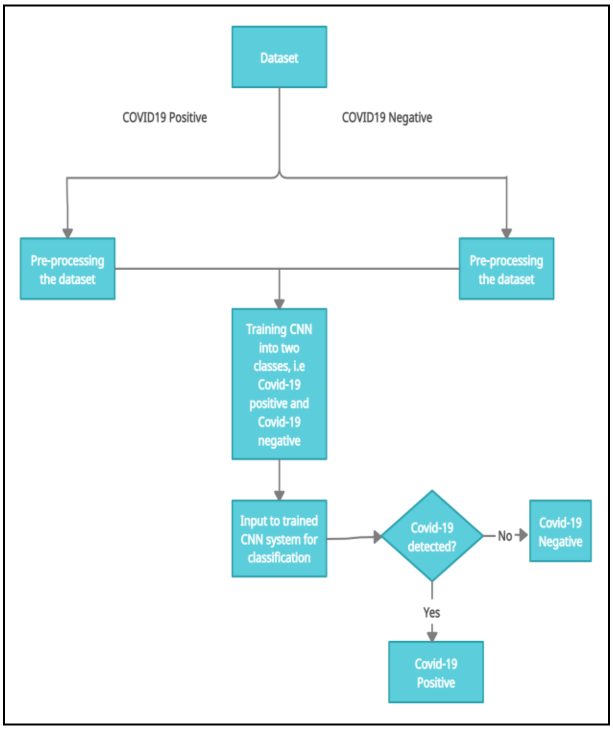



Fig 5: Dataflow of Proposed work

*D. Algorithm*

i. Preprocess the dataset images and convert them into a numpy array.
ii. Split the numpy array dataset into training and test data using sklearn.
iii. Apply CNN filters parallelly for 3, 5 and 7 kernel sizes respectively to extract the features.
iv. Import layers from tensorflow keras API to create the CNN model with parallelization from scratch and insert the training dataset.
v. Now train the models for required number of epochs with batch_size=2, until the loss is less than 0.5.
vi. Cross validate the model using the test data. If the model is overfit or underfit, retrain the model with a different set of layers or different dataset.
vii. By using model checkpoint, save the best_epoch model which has higher accuracy and less loss rate by specifying save_best_only=True.
viii. Then load the saved model to the web app for predicting Covid-19.

*E. Web App Implementation*

We use flask as a backend and loaded in the model with the least validation loss from our training. We use React to make the front end of the application. React is a JavaScript library for building user interfaces. There are 3 separate components in the webpage that each performs different actions.

The main App component manages the state of the app and performs the API requests to the backend. Additionally, there is an ImageSelector component that manages to get the image from the user and communicate it to the App. There is also an Info component that uses the information that the App component got from the API and presents it in a usable way on screen.


Fig 7: Output Screen

## IV. RESULTS AND DISCUSSIONS

The important goal of this proposed model is to build an effective deep learning model which predicts COVID-19 disease with higher accuracy. Here, in this proposed model, we've collected a total of 2737 X-ray images of COVID and Non-COVID chest X-Rays. The websites GitHub and Kaggle.com were used for retrieving images.


Fig 8: Dataset sample images

We have used three different CNN models. We compared two pre-trained models with our proposed CNN model with a parallelization strategy. The two pre-trained models are VGG-16 and MobileNet. VGG-16 consists of 16 layers with millions of parameters. We imported these two pre-trained models and used them for training purposes separately.

In fig 5, the first two images are the accuracy with epoch graph of the pre-trained VGG-16 and MobileNet models. The image at the bottom is the proposed parallel CNN model. This clearly shows that the parallel CNN model has shown a great result with a higher accuracy rate than the other two.

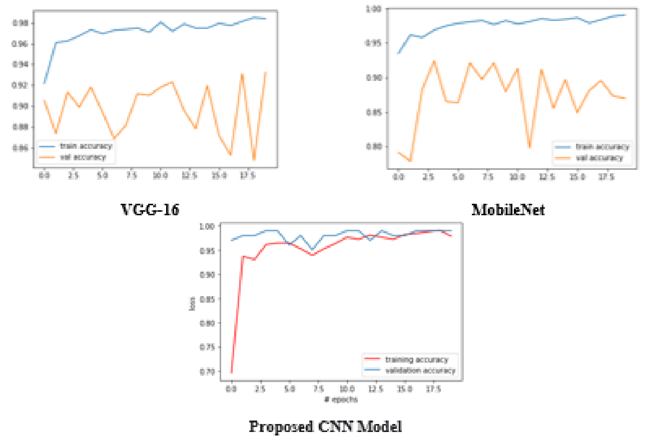



Fig 9: Accuracy vs Epoch graph of VGG-16, MobileNet, and Proposed CNN

In fig 6, the first two images are the loss to the epoch graph of the two pre-trained VGG-16 and MobileNet models. The image at the bottom is the loss graph of our proposed parallel CNN model. This shows that the loss rate of our proposed model is less than the loss rate of the other two pre-trained models. The loss rate of the best epoch in parallel CNN is 0.04 whereas the loss rate of VGG-16 and MobileNet is 0.25 and 1.53 respectively.

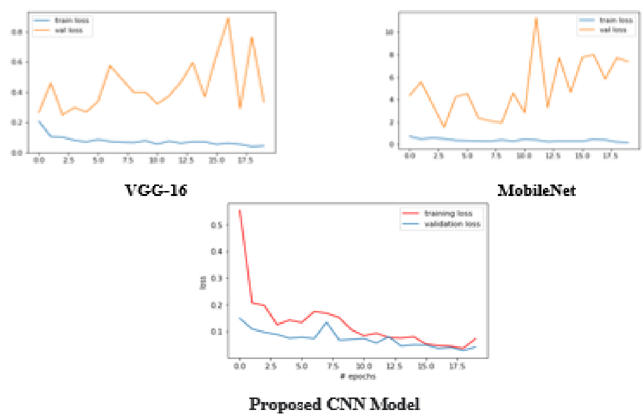



Fig 10: Loss vs Epoch graph of VGG-16, MobileNEt, and Proposed CNN

For better understanding, a confusion matrix is generated for each convolutional neural network used in the prediction. The confusion matrix is a metric for evaluating machine learning classification results. It is a two-dimensional matrix with four characteristics true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The precision values, accuracy, recall, and F-measure are all calculated using the confusion matrix.

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$

The above is the formula for calculating precision using confusion matrix.

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision x Recall}}{\text{Precision + Recall}}$$

The above are the fomula for calculating recall and f1 using confusion matrix. The following confusion matrix shows the values of precision, recall and f1.

| Predicted/ Actual | Covid | Normal | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Covid | TP(29) | FP(1) | 0.966 | 0.906 | 0.935 |
| Normal | FN(3) | TN(27) | 0.900 | 0.964 | 0.930 |

Table: 1 – Confusion Matrix of VGG-16

| Predicted/ Actual | Covid | Normal | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Covid | TP(27) | FP(3) | 0.900 | 0.931 | 0.915 |
| Normal | FN(2) | TN(28) | 0.933 | 0.903 | 0.917 |

Table: 2 – Confusion Matrix of MobileNet

| Predicted/ Actual | Covid | Normal | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Covid | TP(30) | FP(0) | 1.00 | 0.967 | 0.983 |
| Normal | FN(1) | TN(29) | 0.966 | 1.00 | 0.983 |

Table: 3 – Confusion Matrix of our Proposed CNN

From the above confusion matrix of different models, accuracy can be calculated using the following formula:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{(\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})}$$

After calculating accuracy from the above formula, we got an accuracy of 93.3% and 91.6% for VGG-16 and MobileNet respectively which is lesser than that of our proposed CNN model which has an accuracy value of 98.3%.

From the CNN, we use the best epoch model which is saved, to predict the X-ray images. In the web app, using the image selector option anyone can upload their X-ray image. By clicking on the predict button, we can get the result as Covid Positive or Covid Negative with probability percentage.

## V. CONCLUSION

COVID-19 detected from the diagnostic test and antibody test takes a lot of time. According to the facts, even RT-PCR, the gold standard diagnostic test, has a high chance of generating false-negative results. As an auxiliary diagnostic method, using AI-based chest X-Ray detection can assist doctors in providing a more thorough patient assessment. Because the symptoms of COVID-19 and all causes of pneumonia overlap, early diagnosis of the disease is still the most difficult decision to make. To solve this, a deep learning-based model was trained in the proposed work to predict COVID-19 with a higher accuracy rate, lower loss rate, and low time duration. For a high-speed and accurate diagnosis of chest X-rays of patients, this model can be used. Primary health workers in remote areas can use this automated COVID-19 detection system. This proposed method can be improved by building a larger dataset of Covid X-rays and X-rays of other lung diseases. It also can be extended to IoT-related work like connecting them to the X-ray machine to predict the COVID-19 in real-time.

## References

[1] El-Sayed M. El-Kenawy, Abdelhameed Ibrahim, "Novel Feature Selection and Voting Classifier Algorithms for COVID-19 Classification in CT Images", IEEE Access(2020), pp.179317-179335.

[2] H. Hui, X.Wu, L. Li, L. Wang, M. Niu, B. He, L. Li, H. Li, X. Yang, J. Tian, and Y. Zha, "Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study," ScienceDirect(2020).

[3] S. Hu, Z. Niu, Y. Gao, Y. Jiang, M. Wang, L. Li, X. Xiao, E. F. Fang, J. Xia, H. Ye, W. Menpes-Smith, ad G. Yang, "Weakly supervised deep learning for COVID-19 infection detection and classification from CT images", IEEE Access, vol. 8, pp. 118869-118883, 2020.

[4] Sammy V. Militante, Nanette V. Dionisio, Brandon G. Sibbaluca, "Pneumonia and COVID-19 Detection using Convolutional Neural Networks", IEEE Xplore(2020).

[5] Sunwoo Lee, Dipendra Jha, Ankit Agrawal, Alok Choudhary, and Wei-Keng Liao, "Parallel Deep Convolutional Neural Network Training by Exploiting the Overlapping of Computation and Communication", IEEE Conference(2017).

[6] Mehmet Sevi, İlhan AYDIN, "COVID-19 Detection Using Deep Learning Methods", IEEE Xplore(2020).

[7] Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byung-Won On, Waqar Aslam, Gyu Sang Choi, "COVID-19 Future Forecasting Using Supervised Machine Learning Models", IEEE Access(2020).

[8] Muhammad E. H. Chowdhury, Amith Khandakar, Tawsifur Rahman, Rashid Mazhar, Mamun Bin Ibne Reaz, Muhammad Abdul Kadir, Nasser Al Emadi, Zaid Bin MAhbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Mohammad Tariqul Islam, "Can AI Help in Screening Viral and COVID-19 Pneumonia?", IEEE Access(2020).

[9] "Chest X-rays can give COVID-19 test results in just 10 minutes: Expert", The New Indian Express.

[10] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv.org(2015).

[11] Harsh Sharma, Jai Sethia Jain, Priti Bansal, Sumit Gupta, "Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia", IEEE(2020).

[12] H. Qiu, J. Wu, L. Hong, Y. Luo, Q. Song, and D. Chen, Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. The Lancet Infectious Diseases. 2020.

[13] Kang Zhang, Xiahong Liu, Jun Shen, Zhihuan Li, Ye Sang, Xingwang Wu, Yunferi Zha, Wenhua Liang, Chengdi Wang, Ke Wang, Linsen Ye, Ming Gao, Zhongguo Zhou, Liang Li, and others, "Clinically Applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography", doi:10.1016/j.cell.2020.04.045, Cell(2020).

[14] D. Dong, Z. Tang, S. Wang, H. Hui, L. Gong, Y. Lu, Z. Xue, H. Liao, F. Chen, F. Yang, R. Jin, K. Wang, Z. Liu, J. Wei, W. Mu, H. Zhang, J. Jiang, J. Tian, and H. Li, "The role of imaging in the detection and management of covid-19: a review," IEEE Reviews in Biomedical Engineering, pp. 1–1, 2020.

[15] M. Ilyas, H. Rehman, and A. Nat-Ali, "Detection of Covid-19 from chest x-ray images using artificial intelligence: An early review", arXiv(2020).

[16] X. Ouyang, J. Huo, L. Xia, et al. "Dual-Sampling Attention Network for Diagnosis of COVID-19 from Community-Acquired Pneumonia", IEEE Transactions on Medical Imaging, 2020.

[17] Mugunthan. S. R. "Decision Tree Based Interference Recognition for Fog Enabled IOT Architecture" Journal of trends in Computer Science and Smart technology (TCSST) Vol.02, No. 01 (2020), Pages: 15-25.

[18] M. Abdel-Basset, R. Mohamed, M. Elhoseny, R. K. Chakrabortty, and M. Ryan, "A hybrid covid-19 detection model using an improved marine predators algorithm and a ranking-based diversity reduction strategy," IEEE Access, vol. 8, pp. 79 521–79 540, 2020.

[19] Suma. V, and Shavige Malleshwara Hills, "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics" Journal of Soft Computing Paradigm (JSCP) 2, No. 02 (2020): 101-110.

[20] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus COVID-19", PLOS ONE, vol. 15, no. 3, Mar 2020.