Analysis of COVID-19 and Pneumonia Detection in Chest X-Ray Images using Deep Learning

Kanakaprabha. S.
PG Scholar, Dept. of Computer Science and Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
skanakaprabha@gmail.com

D. Radha
Asst. Professor, Dept. of Computer Science and Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
d_radha@blr.amrita.edu

2021 International Conference on Communication, Control and Information Sciences (ICCISc) | 978-1-6654-0295-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCISc52257.2021.9484888

***Abstract*— The on-going COVID-19 outbreak made healthcare systems across the globe to be in the edge of the battle. Recent stats indicate that more than 140+ million confirmed cases are diagnosed globally as of April 2021. The cases are increasing day by day. The early and auto diagnosis helps people to be precautious. The proposed work aims to detect COVID-19 patients and Pneumonia patients from X-Rays which is one of the medical imaging modes to analyse the health of patient's lung inflammation. The suitable Convolutional Neural Network Model is selected for the identified dataset. The model detects COVID-19 patients and Pneumonia patients on the real-world dataset of lung X-Ray images. Images are pre-processed and trained for various classifications like Normal, COVID-19 and Pneumonia. After pre-processing, the detection of the disease is done by selecting the appropriate features from the images in each of the datasets. The result indicates that accuracy of detection of COVID vs Normal and COVID vs Pneumonia. Among those two, COVID vs Normal is with better accuracy than COVID vs Pneumonia. This method detects not only COVID or Pneumonia, but also the subtypes of Pneumonia as bacterial or Viral Pneumonia with 80% and 91.46% respectively. The detection of COVID, Bacterial Pneumonia and Viral Pneumonia using the proposed model helps in rapid diagnosis and to distinguish COVID from Pneumonia and its types which facilitates to use appropriate and fast solutions.**

Keywords—COVID-19, Bacterial Pneumonia, Viral Pneumonia, Convolution Neural Network, lung X-rays, accuracy.

I. INTRODUCTION

At first, the COVID-19 outbreak is declared by Wuhan, China in December 2019.The virus was highly widespread by March 2020 and was asserted by the United Nations Agency (UN agency). By April 2021, over 4 million confirmed cases with 127,040 deaths in UK alone is recorded [1]. The cumulative number of confirmed cases across the world was increasing due to COVID-19, over 140 million, and the mortality rate is 3 million. This epidemic was continuing to spread, according to the WHO assumption incubation period of the virus range 2 to 10 days and can be transmitted by the patients during that incubation period. Identification of this disease in the early stage will care for patient's lung inflammation.

The work aims to detect COVID-19 and Pneumonia patients using deep learning techniques as Normal, Infected, Pneumonia patients. Transfer learning techniques can be used for COVID-19 detection [2]. Image modes are with high resolution and used as painless studies for medical diagnosis. Convolution neural network is widely used in deep learning techniques and helpful to identify diseases in short time [3]. Even with any abnormal conditions of the images, prediction can be done with deep learning methods. Lung lesion is one of the common diseases for all age group of humans. Currently, this disease may rise with various other symptoms such as COVID-19, Pneumonia etc. Golden standard diagnosis in COVID-19 is the reverse transcription polymerase chain reaction (RT-PCR) method. This is one of the most widely used laboratory method for detecting COVID-19 virus. Some studies identified that it showed false-positive results also. Test carried out too early might lead to false negative results.

X-Ray images are very familiar and utilized in lung infection detection, COVID-19 as well as other investigation. Generally, X-ray images of the victim can be classified into three types image such as Normal, COVID-19, and Pneumonia Infected. Bacterial and viral Pneumonia can cause several damages to human respiratory symptoms such as irregular heartbeat, cardiovascular shock, muscle pain, fatigue and it can lead to death in some cases. So, identifying Pneumonia subtypes is a crucial requirement and very helpful for saving time, reduce error in outpatient health conditions [4]. Main motivation is from difficulty for an experienced specialist to identify a subtype of Pneumonia in affected patients. Subtype may lead to complicated health issues for patients.

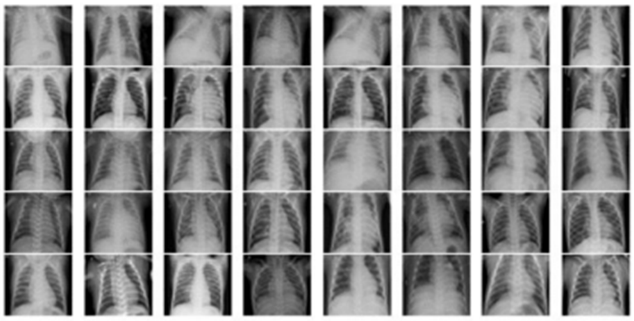

Fig. 1. Chest X-Ray Images

Fig 1. signifies the X-Ray image of the Chest from which the structure of an image classification of Normal, Infected, COVID-19 Infected, Pneumonia Infected and subtype of Pneumonia are done. The swarm with COVID-19 indications tried would be sensible onset evidence. Test movement needs to be exceptionally quick, adaptable, analyse and conclude accurately within shorter time. The current reality of coronavirus infection is exceptionally high and need mechanisms to support people, health organizers, volunteers, employees wherever possible. The official advices of the government, health advises from world-wide health organization like WHO are to be followed. The workplaces are made online instead of onsite places to avoid the spread of infection. The world-wide scenario is rising lockdowns in circumstances of wide spread of COVID-19. One of the major solutions to quickly diagnose the severity is to analyse a patient's lung inflammation. The screening procedure for identifying COVID-19 is by utilizing X-ray images of lungs. Visual pointers are clearly found outstandingly in radiography image of infected patients such as COVID-19, or Pneumonia patients.

The paper is organised as Section II with literature survey, Section III with problem statement, Section IV with System Architecture, Section V with Result and Analysis. And followed by Section VI with conclusion and future scope

II. LITERATURE SURVEY

In the process of finding the diseases and its existing system of literature is discussed by. R. Sethi et al [5], which says that real-time standard diagnostic tool used is RT-PCR test. It is mainly to detect COVID-19 virus or any other pathogen. The process starts with nasopharyngeal swab which is collected either in nose or throat of patients. It helps the medical community to diagnose COVID-19 during the pandemic. Some of the cons in the test are the time taken to get the results, false reports, and inconvenience in test procedure for the patients. An alternative approach for screening method is by using X-Ray images which will reduce the diagnosis burden of the physician. The X-Ray image classification is done using Convolution neural network (CNN) with deep learning techniques. Different lung diseases are identified from Chest X-Ray images and around 100,000 images are used to train the CNN model for diagnosing 14 diseases. However, seven different existing architectures are used for comparing the accuracy. Finally, accuracy of the best model is improved by fine-tuning. New architectures are introduced based on CNN detection model for COVID-19 as well as other diseases. Diabetics can also be detected using CNN model [6].

T. Xia et. al. analyses the complete blood cell counts (CBCs) of COVID-19 hematologic. CBCs include red blood cells (RBC), white blood cells (WBC) and platelets of blood corpuscles. The decrease in red blood cell count, white blood cell count, unusual shapes of blood cell, platelets count diagnosed as anemia, haemophilia, infection or tumor. Expensive hematology of blood cell counts is time-consuming, but easier, faster, and avoids cost burden. The real-time object detection algorithm is YOLOv3.The new feature is introduced by CNN darknet-53, YOLOv3. Dataset used to microscopic blood cell contains 364 images. Trained with unique classifier to identify performance, gradient but shown as false YOLOv3 weights. One of the advantages is to identify blood cell samples of COVID-19 symptoms or blood clots any other symptoms of COVID-19. Further progression is by improving precision and recall of the test results using an ensemble method. The outbreak COVID-19 situation uses a Point of Care (POC) device for early detection and helpful in control measures. A mobile application is faster and simple mode access with microfluidic and microscope is also a practical use of the COVID-19 detection [7].

O. A. Ramwala et al [8] explores the severe acute respiratory syndrome of coronavirus (SARS-COV-2) with main focus on RT-PCR test detection of nucleic acid and symptoms of any virus present in the DNA. Hence, they proved that radiological assessment is highly sensitive than RT-PCR test. Furthermore, COVID-19 detection is with 80% sensitivity accuracy. Dataset used is with two sets of images such as 182 posteroanterior (PA) collect from IEEE data port and other images from pneumonia dataset in Kaggle. In this case the imbalance problems are encountered because of weight cross-entropy. The Residual network model is trained on the dataset; a higher value is obtained by loss function and a lower value for the other classes. The residual network is solving this accuracy detection issue and increases with the vanishing gradient approach. Image classification models are used by heat map which clearly identifies network region of interest to extract correct features, and makes classification decision

L.Xu et al proposed lung lesion detection using CT images in RCNN to track the patient's conditions and for timely treatment K-Means algorithm which carries out clustering of images and Faster RCNN models using model VGG-16 and ResNet 50 classification are compared. The algorithm is divided into two stages in which the first stage extract features from input images and obtain anchors and the second stage converts the extracted features into pooling layers. Finally, they combine the classification models. VGG-16 gives better result in augmented dataset and ResNet50 could be a similar result within the original dataset but when the model is trained with augmented dataset it is prone to overfitting. The current situation is that the accuracy of any model needs to be improved. So, the accuracy of Faster CNN is to be increased for dataset with small image size, which will increase the performance. They further work in improving detection accuracy and model structure [9].

[10] Z. Muftuoglu et al showed an important application of data collection, data size, data quantity using COVID-19 radiological images. The most important aspect of the study was the proposal of differential privacy (DP) practice for such applications to be reliable in real-life use cases. Experiments were repeated with DP applied images and Private Aggregation of Teacher ensembles approach was used to ensure privacy assurance. There are some disadvantages of COVID-19 test cases. It shows false positive results and takes few hours for the results. It also uses high-cost equipment's with proper infrastructure. But X-Ray images can overcome these drawbacks.

III. PROBLEM STATEMENT

The idea is to identify COVID-19 and Pneumonia Infected patients using chest X-Ray images with the help of CNN. The identification can be classified as

- Identification of affected lungs
- Identification of Pneumonia affected lungs.
- Identification of COVID-19 affected lungs.
- Identification of Pneumonia and COVID-19 affected lungs

IV. SYSTEM ARCHITECTURE

The system architecture can be represented as software and hardware type. The proposed system uses the software type which involves data collection, data labelling, and object classification models.

A. Software Model

Fig. 2 shows the trained model for pre-processing of images. TensorFlow framework is mostly used in any deep learning techniques in a vast dataset. Kera's libraries are good enough for all imagery processing systems, classification of image partition and detection image modes using low-cost and high throughput imaging technique [11]. TensorFlow libraries are supported in the Google Collaboratory backend.

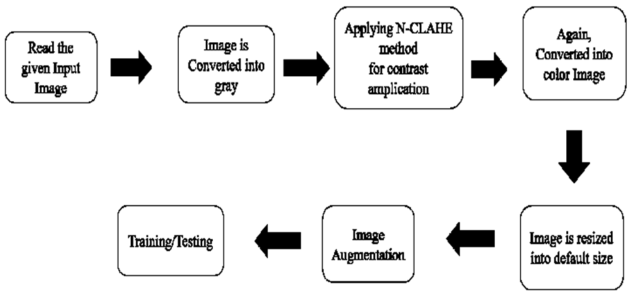

Fig. 2. Pre-processing of Image

The image classification is generally built with Convolution Neural Net Model. Human perception and machine perception of the images is entirely different. A machine can see only numbers of that image in common cases. Every pixel image range is 0 to 225. Finally, a machine can classify with any features or patterns of the image after pre-processing stage.

B. Dataset

Dataset is one of the important aspects of any software related projects. One of the COVID-19 datasets is awarded by the Kaggle community. This database is made by medical physicians. Following Table 1 gives the dataset details.

TABLE 1. DATASET DETAILS

Dataset Names	Number of Images
COVID-19	219
Normal	1341
Pneumonia	1345

The dataset is developed by the Italian Society of Medical and Intervention of Radiology (SIRM) for COVID-19 datasets. Images are extracted from 43 different publications and references are provided in their metadata. Chest X-Ray images database adopt two set images such as Normal, Infected images. All images are Portable Network Graphics (PNG) format which is in 1024 x 1024 pixels resolution. Images are easily converted by 224 x 224 pixels as well as 227 x 227 pixels by Convolution neural network (CNN). A major aim of this dataset is used to produce researcher scholar's work on COVID-19, which can easily find any disorder in the images. A CNN model is used to build three classes to identify Normal lungs, Infected lungs and Pneumonia Infected lungs with Chest X-Ray of images. Fig. 3 shows the samples of all the three types of problems in the chest X- Ray images.

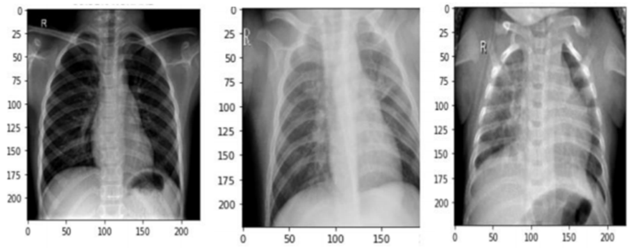

Fig. 3. Normal, COVID-19, Pneumonia Images

C. Preprocessing

Every dataset can be partitioned into training and testing images. Table 2. shows the number of images for training and testing for each comparison type.

TABLE 2. NUMBER OF TRAINING AND TESTING IMAGES FOR NORMAL AND INFECTED LUNGS

Image category	Number of training and Testing
COVID-19 and Pneumonia	199 and 160
Pneumonia and Normal	388 and 160
COVID-19 and Normal	189 and 160

D. Convolution Neural Network

CNN mainly sketches out computer vision and artificial intelligence applications such as image Classification, Text Classification, video recognition, Time series analysis for medical images etc. CNN can support the various pixel arrangement of images. Computer vision algorithm can directly work with the images without any pre-processing steps. Applications using CNN network such as face detection for social media applications in real-time, object detection for self-driving cars, medical image analysis and detection of diseases are very popular ones.

E. Main layers of Convolution Neural Network

Every image can be divided into different layers such as Convolution layer, Max-Pooling layer and Fully Connected layer. The commonly used activation functions are ReLU and Sigmoid which is used in Convolution layer and Fully connected layer respectively. Max-Pooling layer uses both the activation functions. The very initial steps are extracting the features of images and filtering operations that are used to perform the convolution layer functions. The Max-Pooling layer reduces the dimensionality of the image. ReLU activation function in the layer extracts pixel information from images and moved to the next layer. Fully connected layer generates the rectified feature of the images. Finally, flattening process is done to convert two-dimensional vector of the image into single dimensional vector. The flattened values in the fully connected layer help to classify the given images.

The deep learning architecture [12] contains input layer, deep learning layers, output layer. Those deep learning layers are shown in Fig. 4 as three layers of CNN in which first layer is convolution layer following the input layer, and then followed by Max-Pooling and Fully connected layers. The 3x3 kernels in convolution layer will generate 64, 64, 128 and 256 channels in subsequent steps as shown in Fig. 4.

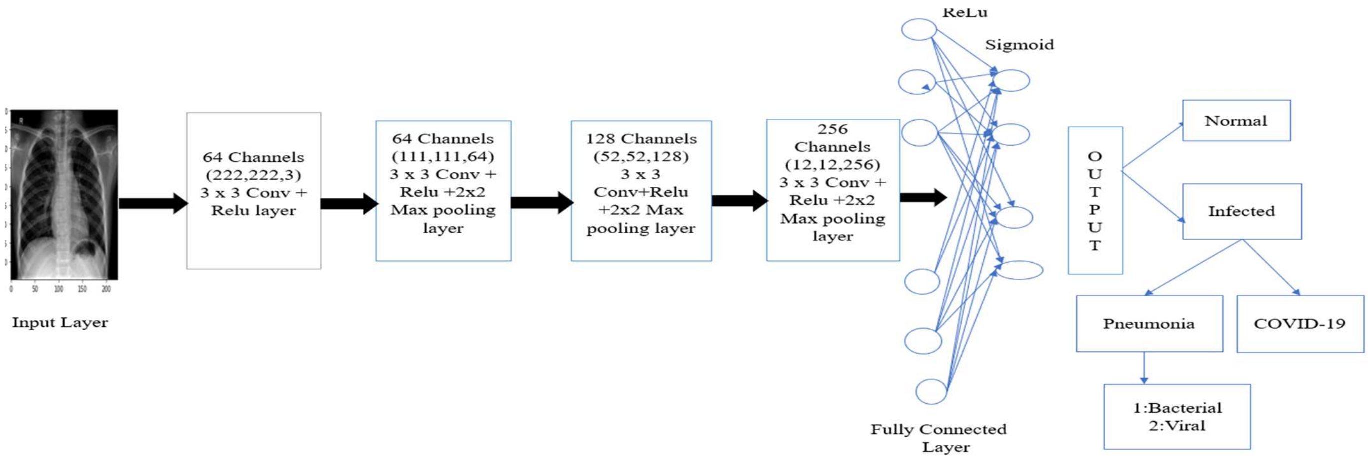

Fig. 4. Architecture of CNN Model

The initial lap is always very small. Hence, the small size of the images can be detected very deeper. The default kernel size is (3,3). The size of input image is taken as (224,224,3). The remaining layers are Convolution and Max-Pooling layer with kernel size as 2x2. The image is classified as Normal and Infected which in turn will classify as COVID-19 or Pneumonia. The same way, if it is classified as Pneumonia, then the subtypes of Pneumonia such as bacterial Pneumonia and viral Pneumonia are found out.

V. RESULTS AND ANALYSIS

Deep Learning process for Convolution Neural Network Model is using the image classification for training and testing over the Normal, and Infected lung X-Ray images. This model with perform the steps for image analysis and finally diagnose the disease.

F. Convolution Neural Network Using Radiography Database

TensorFlow is used to identify the CNN model for the given image set. It will take more than 3 hours for training if not done with GPU. ReLU and sigmoid activation functions are the hyperparameters used for this model. The loss used is binary cross_entrophy, and the optimizer is RMS Prop with an initial learning rate of 0. 001. The total epoch considered is 50 and the batch size is 32. After 50 epochs, the Normal X-Ray image accuracy is 88% and the Infected X-Ray image accuracy is 90%.

Fig. 5 and Fig. 6 show COVID-19 and Pneumonia infected X-Ray images respectively.

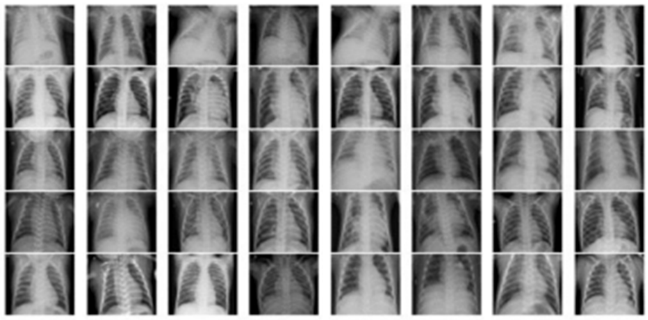

Fig. 5. COVID-19 chest X-Ray Images

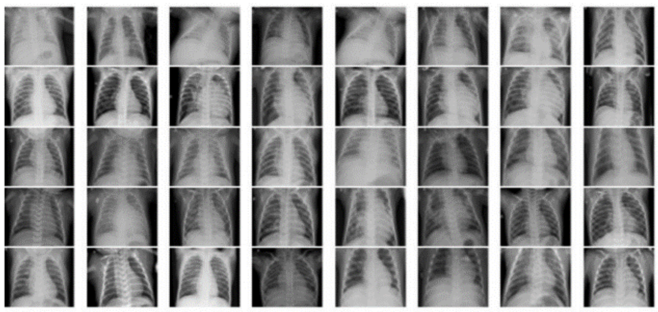

Fig. 6. Pneumonia Infected chest X-Ray Images

Fig. 7 shows the Normal chest X-Ray images. The normal chest X-Ray images have no abnormality of the upper lobe and lower lobe of lungs. It shows the clear picture of lungs.

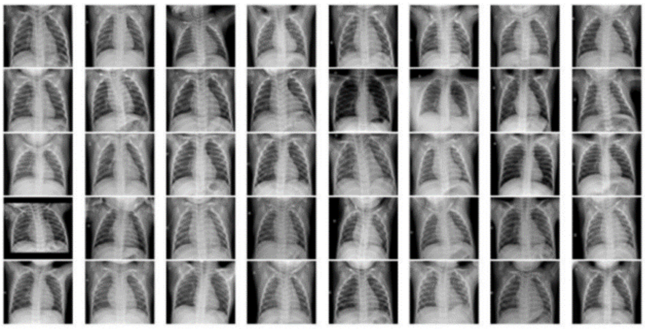

Fig. 7. Normal chest X-Ray Images

Fig. 8 shows the images of lungs affected by COVID-19 and Pneumonia. The percentage shown is for maximum infection. For example, Fig. 8(a) shows the lungs affected by 91.53% of COVID-19 which can be seen in the lower lobe as well as the upper lobe of lungs. The second image Fig. 8(b) shows the lungs affected by 100% of COVID-19 infection. The upper lobe of the COVID-19 Infected lungs is completely invisible. Fig. 8(c) shows the lung image with 52.25% Pneumonia infection and air bronchogram in the lower lobe of the right-side lung. In the same way Fig. 8(d) shows 62.30% of Pneumonia infection in the middle lobe of the lungs.

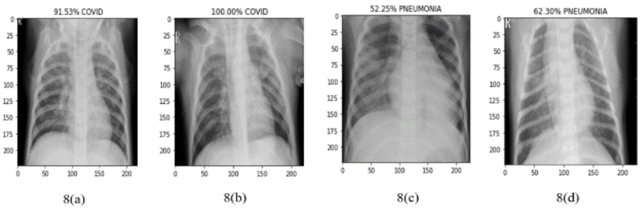

Fig. 8. Images with COVID-19 and Pneumonia infection

Fig.9 shows images with subcategories of Pneumonia infection as bacterial or viral Infected lungs. The percentage shows the severity of the bacterial/ viral infection. Fig 9(a) and 9(c) shows Pneumonia with viral infections of 96.80% and 97.90% respectively. Fig 9(b) and 9(d) shows the Pneumonia with bacterial infection of 70.35% and 81.33% respectively. Fig 9(b) and Fig. 9(d)shows the consolidation of the right middle lobe which is located below the horizontal fissure. Fig 9(a) and Fig. 9(c) shows consolidation of the right upper lobe which is confined inferiorly by the horizontal fissure.

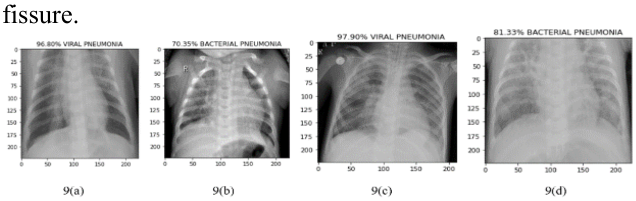

Fig. 9. Images with bacterial/viral category of Pneumonia infection

Fig. 10 shows the images with viral Infected Pneumonia lungs. Both the lobes and the diaphragm of the lungs in the images are very weak compared to Normal Pneumonia images. Fig. 10(a) and 10(b) represent lung images with dense material. Fig. 10(c) and 10(d) represent lung images with low density material such as air

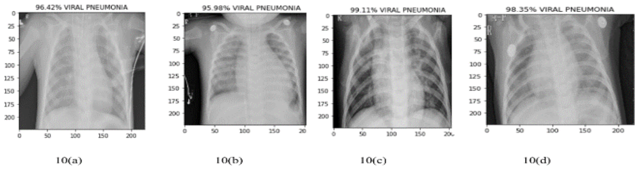

Fig. 10. Lungs with Viral infected Pneumonia X-Ray Images

Fig. 11(a), 11(b), 11(c) and 11(d) show the X-Ray images of Normal lungs. The images are very clear in right and left lungs. The image shows no abnormality at the lungs base. The black color of lung images indicates less dense material, and the white color indicates more denser material. Lungs that appear in the black are spongy with more air content, and in white are very hard with less air content.

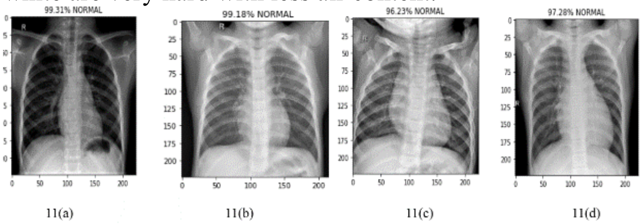

Fig. 11. Classification of Normal Images

The classification of Infected lungs and Normal lungs from X-Ray images is performed with accuracy of 93%. The accuracy in detection of Normal lungs and Infected lungs from X ray images is 93.9%. The accuracy in detection of Normal lungs and COVID-19/Pneumonia Infected lungs is 95/ 80%. And once Pneumonia is detected, the detection accuracy of bacterial infection is 80% and viral infection is 91.46% respectively. Table 3 shows the confusion matrix for Normal and Infected patient's images.

TABLE 3. CONFUSION MATRIX

	Normal	**Infected COVID & Pneumonia**	
Normal (1341)	1271 (TP) 95%	70 (FN) 5%	Sensitivity 95%
Infected (1564)	110 (FP) 7%	1454 (TN) 93%	Specificity 93%
	Positive Prediction 92%	Negative Prediction 95%	Accuracy 93.9%

The proposed method can be compared with Gianchandani et al [13] work in which ensembled model of VGG16, ResNet152V2, Inception ResnetV2, and DenseNet201 models are used to detect Normal lungs, COVID Infected lungs and Pneumonia Infected lungs. The proposed model detects Normal lungs, COVID infected lungs, Pneumonia Infected lungs and even the subtype of Pneumonia can be detected as bacterial Pneumonia or viral Pneumonia Infected lungs. It is necessary to identify pathologies with a similar clinical analytical presentation to define the differential diagnosis [14]. Any successful research work on different diagnostic methods will be an aid to the current pandemic situation.

VI. CONCLUSION

The proposed work aims to detect the Normal lungs, COVID-19 Infected lungs, Pneumonia Infected lungs from the X-Ray images. It's not just detects Pneumonia Infected lungs, but also identifies the subtype of it as bacterial or viral Pneumonia with 80% and 91.46% accuracy, respectively. CNN is the model used to detect Normal and Infected patients like COVID-19 Infected lungs, Pneumonia Infected lungs. The accuracy of detection of COVID -19 is 95 % which a plus about this model and the current requirement of this pandemic with limited cost and computations. This model is more appropriate for medical practitioners, researchers etc. The work can be extended with ensemble of different models to improve the accuracy of detection of different classes but with the cost of time and computation.

REFERENCES

[1] "COVID-19 situation reports",2020. [Online]. Available: Coronavirus Update (Live): 135,479,027 Cases and 2,932,008 Deaths from COVID19 Virus Pandemic - World meter (worldometers.info)

[2] M. J. Horry et al., "COVID-19 Detection Through Transfer Learning Using Multimodal Imaging Data," in IEEE Access, vol.8, pp.149808149824,2020,doi:10.1109/ACCESS.2020.3016780. 2020

[3] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks" 2017 International Conference on Communication and signal Processing (ICCSP), Chennai, India,2017, pp.0588-0592, doi:10.1109/ICCSP.2017.8286426.

[4] R. Anand et al.," Modified Vgg Deep Learning Architecture for COVID19 Classification Using Bio-Medical Images." IOP Conference Series: Materials Science and Engineering, Vol.1084, First International Conference on Circuits Signals, System and Securities (ICCSSS 2020) 11th12th December2020, Tamil Nadu, India. doi:10.1088/1757- 899X/1084/1/012001.

[5] R. Sethi, M. Mehrotra and D. Sethi.," Deep Learning based Diagnosis Recommendations for COVID-19 Using Chest X-Rays Images", 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India,2020, pp1-4, doi:10.1109/ICIRCA48905.2020.918327

[6] S. G., R., V., and Dr. Soman K. P., "Diabetes Detection using Deep Learning Algorithms", ICT Express, Volume 4, Issue 4, December 2018, Pages 243-246

[7] T. Xia, Y.Q. Fu, N. Jin.P. Chazot, P. Angeloy and R Jiang, "AI-enabled Microscopic Blood Analysis for Microfluidic COVID-19Hematology," 2020 5th International Conference on Computational Intelligence and Applications (ICCIA), Beijing, China,2020, pp.98-102, doi:10.1109 /ICCIA49625. 2020.00026.

[8] O.A. Ramwala, H. Mulchandani, P. Dala, M.C. Paunwala and C.N Pinwale," COVID-19 Diagnosis from Chest Radiography Images Using Deep Residual Network", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India,2020.

[9] L. Xu, X. Mao, M. Sun, W. Liu, Y. Wang Y. Tang., "Lung lesions detection from CT images based on the modified Faster R-CNN" 2020 International Conference on Computer, Information and Telecommunication Systems (CITS), Hangzhou, China,2020, pp.1-5, doi: 10.1109/CITS49457.2020.9232611

[10] Z. Muftuolgu, M.A. Kızrak and T. Yildlnm, "Differential Privacy Practice on Diagnosis of COVID-19 Radiology Imaging Using Efficient Net",2020 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Novi Sad, Serbia,2020. pp.1-6, doi:10.1109/INISTA49547.2020.

[11] Gopakumar G, K, H. Babu, D, M., SS, G., and GR, S. Subrahmany, "Cytopathological image analysis using deep-learning networks in microfluidic microscopy", Journal of the Optical Society of America A, vol. 34, no. 1, pp. 111-121, 2017.

[12] Swapna G, Soman K P, Vinayakumar R, "Automated detection of cardiac arrhythmia using deep learning techniques", Procedia Computer Science, Volume 132, 2018, Pages 1192-1201, ISSN 1877-0509,K. Elissa, "Title of paper if known," unpublished.

[13] Gianchandani, N., Jaiswal, A., Singh, D. et al. Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images. J Ambient Intell Human Comput (2020). https://doi.org/10.1007/s12652-020-02669-6

[14] Riaz Ahamed D,Vivek U,Manna Maria, Girish Kumar K. P, Sreekrishnan T. P, Sabarish B Nair, B.S Dhanasekaran, Naveen Mohan, Bharath Prasad, Arun Kumar, Kathyayini V R "Mycoplasma Pneumoniae Pneumonia during COVID Pandemic", International Journal of Health Sciences and Research Vol.11; Issue:1; January 2020