Diagnosis of COVID-19 and Pneumonia using Depthwise Separable Convolutional Neural Network

Shrabana Saha
Dept. of EE
Future Inst. of Engg. and Mgmt.
Kolkata, India
shrabanasaha97@gmail.com

Rajarshi Bhadra
Dept. of EE
Future Inst. of Engg. and Mgmt.
Kolkata, India
rajarshibhadra11@gmail.com

Subhajit Kar
Dept. of EEE
Inst. of Engg. and Mgmt.
Kolkata, India
sksubhajit@gmail.com

2021 IEEE Bombay Section Signature Conference (IBSSC) | 978-1-6654-1758-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IBSSC53889.2021.9673215

***Abstract*—COVID-19 disease is a consequence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus that came to light as an epidemic over the planet. The long-established diagnostic systems are confronting difficulties in identifying the virus expeditiously in the initial stages. In these circumstances, chest X-ray scans can be beneficial for the identification of COVID-19 as well as pneumonia. On that account, in this research, a deep convolution neural network having depthwise separable convolutions has been put forward to look over the chest X-ray scans for identifying COVID-19 and pneumonia precisely. The propounded model with only 0.18 million parameters has been employed on various standard datasets and performs significantly faster than other state-of-the-art models and the exploratory results explain the potency of the propounded approach.**

***Keywords*—COVID-19 diagnosis, Pneumonia detection, Chest X-ray scans, Deep Convolution Neural Network (DCNN), Depthwise separable convolutions, L2 Regularization**

I. Introduction

The outbreak of COVID-19 escalated instantaneously on every side of the globe by early 2020. World Health Organisation (WHO) asserted coronavirus as COVID-19 and announced it as an infestation by January 2020 [1]. Up until May 6, 2021, the disease has been outstretched in the middle of 21 crore individuals in the world and 3 crore human beings in India, and the counter keeps going up exponentially. Conventional testing methodologies use ribonucleic acid (RNA) from nasal and throat swabs to verify the availability of SARS-CoV-2 in the human body. This evaluation methodology, known as real-time reverse transcription-polymerase chain reaction (RT-PCR) has been broadly used as a decisive testing mechanism for morbific t est [2]. A t p resent, t he g enome s equence of SARS-CoV-2 has interchanged because of diverse mutations. In consequence, RT-PCR tests result in large numbers of false classifications [3].

Chest X-rays (CXR) and Computed Tomographical scans (CT scans) shows considerable contributions in error-free diagnosis of pulmonary diseases [4]–[11]. Chen et al. showed that chest X-ray scans take an important part in examining patients who are clinically suspected to be affected by the virus, especially those whose initial RT-PCR test results are negative [4]. Minaee et al. used 4 CNN models i.e. ResNet18, ResNet50, SqueezeNet, and DenseNet-121 for diagnosis of COVID-19 from chest X-ray scans. These 4 CNN models resulted good classification accuracy [5]. Jain et al. showed a comparison between InceptionV3, Xception, and ResNeXT models on chest X-ray scans. Xception resulted accuracy of 97.9% [6]. Rawat et al. used InceptionV3, MobileNet, Xception, and DenseNet121 out of which MobileNet resulted in good classification accuracy [7]. Choudhuri et al. proposed a classical convolutional neural network that is compared with VGGNet, AlexNet, GoogleNet, DenseNet, SqueezeNet, ResNet. Out of these models, VGG16 scored 98.3% accuracy [8]. Priyatharshini et al. put forward a method of COVID-19 detection from chest X-ray scans where image data has been segmented using U-Net and the segmented data has been trained using InceptionV3 with a good classification accuracy of 97% [9]. Narin et al. proposed five pre-trained CNN models i.e. ResNet50, ResNet101, ResNet152, InceptionV3 and Inception-ResNetV2 and implemented on 3 datasets. This study resulted maximum 99.7% test accuracy [10]. Bhadra et al. came up with a CNN model with 0.9 million parameters which resulted in 99.1% blind test accuracy (on 399 samples) in the identification of pneumonia and coronavirus from chest X-ray scans [11].

The techniques portrayed in the research [5]–[11] carry out the diagnosis of COVID-19 using chest X-ray scans with the help of several compound CNN architectures. These architectures having millions of parameters may result satisfactorily in the classification of diseases but they demand prolonged computation time. On the contrary, a deep CNN architecture with depthwise separable convolutions consisting of only 0.18 million parameters has been propounded in this research. Apart from the fact that it requires shorter computation time but also outcomes in great classification accuracy. L2 regularization has been implemented on the proposed model to gain more control over overfit optimization. The propounded method has been trained on a dataset consisting of CXR scans of COVID-19, pneumonia, and healthy cases. The blind testing has been assessed on 399 CXR scans. Experimental outcomes epitomize that the propounded model executes decisively in the diagnosis of COVID-19.

In the remaining paper, section II narrates the proposed method. The datasets have been detailed in Section III. Section IV explains experimental outcomes. Section V portrays the

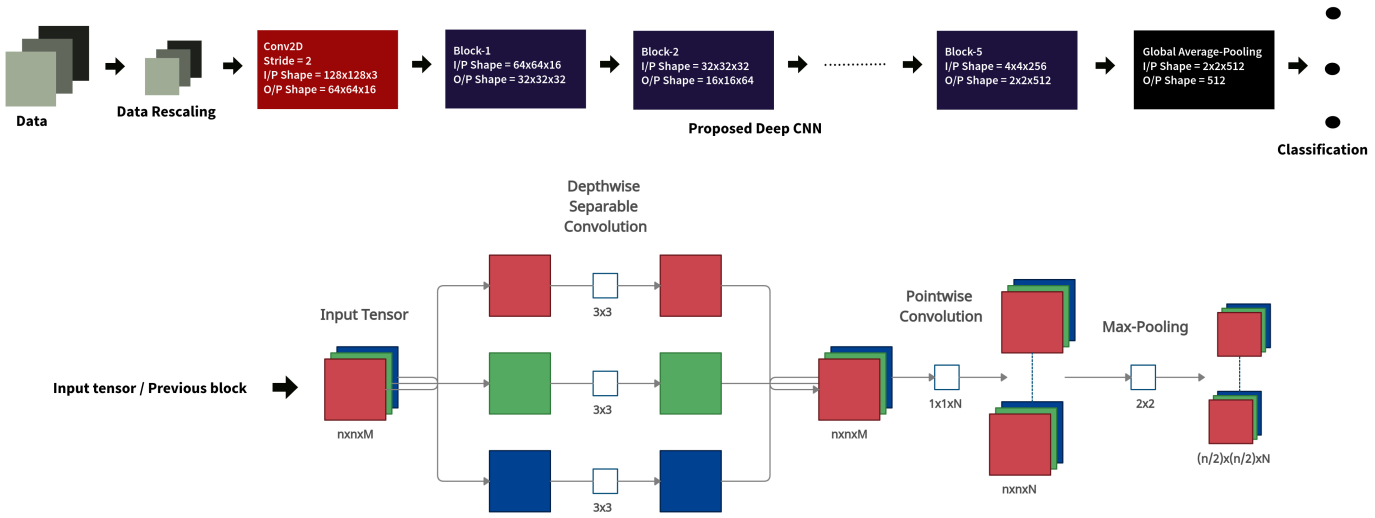


Fig. 1 Block diagram of the proposed approach.

distinctiveness of the experimentation and lastly, the paper is concluded by section VI.

II. Proposed Approach

Primarily, the unevenly shaped CXR scans have been re-sized to retain the uniformity of the images in the dataset. These re-scaled CXR scans have been fed to the propounded deep CNN model. Softmax classifier classifies the CXR scans into their respective classes. Fig. 1 depicts the block diagram of the proposed method.

A. Proposed Deep Convolution Neural Network

Table I details the propounded deep convolution neural network architecture for the diagnosis of CXR scans. The model is having 5 blocks, where each block consists of one depthwise separable convolution layer where, a single 3x3 convolution filter is applied for each input channel, one pointwise convolution layer where n number of 1x1 filters are applied to increase the number of output channels and one maxpool layer for spatial dimension reduction. Using depthwise separable convolutions, the computation cost and the parameters of the model can be significantly reduced which benefits in reduced computation time and also ensures perfect fitting on data distribution as well.

Considering n_H and n_W are the height and width of input image respectively, f_H and f_W are the height and width of filter, M and N are numbers of input and output channels respectively. The computation cost for depthwise separable convolution combined with pointwise convolution is,

$$\text{Cost}_1 = (n_H \cdot n_W \cdot M) \cdot (f_H \cdot f_W) + M \cdot N \cdot f_H \cdot f_W \quad (1)$$

The computation cost for standard convolution is,

$$\text{Cost}_2 = n_H \cdot n_W \cdot M \cdot N \cdot f_H \cdot f_W \quad (2)$$

Hence, the reduction of computation using depthwise separable and pointwise convolution is

$$\frac{\text{Cost}_1}{\text{Cost}_2} = \frac{1}{N} + \frac{1}{n_H \cdot n_W} \quad (3)$$

Apart from the fact that it uses lesser computations which results in significantly lesser computation time but also uses relatively very few parameters which diminish over-fitting issues. After extracting the features using these multi-dimensional layers, the global average-pool with a filter of size 2x2 converts the 3-D tensor to a 1-D feature vector having 512 elements. Thereafter, an output layer having a softmax classifier is used for classifying CXR scans into 3 respective classes i.e. COVID-19, Pneumonia, and Healthy. LeakyReLU [12] activation has been used in hidden layers which are defined by,

$$g(Z^{[l]}) = \begin{cases} \alpha Z^{[l]} & \text{for } Z^{[l]} < 0 \\ Z^{[l]} & \text{for } Z^{[l]} \geq 0 \end{cases} \quad (4)$$

where,

$$Z^{[l]} = W^{T[l]} g(Z^{[l-1]}) + b^{[l]} \quad (5)$$

where, $W^{[l]}$ and $b^{[l]}$ are weight matrix and bias vector assigned to the neurons of l^{th} layer. $g(Z^{[l-1]})$ is activation function of the previous layer i.e. $(l-1)^{th}$ layer. $g(Z^{[l]})$ is output of the activation of l^th layer and $\alpha = 0.3$ i.e. is the gradient assigned to $Z^{[l]}$ which restrains dying ReLU while backpropagation [13].

SoftMax classifier is described as,

$$\hat{y}_i = \frac{e^{Z_i^{[l]}}}{\sum_{j=1}^{k} e^{Z_j^{[l]}}} \quad (6)$$

where, $Z_i^{[l]}$ is the pre-activation of i^{th} node of output layer (l), k is total nodes in output layer (l), $\hat{y}_i$ symbolizes predicted probability of a neuron which also corresponds to a class.

TABLE I SUMMARY OF PROPOSED CNN ARCHITECTURE

Block	Layer Type	Stride	Kernel-size	Output Size	Parameters
Input	Input Layer	2	-	128x128x3	0
	Conv2D	2	3x3	64x64x16	448
	Batch-Normalization	-	-	-	64
Block-1	Depthwise-Conv2D	1	3x3	64x64x16	160
	Batch-Normalization	-	-	-	64
	Pointwise-Conv2D	1	1x1	64x64x32	544
	Batch-Normalization	-	-	-	128
	Max-pooling2D	2	2x2	32x32x32	0
Block-2	Depthwise-Conv2D	1	3x3	32x32x32	320
	Batch-Normalization	-	-	-	128
	Pointwise-Conv2D	1	1x1	32x32x64	2112
	Batch-Normalization	-	-	-	256
	Max-pooling2D	2	2x2	16x16x64	0
Block-3	Depthwise-Conv2D	1	3x3	16x16x64	640
	Batch-Normalization	-	-	-	256
	Pointwise-Conv2D	1	1x1	16x16x128	8320
	Batch-Normalization	-	-	-	512
	Max-pooling2D	2	2x2	8x8x128	0
Block-4	Depthwise-Conv2D	1	3x3	8x8x128	1280
	Batch-Normalization	-	-	-	512
	Pointwise-Conv2D	1	1x1	8x8x256	33024
	Batch-Normalization	-	-	-	1024
	Max-pooling2D	2	2x2	4x4x256	0
	Dropout (20%)	-	-	-	0
Block-5	Depthwise-Conv2D	1	3x3	4x4x256	2560
	Batch-Normalization	-	-	-	1024
	Pointwise-Conv2D	1	1x1	4x4x512	131584
	Batch-Normalization	-	-	-	2048
	Max-pooling2D	2	2x2	2x2x512	0
3-D to 1-D	Global-Average-Pooling2D	-	2x2	512	0
Output	Output Layer	-	-	3	1539
Total Parameters:	**1,88,547**				

To evaluate cost while training, categorical cross-entropy loss has been used. The overall cost function $J(w, b)$ can be described as:

$$J(w, b) = \sum_{i=1}^{M} L(\hat{y}_i, y_i) \tag{7}$$

where, categorical cross-entropy loss is defined as:

$$L(\hat{y}_i, y_i) = -y_i log(\hat{y}_i) \tag{8}$$

where, M symbolizes number of training examples, y_i denotes true output, $\hat{y}_i$ denotes predicted output.

$L2$ regularization has been used in all convolution and dense layers to get rid of overfitting issues as well. $L2$ regularization can be defined by:

$$J(w, b) = \sum_{i=1}^{M} L(\hat{y}_i, y_i) + \frac{\lambda}{M} ||W||_2^2 \tag{9}$$

where, $L(\hat{y}_i, y_i)$ denotes categorical cross-entropy loss, M = number of training examples; $\lambda = 0.00001$ i.e. regularization factor and $||W||_2$ is $L2$ norm of weight matrix W.

The loss function has been optimized using Adam optimizer [14] is defined in Algorithm 1.

III. Dataset used

Fig 2 depicts the dataset using which the experimentation has been. The database consists of chest X-ray scans collected from patients suffering from COVID-19 and pneumonia and from health cases also. From [15] (180 Posterior-Anterior (PA) CXR scans), [16] (238 CXR scans) and [17] (912 augmented CXR scans) COVID-19 affected CXR scans have been cumulated. From [18], 4,273 pneumonia affected CXR scans and 1,583 CXR scans of healthy people have been collected. The database is consists of 3 classes. 1,330 images from each class have been collected to create a balanced dataset.

IV. Exploratory Result

A. Exploratory setup

The experiment has been performed on 3,999 CXR scans, out of which 3,393 CXR scans have been used to train the propounded deep CNN model, 198 CXR scans have been used to validate the model and 399 CXR scans have been used to evaluate the blind test classification performance of the model. The propounded algorithm for training has been embellished in Algorithm 2. Tensorflow 2 has been used to create the proposed CNN model on 6 cores and 12 threads CPU, NVIDIA RTX 2060 GPU having 6GB of DDR6 VRAM, 8GBX2 DDR4 memory clocked at 3200mHz and 240 GB

Algorithm 1: Adaptive Moment Estimation (ADAM)

- Calculate moment estimates
  $v_{dw} = \beta_1 v_{dw} + (1-\beta_1)dw$
  $v_{db} = \beta_1 v_{db} + (1-\beta_1)db$
  $s_{dw} = \beta_2 s_{dw} + (1-\beta_2)dw^2$
  $s_{db} = \beta_2 s_{db} + (1-\beta_2)db^2$
- Apply bias correction
  $v_{dw}^{corrected} = \frac{v_{dw}}{(1-\beta_1^t)}$
  $v_{db}^{corrected} = \frac{v_{db}}{(1-\beta_1^t)}$
  $s_{dw}^{corrected} = \frac{s_{dw}}{(1-\beta_2^t)}$
  $s_{db}^{corrected} = \frac{s_{db}}{(1-\beta_2^t)}$
- Update parameters
  $w_{t+1} = w_t - \alpha \frac{v_{dw}^{corrected}}{\sqrt{s_{dw}^{corrected}}+\epsilon}$
  $b_{t+1} = b_t - \alpha \frac{v_{db}^{corrected}}{\sqrt{s_{db}^{corrected}}+\epsilon}$
- where,
  - $v_{dw}, v_{db}, s_{dw}, s_{db}$ are moving averages of gradient and squared gradient,
  - β_1 and β_2 are moving average parameters, where $\beta_1 = 0.9$, $\beta_2 = 0.999$,
  - w_t, b_t and w_{t+1} and b_{t+1} are initialized and updated parameters respectively,
  - α is initial learning rate i.e. 0.001 with staircase decay where steps = 10,
  - ϵ is a very small number (10^{-7}) which prevents division by 0.

(a)

(b)

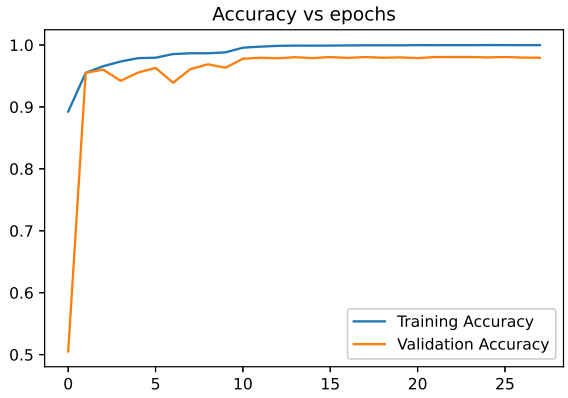
(c)

Fig. 2 Chest X-ray scans (a) COVID-19 (b) Pneumonia (c) Healthy

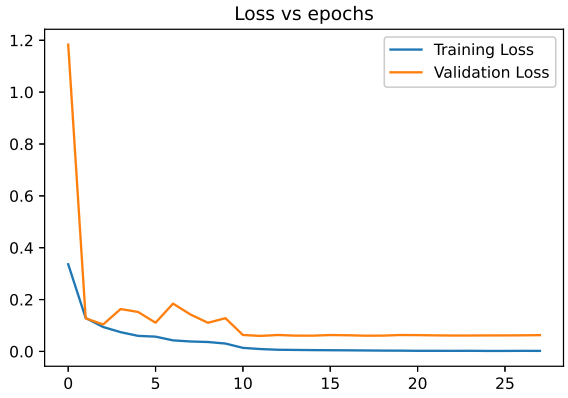
NVMe SSD. A Mini-batch size of 8 has been chosen based on compute capability of GPU.

Algorithm 2: Proposed algorithm for training

- Create a deep CNN architecture using depthwise separable convolution layers
- Optimize using Adam Optimizer.
- Implement staircase learning rate decay.
- Use early stopping where training stops if training loss does not improve for 10 consecutive epochs
- Save the model where training loss is properly optimized.



(a)



(b)

Fig. 3 (a),(b) Training and Validation Characteristics

B. Survey of Results

In Fig. 3 the training and validation characteristics curves have been shown. This proposed technique has achieved a training accuracy of 99.9±0.1%, validation accuracy of 98.3±0.9%, and blind test accuracy of 99.8±0.2%. The classification results reassured by evaluating precision, recall, and F1 score have been shown in Table III. In addition, the confusion matrix, shown in Fig. 4, highlights the amount of accurately classified samples for each class. Thus, the implementation of the proposed approach is satisfactory in the context of coronavirus detection.

TABLE II CLASSIFICATION ACCURACIES

10-fold Cross Validation Accuracies				
Training	**Validation**	**Testing**	**Sensitivity**	**Specificity**
99.9±0.1%	98.3±0.9%	99.8±0.2%	99.25%	100%

TABLE III CLASS-WISE CLASSIFICATION REPORT

Class	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
COVID-19	1.00	0.99	1.00	1.00
Healthy	0.99	1.00	1.00	1.00
Pneumonia	1.00	1.00	1.00	1.00

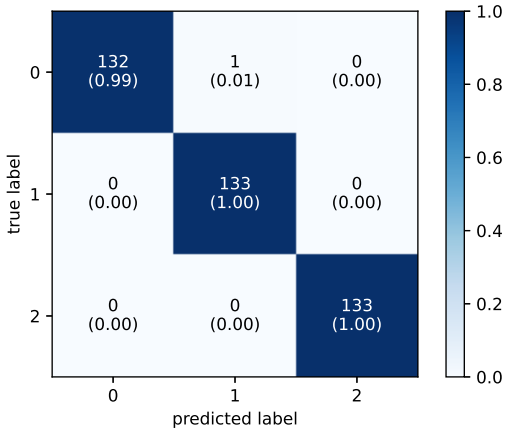


Fig. 4 Confusion Matrix (0) Healthy (1) Pneumonia (2) COVID-19

TABLE IV MODEL COMPARISON

Model	**No. of Parameters**	**Blind Test Accuracy**	**Training Time (seconds)**
VGG16 [19]	14 million	96.6%	3075
VGG19 [19]	20 million	96.5%	2958
MobileNet [20]	3 million	97.1%	888
InceptionV3 [21]	22 million	90.2%	2305
ResNet50 [22]	24 million	96.2%	2901
DenseNet121 [23]	7 million	97.8%	2881
Proposed	**0.18 million**	**99.8%**	**174**

V. Discussion

Different transfer learning models like MobileNet, InceptionV3, ResNet50, VGG16, VGG19, DenseNet121, and Xception have also been trained using the dataset. The comparison has been elaborated in Table IV. The propounded model has fewer parameters for which it took lesser time to get trained, and resulted in good blind test accuracy.

VI. Conclusion

In this research, a deep CNN architecture with only 0.18 million parameters has been put forward to diagnose coronavirus and pneumonia from chest X-ray scans which not only takes lesser training time but also results in good classification accuracy. In this proposed approach, blind testing has been performed to increase the robustness of the system. 99.8% blind test accuracy along with 99.25% sensitivity and 100.00% specificity has been procured by the proposed approach. Therefore, the proposed approach can be used to aid physicians with their diagnostic prediction of coronavirus.

References

[1] A. F. Siddiqui, M. Wiederkehr, L. Rozanova, and A. Flahault, "Situation of india in the covid-19 pandemic: India's initial pandemic experience", *Int. J. Environ. Res. Public Health*, vol. 17, no. 23, p. 8994, 2020.

[2] S. L. Emery, D. D. Erdman, M. D. Bowen, B. R. Newton, J. M. Winchell, R. F. Meyer, S. Tong, B. T. Cook, B. P. Holloway, K. A. McCaustland *et al.*, "Real-time reverse transcription–polymerase chain reaction assay for sars-associated coronavirus", *Emerging infectious diseases*, vol. 10, no. 2, p. 311, 2004.

[3] A. Gupta-Wright, C. K. Macleod, J. Barrett, S. A. Filson, T. Corrah, V. Parris, G. Sandhu, M. Harris, R. Tennant, N. Vaid *et al.*, "False-negative RT-PCR for covid-19 and a diagnostic risk score: a retrospective cohort study among patients admitted to hospital", *BMJ open*, vol. 11, no. 2, p. e047110, 2021.

[4] A. Chen, Dandan Jiang, Xinqing Hong, Yong Wen, Zhihui Wei, Shuquan Peng, Guangming Wei, Xinhua,*et al.*, ,"Can chest CT features distinguish patients with negative from those with positive initial RT-PCR results for coronavirus disease (COVID-19)?", *American Journal of Roentgenology*, vol. 216 no. 1, p. 66-70, 2021.

[5] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest X-ray images using deep transfer learning", *Medical image analysis*, vol. 65, p. 101794, 2020.

[6] R. Jain, M. Gupta, S. Taneja, and D. J. Hemanth, "Deep learning based detection and analysis of covid-19 on chest X-ray images", *Applied Intelligence*, vol. 51, no. 3, pp. 1690–1700, 2021.

[7] R. M. Rawat, S. Garg, N. Jain, and G. Gupta, "Covid-19 detection using convolutional neural network architectures based upon chest X-rays images", in *2021 Proc. 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, 2021, pp. 1070–1074.

[8] R. Choudhuri and A. Paul, "Multi class image classification for detection of diseases using chest x ray images", in *2021 8th Int. Conf. Comput. Sust. Global Dev. (INDIACom)*, 2021, pp. 769–773.

[9] R. Priyatharshini, R. A. S. Aswath, M. N. Sreenidhi, S. S. Joshi, and R. Dhandapani, "An efficient approach for automatic detection of covid-19 using transfer learning from chest X-ray images", in *2021 3rd Int. Conf. Signal Process. Commun.(ICPSC)*, 2021, pp. 741–746.

[10] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks", *Pattern Analysis and Applications*, pp. 1–14, 2021.

[11] R. Bhadra and S. Kar, "Covid detection from cxr scans using deep multi-layered cnn", in *2020 IEEE Bombay Section Signature Conference (IBSSC)*, 2020, pp. 214–218.

[12] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network", *arXiv preprint arXiv:1505.00853*, 2015.

[13] L. Lu, "Dying relu and initialization: Theory and numerical examples", *Communications in Computational Physics*, vol. 28, no. 5, p. 1671–1706, Jun 2020. [Online]. Available: http://dx.doi.org/10.4208/cicp.OA-2020-0165

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[15] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection", 2020.

[16] "Actualmed covid-19 chest X-ray dataset initiative", https://github.com/agchung/Actualmed-COVID-chestxray-dataset, 2020.

[17] S. Alqudah, Ali Mohammad; Qazan, "Augmented Covid-19 X-ray images dataset", *Mendeley Data, V4*, 2020.

[18] D. Kermany, K. Zhang, M. Goldbaum *et al.*, "Labeled optical coherence tomography (oct) and chest X-ray images for classification", *Mendeley data*, vol. 2, no. 2, 2018.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*, 2014.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint arXiv:1704.04861*, 2017.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* , 2016, pp. 2818–2826.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.