A case study on the application of Deep learning for Identifying Covid-19 from chest radiography images

Shweta Dahale, Biswajeet Rout, Sai Munikoti, William Hsu

I. Introduction

Covid-19 or SARS-CoV-2 is an ongoing pandemic the world is going through for more than a year. Although, the vaccines have been developed and started to reach common people, the risk associated with the disease has not decreased due to mutation of the virus. To date, more than three variants (strains) of the covid-19 have been spotted by medical researchers [1]. Under this situation, a critical and necessary step in mitigating the pandemic is the screening procedure. Presently, the primary screening method used for detecting covid-19 is real-time reverse transcription-polymerase chain reaction (RT-PCR). However, this method takes several hours to 2 days to generate a final report and because of high demand, the testing kits are not available in sufficient numbers at many places especially in rural areas of developing countries. In addition, manual interpretation by radiologists is tedious and easy to be influenced by inter-expert and intra-expert factors (such as fatigue, emotion, etc.). An alternative method to RT-PCR is based on chest radiography images. Researchers found that the lungs of patients with covid-19 symptoms have some visual marks like ground-glass opacities—hazy darkened spots that can differentiate covid-19 infected patients from non covid-19 infected ones [2], [3]. The success of Machine learning (ML) based techniques in the automatic diagnosis of multiple prior diseases have motivated researchers to extend these approaches to covid-19. In this regard, several works have been published where deep learning-based methods have demonstrated acceptable accuracy in identifying covid-19 positive cases from radiology images [4]–[6].

In this study, we are exploring various deep learning methods to identify COVID-19 from chest radiography images. In this regard, we will explore various CNN architectures and compare their performance with a state of the art algorithm.

Background: Scientists from all over the world are working day and night to develop highly accurate and reliable methods for the detection and management of Covid-19 disease. In the recent past, deep learning models for covid-19 detection [COVID-Net] were proposed by [4]. These authors have used radiology images. Convolutional neural networks along with SVM classifiers were used by [5] for covid-19 classification. CoroNet based on a pre-trained CNN model was developed in [6] which had 4 classification labels - covid-19, pneumonia bacterial, pneumonia viral, and normal. Hemdan et al [7] used various deep learning models to diagnose COVID-19 from chest X-ray images and proposed a COVIDX-Net model comprising seven CNN models. The CovidGAN approach proposed by [8] was based on a generative adversarial network (GAN). Regarding graph-based models, the authors in [9], have used GNN for image classification by exploring various representations of images.

II. Dataset

Although COVID-19 has impacted a vast number of people, its data in the form of chest radiography images are not readily available from a single source. Therefore, data is collected from multiple sources. Covid positive images are obtained from [10], where data is open-sourced in Github. The authors have compiled the images from various sources that include the Radiology Society of North America, Radiopedia, etc. On the other hand, the images for healthy and Pneumonia cases are obtained from Kaggle dataset [11]. In total, the dataset consists of 320 covid positive images, 1203 healthy, 1590 Pneumonia cases. Pneumonia is further segregated into two types, i.e., 660 bacterial and 931 viral cases. As covid positive images are relatively less in number, random over sampling of the minority classes are needed to overcome the unbalanced data problem. Hence, the final count of images settles at 410 covid positive, 410 healthy, 430 Pneumonia bacteria, and 427 Pneumonia viral cases.

The images are pre-processed before feeding into the network. Since data is collected from different sources, there is a non-uniformity among image sizes. Therefore, all the images resized to 224×224 pixels with a resolution of 72 dpi. Figure 1 displays samples of chest radiography images from each class of the infections.

III. Methodology

A. Baseline/State of the art

The baseline algorithm of classifying covid-19 from other infections is CoroNet [6] which is inspired by the Xception network. Primarily, the feature extraction base of the network is the same as that of Xception [12]. This architecture is based on depthwise separable convolution layers i.e the mapping of the spatial correlations in the feature maps of the CNN is entirely decoupled. Depthwise Separable Convolution replaces classic $n \times n \times k$ convolution operation with $1 \times 1 \times k$ point-wise convolution operation followed by channel-wise $n \times n$ spatial convolution operation. This way, the number of operations is reduced by a factor proportional to 1/k.

The Xception architecture has 38 convolutional layers which are structured into 14 modules, all of which have linear residual connections around them. Residual connections are 'skip connections' which allow gradients to flow through a

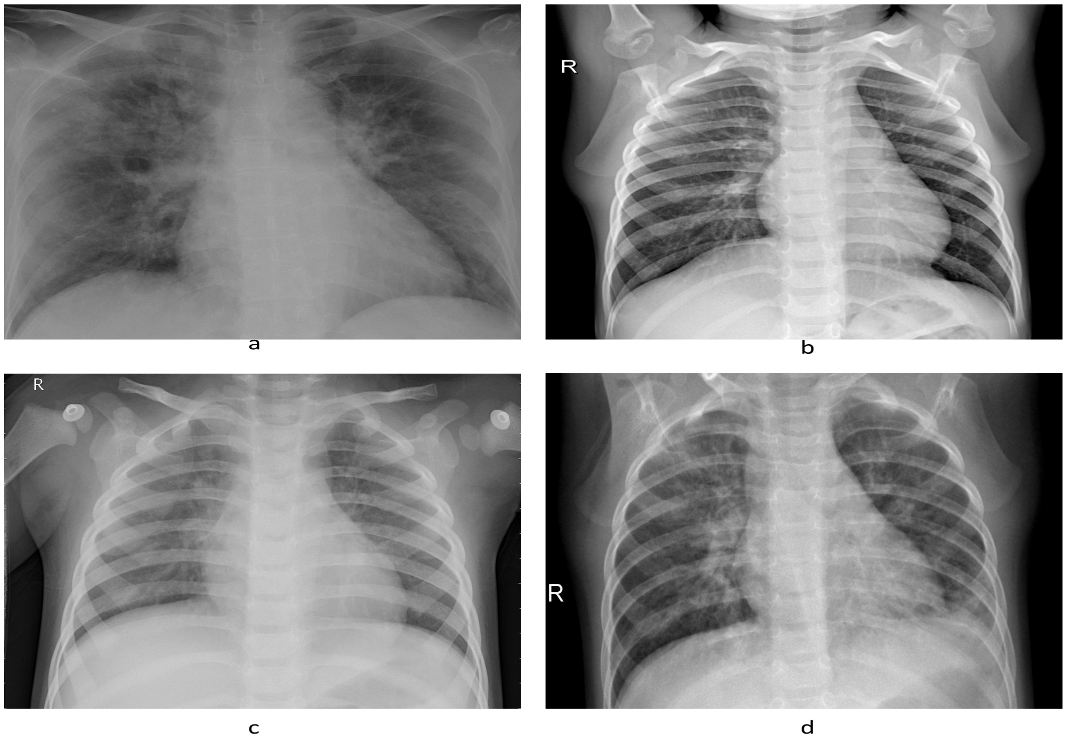

Fig. 1: (a) COVID positive, (b) Healthy, (c) Pneumonia bacteria, (d) Pneumonia viral

network directly, without passing through non-linear activation functions and thus avoiding the problem of vanishing gradients. In residual connections, the output of a weight layer series is added to the original input and then passed through a non-linear activation function. Thus, Xception serves as a base model with a dropout layer and two fully-connected layers added at the node. There are 256 and 4 neurons in the last two dense layers with softmax as the last activation.

Along with Xception, we are also implementing Resnet architecture. ResNet50 is a variant of the ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer [13]. It is based on residual learning to tackle the issue of a decrease in accuracy as the deep layers are increased. Deep networks are hard to train because of the notorious vanishing gradient problem — as the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient extremely small. As a result, as the network goes deeper, its performance gets saturated or even starts degrading rapidly.

In residual learning, instead of trying to learn some features as it learns in the deep convolutional neural network, we try to learn some residual. ResNet does this using shortcut connection (directly connecting the input of nth layer to some (n+x)th layer. It has proved that training this form of networks is easier than training simple deep convolutional neural networks and also the problem of degrading accuracy is resolved.

B. CNN based architectures

In the baseline CoroNet, there are more than 30 million trainable parameters. Therefore, we are also exploring simple convolutional-based architectures that can achieve similar performance with relatively fewer parameters. In this regard, various architectures are implemented which is detailed in Table I. Table I provides all the different deep neural network architectures that are explored in this study along with the baseline. Basically, in CNN based architectures, the number of convolutional layers are kept fixed at three but the neurons count and filter size are varied to study their effect on final model performance. Figures 2 and 3 represent the CNN and Resnet-50 based architectures, respectively for classification of COVID-19 images.

Table I: Experiment architectures

Experiment No.	Conv. layers	Filter size	Dense layers
Exp 1	Xception		256,4
Exp 2	Resnet50		256,4
Exp 3	128,64,32	3x3	128,4
Exp 4	128,64,32	3x3	256,4

C. Training Models

Three different classification models have been studied in this work. The first is binary classification (COVID-19/others), then a three-class (COVID-19/Normal/Pneumonia) model, and finally, four classes (COVID-19/Normal/Pneumonia bacteria/ Pneumonia viral) case.

The models are trained using the Tensorflow framework in a machine accelerated with Nvidia GPU. The optimizer for all the models is Adam with the default learning rate. For illustration, the training accuracy and loss of Resnet50 is shown in figure 4. The figures denote that the model needs

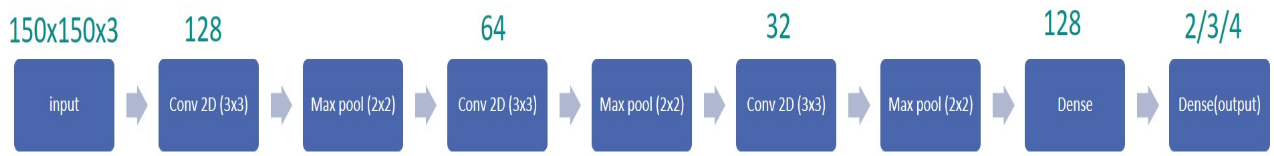

Fig. 2: CNN Architecture: Input image is RGB with 150X150X3 dimension. Numbers in Blue color indicate number of convolutional filters or neuron units

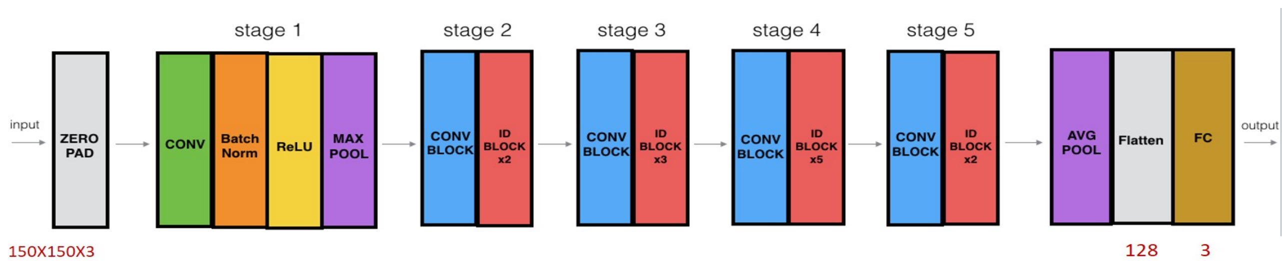

Fig. 3: Resnet-50 Architecture: Input image is RGB with 150X150X3 dimension. Numbers in Red color indicate number of neuron units.

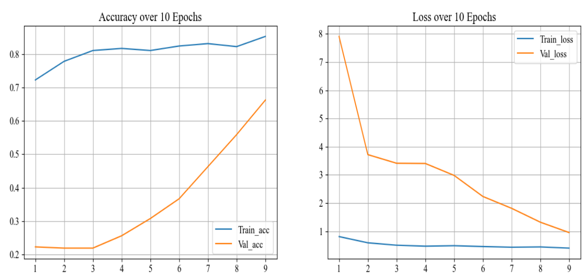

Fig. 4: Resnet50 results

further training as loss and accuracy curves doesn't saturate. Hence, the number of epochs has been increased such that learning curves (loss and accuracy) settles. The categorical cross entropy loss functions for m number of samples can be expressed as,

$$Loss = \frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\mathbb{1}(y_i = k)\log\Big(\frac{exp(\mathbf{w_k^T x_i})}{\sum_{j=1}^{K} exp(\mathbf{w_j^T x_i}))}\Big) \quad (1)$$

where $\mathbb{1}(y_i = k)$ is an indicator function (i.e. $\mathbb{1}(y_i = k) = 1$ if the example belongs to class k and 0 otherwise).

IV. Experiments

A. Evaluation metrics

The identification of Covid-19 infections from chest radiography images is posed as a classification problem. Hence, the relevant evaluation metrics are Accuracy, precision, recall, F1 score, Roc curve, Confusion matrix. Accuracy determines the total no of images correctly classified irrespective of class.

$$\text{Accuracy} = \frac{\text{No of images correctly classified}}{\text{Total no of images}}$$

To further evaluate model in generating false positives and true positives, precision, recall and F1-score is given as

$$\text{Precision} = \frac{\text{Sum of all true positives (TP)}}{\text{Sum of all true positives (TP) + Sum of all false positives (FP)}}$$

$$\text{Recall} = \frac{\text{Sum of all true positives (TP)}}{\text{Sum of all true positives (TP) + Sum of all false negatives (FN)}}$$

$$F_1 = \frac{2TP}{2TP + FP + FN}$$

d capability in discriminating classes is computed through the area under the ROC curve. ROC is plotted with True positive rate (TPR) against the False Positive Rate (FPR) where TPR is on the y-axis and FPR is on the x-axis. Higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s. A confusion matrix is a table with 4 different combinations of predicted and actual values. It is useful for measuring Recall, Precision, Accuracy, and most importantly AUC-ROC Curve. The models. The trained models are evaluated using the above-mentioned metrics and results are reported in the next sub-section.

B. Experiments metrics

This section provides the performance of different model architectures listed in earlier experiment architecture sections and reports mean metrics and execution times on validation data (K-fold). The input data is split into training and testing sets in the ratio of $80:20$, respectively. The validation is done

Table II: Covid19/Healthy/Pneumonia-bacteria/viral

Experiment	Accuracy	Precision	Recall	AUC	# parameters
Exp 1	87.24	88.23	87.15	97.50	33,915,436
Exp 2	71.23	75.68	71.71	93.53	36,696,196
Exp 3	87.19	86.21	86.21	95.62	1,280,228
Exp 4	88.02	88.1	88.1	95.15	2,464,612

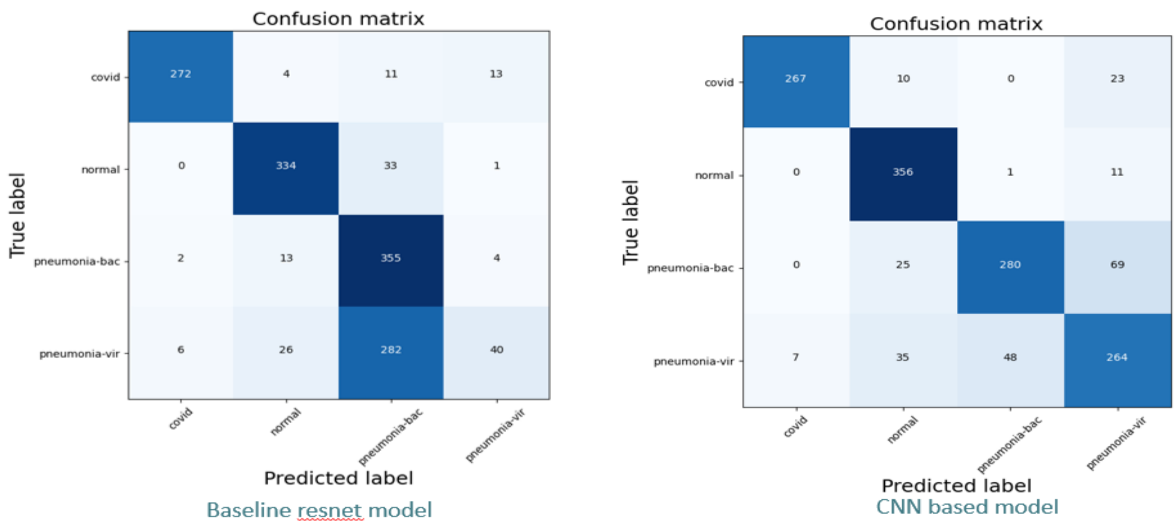

Fig. 5: Four class confusion matrix

Table III: Covid19/Healthy/Pneumonia

Experiment	Accuracy	Precision	Recall	AUC	# parameters
Exp 1	96.34	96.66	96.34	97.74	33,915,436
Exp 2	94.67	94.69	94.69	98.97	36,696,196
Exp 3	89.20	89.30	88.18	97.05	1,280,228
Exp 4	90.50	87.50	86.36	94.85	2,464,612

with 5 fold cross-validation while updating the models during iterations.

Table II tabulates the results of different architectures for the four-class classification model. The confusion matrices for Exp-2 (baseline resnet-50) and Exp-3 (CNN) are shown in figure 5. The class index is as follows: 0: Covid-19, 1: Healthy, 2: Pneumonia bacteria, 3: Pneumonia virus. It can be inferred from the table that the CNN-based model outperforms resnet-50 although having one-thirtieth of the parameters. However, Xception performs at par or slightly better than the CNN-based model. Thus, although having fewer parameters, Xception performs better than Resnet-50 for the classification of chest radiography images.

Table III depicts the performance of the three-class model. Here, both baselines perform better than CNN-based models. Further, Resnet-50 accuracy has increased significantly compared to that of four class cases and indeed lies close to that of Xception. For this class, the confusion matrices for Exp-1 (Xception), Exp-2 (baseline resnet-50) and Exp-3 (CNN) are shown in figures 6 and 7.

Finally, Table IV tabulates the performance score of the two-class model, i.e., healthy or non-healthy. Non-healthy comprises of all the three infectious cases namely Pneumonia bacteria, Pneumonia viral, and Covid-19. Figure 7 depicts the confusion matrices for Exp 1 (Xception) and Exp-3 (CNN based) models. Again, both the baselines perform better than that of CNN-based models. As as we decrease the number of classes, the complexity of the problem is decreasing and consequently, the model is able to find a better separating hyper-plane leading to an increase in performance.

C. *Discussion*

Several inferences can be drawn based on the experiments conducted with large deep neural network architectures and simpler CNN-based models. Overall, large architectures perform better than simple CNN-based models and it is very obvious due to a significantly large number of parameters. Thus, the large models can be deployed for applications that are critical and pose no computational restriction. However computationally restricted applications (e.g., edge devices) and/or secondary importance, CNN-based model will be more appropriate. Further, Xception consistently performs better than Resnet-50 although having fewer parameters. This might be because of the depth-wise separable convolution architecture.

V. Conclusions and Future work

This project uses radiography chest X-ray images to detect the Covid-19 in the human body. The goal is to classify

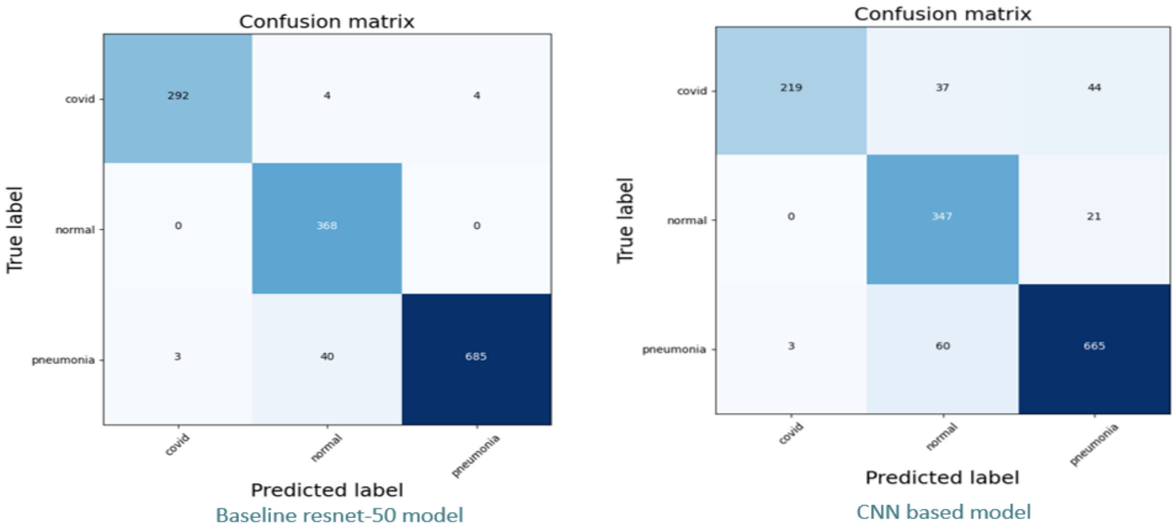

Fig. 6: Three class confusion matrix

Table IV: Healthy/Non-Healthy

Experiment	Accuracy	Precision	Recall	AUC	# parameters
Exp 1	98.34	98.24	98.14	97.65	33,915,436
Exp 2	95.13	95.18	95.20	97.31	36,696,196
Exp 3	90.31	89.14	89.25	96.21	1,280,228
Exp 4	91.23	90.00	90.63	96.20	2,464,612

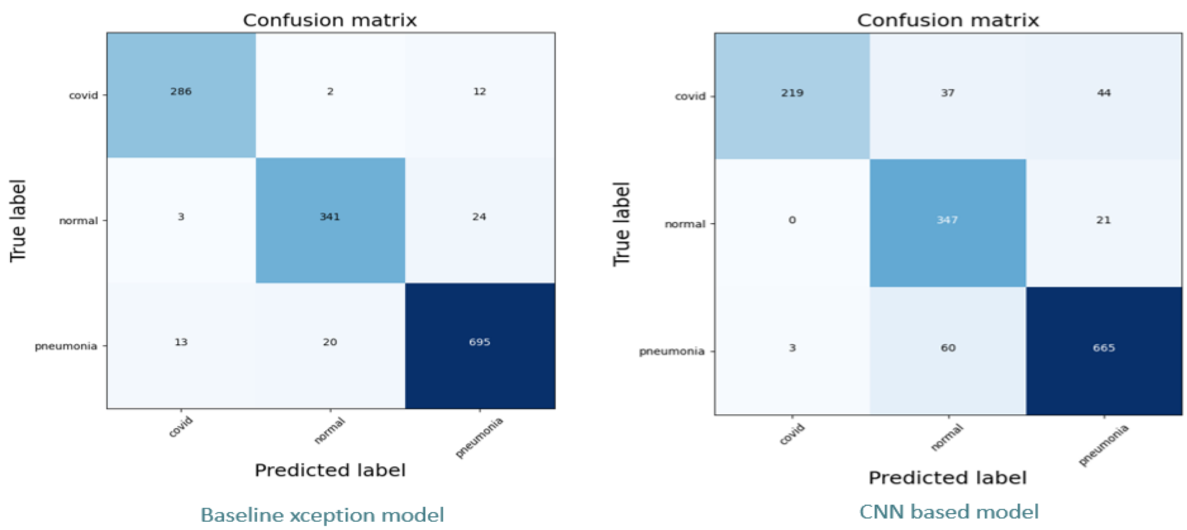

Fig. 7: Three class confusion matrix

the images into 2,3, or 4 classes using Convolutional neural networks. We have compared the CNN approach against the baseline approaches - Resnet-50 and Xception. Our model is observed to correctly classify the images while using very few parameters. CNN model achieves more than 90% accuracy in most of the experiments with relatively few parameters and less complexity. As a part of future work, more CNN-based architectures will be explored such as skip connections to further study the differences in shallow and large deep networks in medical image classification. Apart from CNN-based analysis, we will also explore Graph neural network as an alternative framework for image classification.

References

[1] [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html

[2] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, vol. 296, no. 2, pp. E115–E117, 2020.

[3] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest ct for typical coronavirus disease 2019 (covid-19) pneumonia: relationship to negative rt-pcr testing," *Radiology*, vol. 296, no. 2, pp. E41–E45, 2020.

[4] W. LINDA, "A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images," *Journal of Network & Computer Applications*, vol. 20, pp. 1–12, 2020.

[5] P. Sethy and S. Behera, "Detection of coronavirus disease (covid-19) based on deep features. preprints," *Preprint posted online March*, vol. 19, 2020.

[6] A. I. Khan, J. L. Shah, and M. M. Bhat, "Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105581, 2020.

[7] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[8] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection," *Ieee Access*, vol. 8, pp. 91 916–91 923, 2020.

[9] G. Nikolentzos, M. Thomas, A. R. Rivera, and M. Vazirgiannis, "Image classification using graph-based representations and graph neural networks," in *International Conference on Complex Networks and Their Applications*. Springer, 2020, pp. 142–153.

[10] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv 2006.11988*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[11] P. Mooney. Chest x-ray images (pneumonia). [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.