

# Optimal Classification of COVID-19: A Transfer Learning Approach

Aditya Kakde  
School of Computer Science  
UPES, Dehradun, India

Durgansh Sharma  
School of Computer Science  
UPES, Dehradun, India

Nitin Arora  
Electronics & Computer Discipline  
IIT, Roorkee, India

## ABSTRACT

Till now COVID-19 has affected 196 countries and resulted over 446,946 cases in which 19,811 are deaths, 112,058 got recovered and still many are left to be recovered. This is a viral pneumonia and thus no antiviral drug will work to reduce these cases. During the recovery, only immune system has played a major role. Analyzing and then diagnosing is currently a major challenge. This paper focuses on the classification which can help in analysis of COVID-19 with normal chest X-ray using deep learning technique. An optimal solution has been provided using transfer learning approach keeping in mind the limitation of the dataset. The performance has been determined by train and test loss and accuracy, sensitivity, specificity and p-score. The dataset used for the classification are the x-ray images of the chest as it can help to detect novel coronavirus in patients before RT-PCR lab testing.

## Keywords

COVID-19, deep learning, transfer learning

## 1. INTRODUCTION

A novel virus named as COVID-19 emerged from the part of coronavirus family came into existence at the time of transmission from bat to a person but some other sources according to [1] claimed that this virus was first practiced at the Wuhan Institute of Virology in China as a bioweapon and then due to some lack of protocols, an intern got infected by the virus and it gradually spread from the Wuhan wet market. At first the people considered it to be a normal flue which turns out to be a big mistake and till date no vaccine has been created.

It is then first reported in Wuhan's WHO (World Health Organization) country office on 31<sup>st</sup> December 2019. Its outbreak was declared a public health emergency of international concern on 30<sup>th</sup> January 2020 and the official name was announced on 11<sup>th</sup> February 2020. It is also called as 2019-nCov where 2019 is the year in which it was first determined, n stands for "novel" and Cov stands for "Coronavirus" family name.

It can be spread by touching an infected person or by inhaling near an infected person. It can also spread by touching the surface because when a person sneeze or discharge from the nose, that droplets comes in contact with the surface making it contaminated. So, after touching the surface and then the face can cause this disease. It spreads from one person to another 10 times faster than any other disease. The disease is recognized three days after the person came in contact with it.

According to WHO [2] (World Health Organization), a person initially can experience fever, tiredness or a dry cough. Some can also have runny nose, sore throat, nasal congestion, aches and pains or diarrhea. Old person or those who have some

medical problems like high blood pressure, heart problems or diabetes or chronic respiratory conditions are at the top on the list of COVID-19. It is a viral pneumonia and thus no antiviral drug which is used to treat other flu will not work.

In UK, NHS [3] (National Health Service) has identified some specific symptoms of this virus such as:

- A person can feel hot when touched on his chest or back.
- A person can experience repeated cough.

WHO (World Health Organization) and FIFA (Federation International de Football Association) also launched a campaign called as "Pass the message to kick out coronavirus" [4] to promote these five steps:

- Hand washing
- Coughing Etiquette
- Not touching your face
- Physical distance
- Staying home if feeling unwell

Till now 446,946 COVID-19 cases have been reported in which 19,811 are deaths and 112,058 got recovered. This virus has affected 196 countries till now including the Diamond Princess Cruise ship harbored in Yokohama, Japan.

The diagnosis practice right now are much relied on identifying COVID-19 positive patients and their immune systems. This paper focuses on enhancement of identifying COVID-19 patients using chest x-rays. This is done using the concept of convolutional neural network.

Below linear graph shows the growth of total cases from 22 January 2020 to 22 March 2020 and total deaths from 22 January 2020 to 22 March 2020 worldwide.

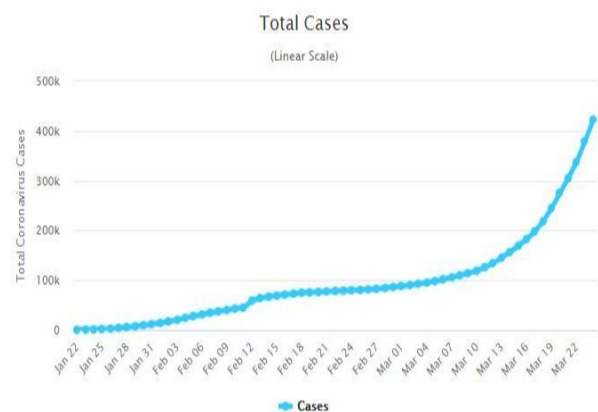
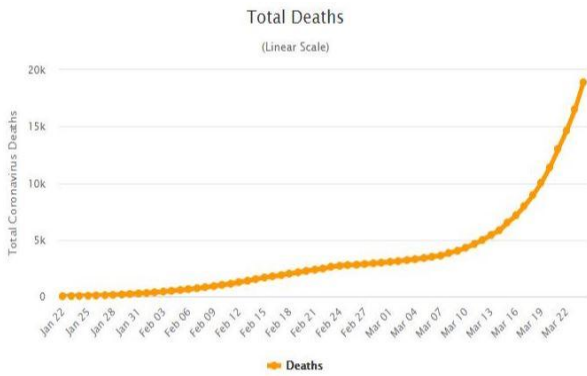


Fig 2 [5]: Total Cases of Novel Coronavirus Worldwide



**Fig 3 [6]: Total deaths due to novel coronavirus worldwide**

From the total cases, active and closed cases details can be seen in the below table.

**Table 1 [7]: Details of active cases which totals to 315,077**

Condition	Number of active cases
Mild Condition	301,502
Serious or critical condition	13,575

**Table 2 [7]: Details of closed cases which totals to 131,869**

Condition	Number of active cases
Mild Condition	301,502
Serious or critical condition	13,575

The statistics provided above is till 25<sup>th</sup> March 2020.

## 2. RELATED WORK

Alex Krizhevsky et al.[8] presented a paper on ImageNet Classification with Deep Convolutional Neural Networks in which they proposed a novel architecture consisting of convolutional layers, dropout function, local response normalization and fully connected layers. A 224x224x3 dimension image is taken as an input. The first convolutional layer has a filter size of 11x11x3 with 96 kernels means 96 number of filters and a strides of 4 pixels. The second convolutional layer used 156 kernels of size 5x5x48. The first two convolutional layers are stacked with an alternate arrangement of max pooling and local response normalization. The third, fourth and fifth convolutional layer are arranged one after the other where third convolutional layer have 384 kernels of size 3x3x256, the fourth convolutional layer has 384 kernels of size 3x3x192 and the fifth convolutional layer has 256 kernels of size 3x3x192. There are 2 fully connected layer for the classification tasks which consists of 4096 neurons. This architecture when tested on ILSVRC-2010 test set, achieved the least top-1 error rate of 37.5% and top-5 error rate of 17.0%. The CNNs described in this paper is tested on ILSVRC-2012 dataset where 7 CNNs achieved the least top-5 errors of 15.3%. Also when the architecture tested on ImageNet dataset, achieved the least top-1 error of 67.4% and top-5 error of 40.9%.

Jiuxiang Gu et al.[9] presented a paper on Recent Advances in Convolutional Neural Networks in which a briefly description about convolutional neural network involving layer design, activation function, loss function, regularization, optimization and fast computation has been described. After the success of Alexnet, further networks based on the similar concepts that

are ZFNet, VGGNet, GoogleNet and ResNet are also described. Some improvements regarding CNN are also mentioned which are Tiled Convolution, Transposed Convolution, Dilated Convolution, Network in Network and Inception Model. Further different categories which lies under pooling layer were also explained that are Lp pooling, Mixed pooling, Stochastic Pooling, Spectral Pooling, Spatial Pyramid Pooling and Multi-scale Order less Pooling. Different types of activation functions which are ReLU, Leaky ReLU, Parametric ReLU, Randomized ReLU, ELU, Maxout and Probout were explained followed by different types of loss functions that are Hinge loss, Softmax loss, Contrastive loss, Triplet loss and Kullback-Leibler Divergence. Different types of regularization techniques are also discussed that are lp-norm regularization, Dropout and DropConnect. In the domain of optimization, a brief description of Data Augmentation, Weight Initialization, Shochastic Gradient Descent, Batch Normalization and Shortcut Connections were given. Fast processing of CNNs and its applications were discussed at the last in this paper.

Sergey Ioffe et al.[10] presented a paper on Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate shift in which a novel method named batch normalization was proposed. It was also seen that sometimes it cuts as a regularizer, thus eliminating the need of dropout. It was also found that there is no need of local response normalization. When tested on LSVRC-2012 data, BN-x30 mean batch normalization with initial learning rate of 0.045 achieved the highest accuracy of 74.8%. Further a comparison of Batch Normalization Inception or BN-Inception is done with different state of the art results where BN-Inception ensemble achieved the least top-1 error of 20.1% and top-5 error of 4.9%.

Djork-Arne Clevert et al.[11] presented a paper on Fast and Accurate Deep Network Learning by Exponential Linear Unit (ELUs) where a new activation function was proposed. The proposed activation function is compared with other activation functions that are ReLU and Leaky ReLU on MNIST dataset where it was seen that ELU lines are average over five runs. Further seep auto encoders with ELU was trained and tested on MNIST dataset with different activation and learning rates where it was seen that ELU results are medians over several runs. Also a comparison of ELU with ReLU, SReLU and ReLU having batch normalization and without it was performed. It was seen that the ELU outperforms ReLU with batch normalization when tested on CIFAR-10 and CIFAR-100 dataset. The average test error achieved by ELU was 28.75(+0.24%). Comparison with other state of the art networks were done on CIFAR-10 and CIFAR-100 dataset in which ELU-Network achieved the least test loss of 24.28% on CIFAR-100 dataset whereas Fract. Max-Pool achieved the least test loss of 4.50% on CIFAR-10 dataset. Final test was conducted on ImageNet dataset where ELU achieved 20% top-5 error at 160K iterations whereas ReLU took around 200K iterations to reach at this loss rate.

## 3. METHODOLOGY

Transfer learning in the ability to apply some knowledge on the given task with the help of memory gained while learning similar task. Suppose that, if a person has learned how to ride a bicycle, then he/she can apply that knowledge to ride a bike as these two tasks are somewhat similar or it is the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related problem. In the domain of deep learning, transfer learning is adding newly learned data into pretrained weights. Basically, it is the feature

addition of new dataset into existing dataset. It is usually performed when dealing with small dataset. In NIPS 2016, a tutorial was given by Andrew Ng in which he said, “After supervised learning, transfer learning will be the next driver of ML commercial use”. This paper also used transfer learning approach and performed optimization in the network to achieve better results.

In this paper, The network architecture used to perform classification with transfer learning is NASNet[13] Large where large signifies dense architecture and ImageNet weights has been used as it has achieved top-1 accuracy of 0.825 and top-5 accuracy of 0.960 which is maximum when compared to Xception[14], VGG16[15], VGG19[15], ResNet50[16], ResNet101[16], ResNet152[16], ResNet50V2[16], ResNet101V2[16], ResNet152V2[16], InceptionV3[17], InceptionResNetV2[17], MobileNet[18], MobileNetV2[18], DenseNet121[19], DenseNet169[19], DenseNet201[19] and NASNetMobile[13].

NASNet also used the concept of convolutional neural network for feature extraction. The word convolutional means that the dot product of the values in the filters with the values of the input image. When dealing with 3-dimensional image, the depth of the filter should always be equal to the depth of the input volume. , the convolutional layer scans the complete image and extract the most important features of it. Then before getting into the pooling layer, it first passes through the activation function. The convolutional layer also consists of other parameters that are number of filters, strides and padding where strides means the distance between the input image and the convolutional layer or also it can be said as the length of receptive field and the padding is divided into subcategories which are called as valid and same. The padding means to have the same dimensions for both input and output. The valid means that there is no padding and same means we are applying some padding. The pooling which is used in this network is used for down sampling so that the dimensions of the input image can be reduced and then it can be easily feed into the fully connected layers for the classification task. It also consists pairs of reduction cell and normal cells where reduction cell means those cells which return a feature map whose height and width is reduced by a factor by 2 and normal cells means those cells which returns a feature map of same dimension.

### 3.1 Batch Normalization

Batch Normalization is the normalization of output of hidden layer. It is used to reduce the dependency of one hidden layer over the other hidden layer. This means that it gains an ability to learn some of the features of the input by itself. Batch Normalization consists of two parameters that are Beta ( $\beta$ ) and Gamma ( $\gamma$ ) which are learned by the network. This helps the network to keep mean and variance fixed so that the hidden layers become somewhat independent for feature learning. It sometimes acts as a regularizer as mentioned by Sergey Ioffe et al.[10] and thus when using batch normalization, there is no need of using dropout function. Thus batch normalization solves the problem of vanishing gradient.

$$\mu_{\beta} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

$$\sigma_{\beta}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\beta})^2 \quad (2)$$

$$x_i \leftarrow \frac{x_i - \mu_{\beta}}{\sqrt{\sigma_{\beta}^2 + \epsilon}} \quad (3)$$

$$y_i \leftarrow \gamma x_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (4)$$

### 3.2 Exponential Linear Unit

The Exponential Linear Unit can also be called as ELU and was first proposed by Djork-Arne Clevert et al.[11] and is given by:

$$elu(x) = \begin{cases} \alpha(e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases} \quad (5)$$

Its gradient is given by:

$$\frac{d\,elu(x)}{dx} = \alpha(e^x - 1) + \alpha \quad (6)$$

For positive value of x it acts as ReLU but unlike ReLU, it also takes negative inputs. In ReLU, for positive values the gradient becomes 1 but for negative part, the gradient automatically converts to 0 thus results in slow learning and sometimes the gradient does not even activate thus causes dead neurons. But in ELU, it saturates in the negative part of the domain. Because of this, the mean activation comes closer to 0. When compared for the positive inputs, in ELU the positive input saturates much faster than ReLU making it more efficient for mapping complex data.

### 3.3 Experimental Setup

NASNet Large architecture is used to conduct transfer learning. The cells are not predefined, instead they are searched by reinforcement learning search method.

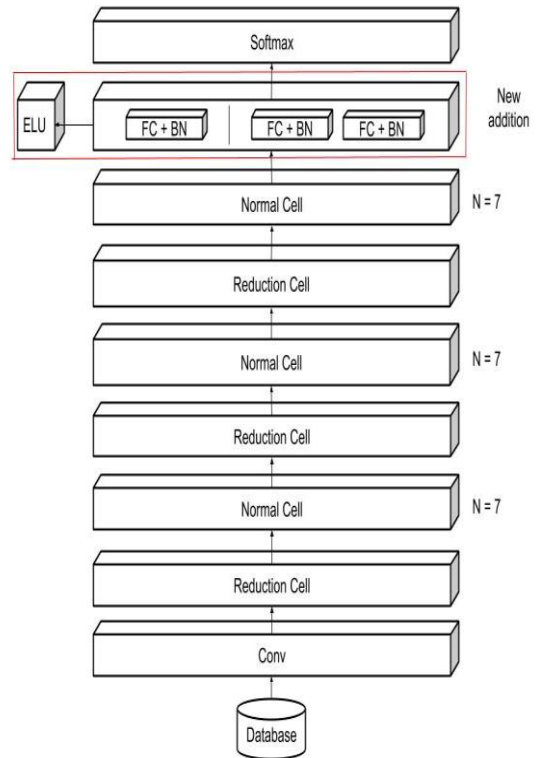


Fig 4: NASNet Large architecture with new additional layers. Here FC means fully connected layers and BN means Batch Normalization

For the classification task, two and three [12] fully connected layers are added at last which can be seen in the above diagram. The first fully connected layers consist of 32 number of filters with ELU activation function, second fully connected layer consists of 64 number of filters and third fully connected layer consist of 128 number of filters with ELU activation function. The fully connected layers are followed batch normalization after each layer to reduce the problem and vanishing gradient and it also reduces some amount of over fitting because according to the authors [10], being an optimizer, it also act as a regularizer. Thus, it forms a tree like structure to reduce the complexity. In the figure,  $N = 7$  signifies the number of cells repeated. 7 is chosen because as per the paper by Barret Zoph et al.[13],  $N$  with 7 has achieved state of the art error rate of 2.40% and ImageNet pretrained weights are used to conduct transfer learning. The experiment was conducted on Nvidia GEFORCE GTX 1050 GPU based machine consisting of 4GB of RAM and Intel core i5 processor.

### 3.4 Performance Metrics

The Performance metrics tells how good our model is working or how better it achieved its target. It yields cleared quantities defined in range and also provides a space for improvement. Overall, it is a basic of achieving a particular target. Some of the performance metrics which are used to evaluate a model which are specially based on the classification of biometric data are sensitivity, specificity and p-score.

#### 3.4.1 Sensitivity

Sensitivity means true positive rate. It is one of a performance metric which identifies the portion of actual positives. It means from all the people, it is used to identify how many of them have disease. Thus in our paper, it gives the percentage that how many people are correctly predicted that they are suffering from COVID-19 disease. It is given as:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (7)$$

#### 3.4.2 Specificity

It is another part of performance metric which is used to identify the correctly predicted negative portion. Means, it tells us that out of total predicted people, how many of them are not suffering from COVID-19 diseases. It can be given as:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (8)$$

#### 3.4.3 P-score

P-score can also be called as Physionet score. It is used to calculate the average of sensitivity and specificity of the model. The performance through p-score is identified by reviewing the value of p-score nearer to 1. Closer the value of p-score to 1, better will be our model. It can be given as:

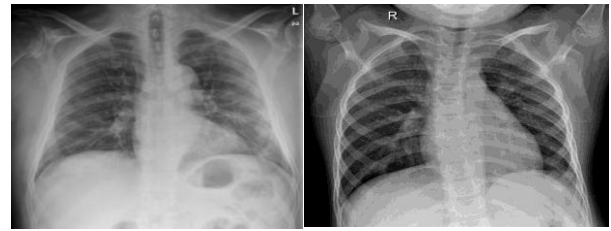
$$P - \text{score} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (9)$$

## 4. RESULTS AND DISCUSSION

### 4.1 Dataset

For the proposed approach, the dataset is referred from [20] and [21]. First part of the dataset consists of 69 images of chest x-ray infected with COVID-19 and other part of the dataset consists of 25 normal chest x-ray images. Both images comes under the category of multispectral images means those images whose features lies beyond visible spectrum. Thus the

proposed approach fit very well in this dataset as the features will be more than the usual.



**Fig 5: Left side of the picture is the infected chest x-ray with COVID-19 and right side in the normal chest x-ray.**

The images of COVID-19 are extracted through the given Meta data consisting of:

- Patient Id
- Offset
- Sex
- Age
- Finding
- Survival
- View
- Modality
- Date
- Location
- Filename
- DOI
- URL
- License
- Clinical notes
- Other notes

The original dimension of the images are around 1000 x 1000 which are then converted to 224 x 224 to reduce the preprocessing time. For training and testing part, the dataset is divided into 80% and 20%, means 80% of the data is used for training and 20% is used for testing.

### 4.2 Experimental Results

The networks were trained up to 30 epochs at 0.001 learning rate and for the optimization, Adam optimizer is used. The results can be seen in the below table.

**Table 3: Performance result on the basis of accuracy, sensitivity, specificity and p-score**

Model	Train L	Train A	Test L	Test A
NASNet+2FC	0.0011	1.0000	0.0477	0.9474
NASNet+3FC	0.0011	1.0000	0.0514	0.9474

NASNet + 2FC means in the original architecture, additions of two fully connected layers and 3FC means three fully connected layers. In Train and Test L, L means loss and in Train and Test A, A means accuracy. Further analysis were also made on the basis of sensitivity, specificity and p-score. In this it can be seen that test loss of NASNet+2FC is lower than the NASNet+3FC and the other parameter values are same.

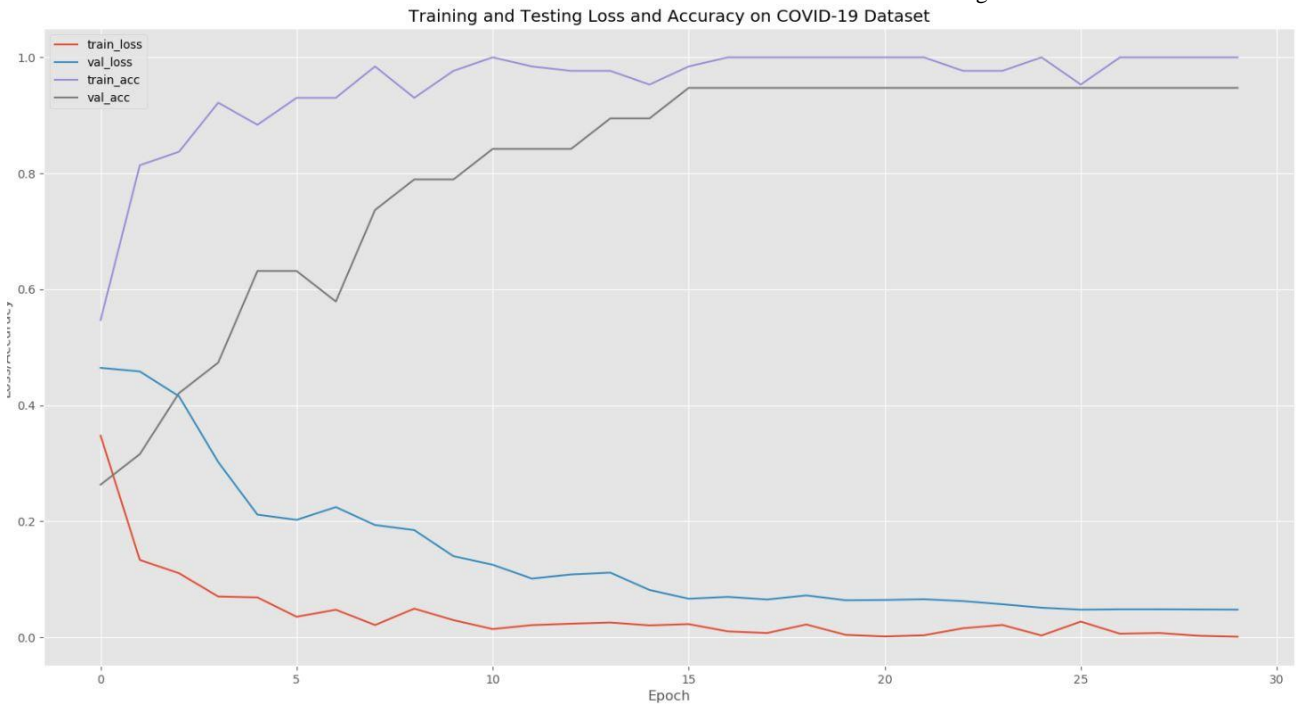
**Table 4: Performance result on the basis of sensitivity,**

**specificity and p-score**

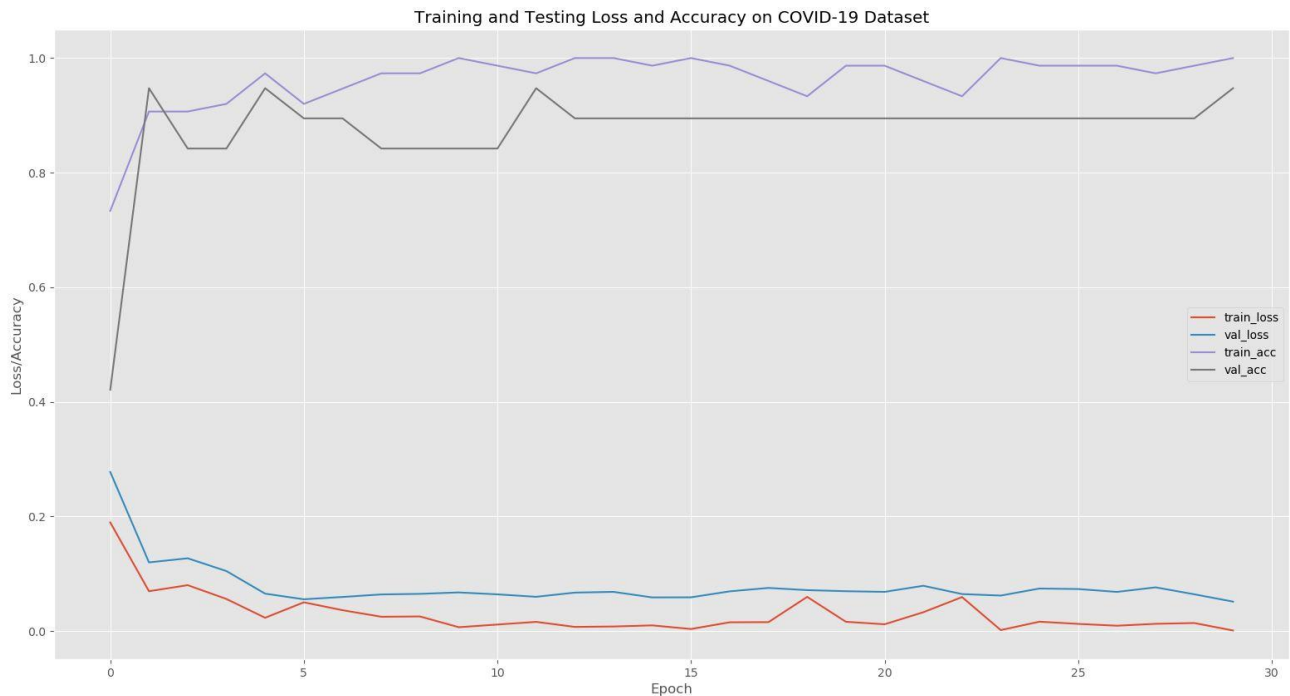
Model	Sensitivity	Specificity	P-score
NASNet + 2FC	0.9826	1.0000	0.9643

NASNet + 3FC	1.0000	0.8000	0.9000
--------------	--------	--------	--------

In the above table, it can be seen that NASNet+2FC and better specificity but less sensitivity than NASNet+3FC, but the p-score for NASNet+2FC is higher than NASNet+3FC.



**Fig 6: Training and testing loss and accuracy of two fully connected layer where red line shows train loss, blue line shows test loss, purple line shows train accuracy and brown line shows test accuracy**



**Fig 7: Training and testing loss and accuracy of three fully connected layer where red line shows train loss, blue line shows test loss, purple line shows train accuracy and brown line shows test accuracy**

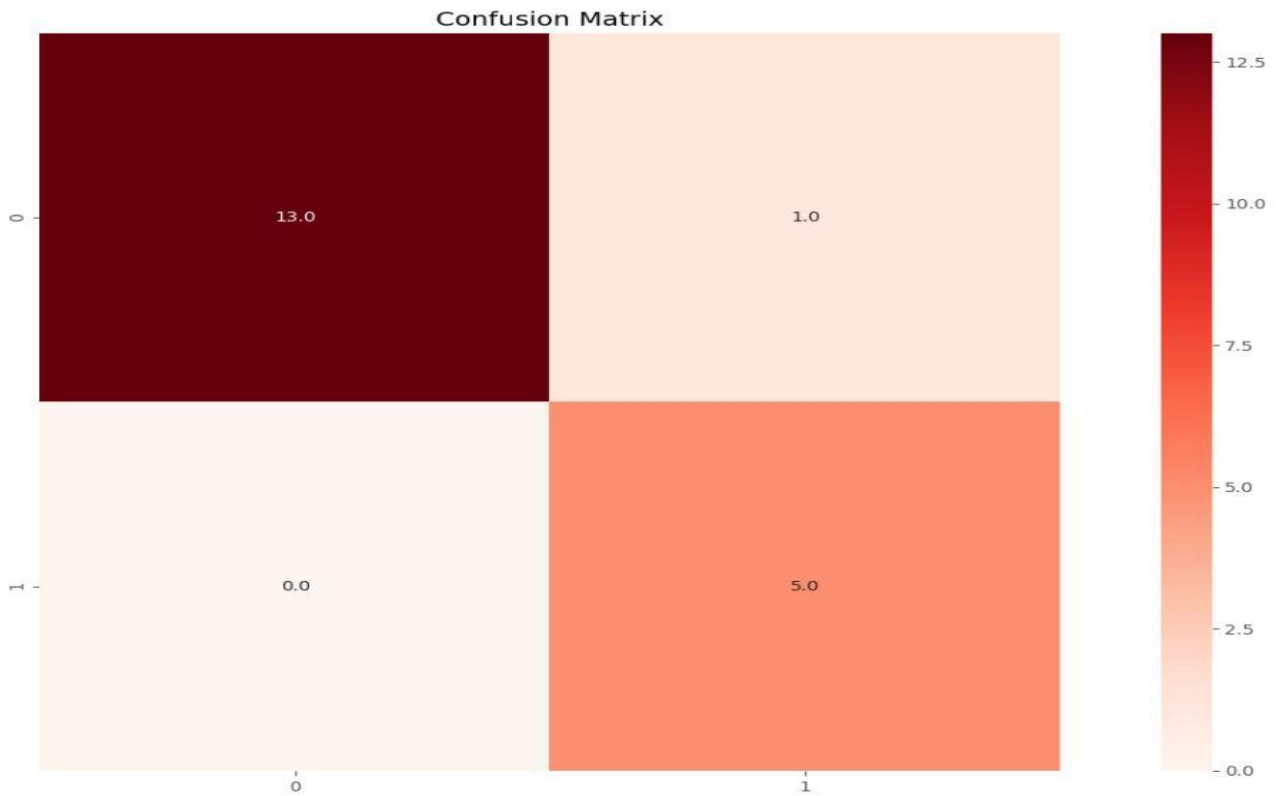


Fig 8: Heat map of confusion matrix of two fully connected layers acquired from the proposed approach. The x-axis shows the predicted labels and y-axis shows the actual labels

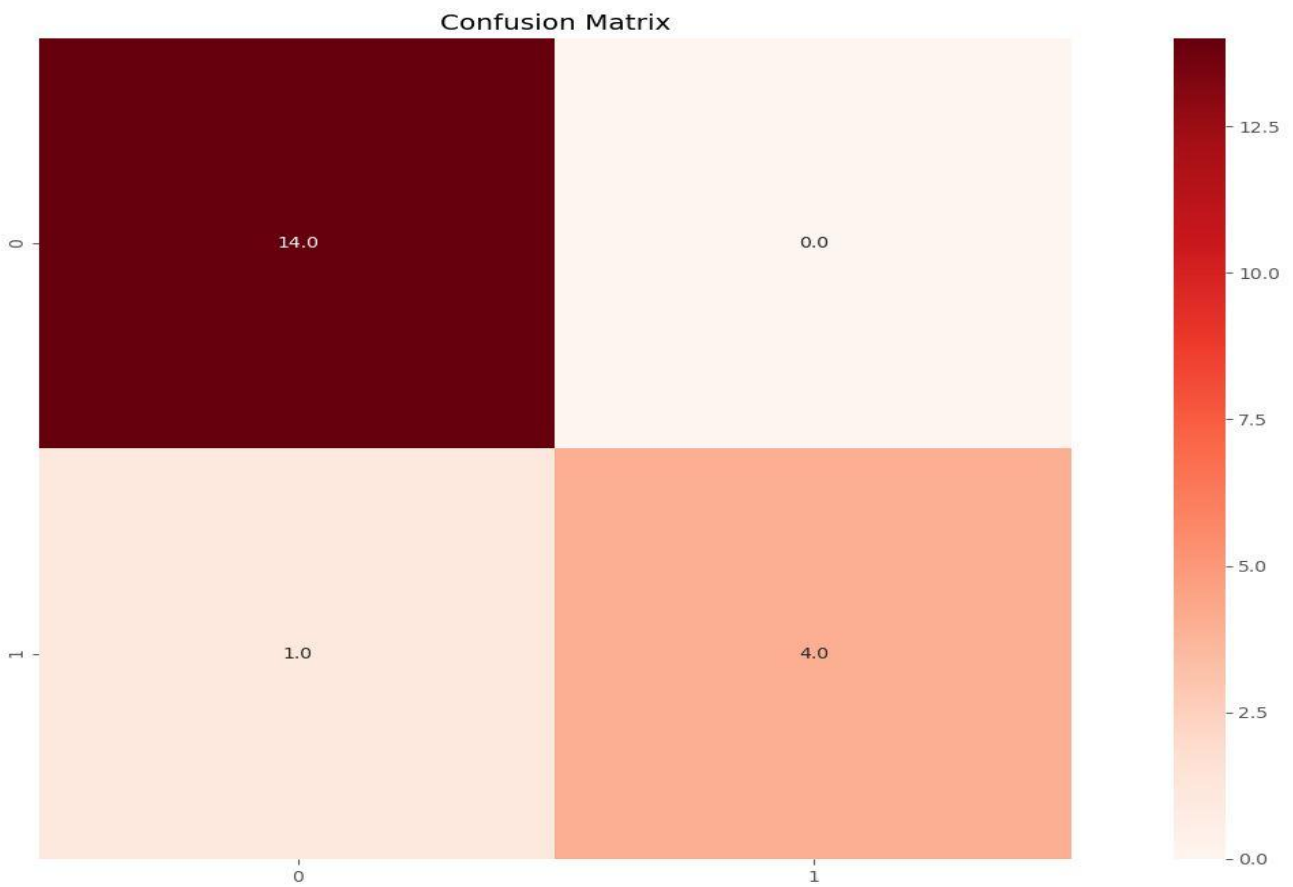


Fig 9: Heat map of confusion matrix of three fully connected layers acquired from the proposed approach. The x-axis shows the predicted labels and y-axis shows the actual labels



## 5. CONCLUSION

As NASNet Large has better performance in terms of top-1 and top-5 accuracy when used ImageNet weights, thus this paper also used that approach and also proposed some addition of classification layers to the network for optimal classification of COVID-19 from normal chest x-rays. The dataset consists of two classes and were less in number, thus transfer learning approach has been used. Analysis with two and three fully connected layers were done which showed that two fully connected layers when used with batch normalization gave better performance on the basis of test loss and p-score which are 0.0477 and 0.9643.

## 6. REFERENCES

- [1] <https://www.hindustantimes.com/world-news/covid-19-virus-accidentally-leaked-by-intern-at-wuhan-lab-says-us-media/story-15OJrBp6t9zdY9baRY8P1N.html>
- [2] [https://who.int/health-topics/coronavirus#tab=tab\\_1](https://who.int/health-topics/coronavirus#tab=tab_1)
- [3] <https://www.theguardian.com/world/2020/apr/17/what-is-coronavirus-what-are-its-symptoms-and-when-should-i-call-a-doctor>
- [4] <https://www.who.int/news-room/campaigns/connecting-the-world-to-combat-coronavirus/pass-the-message-to-kick-out-coronavirus>
- [5] <https://www.worldometers.info/coronavirus/coronavirus-cases/>
- [6] <https://www.worldometers.info/coronavirus/coronavirus-death-toll/>
- [7] <https://www.worldometers.info/coronavirus/>
- [8] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Neural Information Processing Systems*, v1, 2012, pp. 1097-1105.
- [9] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai and Tsuhan Chen, "Recent advances in convolutional neural networks", *Pattern Recognition*, v77, 2018, pp. 354-377.
- [10] Sergey Ioffe and Christain Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *Proceedings of 32<sup>nd</sup> International Conference on International Conference on Machine Learning*, v37, 2015, pp. 448-456.
- [11] Djork Arne Clevert, Thomas Unterthiner and Sepp Hochroiter, "Fast and Accurate Deep Network Learning by Exponential Linear Unit (ELUs)", *Proceedings of the International Conference on Learning Representations (ICLR)*, v5, 2016.
- [12] Aditya Kakde, Nitin Arora and Durgansh Sharma, "Novel Approach towards Optimal Classification of Multilayer Perceptron", *International Journal of Research and Engineering, IT and Social Sciences*, v8, 2018, pp. 29-38.
- [13] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V.Le, "Learning Transferable Architecture for Scalable Image Recognition", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT*, 2018, pp. 8697-8710.
- [14] Francois Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI*, 2017, pp. 1800-1807.
- [15] Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", 2015, arXiv:1409.1556 [cs.CV]
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Deep Residual Learning for Image Recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, 2016, pp. 770-778.
- [17] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens and Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, 2016, pp. 2818-2826.
- [18] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijum Wang, Tobias Weyand, Marco Andreetto and Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", 2017, arXiv:1704.04861 [cs.CV]
- [19] Gao Huang, Zhuang Liu, Laurens van de Maaten and Kilian Q. Weinberger, "Densely Connected Convolutional Networks", *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI*, 2017, pp. 2261-2269.
- [20] Joseph Paul Cohen, "COVID-19 image data collection", <https://github.com/ieee8023/covid-chestxray-dataset>, 2020.
- [21] Kermany, Daniel; Zhang, Kang's Goldbaum, Michael (2018), "Labeled Optical Cogence Tomography (OCT) and chest X-Ray Images for Classification", *Mendeley Data*, v2.