Covid-19 Patients Lung CT Scan Image Analysis using CNN Algorithms

Journal:	*IEEE Journal of Biomedical and Health Informatics*
Manuscript ID	JBHI-01118-2020
Manuscript Type:	Special Issue on Emerging Challenges for Deep Learning
Date Submitted by the Author:	01-Aug-2020

Complete List of Authors:	G S, UTHAYAKUMAR; St Joseph's Institute of Technology, Dept.of ECE K, VENKAT RAMNAN; PES University, Department of ECE

SCHOLARONE™
Manuscripts

Covid-19 Patients Lung CT Scan Image Analysis using CNN Algorithms

[1]Mr.K Venkat Ramnan , [2]Uthayakumar G.S,
[1]PES University, Ring Road Campus, Bangalore-560085
[2]Dept. of ECE, St.Joseph's Institute of Technology, OMR Chennai-600119
[2]E-mail id:guthaya29@gmail.com
[2]Mobile No.+917871895523

***Abstract*:-COVID-19 is the coronavirus family, which includes Severe Acute Respiratory Syndrome(SARS) and Middle East Respiratory Symptoms(MERS) virus. Coronavirus is the virus strains that cause the cold and flu in the human body. COVID-19 affects the lungs and in severe cases death due to Acute respiratory distress syndrome (ARDS). Currently the COVID-19 disease has affected over 17,201,686 people all over the globe and has claimed the lives of over 6,70,463 lakhs people. In order to test if a person has COVID-19 or not the most common way being done today is Reverse transcription polymerase chain reaction (RT-PCR) and blood tests (serology). Another more effective way is using Computed Tomography (CT) scans of lungs and X-RAY scan of the chest. The visible features on computed tomography images include multilobar ground-glass opacities with a peripheral or posterior distribution. In this research paper, an attempt has been made to use different Machine learning algorithms, such as Convolutional Neural Networks(CNN), with different architectures behind them in order to classify if a Computed Tomography scan image contains COVID-19 presence or absence. Here 9 different models of which 8 models use pre-trained deep learning architectures and one model uses basic CNN concept in order to classify the CT scans. The results of all these models are a range of accuracy values. The dataset contains 1252 images of COVID positive cases and 1229 images of COVID negative cases. The range of accuracy varies from 90 percent to 97 percent for different models. The final objective is to analyze how different models can be effectively used to identify COVID-19 in the CT scans of lungs and to prepare a case study on what has been observed. Once the accuracy is perfected by tuning the right model, the model can be applied in real life situation to directly detect COVID-19 using CT scans only.**

Key words: Computed Tomography, Convolutional Neural Networks, Covid-19, MERS, SARS

I.INTRODUCTION

As of July 2020, it has been six months since the World Health Organization(WHO) announced that COVID-19 to be a pandemic of worldwide concern. Almost 16 million cases have now been reported to WHO, and more than 640,000 deaths[1]. And the cases seem to accelerate so badly that more than 100,000 cases are being reported every day. COVID-19 was first detected in the Wuhan region of China as a mystery disease with effects like SARS and started causing pneumonia-like cough and fever situations in the region[2]. It was later then found that this COVID disease was indeed a SARS type of disease as the reproduction number of this virus was identified to be very similar to that of SARS and as of MERS [3]. Now that we have successfully identified that COVID is similar to pneumonia in the region it will affect mainly in the lungs. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis are some of the findings that prove that COVID affects the lungs mainly[4]. Furthermore, study has been done to observe and accumulate and analyze data based on CT scans of lungs of COVID patients after observing them for time course. As they observed the CT scans from the initial diagnosis to recovery, it has been concluded that lung abnormalities on chest CT scans showed the greatest severity approximately 10 days after the initial onset of symptoms[5].

The usage of Machine learning to perform image segmentation as well as conceive information from images is being used in almost a lot of applications from underwater to outer space. The introduction of the concept of Convolutional Neural Network brought a complete paradigm change in how an image can be classified and has brought out the importance of Computer Vision in many applications. One of the most important regions where we have applied CNN is in the field of radiology. CNN is essential to leverage its potential in diagnostic radiology, to augment the performance of radiologists and improve patient care[6]. Multi-scale CNN has been used to quantify lung nodule characteristics and thus classify lung nodules[7]. One of the most impressive models ever made for lung CT scan pneumonia detection is ChexNet and is very helpful in the field of radiology. CheXNet exceeds average radiologist performance on the F1 metric and thus has proved the importance of CNN application in lung CT scan based detection purposes[8].

When coming to the case of COVID-19, there have been many papers that incorporate the use of Artificial Intelligence, Machine learning, and deep learning to analyze the CT scans of lung images as well as using certain models that can predict if the lung scan is COVID positive or not [9][10][11]. This paper specifically shows how supervised deep learning can be used to analyze the CT scan images of lungs and gives a proper insight into this concept[12]. Although RTPCR is the most common method in order the detect if a given person has COVID or not, the application of the above ML models has brought light to the usage of only CT scans to detect this virus in the lungs.

In this paper, we attempt to use already pre-trained models of CNN on these CT scan images of lungs of COVID positive and COVID negative so that we can find the accuracy values for each of these models. We have used the most famous models including VGG16, Alexnet, and many more pre-trained models, and tried to apply these models on the same dataset containing close to 3000 images of both positive and negative images. The result that has been obtained is the accuracy values for each of the models. The accuracy values have been found between 90% to 97% which is a good accuracy value. Using these results we have written a case study looking through all the models and thereby prove the importance of usage of CNN in COVID radiology. These models can be changed or modified to obtain more precise results. We have also presented a simple CNN model which is also capable of giving accuracy in the range of 89 to 95 percent and with the highest accuracy obtained as 96.5% when more layers were added to the CNN model. These models will be highly useful since it can reduce the burden of RTPCR procedure and can be helpful to detect the disease.

II.METHODOLOGY

The concept of machine learning in images is applied using a concept called Convolutional Neural network or CNN or ConvNet. Here we are trying to classify the given images into primarily two classes : COVID positive or COVID negative. Usually most CNN models use a dataset of large number of images. Pre trained models perform well when the dataset is very large. Even though the count of COVID positive patients is very large worldwide ,the data set of CT scan images of COVID patients are small and scattered. Most of the data of images for this paper was obtained from public dataset available in kaggle.

A.DATASET

The dataset used is available for public in kaggle[13]. The datasets are clearly divided into 2 sub directories , COVID and non-COVID. The dataset is available as 231MB zip file in the given website. The dataset contains 1252 CT scans that are positive for SARS-CoV-2 infection (COVID-19) and 1230 CT scans for patients non-infected by SARS-CoV-2, 2482 CT scans in total. These data have been collected from real patients in hospitals from Sao Paulo, Brazil.

All the 9 models which are mentioned in this paper have been implemented using Python3.7 language. The libraries that were extensively used for Machine learning were Tensorflow v2.2.0 , Keras , scikit-learn. Other libraries like matplotlib and OpenCV 2 were used as well. The basic CNN model was implemented using a computer with 8GB RAM, i5-9th gen processor and Nvidia GTX 1050Ti GPU. All the other pretrained models were implemented using Google COLAB with GPU extension[14]. Most of the programming was done with help of Jupyter Notebook feature.

B.MODELS

Among the 9 models that have been implemented in this paper only one of them is not pre-trained while other are very famous pre-trained models that have been used in many applications. The pre-trained models are very complex and they differ in the way they are built. A glance into all these 9 models have been summarized as below:

C.Pre-trained Models:

1.AlexNet: AlexNet is the name of a convolutional neural network (CNN), designed by Alex Krizhevsky, and published with Ilya Sutskever and Krizhevsky's doctoral advisor Geoffrey Hinton. The model was found to perform better with utilization of GPUs[15]. This model has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax.[16] The usage of dropout concept was also used in order to reduce over-fitting. Alexnet is one of the best models that is clearly able to classify a given image. It has been trained over weeks in order to perform object classification. The image dataset used to train this model contained a massive 15 million images with 22 thousand classes. Fig 1 shows the structure of the alexnet architecture.

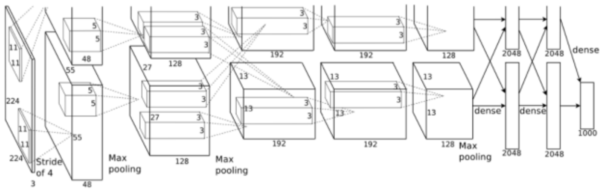


Fig 1.Illustration of AlexNet's architecture. Image credits to Krizhevsky et al., the original authors of the AlexNet paper.

2. VGG16 and VGG19 : VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford[17]. This model is considered to be one of the best models in the field of computer vision. The concept of VGG16 is that instead of having a large number of hyper-parameter as in AlexNet, it has convolution layers of 3x3 filter with a stride 1 and always uses same padding and Maxpool layer of 2x2 filter of stride 2. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture. In the end it has 2 FC(fully connected layers) followed by a softmax for output. The VGG16 refers to 16 layers that have weights while 19 of VGG19 refers to 19 layers with weights. This model has also been trained on the same ImageNet dataset containing millions of images. VGG stands for Visual Geometry Group at University of Oxford. The basic architecture of VGG is illustrated in Fig 2.

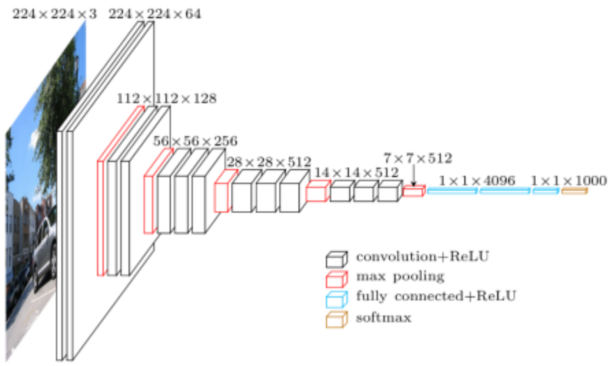


Figure 2. Architecture of VGG

3. MobileNet: MobileNet is one of the latest architectures that has only been released recently in 2017. It is gaining ground in the field of CNN as it is lightweight in nature, as the word mobile suggests. The model was proposed by A. Howard and team. MobileNets are based on an architecture that uses depth-wise separable convolutions to build light weight and faster deep neural networks[18].It uses depthwise separable convolutions which basically means it performs a single convolution on each color channel rather than combining all three and flattening it. This has the effect of filtering the input channels. As this model is light, it is extremely fast compared to the other models and thus is a very good model in the field of CNN. The architecture of MobileNet has been illustrated in Fig 3.

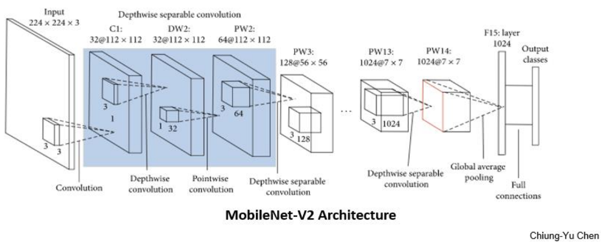


Figure 3. MobileNet architecture

4. Resnet 101 and Resnet 152: The main motivation behind this model is to enable thousands of convolutional layers. Resnet was built to solve the problem of vanishing gradient. Vanishing Gradient Problem is a difficulty found in training certain Artificial Neural Networks with gradient based methods (e.g Back Propagation). In specific, this problem makes it really hard to learn and tune the parameters of the earlier layers in the network. This problem worsens as the number of layers in the architecture increases[19].ResNet stacks up identity mappings, layers that initially don't do anything, and skips over them, reusing the activations from previous layers[20] This reusability feature enables it to be faster while learning and when the network trains over and over , the residual layers are explored more and more , eventually leading to excellent analysis of the given image. Resnet 101 has 101 deep neural layers while Resnet 152 has 152 layers. Resnet 50 is another model with 50 layers. Different Resnet models have different accuracy values in different situation with different set of labeled images. Fig 4 shows clearly the arrows which represent the reusing of the layers in Resnet.

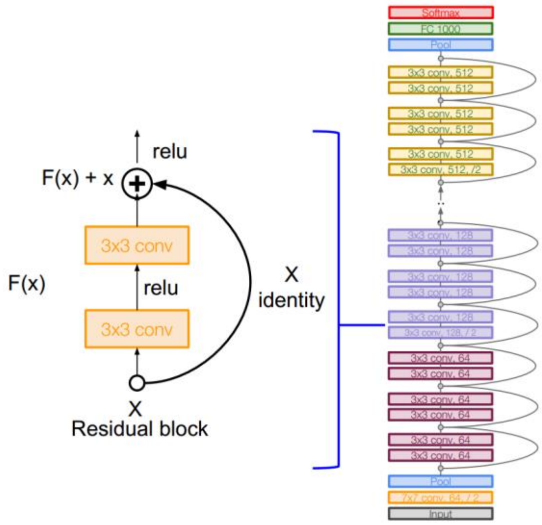


Figure 4: Resnet architecture

5. DenseNet 201: This model was built on the fact of the discovery of the point that shorter connections between layers closer to output and input can increase the overall performance of the model. Each layer is connected in feed forward fashion[21]. This model solves the vanishing-gradient problem, strengthens feature propagation, encourages feature reuse, and substantially reduces the number of parameters. DenseNet 201 is 201 layers deep. This model is extensively used in many classification problems in medical imaging field as well[22]. The Fig 5 illustrates a basic architecture of DenseNet.

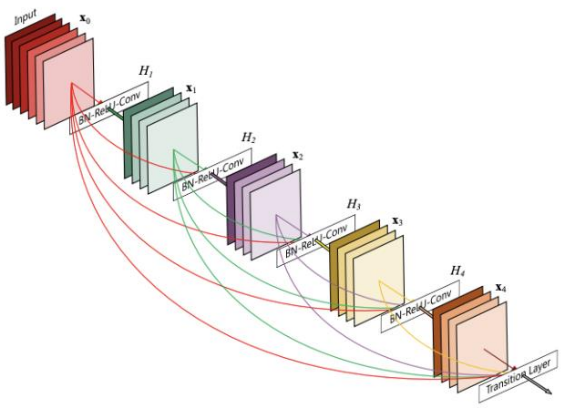


Figure 5. DensNet architecture

6. InceptionV3: This model created by Google has fewer parameters and is 42 layers deep. This model was built on the concept that overfitting and more computational resources are one of the major problems in deep learning. The solution proposed was to move on to sparsely connected network architectures which will replace fully connected network architectures, especially inside convolutional layers. The key innovation on the inception model is called the inception module. This is a block of parallel convolutional layers with different sized filters (e.g. 1×1, 3×3, 5×5) and a and 3×3 max pooling layer, the results of which are then concatenated[23].Thus this simple and powerful model has the ability to learn in multiple scales with the feature of parallelism. 1×1 convolutions are used to compute reductions before the expensive 3×3 and 5×5 convolutions. The use of ReLU layers also helps the performance of the model. Fig 6 shows the architecture image as showed in Google Cloud.

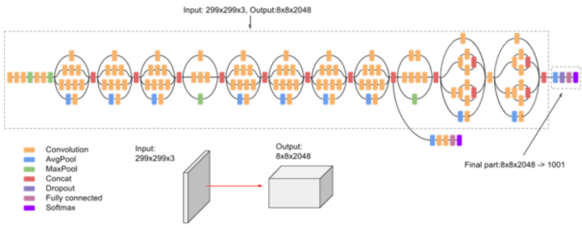


Figure 6: Inception Architecture

B. Basic CNN model:

As we have seen above the best models in the field of CNN and Image processing , we will now deep dive into the model created with basic CNN concepts in mind. In this model we convert the input images into 32x32 size and feed it to the model. To clearly understand the model the following terms that are used have been explained as below:

Conv2D: Keras Conv2D is a 2 dimensional Convolution Layer, which creates a convolution kernel that is wind with layers input which helps produce a tensor of outputs.This layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs.

Dense: These layers are dense connected layers with different activation functions.

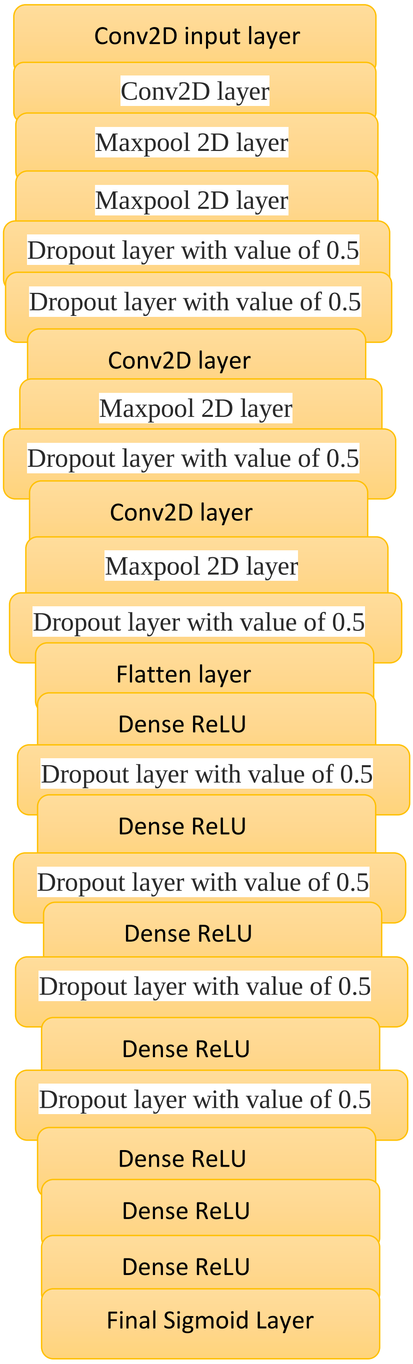
MaxPooling: Max pooling is a discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality, i.e,its size and allows for assumptions to be made about features contained in the sub-region.

Dropout: Dropout refers to ignoring units (i.e. neurons) during the training phase of certain set of neurons which is chosen at random. This method is used for preventing overfitting.

This model uses 3 Conv2D layers which are followed by Maxpooling layer and dropout layers respectively. Then it is followed by 4 sets of ReLU activated dense layers and Dropout layers. This set has 512 nodes each. Then it is followed by 3 ReLU layers with 512 nodes each. The final output layer is a sigmoid layer to give classification output. Given below is the flow of architecture of the proposed CNN model. The layers are arranged in successive manner as below:

D.PREPROCESSING

The Lung CT scan images were re-sized to different shapes depending on the model to which it was given as input. AlexNet model was given input of images with shape 227 x 227 pixels. MobileNet , VGG16,VGG19, Resnet 101, Resnet 152, InceptionV3 and DenseNet201 models were given input images with shape of 224 x 224 pixels. The proposed CNN was given image size of 32 x 32 pixels. The data was shuffled first and the models were experimented with a train test split scheme of having 80% for training and 20% for testing purposes. Most of the models used the concept of Image Data Generator to be fed as input to the model while 3 others , Proposed CNN, MobileNet and AlexNet , used the concept of directly converting images to arrays and then input them as tensors.



E.OPTIMIZER AND LOSS FUNCTION

All the experiments on different models used ADAM optimizer. ADAM was chosen over other optimizers because it gave the highest accuracy compared to the others. DenseNet201,Resnet101,Resnet152,VGG16,VGG19 and InceptionV3 were experimented with 'categorical crossentropy' function as loss function. AlexNet and MobileNet were experimented with "sparse categorical crossentropy" as loss function. Proposed Basic CNN model used ""Binary Crossentropy" as the loss functions. These specific functions were selected for each of these models, as the given loss function gave better accuracy for the given model as compared to the other loss functions.

F.TRAINING AND INVESTIGATIONS

Depending on the complexity and the way the models are built , different models were trained in different manners in order to obtain the best results. Some models used less number of epochs while other required more number of epochs to obtain better results. The batch size and number of epochs were also selected based on the model in order to get good results in a sufficient amount of time. Number of epochs varied from 20-30 epochs for certain models like VGG , Resnet , MobileNet and DenseNet. AlexNet and proposed Basic CNN were trained with 1000 epochs.

To clearly understand what was being identified in each of the image inside each CNN model, the concept of class activation mapping was used. Using this concept , it can observed which regions of the lung CT scan was each model discriminating to identify the category. It was possible to identify the importance of the image regions by projecting back the weights of the output layer on the convolutional feature maps obtained from the last Convolution Layer. Class activation mapping was used on each layer to understand which part of the lungs was being affected by the COVID-19. These activation maps color the region where the effect is profound. As it is being done in each layer of the model, it can be clearly observed how the model builds many copies of a single image with various different features highlighted in each of the copy. We can also observe how the model comes to the conclusion that is the given image positive or negative. As the accuracy obtained is not 100% we can also observe where the model fails. Heat maps of the images activation mapping in ach layer was also observed to get a clear idea of where the model was giving importance and how it was detecting the differences or regions of interest in each image. Attached here in the below figures we can see the activation mapping as well as the heat maps of one certain Positive confirmed image of lung CT scan for the Basic CNN model proposed.

III RESULTS AND DISCUSSION

As mentioned before all the models were experimented on a dataset of shuffled images which were split into 80 percent for training images and 20 percent for testing images. Each of the images were mapped with labels 0 or 1 for COVID or NON-COVID respectively.

A.DETECTION OF INFECTION

1) Class Activation Mapping

Because of the presence of multiple layers in the proposed Basic CNN model , Class activation maps for the all conv2D, maxpool2D and dropout layers were obtained in order for analysis. The Class activation maps reveal the way the model sees each image. Given below are a set of images containing class activation of all the mentioned type of layers.The Class activation maps color the region where profound effects are found[24]. It can be observed that as the image passes successively through all the layers of the model , there are regions with green color which are regions with small effects of infection and red regions which clearly show the confirmed area of infections. The activation maps of all the dropout layers specifically show the areas of infection. This way the regions which are being affected are clearly identified by the model. Another observation shows that how the model actually plays around with the images by applying filters in the given image to create multiple copies, each image implying the way the model learns and analyses the images. Further for the same image , the saliency map feature was also performed to confirm the regions of infection identified. As the layers work with each other, they are clearly able to pinpoint the regions of infection.

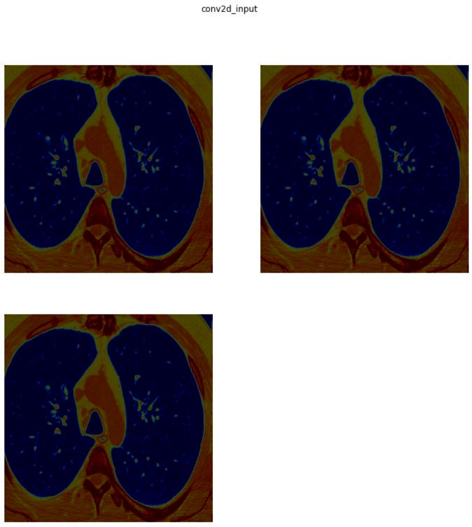
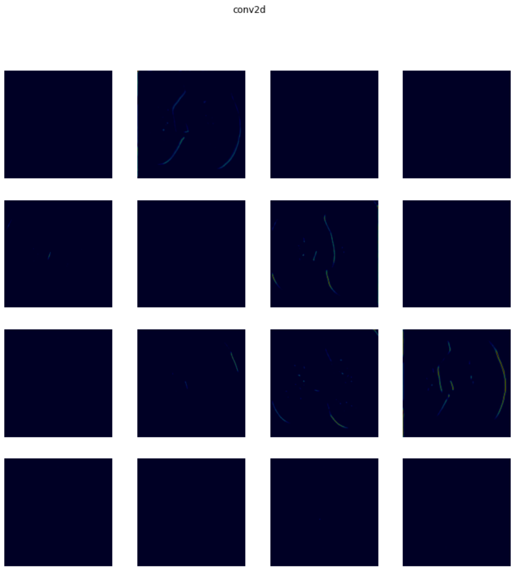
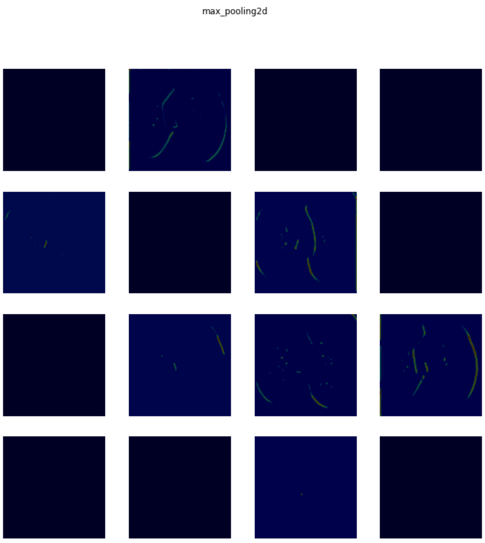
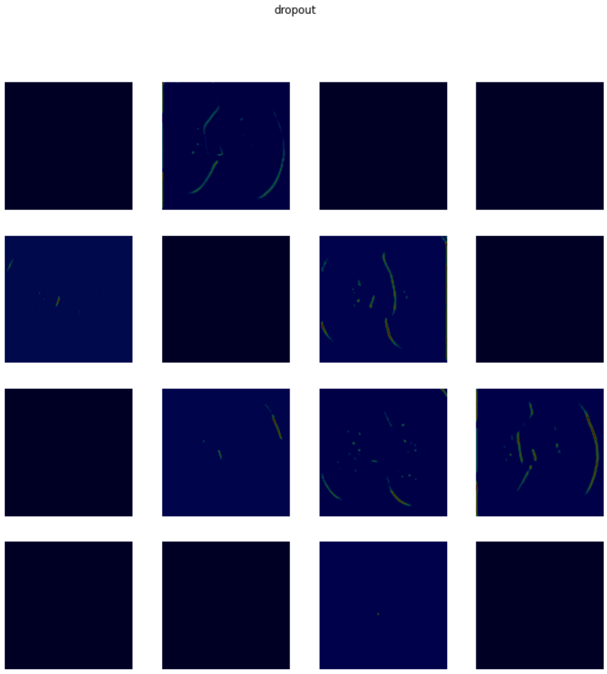
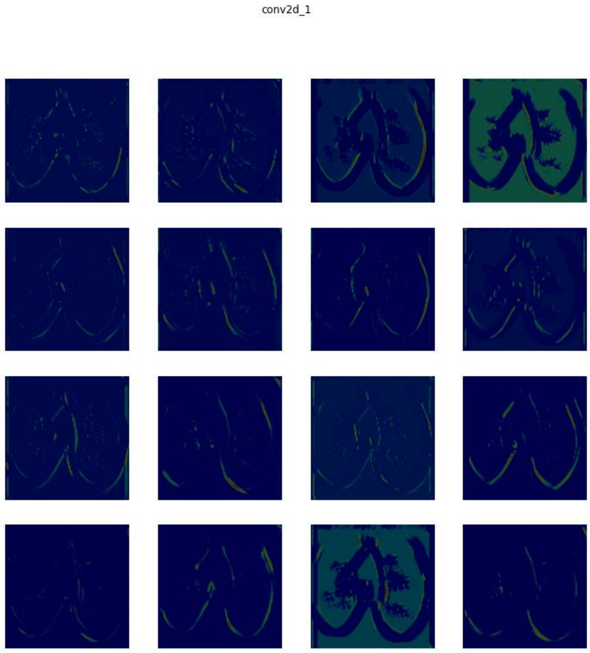


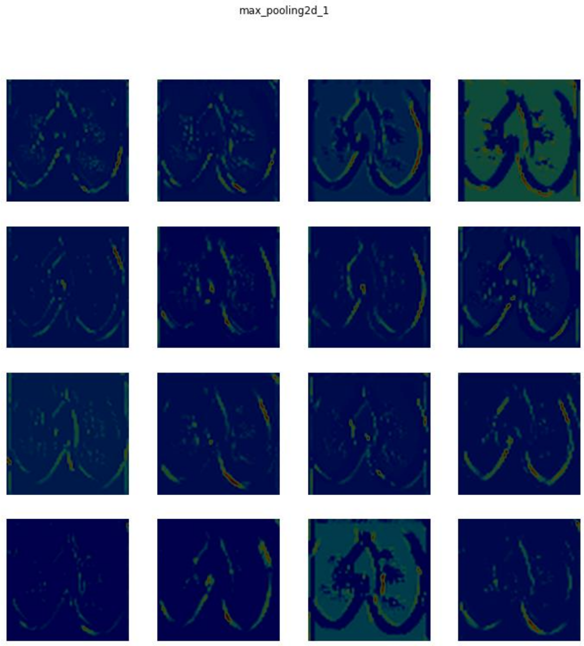
Figure 7. Input Conv2D layer activation map











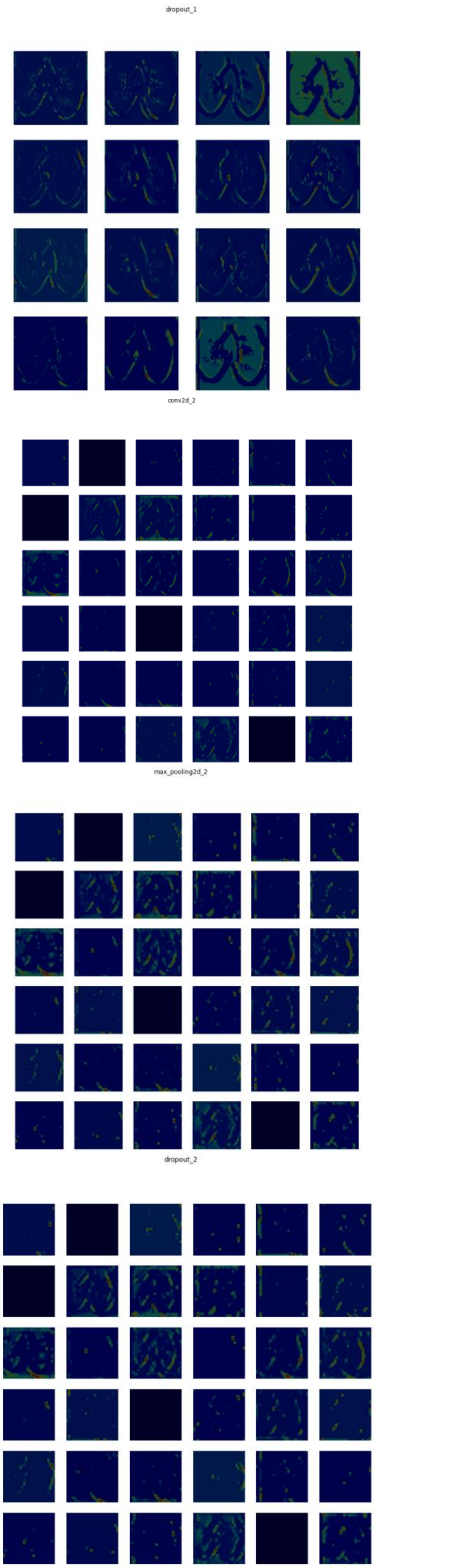


Figure 8. Activation maps of all 3 sets Conv2D, Maxpool2D and dropout layers in successive manner

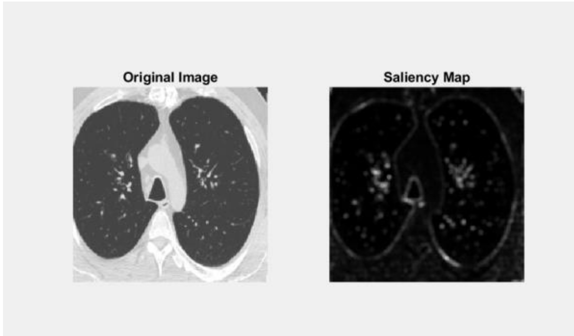


Fig. 9: Saliency map

Figure 9 shows the saliency map which is visualization technique which shows region that the CNN model identifies as region of infection.

2) Heat Maps

Similar to the class activation mapping, heat maps for each layer for the same input image was also obtained and analysed. Heat maps are a type of visualization methods captured the information in form of array of colors and has been used in medical imaging CNNs[25]. Heat maps were observed again for Conv2D, Maxpool and dropout layers only for the same proposed Basic CNN model. In the case of heat maps, it was observed that the regions that was infected was colored in yellow pixels. Although heat maps are not as precise as Class activation mapping, it was able to differentiate the regions infected and non infected. Heat maps also are one of the features which shows the models' perspective of looking at an image and draw conclusions from it. Below attached are the images of heat maps obtained from different layers of the model.

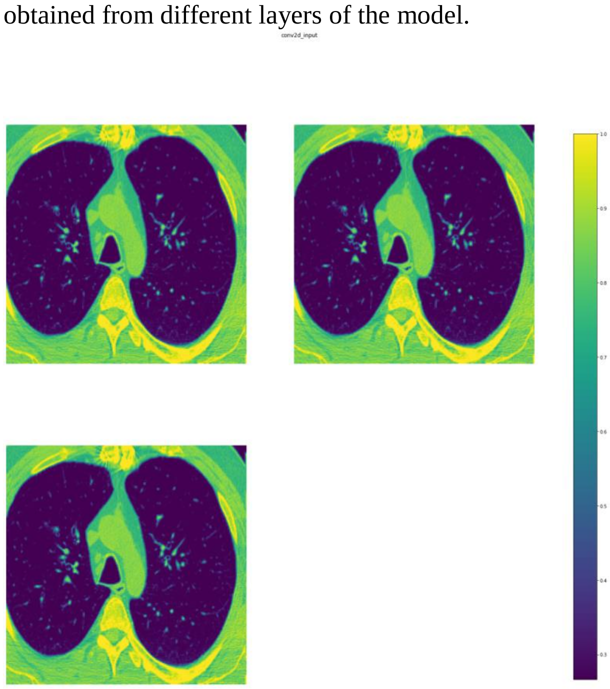
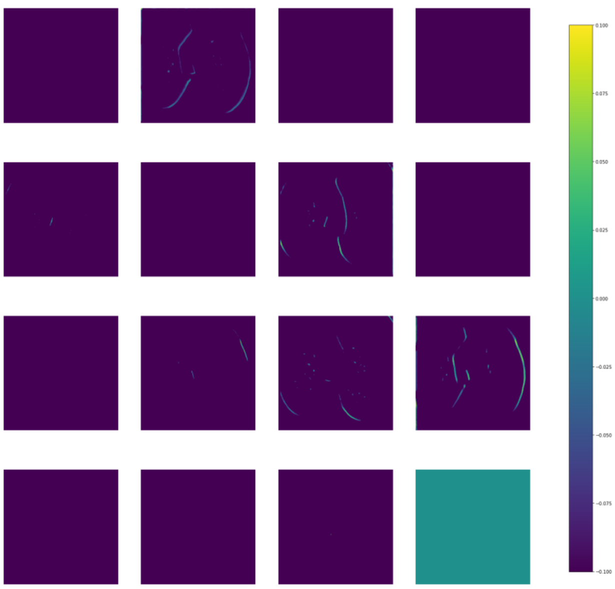
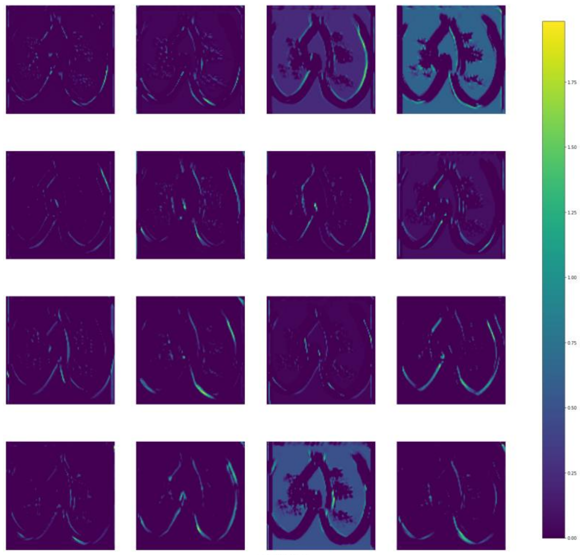


Figure 10.Input Conv2D layer heat map

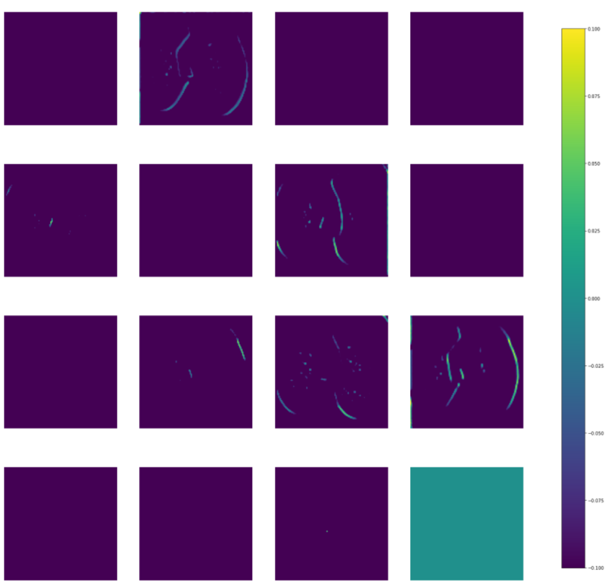
conv2d



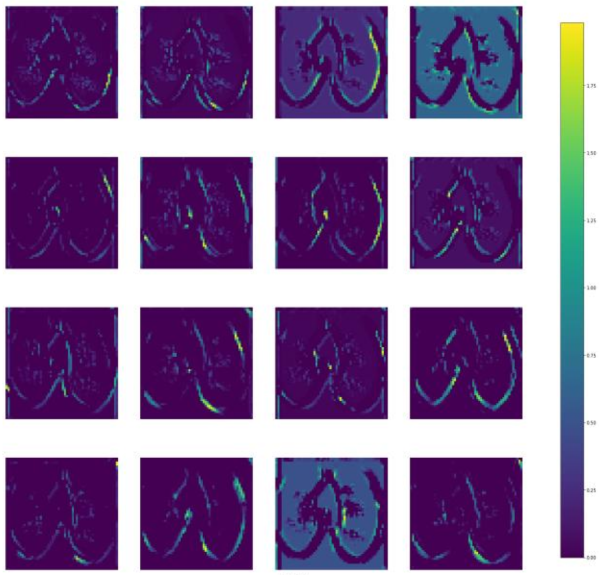
conv2d_1



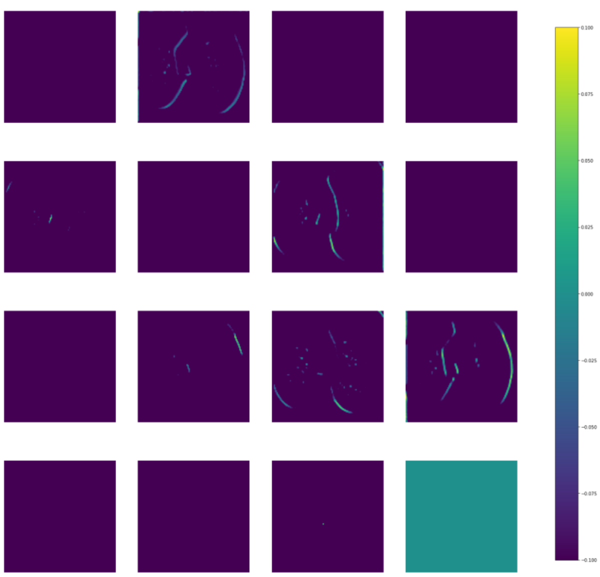
max_pooling2d



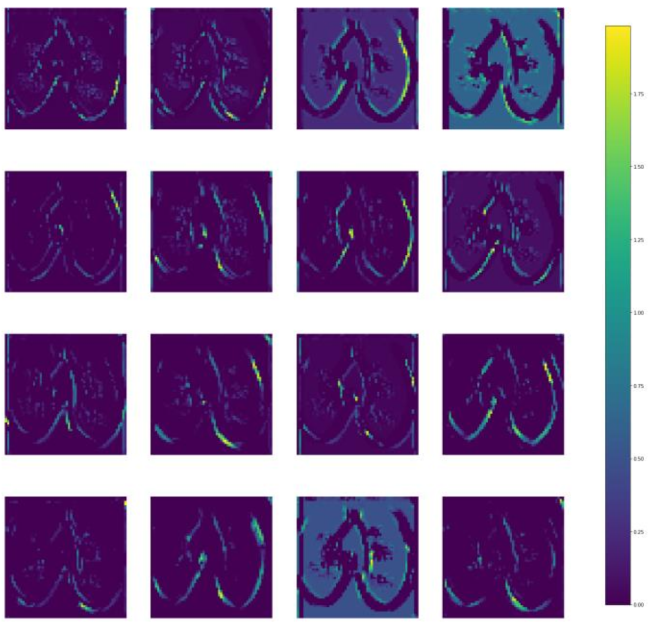
max_pooling2d_1



dropout



dropout_1



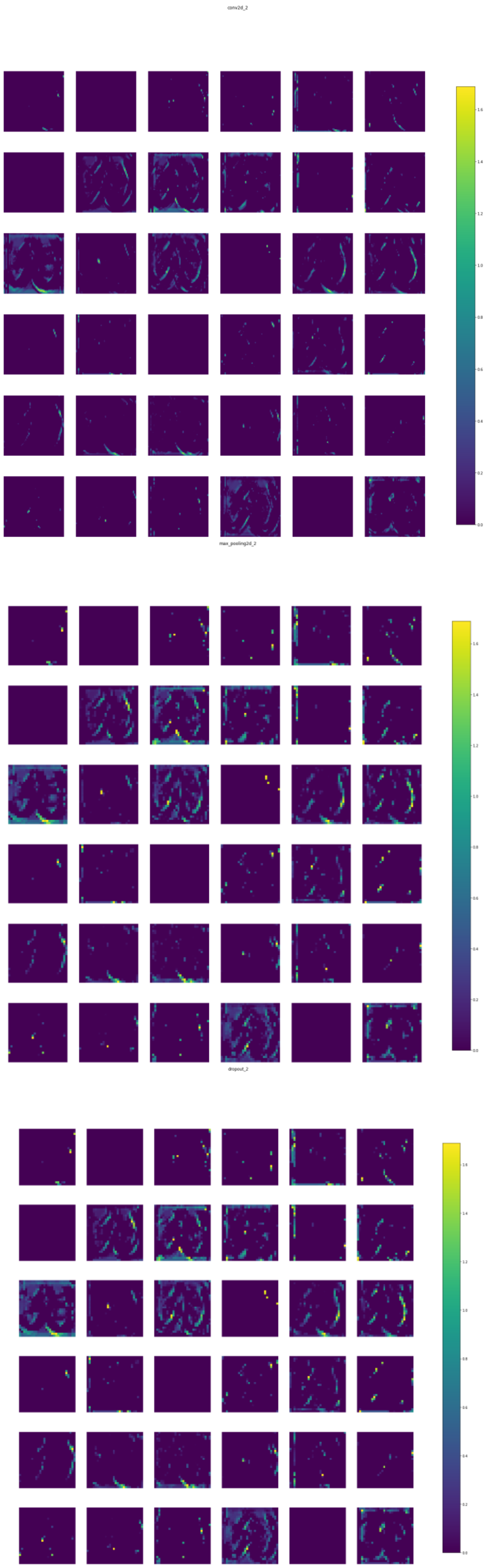


Figure 11. Heat maps of all 3 sets Conv2D , Maxpool2D and dropout layers in successive manner

B.CLASSIFICATION RESULTS

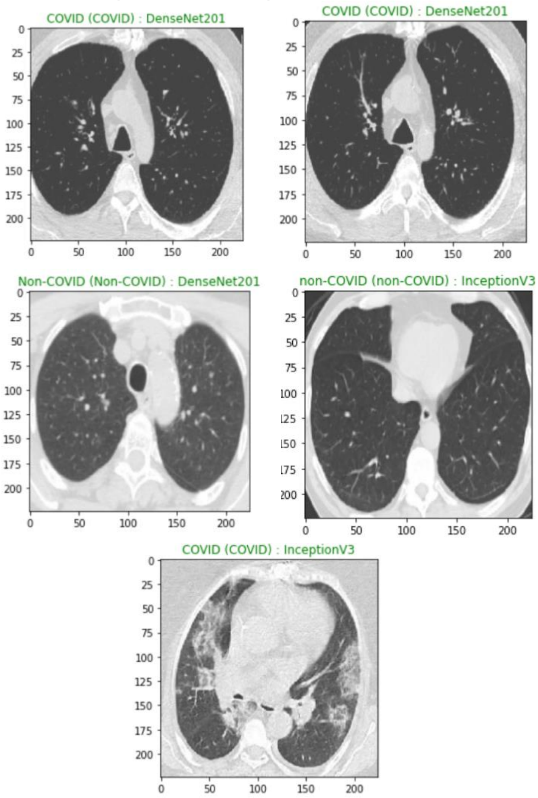
One of the primary objectives of this paper is to find the accuracy values obtained on the same CT scan image dataset and perform a comparative analysis on the results obtained. For this purpose as mentioned before , 8 pre-trained models and 1 Basic CNN model was built for this purpose.

Different models were given different sizes of images and were trained with different number of epochs and batch sizes. All these factors were changed accordingly and experimented multiple time for each model so that they could achieve the best range of accuracies.

Table 1 : Performance metrics of various models

MODELS	ACCU	PRE	F1-S	RECALL	AUC
Basic CNN	96.57	96.4	96.3	96.2	98.65
AlexNet	97.4	97.4	97.4	97.39	99.0
VGG16	97.18	97.17	97.13	97.09	98.88
VGG19	97.00	97.00	96.64	96.9	98.81
MobileNet	96.4	96.4	96.4	96.3	97.99
ResNet101	96.78	96.78	96.78	96.78	98.01
ResNet152	95.5	95.5	95.5	95.5	97.64
DenseNet201	93	93	93	93	95.54
InceptionV3	95.6	95.6	95.6	95.6	97.3

Each model after training were tested with both COVID and non-COVID images. The following figures show the correct detection of the given test images in some of the models.



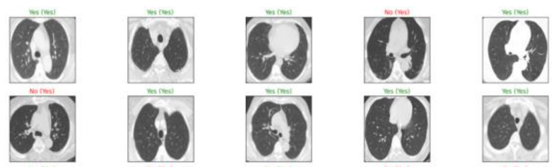


Fig 13:MobileNet results

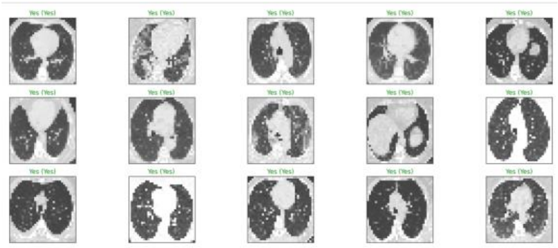


Figure 14: Basic CNN model results

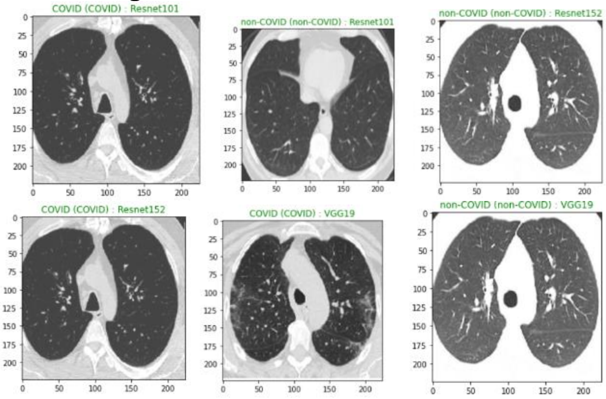


Figure 15: Outputs of Resnet152,Resnet101 and VGG19 models

In the above images the value inside the brackets is the original value and the one outside is predicted value. Further more if the predicted value matches original value, the text color is green or else it is red. These models will perform much better when the data set of images are increased to numbers of 50,000. Nevertheless the average accuracy range was found to be 96.16% which is a good value of accuracy.

IV.CONCLUSION

The result metrics that were focused were Accuracy(acc) , precision ,recall(sensitivity),F1-score and AUC score. All the models gave accuracies ranging from 90% to 97%. AlexNet gave the highest accuracy value of 98% in one of the multiple trials. The Basic CNN model proposed also performed well giving an accuracy value of 96.5%. The most impressive models that performed consistently were the VGG16 and MobileNet models. The table 1 gives the best results obtained in multiple trials of different models have been recorded.

References

[1] https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---27-july-2020

[2] Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle

[3] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, Joacim Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus, *Journal of Travel Medicine*, Volume 27, Issue 2, March 2020, taaa021, https://doi.org/10.1093/jtm/taaa021

[4] Ackermann, Maximilian, et al. "Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19." *New England Journal of Medicine* (2020).

[5] Pan, Feng, et al. "Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia." *Radiology* (2020): 200370.

[6] Yamashita, Rikiya, et al. "Convolutional neural networks: an overview and application in radiology." *Insights into imaging* 9.4 (2018): 611-629.

[7] Shen, Wei, et al. "Multi-scale convolutional neural networks for lung nodule classification." *International Conference on Information Processing in Medical Imaging*. Springer, Cham, 2015.

[8] Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).

[9] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, and S. Hu, ``Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study," *Medrxiv* Jan. 2020. [Online]. Available:https://www.medrxiv.org/content/10.1101/2020.02.25.20021568v2

[10]F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, ``Lung infection quantication of COVID-19 in CT images with deep learning," 2020, *arXiv:2003.04655*. [Online]. Available: http://arxiv.org/abs/2003.04655

[11]L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia, ``Articial intelligence distinguishes COVID-19 from community Acquired

[12]S. Hu et al., "Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification From CT Images," in IEEE Access, vol. 8, pp. 118869-118883, 2020, doi: 10.1109/ACCESS.2020.3005510.

[13] https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset

[14] http://colab.research.google.com/

[15] https://en.wikipedia.org/wiki/AlexNet

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 1097–1105.

[17] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[18] *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, Howard et al, **2017**.

[19] https://en.wikipedia.org/wiki/Vanishing_gradient_problem

[20] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2015-12-10). "Deep Residual Learning for Image Recognition". arXiv:1512.03385 [cs.CV].

[21] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[22] Zhou, Yufan, et al. "Holistic brain tumor screening and classification based on densenet and recurrent neural network." *International MICCAI Brainlesion Workshop*. Springer, Cham, 2018.

[23] https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html

[24] Zhou, Bolei, et al. "Learning deep features for discriminative localization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[25] Payer, Christian, et al. "Integrating spatial configuration into heatmap regression based CNNs for landmark localization." *Medical image analysis* 54 (2019): 207-219.