

# Artificial Intelligence for the Detection of Coronavirus Disease (COVID-19) from Chest X-Ray Images

V.N. Sukanya Doddavarapu<sup>1\*</sup>, Giri Babu Kande<sup>2</sup>, B. Prabhakar Rao<sup>3</sup>

<sup>1\*</sup>Assistant Professor, Department of ECE, SACET, Chirala, Andrapradesh, India.

E-mail: sukanyadvn@gmail.com

<sup>2</sup>Professor, ECE Department, VVIT, Nambur, Andrapradesh, India.

E-mail: giribabukande@gmail.com

<sup>3</sup>Professor, ECE Department, JNTUK, Kakinada, Andrapradesh, India.

E-mail: drbrprjntu@gmail.com

**Abstract:** *The COVID-19 pandemic keeps on devastatingly affecting the wellbeing and prosperity of the worldwide populace. To reduce the rapid spread of the COVID-19 virus primary screening of the infected patient repeatedly is a need. Medical imaging is an essential tool for faster diagnosis to fight against the virus. Early diagnosis on chest radiography shows the Coronavirus disease (COVID-19) infected images shows variations from the Normal images. Deep Convolution Neural Networks shows an outstanding performance in the medical image analysis of Computed Tomography (CT) and Chest X-Ray (CXR) images. Therefore, in this paper, we designed a Deep Convolution Neural Network that detects COVID-19 infected samples from Pneumonia and Normal Chest X-Ray (CXR) images. We also construct the dataset that contains 6023 CXR images in which 5368 images are used for training and 655 images are used for testing the model for the three categories such as COVID-19, Normal, and Pneumonia. The proposed model shows outstanding performance with 97.74% accuracy and 96% average F-Score. The results prove that the model can be used for preliminary screening of the COVID-19 infection using radiological Chest X-Ray (CXR) images to accelerate the treatment for the patients under investigation (PUI) who need it most.*

**Keywords:** *Coronavirus Disease (COVID-19), Convolution Neural Network (CNN), Deep Learning (DL), Chest X-Ray (CXR) Images, Pneumonia.*

## 1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an irresistible infection brought about by Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2). The ailment was first recognized in December 2019 in Wuhan, the capital of China's Hubei territory, and has since spread all-inclusive, bringing about the continuous 2019–20 coronavirus pandemic. To control the spread of the infection, screening huge quantities of suspected cases for fitting isolate and treatment measures is a need. The most common symptoms include dry cough, fever, shortness of breath, fatigue, loss of smell, and sputum production. While most cases bring about mellow indications, some advancement to viral pneumonia, and multi-organ failure. The diagnosis of COVID-19 depends on many factors such as clinical symptoms, the epidemiological history of the patient, and pathogenic testing.

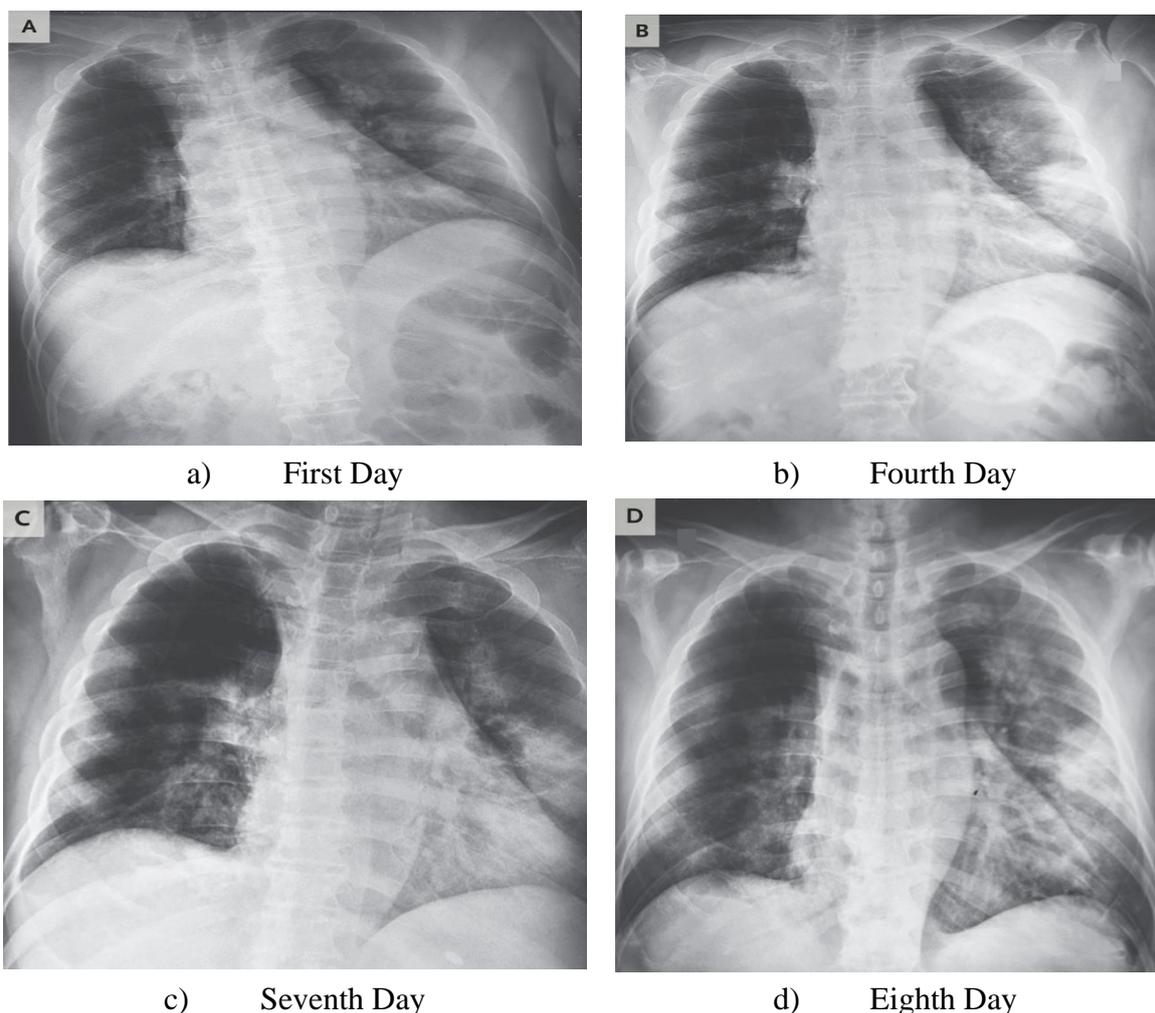
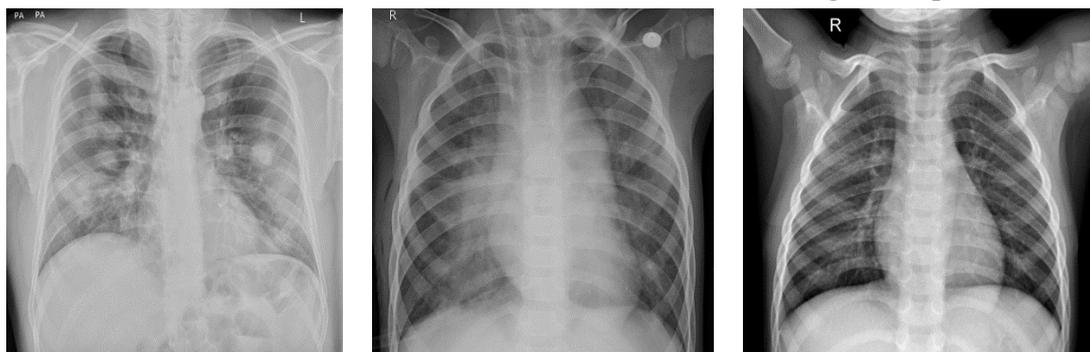


Figure 1 Visual aspect of CXR images of 65 years old male survived Patient under Investigation (PUI)

The clinical characteristics are non-specific for example [1] an asymptomatic family chest scan results in pneumonia where the pathogenic testing shows positive for the virus. The pathogenic testing Real-Time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) turned as a standard method for the diagnosis of the virus. For this pathogenic test, the respiratory specimens such as Broncho alveolar lavage, sputum are collected from the person under test (PUI). The efficiency of this test depends on several factors including availability of kits in the most affected areas, reproduction of kits, specimen collection methods, nucleic acid extraction methods, and amplification system. To fight against COVID-19 keeping time as constraint another tool for the diagnosis of the virus with rapid triaging, availability, and portability, radiography examination such as CXR or CT imaging can be used for primary screening. Figure 1 shows the chest X-Ray in the posterior-anterior (PA) view of a 65 years old male survived Patient Under Investigation (PUI). Initially in figure 1(a) infiltrate in the upper lobe of the left lung is seen. Figure 1(b), 1(c), and 1(d) show progressive infiltrate in the upper lobe of the left lung and turns to consolidation. Currently, all the COVID-19 Chest X-Rays are collected as discussed in section 4.

COVID-19 cases have similar features with Ground Glass Opacity (GGO) in the initial days to the pulmonary Consolidation in the final stages of infection [2]. Also, CT images of

various viral Pneumonia are similar and they exhibit the same characteristics with other classes and different characteristics with in the same class of the lung tissue patterns.



a) COVID-19                      b) Pneumonia                      c) Normal

Figure 2 Examples of the CXR images for the three classes for the present work

Considering all these for diagnosing the CXR or CT images of COVID-19 cases artificial intelligence system with deep convolution neural networks is used. In the present work, we designed a novel deep convolution neural network for the classification of three categories of lung tissues such as COVID-19, Pneumonia, and Normal. Figure 2. Shows the visual aspect of the three categories of lung tissue patterns.

## 2. RELATED WORK

Now-a-days deep learning techniques for medical image classification are widely used because of their outstanding performance in extracting the spatial features of the medical images. Rigorous works are carried for the classification of lung tissue patterns from Chest CT images using deep convolution neural networks. Shin [4] employees deep convolution neural networks such as CifarNet [5], AlexNet [6], and GoogleNet [7] for the Computer-aided Diagnosis of two problems namely Interstitial Lung Diseases (ILD) classification and thoracoabdominal lymph node (LN) detection. Anthimopolous [8] introduces a model with five convolution layers and three dense layers, Wang [9] introduces a Multi-Scale Rotation-Invariant model (MRCNN) model for the classification of five categories of Lung tissue patterns, Joyseeree [10] introduces fusing learned characteristics from Riesz filters and deep CNN for the lung tissue classification.

With the advent of deep learning methods, very few works are carried for the detection of CoronaVirus Disease (COVID-19) infection from Chest X-Ray (CXR) or Computed Tomography (CT) images. Shuai Wang [11] collected the 453 CT images that contain COVID-19 cases and viral Pneumonia and trained the Inception model with 217 images. The accuracy obtained with the internal testing dataset is 82.9% and the external testing dataset is 73.1%. Linda Wang [12] generated a COVIDx dataset with 13,800 CXR images obtained from 13,725 patients and introduced COVID Net, that is available publicly. Tulin [13] designed a DarkNet and attained a classification accuracy of 87% for multiple classes.

### Our Contribution

We designed a novel deep Convolution Neural Network for the detection of COVID-19 infected Chest X-Ray (CXR) images that are available publicly. The flow diagram representing the classification model is as shown in Figure 3. The major contributions are

1. Collection of COVID-19 cases, Normal, and Pneumonia CXR images which are available publicly.

2. Combining the CXR images for the three classes and construct a dataset for training and testing the proposed model.
3. Evaluating the performance of the model in terms of the state-of-the-art.

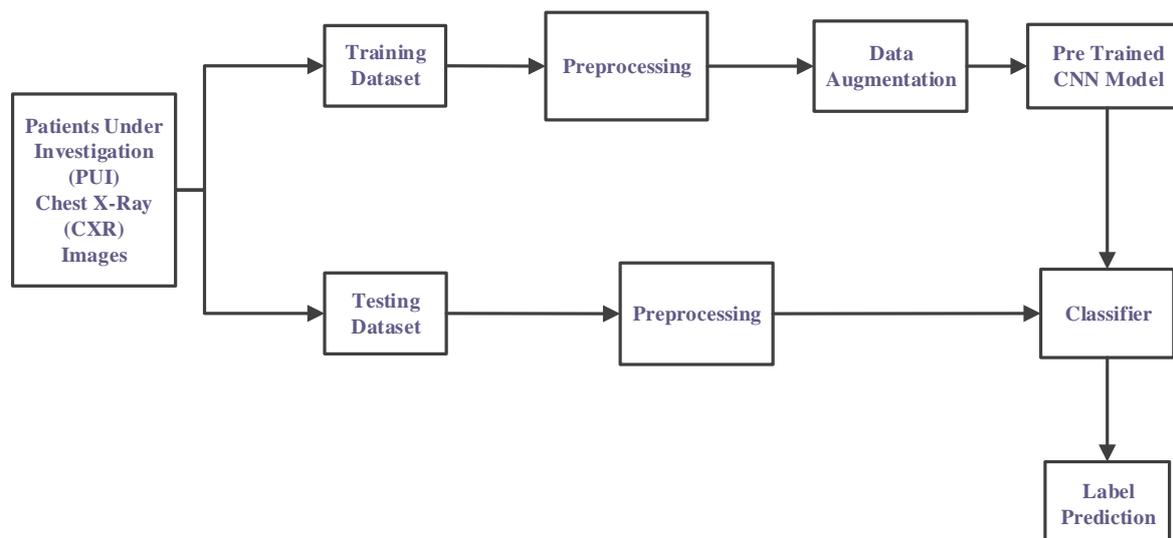


Figure 3 The flow diagram representing the detection of COVID-19 cases from CXR images of PUI

### 3. METHODS

In this paper, we introduced a novel deep convolution neural network for the classification of three categories of lung tissue patterns such as the Normal, Pneumonia, and COVID-19 cases from the CXR images. The overall framework of the proposed model is as shown in Figure4. The network used is pre-trained AlexNet [6] which consists of five convolution layers, three maximum pooling layers, and three Fully Connected layers with approximately sixty million parameters. In the network, non-saturated ReLu activation functions are introduced instead of tanh and sigmoid functions for speeding the process of convergence of the network during training. The input layer in the AlexNet [6] is designed with size  $227 \times 227 \times 3$  i.e color images. So, the input CXR images of the three classes of lung tissue patterns are resized to the size of the input layer of the network. Data Augmentation is done in the networks to increase or decrease the size of the feature maps by translation or rotation operations. The last Fully Connected layer maps the feature maps to the corresponding number of classes namely COVID-19, Normal, and Pneumonia. The training of the Convolution Neural Network (CNN) is done in two steps. First, we define a loss function and second we define an algorithm for optimizing the defined loss function. In the present work, the Stochastic Gradient Descent Method (SGDM) algorithm is used by properly tuning the hyperparameters.

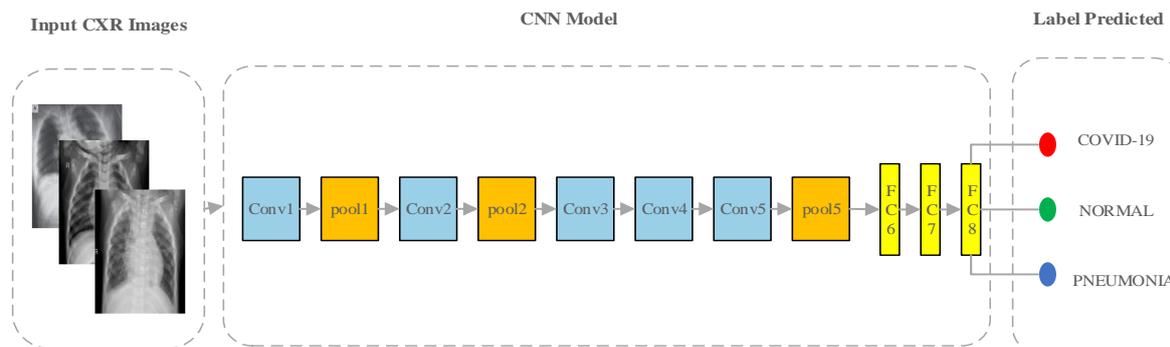


Figure 4 Deep Convolution Neural Network frame work for the detection of COVID-19cases from the CXR images

#### 4. EXPERIMENTAL SETUP

##### Dataset Generation

The pre-trained model is evaluated on the publicly available chest x-ray dataset constructed from the following URL's. COVID-19 cases are available from <https://github.com/ieee8023/covid-chestxray-dataset.git> and <https://github.com/agchung/Figure1-COVID-chestxray-dataset> [3]. The dataset consists of 369 images collected nearly from 191 patients in different views. For the present work, posterior-anterior (PA) view chest x-ray images in jpeg format are considered. Normal and Pneumonia chest x-ray images are downloaded from the following link <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>. The dataset consists of 5863 chest x-ray (CXR) images of two categories namely Normal and Pneumonia. By using the above three links the dataset is constructed with 6023 images.

Table 1: Overview of the Dataset

Type	Normal	Pneumonia	COVID-19	Total
Train	1341	3875	152	5368
Test	234	390	31	655

The Table 1 gives an overview of the dataset constructed and the distribution of training and testing datasets for the classification of the lung tissue patterns. Clearly from the Figure 5, the availability of the COVID-19 cases is less. In the database, the images are named in the format filename\_image format. The filename contains the patient under investigation (PUI) name, year, and location. The age of patients are varying nearly  $55 \pm 15$  and more males gets infected compared to female.

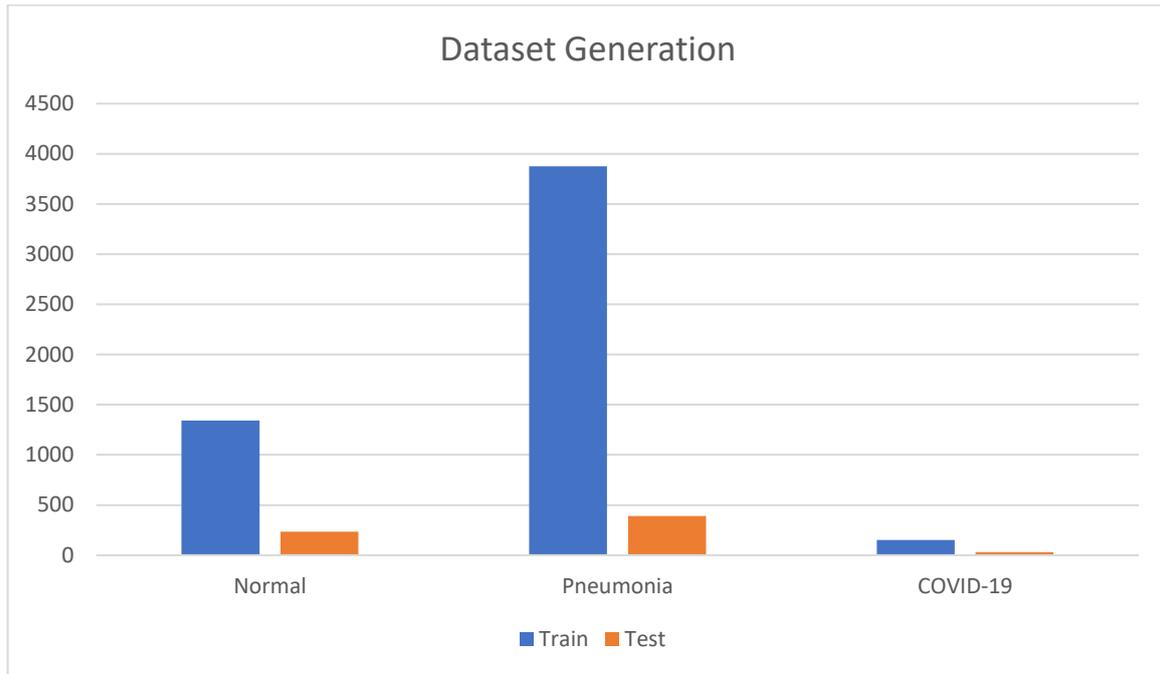


Figure 5 Chest X-Ray (CXR) image distribution for each disease. Left side corresponds to training data and Right side corresponds to testing data

### Implementation

The proposed deep convolution neural network has an input layer size of  $227 \times 227 \times 3$ , as a result, we have to resize the CXR images in the dataset. Data Augmentation is done through reflection and translation operation to decrease the feature maps. The hyperparameters are tuned before training the model. Stochastic Gradient Descent Method (SGDM) algorithm is used for minimizing the loss function, the minimum batch size is set as 15, initial learning rate  $1 \times 10^{-4}$ , and number iterations per epoch as 269. The proposed model is executed on Matlab R2019a software with an I7 processor CPU and 1TB RAM.

### Performance Metrics

The proposed Deep Convolution Neural Network for the classification of the three cases of lung tissue patterns is evaluated with the following metrics for a better understanding of the performance of the model.

- a) **Accuracy:** It is defined as the ratio of the total correctly predicted classes to the total actual classes.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- b) **Sensitivity:** It is defined as the ratio of the correctly predicted cases per class to the total actual cases per class.

$$sensitivity = \frac{TP}{TP+FN} \quad (2)$$

- c) **Specificity:** It defines the ratio of the negative classes that are correctly discriminated.

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

- d) **Positive Prediction Value (PPV):** It is defined as the ratio of the correctly predicted cases per class to the total predicted cases per class.

$$PPV = \frac{TP}{TP+FP} \quad (4)$$

e) **F-Score:** It is used to compare the similarity and diversity of the model and is defined as follows.

$$F - Score = \frac{2TP}{2TP+FP+FN} \quad (5)$$

Where TP- True Positive, TN- True Negative, FP-False Positive, and FN- False Negative values obtained from the confusion matrix.

## 5. RESULTS AND DISCUSSIONS

COVID-19	94.28	0	5.71
Normal	0.24	97.26	2.48
Pneumonia	0.17	1.54	98.27
	COVID-19	Normal	Pneumonia

Figure 6 Confusion Matrix for the proposed model evaluated on the dataset constructed from CXR images

Here we analyze the lung tissue classification results using our proposed model. The figure shows the confusion matrix for the three classes COVID-19, Normal, and Pneumonia. The model shows outstanding performance with an overall accuracy of 97.74%.

The error rate per class is also less than 2.5%. There exists misclassification among COVID-19 cases and Pneumonia, Normal and Pneumonia. 5.7% cases of Pneumonia falsely classified as COVID-19 cases as we know some the viral Pneumonia cases exhibit similar features with the COVID-19 cases. Opacity (GGO) in initial stages turns to consolidation in final stages in COVID-19 cases that Ground Glass. From the confusion matrix, we can conclude COVID-19 infection shows large variation to the pneumonia.

Table 2: Performance of the model in terms of state-of-art

Performance Measure	Normal	Pneumonia	COVID-19
<b>Sensitivity</b>	0.9726	0.9827	0.9428
<b>Specificity</b>	0.9846	0.9725	0.9981
<b>Positive Prediction Value</b>	0.9559	0.9896	0.9167
<b>Negative Prediction Value</b>	0.9905	0.9550	0.9987
<b>F-Score</b>	0.9642	0.9861	0.9295

The Table 2 summarizes the classification results of the proposed model in terms of the performance metrics as mentioned in section4. The model shows the average true positive rates of 96.6%. Clearly, from the table the sensitivity of COVID-19 cases is less compared to the remaining two this is because of the fewer CXR images availability.

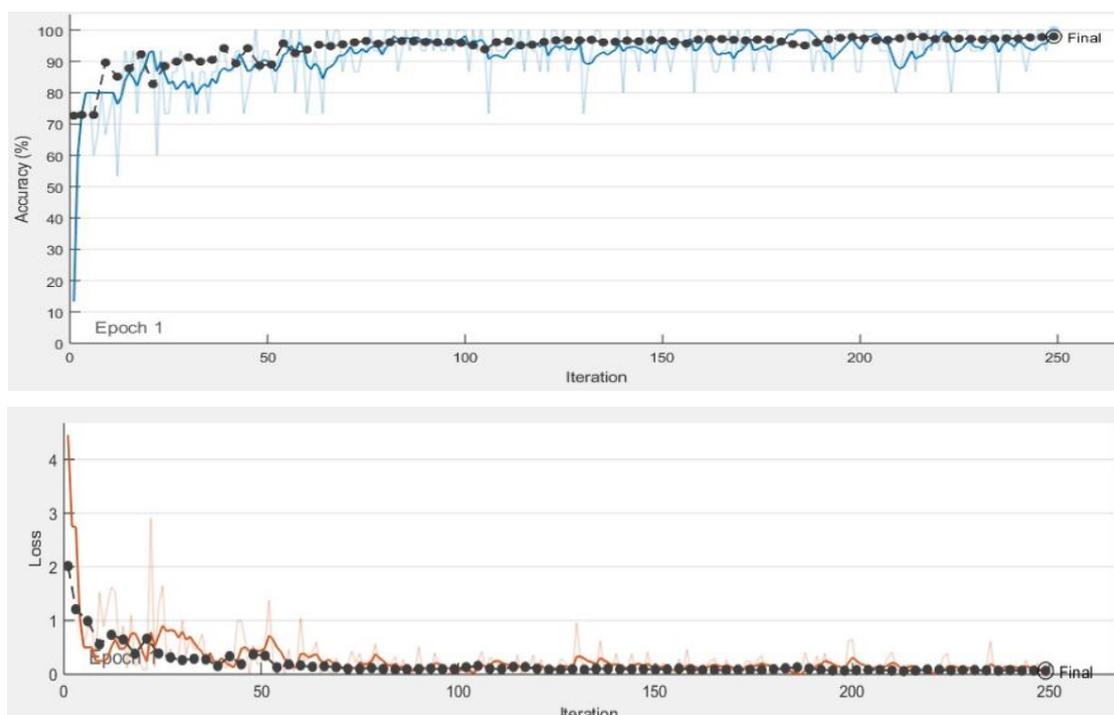


Figure 7 Loss and Accuracy during training the model

The positive predictive value of Normal and Pneumonia are high compared to COVID-19 cases because of the availability of more number of CXR images. Some additional information in terms of Loss and Accuracy curves is as shown in Figure 7. The red descending curve shows the loss curve and the blue ascending curve shows the accuracy curve corresponding to the loss function values during training the proposed model.

## 6. CONCLUSION

In this paper, we designed a deep Convolution Neural Network for detecting COVID-19 cases from Pneumonia and Normal Chest X-Ray (CXR) images. The model used is pre-trained AlexNet with five convolution layers, three maximum pooling layers, and three Fully Connected layers. We also described the dataset constructed for training and testing the model. The dataset consists of 6023 Chest X-Ray (CXR) images in which 5368 images are used for training and 655 images are used for testing. The model shows an outstanding performance in terms of the state-of-the-art when evaluated on the publicly available dataset constructed. The real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) is a standard method for the diagnosis of the Coronavirus disease (COVID-19), we hope the proposed Convolution Neural Network can be used for the primary screening of the infected patients for faster diagnosis since the COVID-19 infection exhibits large variation to the Pneumonia. The sensitivity of the COVID-19 cases is less compared to Normal and Pneumonia because of the availability of fewer samples. In the future, we will extend to more number of images for better classification results.

## 7. REFERENCES

- [1] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. “ Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China”, *Jama* 2020.

- [2] J.P. Cohen, P. Morrison, and L. Dao, "COVID-19 Image Data Collection," 2020.
- [3] Chung et. al. Figure 1 COVID-19 chest x-ray data initiative. <https://github.com/agchung/Figure1-COVID-chestxray-dataset>, 2020.
- [4] H. Shin, H.R. Roth, M. Gao, "Deep Convolution Neural Network for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", *IEEE Trans. Medical imaging*, 2016.
- [5] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.
- [6] A. Krizhevsky, "Learning multiple layers of features from tiny images," in *Master's Thesis. Dept. of Comp. Science, University of Toronto*, 2009.
- [7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conf. on CVPR*, 2015.
- [8] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207-1216, 2016.
- [9] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin, "Multi-Scale Rotation - Invariant Convolutional Neural Networks for Lung Texture Classification," *Journal of Biomedical and Health Informatics*, 2017.
- [10] R. Joyseeree, S. Otalora, H. Muller, "Fusing Learned Representations from Riesz filters and Deep CNN for lung tissue classification", *Medical Image Analysis*, 56 (2019), pp. 172-183.
- [11] S. Wang, B. Kang, J. Ma, X. Zeng, and M. Xiao, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *pp. 1-19*, 2020.
- [12] L. Wang, Z.Q. Lin, and A. Wong, "COVID-Net : A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images."
- [13] T. Ozturk, M. Talo, E. Azra, U. Baran, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, no. April, p. 103792, 2020.

## AUTHOR BIBLIOGRAPHY



**V.N. Sukanya Doddavarapu** is a Ph.D. student of Jawaharlal Nehru Technological University, Kakinada, and working as Assistant Professor of Electronics and Communication Department in St. Ann's College of Engineering and Technology, Chirala. She received her Master's degree in VLSI & ES in the year 2012 from Jawaharlal Nehru Technological University, Kakinada, and Bachelor's degree in Electronics and Communication Engineering in the year 2008 from Acharya Nagarjuna University, A.P, India. Her research interests focus on medical image processing, with specific emphasis on the classification of various lung tissue patterns from CT scan images.



**Dr. Giri Babu Kande** is a Professor and Principal i/c of Electronics and Communication Department in Vasireddy Venkatadri Institute of Engineering and Technology, Guntur. He has teaching experience of about 21 years. His research interests include digital image processing, Machine Learning, and Computer Vision. He received his Ph.D. degree in Digital Image Processing from Jawaharlal Nehru Technological University,

Hyderabad in the year 2010. About 2 scholars were awarded Ph.D. under his tutelage. He is a member of various professional chapters and published many research papers in various SCI journals, national and international conferences.



**Dr.B. Prabhakara Rao** was born in the year 1955 at Buchireddypalem, Nellore Dist., A.P India. He obtained B.Tech & M. Tech from S.V. University, Tirupathi with Specializations in Electronics and Communications Engineering (ECE), Electronic Instrumentation, and Communications Systems in the years 1979 and 1981 respectively. Dr. Rao received Doctorate from Indian Institute of Science, Bangalore in the area of Sonar Signal processing in the year 1995. Dr Rao joined as Assistant Professor in the JNT University in the Year 1982 and became Professor in ECE Dept. in the year 2003. He worked as Director (Institute of Science & Technology), Rector, and Vice-Chancellor i/c at JNT University, Kakinada. His current areas of research interest are Optical, Wireless, Microwave and Digital Communications & Imageprocessing. About 35 scholars were awarded Ph.D. under his tutelage. Rao published many research papers in various SCI journals and National and International conferences and is a fellow of many Professional Organizations. Presently he is Vice president of IETE, New Delhi.