COVID-19 pneumonia: CT Imaging in Early Diagnosis and Key Differences of COVID-19 and Non-COVID-19 Pneumonia

Shunyu Yao, Yanxi Veritas Chen

November 5, 2022

Abstract

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus which was firstly found in Wuhan, China in December 2019, and was later declared as pandemic by the World Health Organization (WHO) on 11th March. Most patients infected with the virus have symptoms similar to cold or flu, including high temperature, sore throat, coughing, and fatigue. The virus spreads in particles from saliva, and can easily transmit while coughing, sneezing and speaking. However, the symptoms differ in people, and are frequently mistaken as normal cold or flu, hence likely to be ignored, leading to wider spread of the disease and delay in necessary treatments. Certain populations are more in risk of developing serious illness due to underlying medical conditions, including diabetes, cardiovascular diseases, respiratory diseases, cancer, etc. Thus it is essential to identify COVID-19 and treat accordingly.

In this article, we will demonstrate the model of identifying COVID-19 based on CT images, as well as methods and models of further distinguishing COVID-19 pneumonia and other types of pneumonia.

Contents

1 Introduction

Coronavirus disease is a severe acute respiratory syndrome caused by SARS-CoV-2 virus. The transmission of SARS-CoV-2 virus can be both direct and indirect, and is dependent on the respiratory secretions which cause both the direct and indirect transmission. The symptoms of COVID-19 overlap with those of normal cold and flu, including high temperature, coughing, sore throat, etc. Although some may have more severe symptoms than others, some may also not have symptoms at all. Thus it is very easy to ignore and be left untreated, leading to wider spread of virus.

Due to the high infection rate, it is essential to test immediately once patients start to show some symptoms of COVID. Most common tests include the NAATs(PCR-based tests) and the Antigen tests(Lateral Flow tests), both of which test specimens from the patient's nose or mouth. NAATs are considered the most reliable tests for people showing symptoms or no symptoms at all. The tests are normally performed in a laboratory setting, and they produce results in 1-3 days. However, as the tests detect viral genetic material that might stay in the patient's body for up to 90 days after first being tested positive, it is not recommended to have another test within the 90 days' window after being diagnosed with COVID. Lateral flow tests are the most convenient tests that can be performed at home and produce results in 15-30 minutes. Nevertheless, the Antigen tests are not as reliable as PCR-based tests, especially for testing patients with no symptoms. It is recommended that negative antigen tests should be repeated no less than 48 hours apart, and with another NAAT to further confirm the negative result. The specificity of PCR test is around 95%, which is relatively high compared with the Antigen test. However, due to potential low sensitivity rate caused by faulty kits, shortage in testing kits, and the demand for rapid diagnosis, chest-computed tomography (CT) has been frequently used in the diagnosis of COVID-19 under the pandemic setting.

According to a recent meta-analysis[1], CT imaging has demonstrated effectiveness and efficiency in real life, with a pooled sensitivity of 94% and a specificity of 37% with a reference of PCR diagnosis. In this article, we will aim to propose a model that has higher sensitivity and efficiency.

Although CT imaging has proven to be helpful in detecting abnormal patterns in the lung, which further helps with COVID-19 diagnosis, it is not always the case that the detected pneumonia is caused by SARS-CoV-2 virus. Hence more analysis should be done to distinguish the non-COVID-19 pneumonia with the COVID-19 pneumonia, which will be done in section 4.

2 CT-Dataset

2.1 Dataset Overview

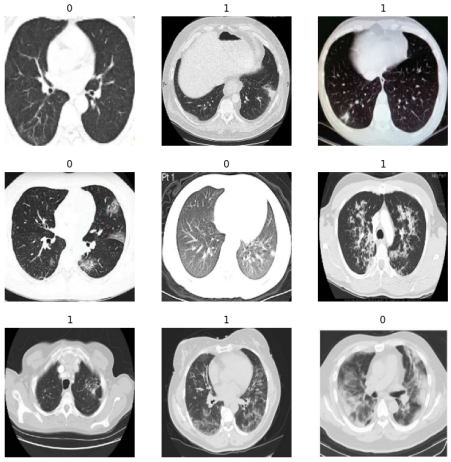


Figure 1: CT images preview

In this section, we will discuss how we built our dataset. We obtained the raw data from https://github.com/UCSD-AI4H/COVID-CT, which consists of 349 COVID-19 positive CT images from 216 patients, and 397 COVID-19 negative CT images. According to a senior radiologist in Tongji Hospital,[2], Wuhan, China, who has been diagnosing and treating COVID-19 patients since the outbreak of the disease in China, degradation of the original CT-images caused by resizing and extraction will not be a problem in diagnosis, and secondly, while it is preferable to read a sequence of CT slices, oftentimes a single-slice of CT contains enough clinical information for accurate decision-making.

2.2 Data Preprocessing

After we obtained the raw data, we did resizing and random sampling. Because the number of samples of non-COVID is 13.75% bigger than the COVID, we randomly removed 48 samples from the non-COVID dataset, in order to keep the balance between them. As a result, when generating dataloader and shuffling the input during each batch, the model will not be of a proclivity of predicting more

samples to be negative. Besides, the size of these images vary, we used openCV to resize the image to [3,224,224], with 3 channels of RGB, and a width x height of 224x224 pixels, so that it is compatible with the model.

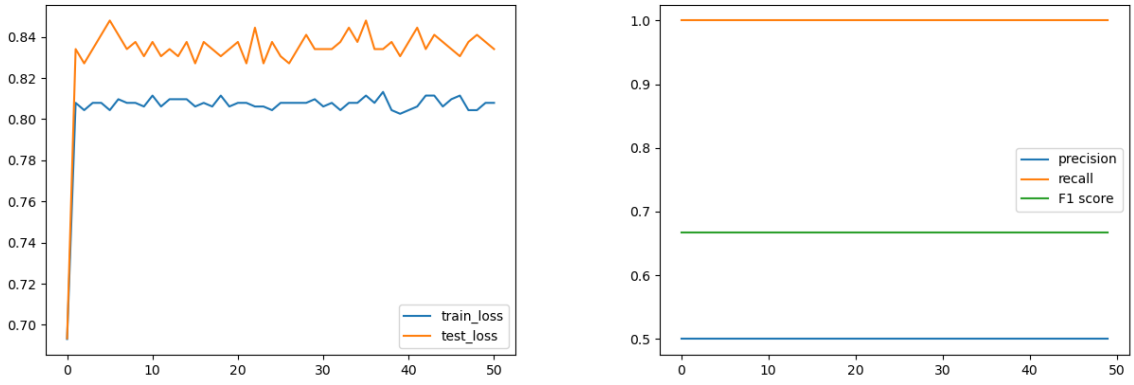
3 Deep Network and Deep Residual Network

3.1 VGGNet

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2: VGGNet

VGG stands for "Visual Geometry Group", which is a research group at the University of Oxford. In their work "Very Deep Convolutional NetWorks for Large-Scale Image Recognition"[3] in 2014, they proved that a significant improvement can be achieved when pushing the depth to 16, 19 layers, which corresponds to configuration D and E in the graph above.



(a) loss of VGG19

(b) precision-recall and F1 score

Figure 3: loss and precision-recall of VGG19

We used VGG19 to do the classification problem at first. However, it turns out to be hard to train and the result was unpleasant. It was time-consuming when training, with 13.86s/epoch on a laptop with RTX2060 and I7-10875. The training loss and testing loss were fluctuating, and cannot converge, with values even bigger than the ones after initialization. It simply classified every sample to be positive, with a precision of 0.5 and recall of 1. Hence we decided to account for residual network.

3.2 Deep Residual Network

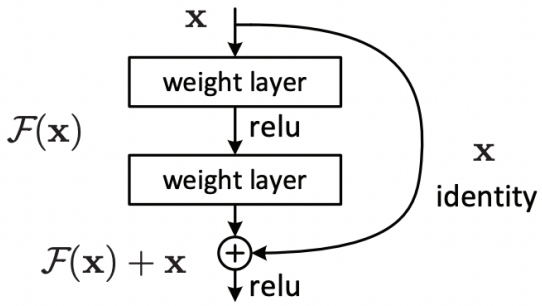


Figure 4: Residual Block

ResNet uses specific residual block to solve problem of vanishing gradients, or to mitigate the degradation problem. On the other hand, it uses convoluted kernel of size 1x1 instead of large full connected linear layer to downsize the output, which is easier to train.[4]A big linear layer like 4096 to 4096 in VGG19 has 4096x4096+4096 parameters, which is 16781312 = 16million parameters.

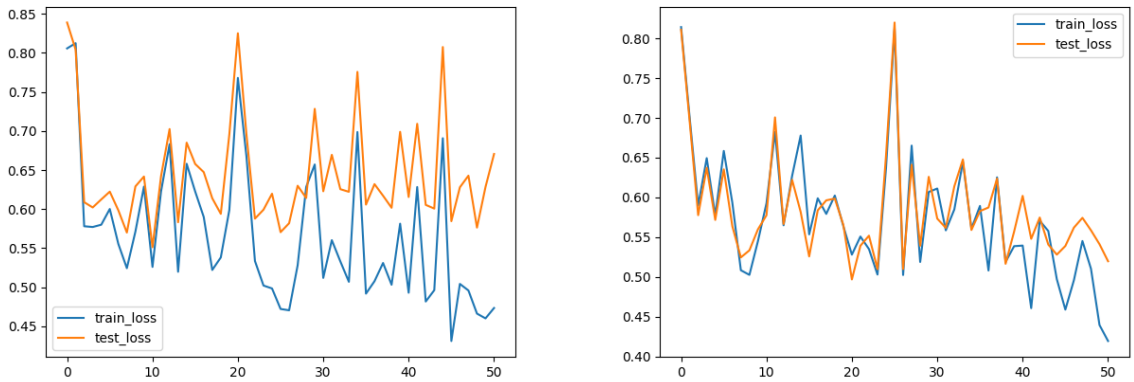
In each residual block, the output equals F(x)+x, with this skip connection or shortcut, the identity can always be promised. Thus the performance of a deeper network with more residual blocks should be no less than the one of shallower network's.

We compared the performance of customized ResNet34 and ResNet50, in hopes of checking whether a deeper ResNet leads to a better performance, as well as the cost accompanied? We carried out investigation on whether a deeper residual network would be harder to train compared with a plain deep network, and the results are listed below.

3.3 Results and Accuracy

We used the same hyper parameters for all models, including the batch size, learning rate, optimizer, learning rate decay and iterations.

3.3.1 Loss



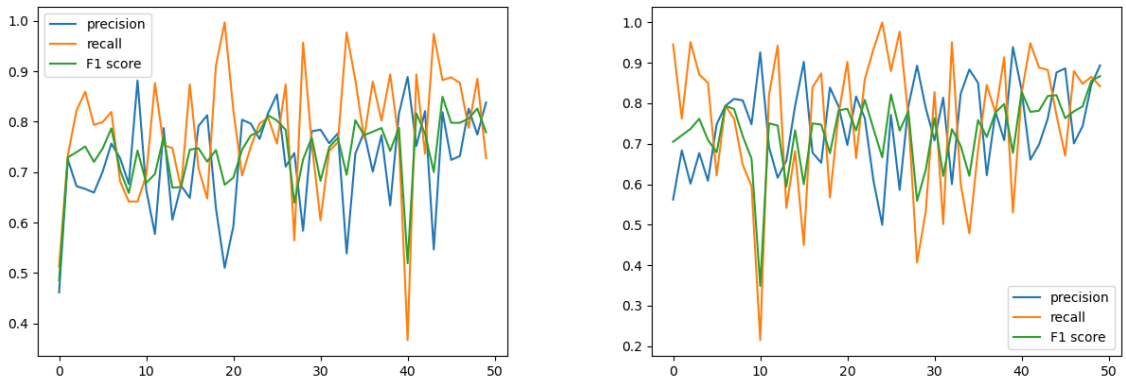
(a) Loss of Res34

(b) Loss of Res50

Figure 5: Loss for Models along Learning

Although both lines fluctuate a lot, the one of ResNet50 was relatively more stable. It is obvious that ResNet34 had overfitting problem on the training set, and the gap between the training loss and the testing loss was much larger in comparison with ResNet50.

3.3.2 Precision-Recall and F1 score



(a) ResNet34 (b) ResNet50

Figure 6: Precision-Recall and F1 score of both models

From the graph we can see that ResNet34 was actually more unstable than ResNet50, with more fluctuation in the graph. With drastic fluctuation at turning points like 18, 28, 33, 40 and 44 epoch. While on the ResNet50 line, the criterion had a more steady growth.

3.4 Model Comparison

Table 1: Model performance

Model Name	Precision	Recall	Accuracy	F1 score(highest)	At epoch
VGG19	0.5	1	0.5	0.6667	Each
Res34	0.8191	0.8825	0.8438	0.8496	44
Res50	0.8936	0.8672	0.8710	0.8672	49

Table 2: Model size

Model Name	Total Parameter	Parameter Size	Training time
VGG19	139,589,442	532.49 MB	13.86s/epoch
Res34	21,801,674	83.17 MB	3.41s/epoch
Res50	25,561,034	97.51 MB	6.01s/epoch

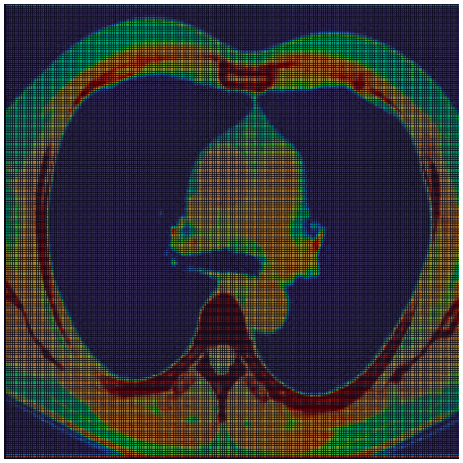
Due to the poor quality of graphs, we made tables to compare these models. The following conclusions were made:

- Deep plain network was rather hard to train, VGG19 could only classify every single sample into one category.
- A deep ResNet with more residual blocks(ResNet50) performed better than a shallower one(ResNet34). An accuracy of 3.22% and an F1 score of 2.07%, with a cost of 76.24% more training time and a parameter size that was 17.24% larger.
- Regardless, for each residual network, the highest F1 score occurred in 44 and 49 out of 50. We believe that with more iterations being trained, a better result could be reached.
- More variation in ResNet should be tested, i.e.: ResNet101 and ResNet 152, in order to observe the benchmark of diminishing marginal effect.
- Because of the imbalance between the number of positive images and the number of negative images, we removed 13.75% of the non-COVID CT images. Whereas given the relatively small dataset, we should have done a data augmentation and fix the problem.

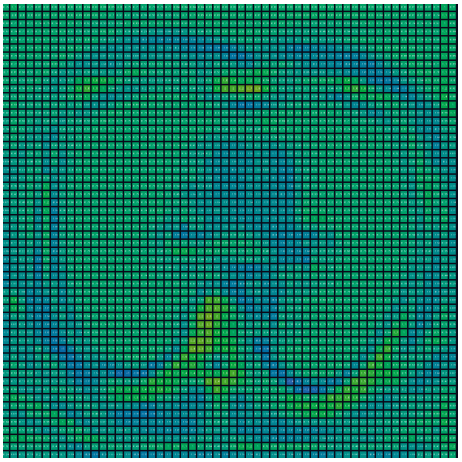
3.5 Adding Residual Blocks Manually

Although Res34, Res50, Res101 and Res152 are as classical deep residual networks, they do not necessarily fit in our dataset the best. We decided to visualize the output of each residual block and exam whether this block was learning something, or it was skipped, leading to the output simply being the input X.

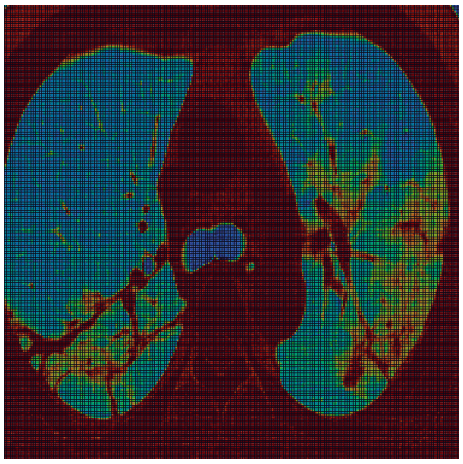
We converted the pytorch model in pth format into onnx format model, and loaded it in Zetane to visualize each layer.



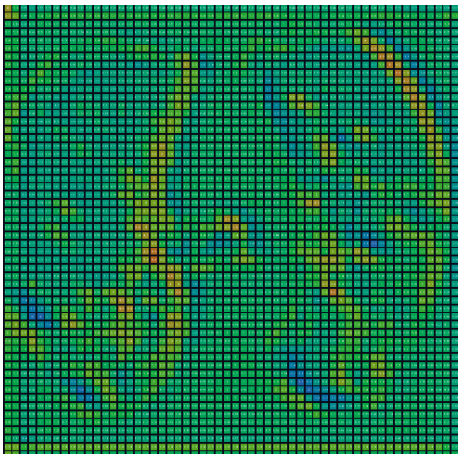
(a) Negative Input



(b) Output of the First Block (One of Several Channels)



(c) Positive Input



(d) Output of the First Block (one of several channels)

Figure 7: Heat Map of Negative/Positive Input and Output of the Residual Block

As shown in the picture, in the heat map, the model was actually learning that the cotton-shape tissues in the lung is imperative, and that the margin of the lung in the picture is reduced. In other words, the model had learnt that the edge of the lung image and the area outside was useless for diagnosis, and tissues in the pulmonary lobe was the most crucial for confirming the diagnosis. Thus, this block should be kept.

4 Distinguishing non-COVID-19 Pneumonia and COVID-19 Pneumonia

4.1 Pneumonia

Pneumonia is a lung inflammation caused by bacteria, fungus or virus, resulting in fluid or pus (purulent material)in the lungs. The symptoms include coughing, difficulty in breathing, fever, and chest pain. Generally, pneumonia differs in types according to the causes. However, it is hard to confirm diagnosis solely based on symptoms, as most of them share similar symptoms, and sometime resemble those of normal cold or flu. Hence CT screening has been commonly implemented in the process of differentiating pneumonia, as well as the decision-making in future medical treatments.

In this section, we will summarise the previous work done in differentiating non-COVID-19 pneumonia and COVID-19 pneumonia.

4.2 CT Imaging Features of COVID-19 Pneumonia

In the study "COVID-19 pneumonia: the great radiological mimicker", [5], typical CT features of COVID-19 pneumonia include bilateral ground-glass opacities (GGOs), predominantly peripheral, consolidations, combination of GGOs with consolidations, and GGOs superimposed with interlobular/intralobular septal thickening, creating a "crazy-paving" pattern and subpleural linear opacities. The following observations were also made: air bronchograms, vascular enlargement, CT halo sign, and reverse halo sign, with more likelihood of presence in non-COVID-19 pneumonia, which will be elaborated in the following subsections.

We will only discuss a few types of non-COVID-19 pneumonia here, focusing on the 3 most common causes of infection: virus, bacteria and fungus.

4.3 Normal Viral Pneumonia vs. COVID-19 Pneumonia

In this study, it was stated that the main difference between normal viral pneumonia and COVID-19 pneumonia is the existence of centrilobular nodular opacities, it is uncommon to observe such feature in COVID-19 pneumonia CT images. It is worth mentioning that similar to the previous outbreaks of SARS virus in 2002 and MERS virus in 2012, the positively marked CT images commonly lacked properties including cavitation, pleural effusion, and lymphadenopathy.

4.4 Bacterial vs. COVID-19 Pneumonia

Unlike COVID-19 pneumonia, it is less common to find GGOs in bacterial pneumonia. However, it is more common to find peribronchial thickening, centrilobular nodular opacities, and pleural effusion in bacterial pneumonia.

4.5 Fungal Pneumonia vs. COVID-19 Pneumonia

Fungal pneumonia is a result of direct infection in the pulmonary tissue when fungi are inhaled in spore form, either a single celled yeast, or a multicelled group mold. CT screening also indicated that centrilobular nodular opacities, cavitation, pleural effusion, and lymphadenopathy are more common in fungus infected lungs.

Although fungal pneumonia is not contagious, and most cases are not severe, it is potentially life-threatening for patients that are immunocompromised, i.e. HIV-positive or having gone through bone marrow transplant. For instance, pneumocystis pneumonia is a severe lung infection that usually occurs in people with weakened immune system, e.g. people infected with HIV. In comparison with COVID-19 pneumonia, features like upper lobe predilection, pulmonary cysts, as well as relative central lung involvement are found in pneumocystis pneumonia.

4.6 Model Comparison in Identifying Different Types of Pneumonia

In this section, we will summarise and compare two machine learning models developed for the purpose of distinguishing non-COVID-19 pneumonia and COVID-19 pneumonia in chest X-rays.

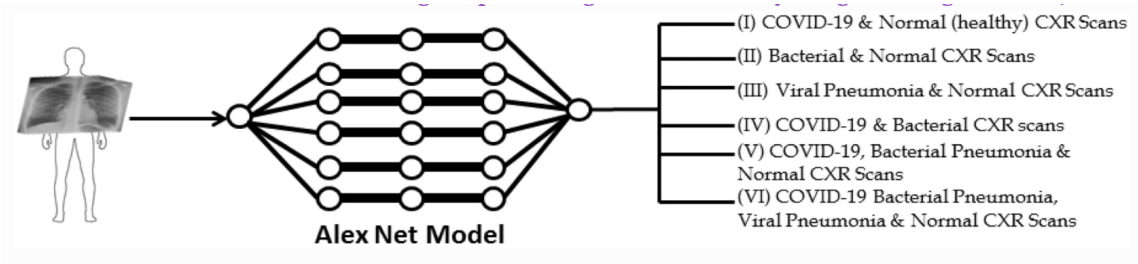


Figure 8: Workflow of the Proposed TL on AlexNet

The first article we studied was "Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19" [6] written by A.U.Ibrahim, M.Ozsoz, et al. In which they implemented

transfer learning on pretrained AlexNet model that uses ReLU as an activation function(Fig. 8). It was then used in classifications of (i) COVID-19 vs. healthy, (ii) bacterial vs. normal, (iii) viral vs. normal, (iv) COVID-19 vs. bacterial, (v) COVID-19, bacterial, and normal (vi) COVID-19, bacterial, viral, and normal.

S/N	Dataset	Training accuracy (%)	Testing accuracy (%)	Sensitivity (%)	Specificity (%)
I	Non-COVID-19 viral pneumonia and healthy	96.43	94.05	98.19	95.78
II	Bacterial pneumonia and healthy	95.28	91.96	91.94	100.00
III	COVID-19 and healthy	99.71	99.16	97.44	100.00
IV	COVID-19 and non-COVID-19 viral pneumonia	99.57	99.62	90.63	99.89
V	COVID-19, bacterial pneumonia, and healthy	97.40	95.00	91.30	84.78
VI	COVID-19, non-COVID-19 viral pneumonia, bacterial pneumonia, and healthy	94.18	93.42	89.18	98.92

Figure 9: Performance Statistics

As shown on the table (Fig 9), this is the performance statistics on the model, whilst classifying amongst different groups. As we can see, both the training and testing accuracy was very high, and it could definitely help a lot with differentiating COVID-19 and non-COVID-19 viral pneumonia. In all, Transfer Learning on Deep Learning models performed efficiently with small amount of dataset, in comparison with DL models built from scratch which normally requires large amount of data to reach optimal performance.

In the second paper we studied, which was "Diagnosing and differentiating viral pneumonia and COVID-19 using X-ray images" written by H.Kör, H.Erbay, A.H.Yurttakal, [7] they proposed a model which was developed from a pretrained NASNet-Mobile Convolutional Neural Network.

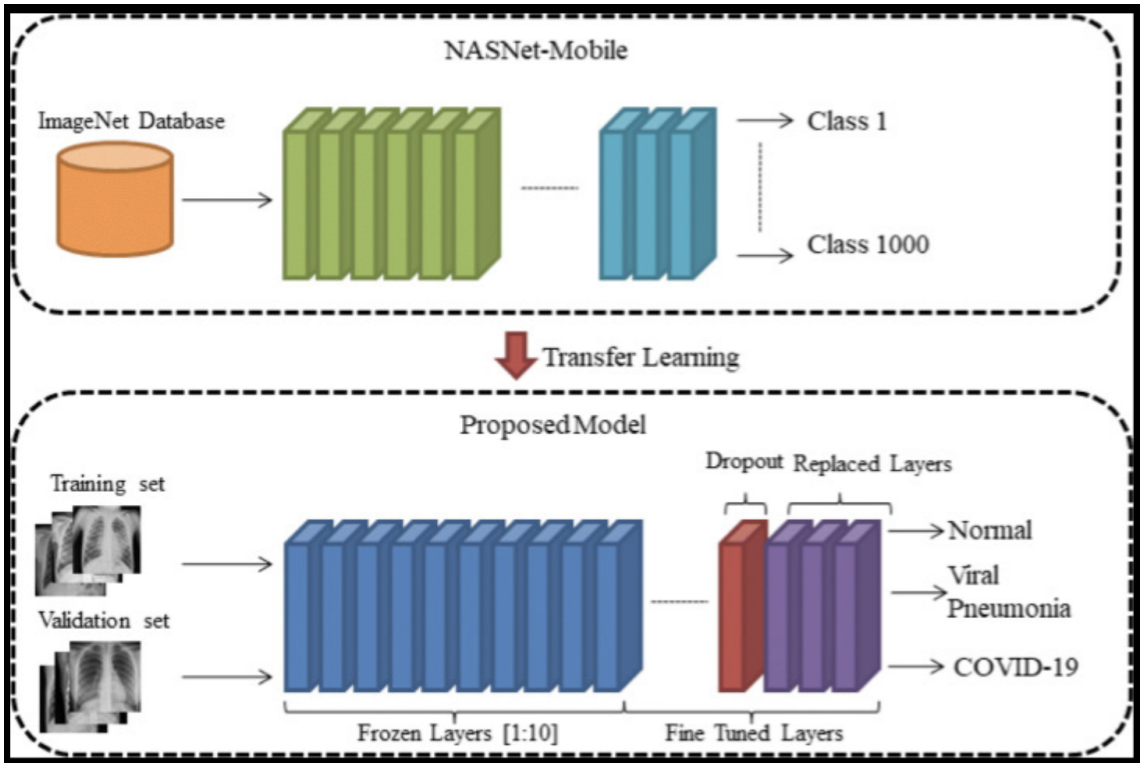


Figure 10: Workflow of the Proposed TL on NASNet

Figure 10 demonstrates the workflow of the model. In order to prevent overfitting, a dropout layer was added to the original network.

NASNet-Mobile performance statistics for the test set

Measure	Normal	Pneumoniav	COVID-19	Macro-Average
Accuracy	0,9667	1.0000	0.9667	0.9778
Recall	0.9500	1.0000	0.9500	0.9667
Precision	0.9500	1.0000	0.9500	0.9667
F1 Score	0.9500	1.0000	0.9500	0.9667
Negative Predictive Value	0.9750	1.0000	0.9750	0.9833
False Positive Rate	0.0250	0.0000	0.0250	0.0167
False Negative Rate	0.0500	0.0000	0.0500	0.0333
False Discovery Rate	0.0500	0.0000	0.0500	0.0333
False Omission Rate	0.0250	0.0000	0.0250	0.0167
Critical Success Index	0.9048	1.0000	0.9048	0.9365
Matthews Correlation Coefficient	0.9250	1.0000	0.9250	0.9500

Figure 11: Performance Statistics on the Test Set

Here is some performance evaluation on the testing set, the accuracy was proven to be really high. The model's macro-average accuracy rate on the training set is 0.9886 and 0.9778, and the macro-average recall is 98.29% on the training set and 96.67% on the test set.

5 Conclusion

In conclusion, our proposed model reached an accuracy of 87% in the validation dataset, with a prediction time within 0.2s, to diagnose covid-19 using processed CT image.

In terms of differentiating non-COVID-19 pneumonia and COVID-19 pneumonia, the research was mostly done in analysing chest X-rays instead of our focus which is CT screening. However, the models could still be studied further, and we could incorporate some key features to identify COVID-19 pneumonia with higher accuracy.

References

[1] Jessica Tsao, Andrea L. Kussman, Cristina Costales, Benjamin A. Pinsky, Geoffrey D. Abrams, and Calvin E. Hwang. Accuracy of Rapid Antigen vs Reverse Transcriptase–Polymerase Chain Reaction Testing for SARS-CoV-2 Infection in College Athletes During Prevalence of the Omicron Variant. *JAMA Network Open*, 5(6):e2217234–e2217234, 06 2022.

[2] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *medrxiv*, 2020.

[3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[4] Shaoqing Ren Kaiming He, Xiangyu Zhang and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[5] Selin Ardali Duzgun, Gamze Durhan, Figen Basaran Demirkazik, Meltem Gulsun Akpinar, and Orhan Macit Ariyurek. Covid-19 pneumonia: the great radiological mimicker. *Insights into Imaging*, 11(1):118, Nov 2020.

[6] Abdullahi Umar Ibrahim, Mehmet Ozsoz, Sertan Serte, Fadi Al-Turjman, and Polycarp Shizawaliyi Yakoi. Pneumonia classification using deep learning from chest x-ray images during covid-19. *Cognitive Computation*, Jan 2021.

[7] H. Kör, H. Erbay, and A. H. Yurttakal. Diagnosing and differentiating viral pneumonia and COVID-19 using X-ray images. *Multimed Tools Appl*, pages 1–17, Apr 2022.