Ashwin Rachha[1]

[1]Affiliation not available

November 11, 2020

Ashwin Rachha

———————————————————————————***———————————————————————————

—

Abstract - *The coronavirus outbreak has caused a devastating effect on people all around the world and has infected millions. The exponential escalation of the spread of the disease makes it emergent for appropriate screening methods to detect the disease and take steps in mitigating it. The conventional testing technique involves the use of Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR). Due to limited sensitivity it is more prone to providing high false negative rates. Also due to a high turnaround time (6-9 hours) and a high cost, an alternative approach for screening is called for. Chest radiographs are the most frequently used imaging procedures in radiology. They are cheaper compared to CT scans and are more readily available and accessible to the public. Application of advanced artificial intelligence (AI) techniques coupled with radiological imaging can be helpful for the accurate detection of this disease. In this paper 8 different architectures of CNN (EfficientNet, ResNet, DenseNet, VGG, Mobilenet) are compared regarding classification performance on the COVID-19 dataset. The proposed models are developed to provide accurate diagnostics for multiclass classification (Covid vs No Findings vs Pneumonia) and the best performing models are used to perform binary classification (Covid vs No Findings). Areas under the receiver operating characteristics curves (AUROC) between 0.9622 and 0.9987 could be achieved for multiclass classification while accuracy scores ranged from 90.32% to 94.93% achieving next to state-of-the-art results. A dataset was created as a mix of publicly available X-ray images from patients with confirmed COVID-19 disease, common bacterial pneumonia and healthy individuals.*

Key Words: COVID-19 detection, X-RAY, Transfer learning, Deep learning, Convolutional Neural Networks, Efficient Net, Resnet, Densenet.

1.INTRODUCTION

Covid-19 is a Severe Acute Respiratory Syndrome (SARS), firstly discovered in the Wuhan region of China in December 2019. Since its inception it has rapidly spread across all countries infecting almost 44 million people worldwide and causing fatality of over 1 million people. It is a highly contagious disease the symptoms of which are characterized by fever, shortness of breath, cough and loss of smell. The incubation period of Covid-19 ranges from 1 to 12.5 days with the median being 5-6 days but can take as long as 14 days. Because of a long incubation period, the infections spread out exponentially since people are unaware of the fact that they have contracted the disease and unknowingly spread it. The detection of COVID-19 disease is critically significant and vital so that those infected can receive immediate treatment and care, as well as be isolated to attenuate the spread of the disease. According to WHO, it is mandatory that the patients be screened by Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) which can detect SARS-CoV2 RNA from respiratory specimens. While RT-PCR has been the gold standard test since it is highly specific, it is time consuming and the sensitivity is unreliable, many reports suggesting poor sensitivity. Hence a needfulness for a quick screening method emerges that might help doctors to rapidly triage to be further assigned to be tested by RT-PCR.

Chest radiographs are the most frequently used imaging procedures in radiology. They are cheaper and more easily available than Computed Tomography and Magnetic Resonance Imaging. Covid-19 reveals some radiological signature that can be easily detected through chest radiographs. Apart from this, there are other advantages that could be leveraged regarding chest radiographs. Chest radiographs enable rapid triaging of patients which can be done in parallel with viral testing to mitigate the high number of infected people in areas most affected where the capacity is overburdened by demand. Along with this chest radiographs can be taken in an isolated room thus reducing the risk of contraction of the disease. The automatic analysis and detection can be employed with deep learning based approaches. Convolutional Neural Networks primarily have been successful due to their ability to learn features automatically from domain specific images unlike classical machine learning algorithms. One of the most popularly adopted methods in the field of medical imaging is transfer learning and to use knowledge derived from training models from one domain applied over to another. This method is particularly effective when the annotated dataset is smaller.

Transfer learning can be employed in one of the following 3 ways:

1. Shallow Tuning - Adapts only last classification layer to cope with new task
2. Deep Tuning - retrain all parameters of the model from end to end.
3. Fine Tuning- aims to gradually train more layers till an improvement in metrics is observed.

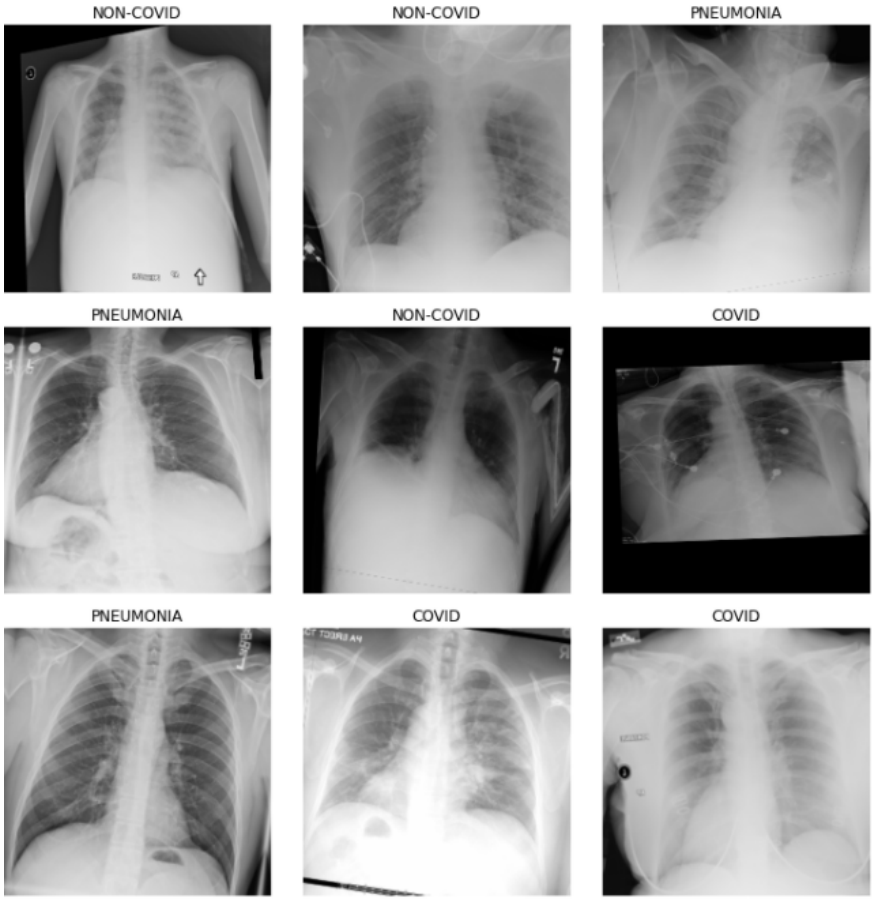


Figure 1 : Sample radiographs from the dataset pertaining to each label.

Motivated by an urgent need for a screening method, portability and easy availability of chest radiographs

and advanced state of the art deep learning models, a study of 8 different CNN architectures viz. EfficientNet, ResNet, DenseNet, MobileNet and VGGNet are presented which are fine tuned to perform classification on a custom dataset comprising 3 classes - Covid, No-Findings and Pneumonia respectively. The respective models show promising performance metrics and high scores with regard to AUROC, accuracy, Top 2 Accuracy, Precision and F1 scores.

2. Related Work

Motivated by the demand for better interpretation of X-Ray images, a number of deep learning methodologies have been proposed and their respective results have been quite promising in detecting COVID-19.

Wang et al proposed a new architecture of CNN called COVID-net which aimed to classify the 3 classes of covid, pneumonia and no findings respectively. They also introduced the benchmark open access COVIDx dataset consisting of 13975 CXR images from which 182 belong to COVID-19 patients. The authors reported an overall accuracy of 92.4% and a sensitivity of 80% for COVID-19.

Farooq and Hafeez, 2020, introduced a paper in which they fine-tuned a ResNet50(Szegedy et al) model to classify chest xray into normal, COVID-19, bacterial pneumonia and viral pneumonia. The authors reported better results compared to COVID-net with an overall accuracy of 96.23% and a 100% sensitivity for COVID-19 images.

Hemdan et al, 2020 presented a comparison among seven different well known CNN architectures to classify COVID CXRs. They performed their training on a smaller dataset with only 50 images in which 25 samples are from healthy patients and 25 from COVID-19. Among their experiments, they found VGG19(Simonyan and Zisserman, 2014) and DenseNet201 (Huang et al, 2019) gave the best results.

In (Khan et al, 2020), the authors propose a Convolutional Neural Network-based model to automate the detection of COVID-19 infection from chest X-ray images, named CoroNet. The proposed model uses the Xception CNN architecture (Chollet, 2017), pre-trained on ImageNet dataset (Deng et al, 2009).

3. Methodology

3.1 Dataset Preparation

In this study, X-ray images obtained from two different sources were used for the diagnosis of COVID-19.The Covid-19 Image Data Collection, which is a free open source dataset that contains chest radiographs for the novel coronavirus SARS-CoV-2 with associated covid-19 pneumonia was developed by JP Cohen et al. This dataset is constantly updated with new images. These images were amalgamated with images from the COVIDx dataset provided by Wang et al. The dataset involves a total of 495 covid images. In order to avoid the unbalanced data problem, 704 no-findings and 500 pneumonia class frontal chest X-ray images were chosen randomly from COVIDx dataset for multiclass classification.

The train and validation split of the respective images is as follows.

Split	COVID-19	NON-COVID	PNEUMONIA
Training Set	454	610	418
Testing Set	41	94	82

Table 1 : Distribution of dataset images.

The train set was applied to train the classifier, while the validation set was applied to choose the most satisfactory performance.

3.2 Model Training

8 Convolutional Neural Networks of 5 different architectures(EfficientNet, ResNet, DenseNet, VGG, MobileNet) were trained on the aforementioned dataset. Training was done using the python programming

language(https://www.python.org) on top of the pytorch (https://pytorch.org) and FastAI (https://fast.ai) frameworks. The training was done on the Google Colab platform (https://colab.research.google.com/) and the GPU in use is the NVIDIA Tesla K80.

For training the images were resized to size (224,224) to conform to most of the pretrained models' size. Image Augmentation was performed with the default transforms viz. Rotation, horizontal flip, vertical flip, zooming, warping, affine transformation etc. Image augmentation increases the train set by performing transformations on images thus enhancing performance by introducing more features to the models and using no memory, since the transformed images are not stored in the memory. A batch size of 32 images was kept constant throughout training of all models. Each model was trained for 50 epochs. Before training, the optimal learning rate was determined. Its value ranged between 1e-6 to 1e-5. Custom tweaks were made while using the transfer learning models. The optimization function used was Rectified Adam whereas a novel Label Smoothing Crossentropy loss function was used which enabled the models to achieve enhanced metrics. Along with Accuracy and Area Under Receiver Operating Characteristics curve, Top 2 Accuracy was also used for evaluation.

4. CNN Architectures

In recent times the use of deep learning models particularly Convolutional Neural Networks has led to many advancements in a range of computer vision applications such as image segmentation, object detection and object recognition. Work by Hubel and Wiesel in the 1950s and 1960s showed that cat and monkey visual cortices contain neurons that individually respond to small regions of the visual field. They also showed that a monkey's visual field can be represented by a topographic mapping.

A model termed 'neo recognition' introduced by Fukushima and Miyake, an extension of the model proposed by Hubel and Wiesel introduced two basic types of layers in CNN: convolutional layers and downsampling layers.

In 1998 Yann LeCun et al proposed LeNet5 a 7 layer convolutional neural network that classified handwritten digits from postal zip codes. Although this work was pioneering in the field of computer vision, due to computational complexities to train higher resolution images, this model was constrained due to unavailability of computing resources. It took 3 days to train the model.

The need for computational resources was addressed by AlexNet, proposed by Krizhevsky et al which won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and was able to successfully reduce the top-5 error rate from 26% of older models to new state-of-the-art 15.3%. AlexNet consisted of eight layers, with five convolutional layers some of which were followed by max pooling layers followed by three fully connected dense layers. It made use of a non-saturating ReLu activation function which showed improvement in performance compared to tanh and sigmoid activation functions.

Following the pioneering results brought by AlexNet, various companies such as Google, Facebook and Microsoft helped in the advancement of deep learning in computer vision by participating annually in the Imagenet challenge. The models in this study are a result of such endeavours.

The name Convolutional is derived from a mathematical operator. A typical Convolutional Neural Network has a convolutional layer which extracts out relevant features from input image by applying certain matrices called kernels. This operation is followed by a pooling layer which condenses the output of the convolution operation and makes it feasible for computational performance and finally the output of which is fed to a fully connected neural network which is the dense layer.

In this study we have trained models from 5 architectures as follows:

4.1 EfficientNet

Tan and Le,2019 introduced a new baseline network and scaled it up to develop a family of models called EfficientNets. They proposed a new scaling method that

Uniformly scales all dimensions i.e depth/width/resolution using a simple yet highly effective compound coefficient. Its main component is the Mobile Inverted Bottleneck Conv(MBConv) depicted in the table below:

Stage	Operator	Resolution	#channels	#layers
1	Conv3x3	224x224	32	1
2	MBConv1, k3x3	112x112	16	1
3	MBConv6, k3x3	112x112	24	2
4	MBConv6, k5x5	56x56	40	2
5	MBConv6, k3x3	28x28	80	3
6	MBConv6, k5x5	14x14	112	3
7	MBConv6, k5x5	14x14	192	4
8	MBConv6, k3x3	7x7	320	1
9	Conv1x1/Pooling/FC	7x7	1280	1

Table 2 : Architecture of EfficientNet baseline.

EfficientNet is based on the grounds of introducing a high quality yet compact baseline and uniformly scale each dimension with a set of constant scaling coefficients.

Starting from the baseline model each dimension is scaled by parameter φ according to

depth = α φ

width = β φ

resolution = γ φ

s.t. α · β 2 · γ 2 [?]

α [?] 1, β [?] 1, γ [?] 1 ,

Where α, β, γ are obtained by grid search.

4.2 ResNet

The ResNet architecture was the winner of 2015 ILSVRC and was proposed by He et al. The main novelty behind the success of ResNets were the use of residual layers and skip connections which solved the vanishing gradients problem that may result in stopping the weights in the network to further change. Skipping connections will help gradients to flow backwards directly from end layers to initial layer filters enabling CNN models to deepen with a number of layers. They stack residual blocks on top of each other to form a complete network e.g. Resnet101 has 101 layers using these blocks. In our study we have implemented fine tuning on Resnet 101 and Resnet 152.

Layer name	Output size	Resnet 101	Resnet 152
conv1	112x112	7x7, 64, stride 2	7x7, 64, stride 2
conv2_x	56x56	3x3 maxpool, stride2 [1x1,64] [3x3,64] X 3 [1x1,256]	3x3 maxpool, stride2 [1x1,64] [3x3,64] X 3 [1x1,256]
conv3_x	28x28	[1x1,128] [3x3,128] X 4 [1x1,512]	[1x1,128] [3x3,128] X 8 [1x1,512]
conv4_x	14x14	[1x1,256] [3x3,256] X 23 [1x1,1024]	[1x1,256] [3x3,256] X 36 [1x1,1024]
conv5_x	7x7	[1x1,512] [3x3,512] X 3 [1x1,2048]	[1x1,512] [3x3,512] X 3 [1x1,2048]

Layer name	Output size	Resnet 101	Resnet 152
	1x1	average pool, 1000-d fc, softmax	average pool, 1000-d fc, softmax
FLOPS		$7.6 \text{ X } 10^9$	$11.3 \text{ X } 10^9$

Table 3 : Architecture of Resnet 101 and 152.

4.3 DenseNet

In DenseNet, each layer obtains additional inputs from preceding layers and passes on its own feature-maps to all subsequent layers. Concatenation is used to achieve knowledge from previous layers. This helps in reducing the risk of gradient-vanishing. In this paper implementation of Densenet169 and Densenet 201 is done.

4.4 VGGNet

VGG architectures were proposed by Oxford university's visual geometry group, whereby they demonstrated that using small filters of size 3-by-3 in all of the convolutional layers throughout the network leads to a better performance. The main intuition behind VGG architectures is that multiple small filters in a sequence can imitate the effect of larger filters. Due to its simplicity in design and generalization power, VGG architectures are widely used. We use VGG16 and VGG19 that consist of 16 and 19 convolutional layers, respectively.

4.5 MobileNet

MobileNetV2 is an extension of MobileNetV1 keeping mobile devices in mind while designing. The ability to run neural networks on mobile devices increases the availability of models and adds additional benefits such as added security and saving energy consumption. MobileNets uses depthwise separable convolutional layers as basic building blocks. However, MobileNetV2 introduces two additional features:1. Linear bottlenecks between layers and 2. Shortcut connections between bottlenecks. As with traditional residual connections, shortcuts enable faster training and better accuracy.

5. Experimental Results

5.1 Accuracy, AUROC and Top-2 accuracy.

The trained models beat the previously recorded accuracy scores on the Cohen dataset with Efficient-net B7 and Resnet 152 achieving over 94% accuracy rates and AUROC scores of 0.9987 and 0.9833 respectively. AUC-ROC curve is a performance measurement for classification at various threshold settings. It tells how much the model is capable of distinguishing classes. AUROC is a very important metric in this study since radiographs of pneumonia, covid and no-findings are similar in visual appeal hence the model needs to be trained so as to distinguish them properly. AUROC scores ranged between 0.9987 and 0.9622 for the models. All the models achieved a top-2 accuracy of 99% and above indicating strong differentiation abilities between pneumonia and covid images.

Training Model	Performance Measure(% Accuracy)	AUROC
Efficient-net B7	94.93%	0.9987
Resnet 152	94.47%	0.9833
Resnet 101	91.24%	0.9947
Efficient-net B4	92.16%	0.9622
Densenet 169	90.32%	0.9997
Densenet 201	90.78%	0.9967
Mobilenet V2	90.32%	0.9966
VGG 19 BN	91.24%	0.9994

Table 4 : Performance Metrics.

5.2 Precision, Recall and F1-score

Although the dataset used in the experiment is balanced, traditional accuracy based performance metrics are not enough to validate an optimal classifier. The use of Precision, Recall and F1-score has been made to evaluate classification accuracy.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

TP = True positive

TN = True negative

FP = False positive

FN = False negative

Figure 2 : Equations of precision, recall and f1 score.

The classification reports for the classifiers are represented below:

EfficientNetB7

	Precision	Recall	F1-score	Support
COVID-19	0.95	1.00	0.98	41
No-Findings	0.89	1.00	0.94	94
Pneumonia	1.00	0.83	0.91	82

Resnet 152

	Precision	Recall	F1-score	Support
COVID-19	0.98	0.98	0.98	41
No-Findings	0.90	1.00	0.94	94
Pneumonia	0.99	0.85	0.92	82

Resnet 101

	Precision	Recall	F1-score	Support
COVID-19	0.98	0.98	0.98	41
No-Findings	0.84	1.00	0.91	94
Pneumonia	0.98	0.77	0.86	82

EfficientNet B4

	Precision	Recall	F1-score	Support
COVID-19	0.95	0.93	0.94	41

	Precision	Recall	F1-score	Support
No-Findings	0.87	1.00	0.93	94
Pneumonia	0.99	0.83	0.90	82

Densenet 169

	Precision	Recall	F1-score	Support
COVID-19	0.95	0.95	0.95	41
No-Findings	0.84	1.00	0.91	94
Pneumonia	0.98	0.77	0.86	82

Densenet201

	Precision	Recall	F1-score	Support
COVID-19	0.97	0.93	0.95	41
No-Findings	0.85	1.00	0.92	94
Pneumonia	0.97	0.79	0.87	82

VGG 19 BN

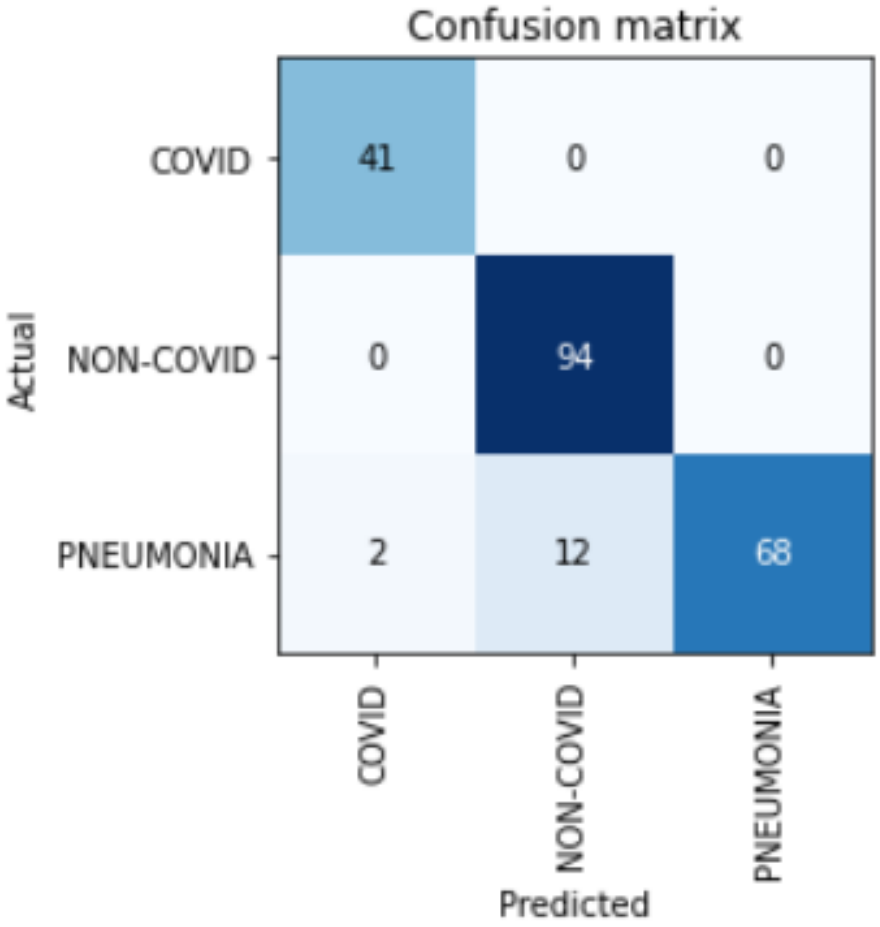
	Precision	Recall	F1-score	Support
COVID-19	0.98	0.98	0.98	41
No-Findings	0.84	1.00	0.91	94
Pneumonia	1.00	0.78	0.88	82

MobileNetV2

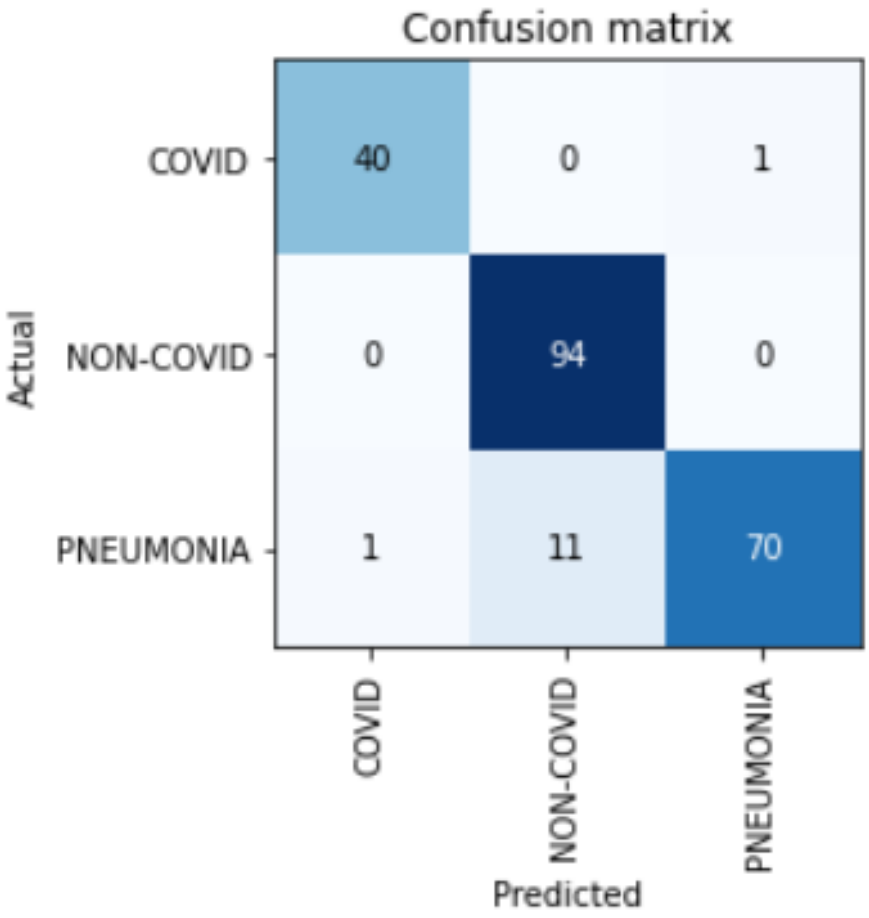
	Precision	Recall	F1-score	Support
COVID-19	0.95	0.93	0.94	41
No-Findings	0.84	1.00	0.91	94
Pneumonia	0.98	0.78	0.87	82

5.3 Confusion Matrix

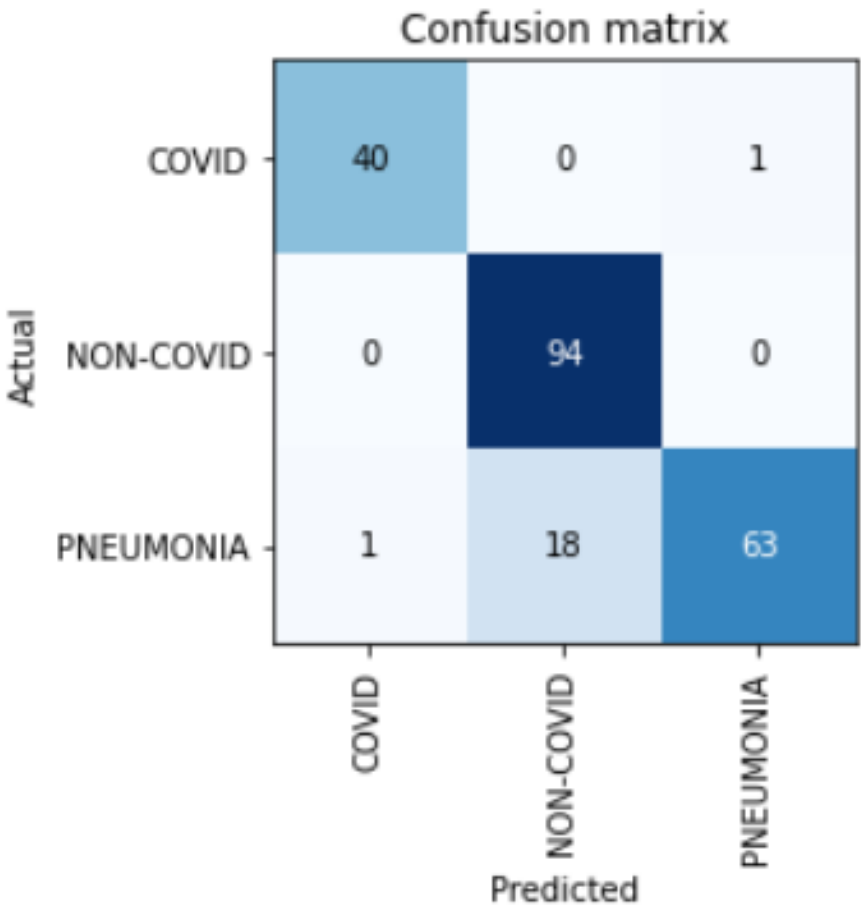
EfficientNet B7



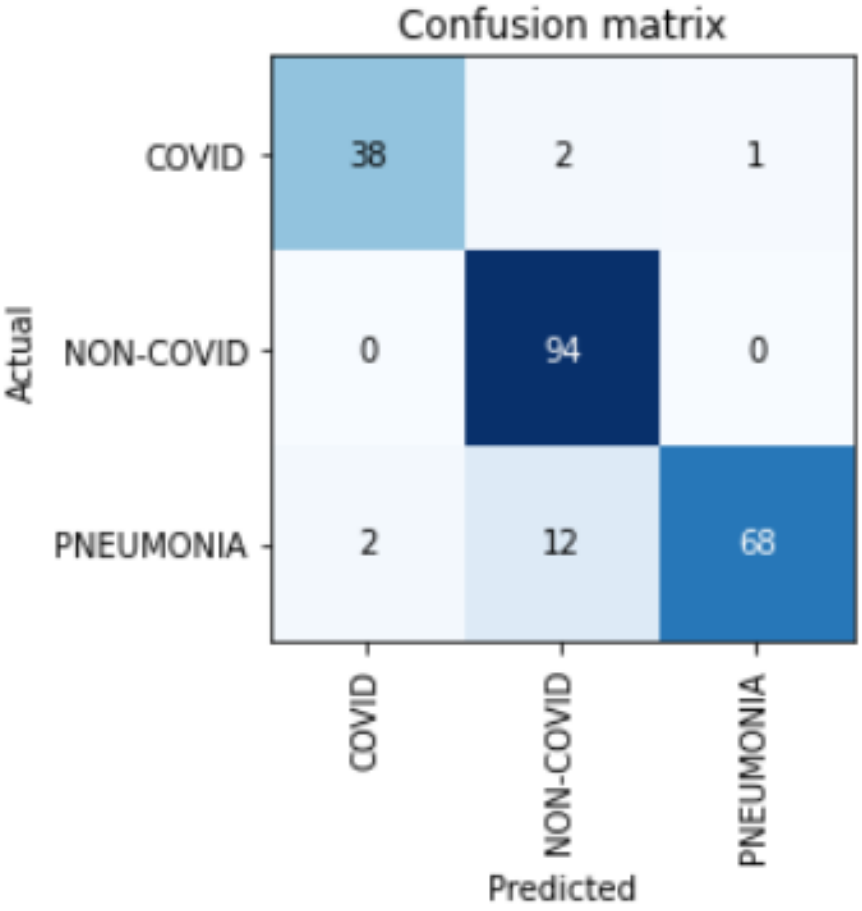
Resnet 152



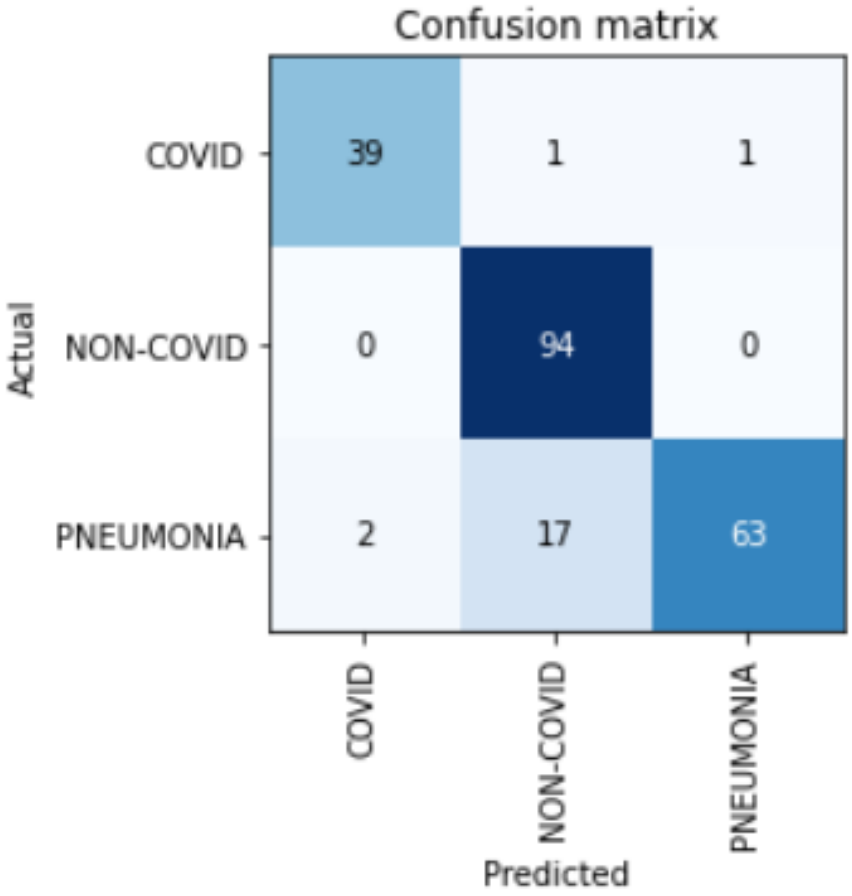
Resnet 101



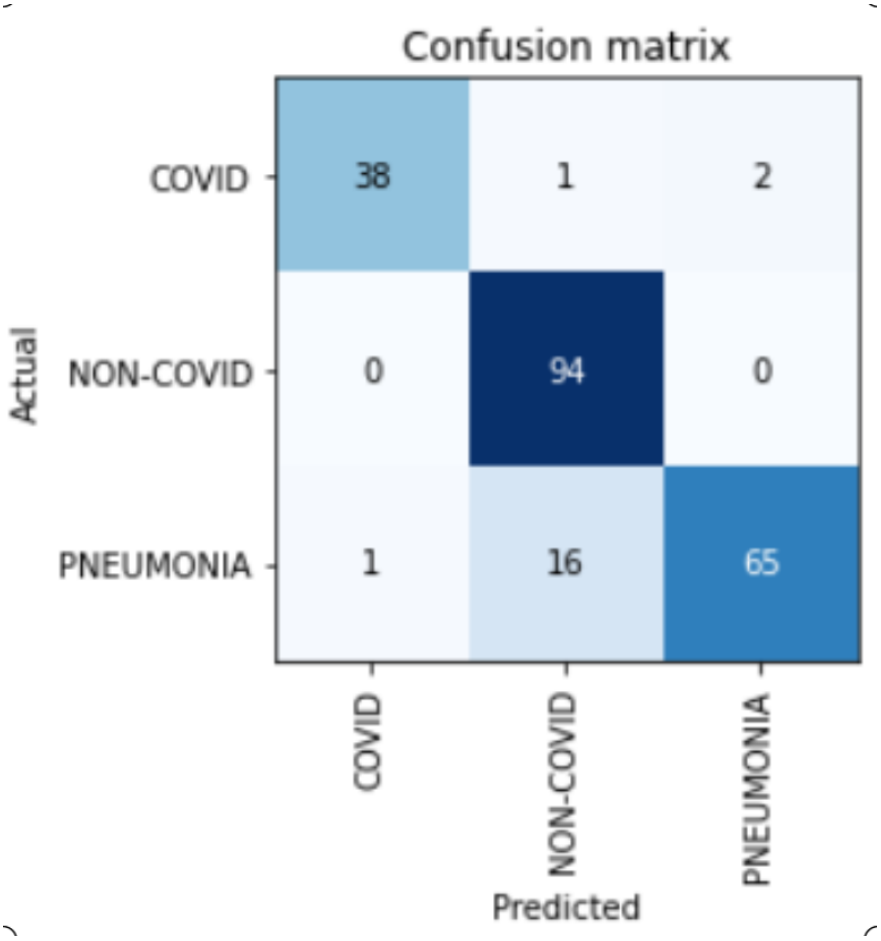
EfficientNet B4



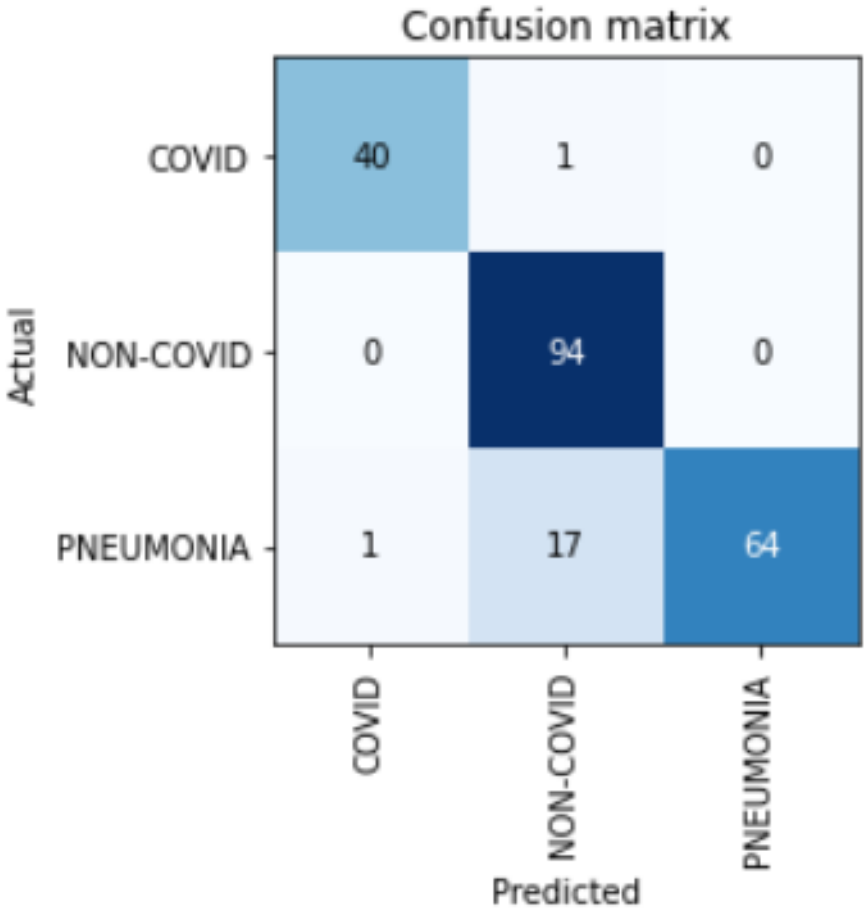
Densenet 169



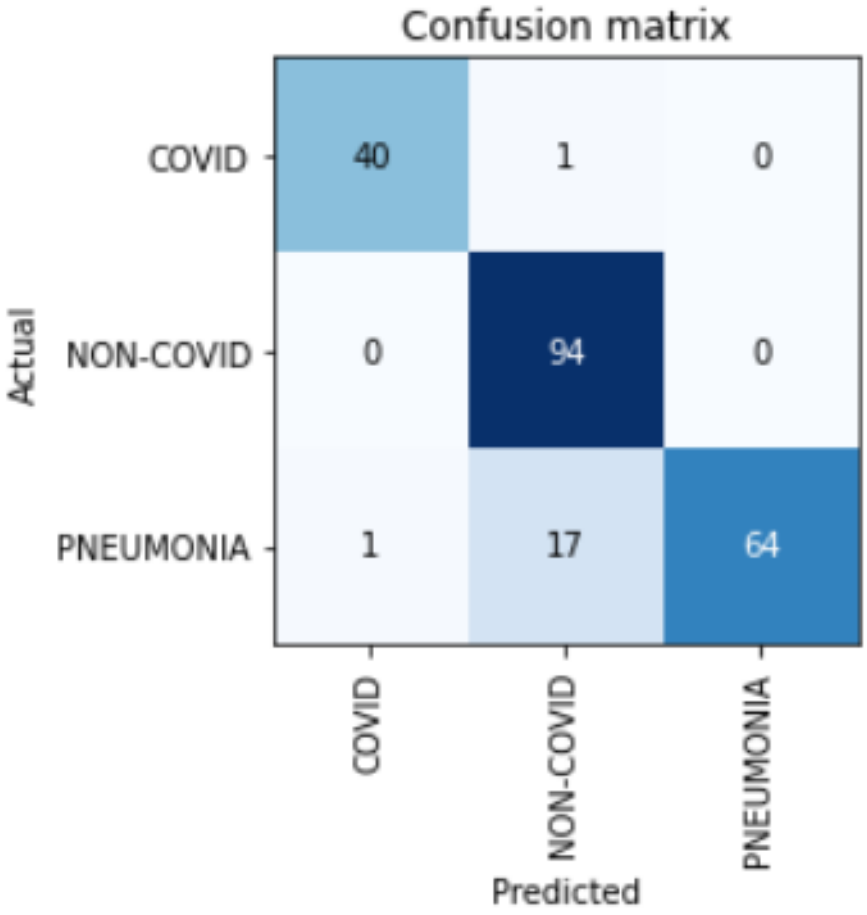
Densenet 201



VGG 19 BN



MobileNet V2



6. CONCLUSIONS

In the present work, it could be shown that increasing complexity and depth of artificial neural networks for the classification of chest radiographs is not always necessary to achieve state of the art results. In contrast to many previous studies, which mostly used deeper models with various layers, models with a fewer layers and appropriate scale can be used to classify chest radiographs suitably.

REFERENCES

1. Cohen, J. P., Morrison, P. & Dao, L. COVID-19 image data collection. arXiv 2003.11597 (2020).
2. Wang, Linda & Wong, Alexander. (2020). COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images.
3. Farooq, Muhammad & Hafeez, Abdul. (2020). COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs.
4. Hemdan, Ezz El-Din & Shouman, Marwa & Karar, Mohamed. (2020). COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images.
5. Khan, Asif & Shah, Junaid & Bhat, Mohammad. (2020). CoroNet: A Deep Neural Network for Detection and Diagnosis of COVID-19 from Chest X-ray Images. Computer Methods and Programs in Biomedicine. 196. 105581. 10.1016/j.cmpb.2020.105581.
6. Tan M, Le QV (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:190511946

7. D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," J. Physiol., vol. 195, no. 1, pp. 215–243, Mar. 1968.
8. K. Fukushima and S. Miyake, "Neocognitron: A SelfOrganizing Neural Network Model for a Mechanism of Visual Pattern Recognition," in Lectute notes in Biomathematics, Springer, Berlin, Heidelberg, 1982, pp. 267–285.
9. Y. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten Digit Recognition with a Back-Propagation Network," in NIPS, 1990, pp. 396–404.
10. A. Krizhevsky, A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Adv. Neural Inf. Process. Syst., p. 2012.
11. O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.
12. J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognit., vol. 77, pp. 354–377, May 2018.
13. K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015.
14. C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Computer
15. Society Conference on Computer Vision and Pattern Recognition, 2015, vol. 07-12-June-2015, pp. 1–9.
16. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
17. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 1800–1807, Oct. 2016.
18. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778 (2016).
19. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K.Q. Densely connected convolutional networks. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 4700–4708 (2017).
20. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249– 256.
21. D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.
22. Liu, Liyuan & Jiang, Haoming & He, Pengcheng & Chen, Weizhu & Liu, Xiaodong & Gao, Jianfeng & Han, Jiawei. (2019). On the Variance of the Adaptive Learning Rate and Beyond.
23. Sandler, Mark & Howard, Andrew & Zhu, Menglong & Zhmoginov, Andrey & Chen, Liang-Chieh. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 4510-4520. 10.1109/CVPR.2018.00474.