Detection of COVID-19 in Neural Network Models

Zixuan Yao
Johns Hopkins University
zyao5@jhu.edu

Nairong Zhang
Johns Hopkins University
nzhang50@jhu.edu

ABSTRACT

The COVID-19 pandemic has already resulted in tons of deaths and infections, not even to mention its side effect towards economy, education and transportation. The problem of immediately detecting COVID-19 is receiving considerable attention nowadays within the entire world. The theoretical modeling of neural networks that would be produced in particular ideal accuracy in its simple classification task, and this task will expedite the diagnosis for COVID-19 and furthermore will provide assistance in respiratory disease. The characteristic respiratory disease will be presented shadow on patients chest x-rays, and by using classification on neural networks, its distinct feature will be detected at a second. In future, this method can be also used to predict the possibility of the respiratory infection and provide protection in a timely manner. Through the nature of neural networks and its superior ability in dealing with images, ideas on distinguishing COVID-19 with other respiratory disease, such as other bacterial and viral pneumonia are developed through the later part of this project. This work presents a 3-step technique to fine-tune a pre-trained ResNet-101 VGG 16 and CNN architecture to improve model performance and reduce training time. This is achieved through progressively resizing of input images to 224x224x3 pixels and fine-tuning the network at each stage. This approach enabled us to achieve state of more than 93.0% recall rate of COVID-19 with only 25 to 50 epochs. This model can help in early screening of COVID-19 cases and help reduce burden on healthcare systems.

KEYWORDS

COVID-19, neural networks, CNN, ResNet, VGG, classification

ACM Reference Format:
Zixuan Yao and Nairong Zhang. 2020. Detection of COVID-19 in Neural Network Models. In *Human-Robot Interaction.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The outbreak was identified in Wuhan, China, in December 2019. The World Health Organization declared the outbreak a Public Health Emergency of International Concern on 30 January, and a pandemic on 11 March. As of 9 May 2020, more than 3.93 million cases of COVID-19 have been reported in over 187 countries and territories, resulting in more than 274,000 deaths. More than 1.31 million people have recovered.

The Reverse Transcription-Polymerase Chain Reaction (RT-PCR) test is commonly used to screen COVID-19 patients, allowing those who get infected to receive treatment quickly. As the hospitals have collectively tested millions of people, some test results take more than a week. Testing of coronavirus remains among the most pressing problems with America's response to the pandemic. Patients suspected of COVID-19 are in urgent need of diagnosis and proper treatment — this calls for quick and accurate coronavirus diagnostics. As such, scientists across the globe are on the search for a more reliable assessment.

Since COVID-19 is a respiratory disease, it can cause a range of breathing problems. This caused a lung infection in which the alveoli are inflamed and doctors can see signs of respiratory inflammation on computed tomography. CT imageries provide high-quality 3D images of our lungs, which are useful for detecting the presence of COVID-19. With most scans only take just a few minutes, healthcare workers and researchers can acquire a large volume of these high-quality imageries. A 3D CT imagery contains 200–400 slices of images; this can take a long time for a specialist to diagnose. As COVID-19 has similar characteristics of other types of pneumonia, it would take a very experienced doctor at least 10 minutes to diagnose one patient. Thus an AI-assisted diagnosis using computer vision is highly desired.

A number of studies has shown the ability of neural networks, especially convolution neural networks to accurately detect the presence of COVID-19 from the CT-scans. However, the challenge is the datasets are not always available to public or those who are less professional on medicine but developed neural network models. Adrian X. were able to achieve over 80% accuracy based on VGG while working with three classes of healthy, bacterial, and Covid-19 images base on. We plan to utilize different architectures to model the X-ray data, in the hope of establishing open access techniques to fight against this pandemic.

Hopkins, May 08–03, 2020, Baltimore, MD
2020. ACM ISBN 978-1-4503-XXXX-X/18/06…$0.00
https://doi.org/10.1145/1122445.1122456

2 METHOD

Datasets

We used open datasets online as our training and testing dataset. We manipulate the datasets to form our own balanced dataset based on several criteria, which will be elaborated below. COVID-19 image data collection on GitHub includes 201 Chest X-rays images of lungs of patients suspected COVID-19. Among them, 155 images of the patients are confirmed positive with COVID-19. Other labels include ARDS, Bacterial Pneumonia, Legionella, MERS, No Finding, Pneumocystis, Pneumonia, SARS, Streptococcus, Viral Pneumonia etc. The sample of these labels are too few in amount to train neural networks. They are basically abandoned due to the complexity of labels and fewness in number. Among the 155 COVID-19 Chest X-rays images, some images are from a single patient shooting from different angles or on different days. So data filtering was performed to select only the front angle Chest X-rays images and only one image from one patient with the earliest possible shooting date. Therefore, only 100 COVID-19 Chest X-rays images are selected.

Figure 1: Data Filtering

Chest X-Ray Images (Pneumonia) on Kaggle includes 5,863 images, which are classified into 2 categories: normal and pneumonia. The pneumonia class is further classified into bacterial pneumonia and viral pneumonia. Again, we have a single patient contributing to multiple Chest X-Ray images from different angles and days. We also performed the data filtering by selecting only the front angle Chest X-rays images and only one image from one patient with the earliest possible shooting date. Another important finding is that some Chest X-Ray images are classified as bacterial pneumonia; while some Chest X-Ray images from the same patient are classified as viral pneumonia. We just abandon those data points again. To maintain a balanced training dataset, only 150 data points classified as pneumonia are selected (half classified as bacterial pneumonia, half classified as viral pneumonia) based on the above data filtering.

Network Architecture

Rather than proposing our own architecture, first we have leveraged the knowledge from a pool of the already existing Convolution Neural Network architectures that have shown excellent results using a wide variety of classification tasks. We designed three stages and each stage will perform a different network model, in the first two steps, we implemented VGG and ResNet with transfer learning, in the third step, we constructed an Convolutional Neural Networks by ourselves based on the experience of the previous two steps.

ResNet: Our first model is a convolutional neural network using ResNet 101 as the backbone. It takes a series of X-rays as input and generates a classification prediction of the X-Ray image. The CNN features from each slice of the X-rays are combined by an average-pooling operation and the resulting feature map is fed to the fully connected layers to generate a probability score for each class.

VGG:. Our second model is a convolutional neural network using VGG 16 as the backbone. It also takes a series of X-rays as input and generates a classification prediction of the X-Ray image. It has 13 convolutional layers, 5 max pooling layers and 3 dense layers. The weights of the first 13 convolutional layers are fixed using transfer learning technique. Then, the CNN features from each slice of the X-rays are combined by an average-pooling operation and fed to the last 3 dense layers, which are trained on our own Chest X-Ray dataset.

For the above two models, we will perform transfer learning by using a pretrained ResNet and VGG on ImageNet, given the small dataset we have. The last three fully connected layers are replaced and trained on the Chest X-Ray dataset, as the following figure shows.

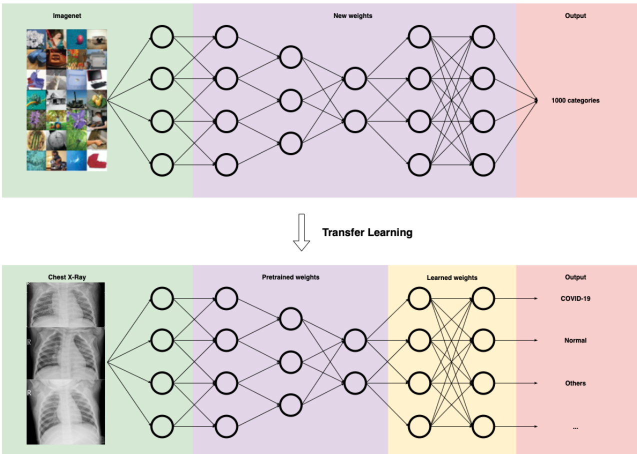

Figure 2: Transfer Learning

Vanilla CNN:. The main motivation for this architecture was to see whether X-ray image data can be modeled in a simpler architecture compared to VGG and ResNet. After multiple tries, our optimal architecture has 3 convolutional layers, max pooling, dropout, and uses ReLu as activation function. The self constructed CNN utilized custom layer design. After multiple tests and changes, we arrived at an architecture that provided great training speed and accuracy.

Training Network

After the data filtering, we further separate the dataset into three different sets of data for training in order to solve 3 different tasks: separating the COVID-19 patients from healthy people (2 classes classification), separating COVID-19 patients, other pneumonia patients and healthy people (3 classes classification) and separating COVID-19 patients, other bacterial pneumonia patients, other viral pneumonia and healthy people (4 classes classification). The first set contains only 200 images with 100 marked normal and 100 marked COVID-19 positive. The second set contains 225 images with 75 marked normal, 75 marked COVID-19 positive and 75 marked with other pneumonia. The last set contains 300 images with 75 marked normal, 75 marked COVID-19 positive, 75 marked with other bacterial pneumonia and 75 marked with other viral pneumonia. The viral and bacteria specifically refer to lung diseases that involve viral or bacteria infections. For each dataset, we trained the three networks mentioned above: VGG with transfer learning, ResNet with transfer learning and self-constructed CNN.

Therefore, we have in total 3 models × 3 tasks = 9 experiments. The results will be shown in the next section.

First, the data points are randomly separated into two parts: 80% for training and 20% for testing. Only training data points can be seen by the model. The training data points are pre-processed by converting to RGB channel ordering, and resizing it to 224×224 pixels so that it is ready for our Convolutional Neural Network. For Resnet and VGG, the convolutional weights are initialized by setting them to be the weights of ImageNet (pretrained). Then, the output of the convolutional layers is collected by an average-polling and fed to 3 dense layers with dropout to generate a probability score for each class. The class is assigned to the one with the highest probability score.

3 RESULT

The training loss with training accuracy and validation loss with validation accuracy for each experiment are shown from Figure 3 to Figure 11. The detailed evaluation table for each experiment is under each figure from Table 1 to Table 9.

For the two classes classification task, CNN and VGG performs the best with 100% and 97.5% testing accuracy, while ResNet can only achieve 90% testing accuracy.

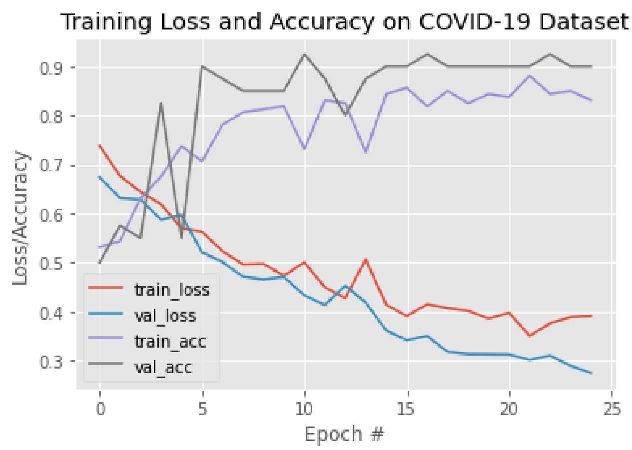

Figure 3: 2 classes ResNet

2 classes ResNet	precision	recall	f1-score	support
covid	0.86	0.95	0.90	20
normal	0.94	0.85	0.89	20
accuracy			0.90	40
macro avg	0.90	0.90	0.90	40
weighted avg	0.90	0.90	0.90	40

Table 1: 2 classes ResNet Evaluation Table

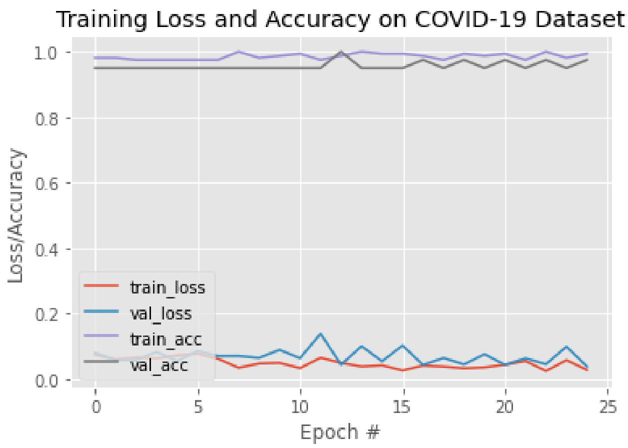

Figure 4: 2 classes VGG

2 classes VGG	precision	recall	f1-score	support
covid	0.95	1.00	0.98	20
normal	1.00	0.95	0.97	20
accuracy			0.97	40
macro avg	0.98	0.97	0.97	40
weighted avg	0.98	0.97	0.97	40

Table 2: 2 classes VGG Evaluation Table

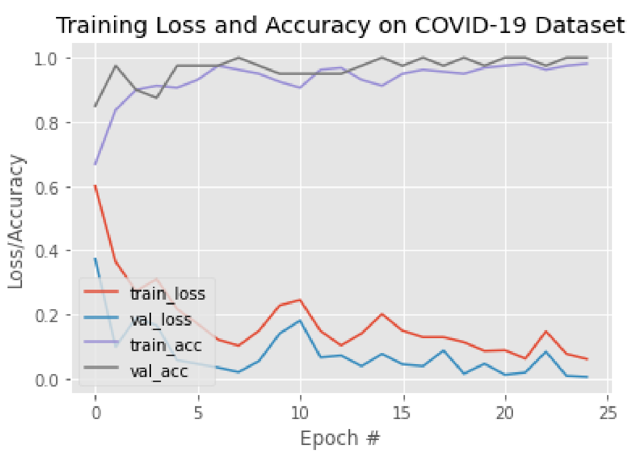

Figure 5: 2 classes CNN

Note: VGG achieved very high validation accuracy at the beginning of the training. At first, we thought we might forget to initialize the model weights. However, we initialized and ran it for sevral times and still go the same result.

2 classes CNN	precision	recall	f1-score	support
covid	1.00	1.00	1.00	20
normal	1.00	1.00	1.00	20
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

Table 3: 2 classes CNN Evaluation Table

Possible explanation would be: for VGG, the two classes classification is too simple, the fully connected layers even do not need to learn much in order to separate the two classes. The phenomenon disappeared for 3 classes and 4 classes tasks.

In the three classes classification task, ResNet performs much worse with only 65% testing accuracy while VGG and CNN still have fairly high testing accuracy: 93% and 91% respectively. The loss is of ResNet is still very high and the accuracy fluctuate a lot.

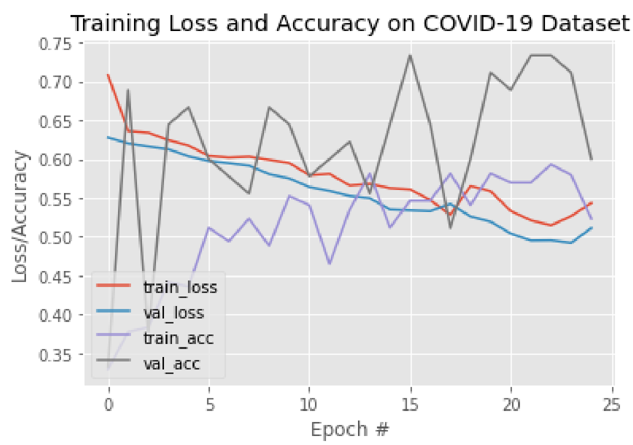

Figure 6: 3 classes ResNet

3 classes ResNet	precision	recall	f1-score	support
covid	0.52	1.00	0.68	15
normal	1.00	0.60	0.75	15
other pneumonia	0.43	0.20	0.27	15
accuracy			0.60	45
macro avg	0.65	0.60	0.57	45
weighted avg	0.65	0.60	0.57	45

Table 4: 3 classes ResNet Evaluation Table

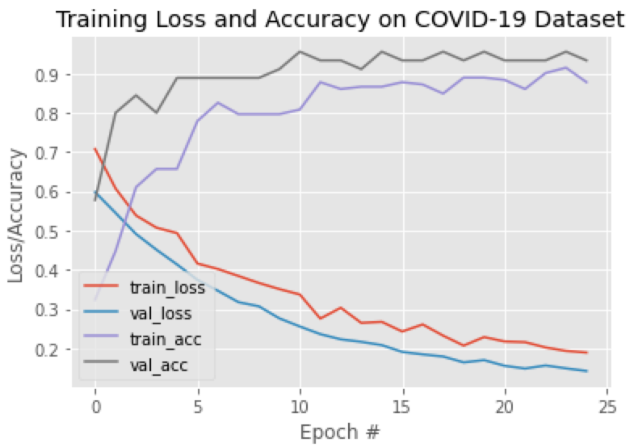

Figure 7: 3 classes VGG

3 classes VGG	precision	recall	f1-score	support
covid	1.00	0.93	0.97	15
normal	0.88	1.00	0.94	15
other pneumonia	0.93	0.87	0.90	15
accuracy			0.93	45
macro avg	0.94	0.93	0.93	45
weighted avg	0.94	0.93	0.93	45

Table 5: 3 classes VGG Evaluation Table

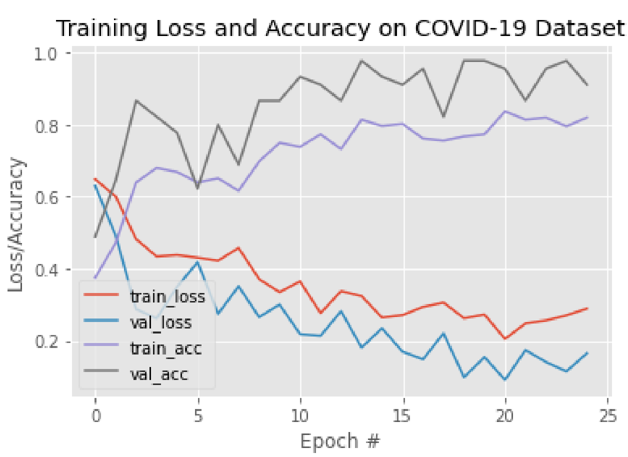

Figure 8: 3 classes CNN

3 classes CNN	precision	recall	f1-score	support
covid	1.00	0.93	0.97	15
normal	0.87	0.87	0.87	15
other pneumonia	0.88	0.93	0.90	15
accuracy			0.91	45
macro avg	0.91	0.91	0.91	45
weighted avg	0.91	0.91	0.91	45

Table 6: 3 classes CNN Evaluation Table

In the four classes classification task, ResNet still performs the worst with only 65% testing accuracy. However, the performance of VGG and CNN are much lower than the previous two tasks with only 79% and 73% testing accuracy respectively.

In the four classes classification task, 25 epochs is not enough. The loss is still dropping after 25 epochs, because the task is harder, more epochs may needed. Therefore, for four classes classification, 50 epochs are trained.

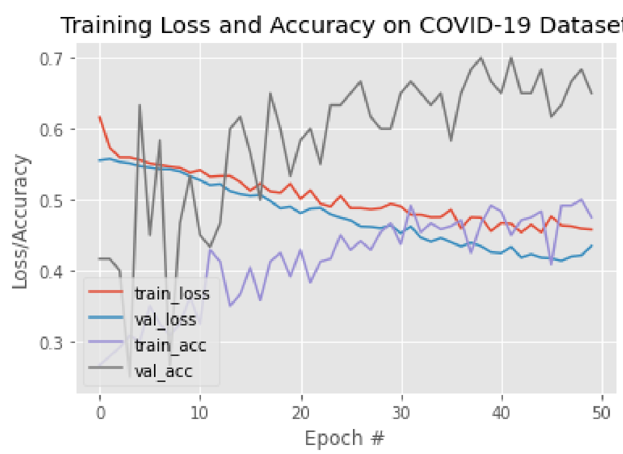

Figure 9: 4 classes ResNet

4 classes ResNet	precision	recall	f1-score	support
covid	0.91	0.67	0.77	15
normal	0.62	1.00	0.77	15
other bacteria	0.40	0.27	0.32	15
other viral	0.67	0.67	0.67	15
accuracy			0.65	60
macro avg	0.65	0.65	0.63	60
weighted avg	0.65	0.65	0.63	60

Table 7: 4 classes ResNet Evaluation Table

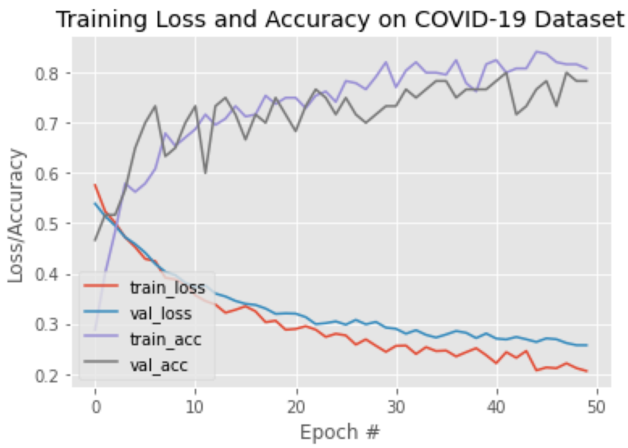

Figure 10: 4 classes VGG

4 classes VGG	precision	recall	f1-score	support
covid	0.83	1.00	0.91	15
normal	0.70	0.93	0.80	15
other bacteria	0.82	0.60	0.69	15
other viral	0.82	0.60	0.69	15
accuracy			0.78	60
macro avg	0.79	0.78	0.77	60
weighted avg	0.79	0.78	0.77	60

Table 8: 4 classes VGG Evaluation Table

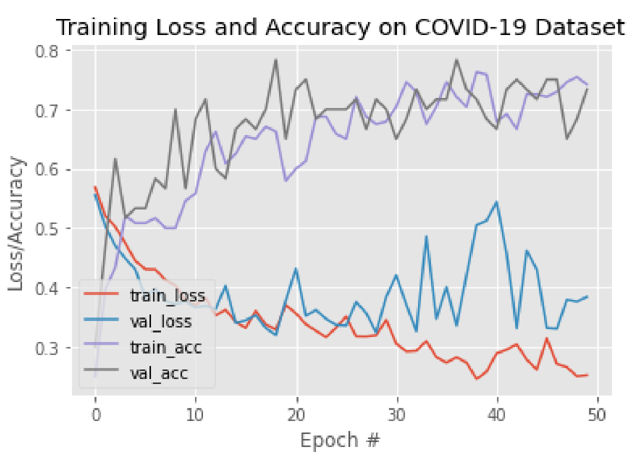

Figure 11: 4 classes CNN

4 classes CNN	precision	recall	f1-score	support
covid	0.82	0.93	0.87	15
normal	0.80	0.80	0.80	15
other bacteria	0.69	0.60	0.64	15
other viral	0.60	0.60	0.60	15
accuracy			0.73	60
macro avg	0.73	0.73	0.73	60
weighted avg	0.73	0.73	0.73	60

Table 9: 4 classes CNN Evaluation Table

4 DISCUSSION

The summary of the 9 experiments can be viewed in Table 10. Ablation Table. The performance of ResNet in these three tasks are all lower than VGG and CNN. According to the training loss and accuracy diagrams, the variance of ResNet's performance was huge, which suggests that the overall performance of ResNet on these tasks are not stable. VGG is the best model overall in the three tasks. It performed the best on 2 out of the 3 tasks. The only task that VGG is weaker than CNN is two classes classification, where CNN achieved 100% accuracy, which could be by chance of the testing data, while VGG also achieved very high accuracy (97.5%) in this task. Moreover, the training loss and accuracy and the validation loss and accuracy on VGG is much more stable than CNN, which suggests that VGG would have an even better performance on larger datasets. Good news is that the recall score of both VGG and CNN are extremely high (93% or above) in all experiments even on this small dataset, which means if the patient does get infected by coronavirus, he/she is very likely to be diagnosed to have COVID-19 according to our neural network.

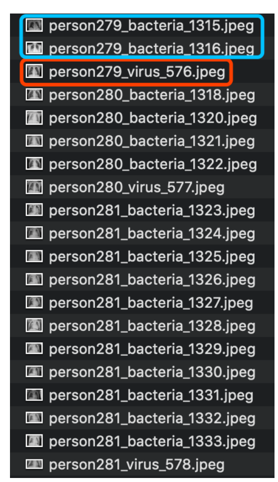

Figure 12: Mixed labels

Compared with the previous two tasks, four classes classification has a much lower accuracy on all three models, which would be caused by reasons other than the model itself, since in the previous two tasks, the model captured the features in the chest X-Ray images fairly well. The dataset is highly suspicious. Therefore, we looked into the dataset, some patients are labeled as both bacterial pneumonia and viral pneumonia, for example, person 279 280 281 in the above figure.

Model	Description	Test Accuracy
ResNet 101 (transfer learning)	Task 1 (100 normal :100 COVID)	90%
VGG 16 (transfer learning)	Task 1 (100 normal :100 COVID)	97.5%
CNN	Task 1 (100 normal :100 COVID)	100%
ResNet 101 (transfer learning)	Task 2 (75 normal :75 COVID : 75 other pneumonia)	65%
VGG 16 (transfer learning)	Task 2 (75 normal :75 COVID : 75 other pneumonia)	93%
CNN	Task 2 (75 normal :75 COVID : 75 other pneumonia)	91%
ResNet 101 (transfer learning)	Task 3 (75 normal :75 COVID : 75 other bacterial : 75 other viral)	65%
VGG 16 (transfer learning)	Task 3 (75 normal :75 COVID : 75 other bacterial : 75 other viral)	79%
CNN	Task 3 (75 normal :75 COVID : 75 other bacterial : 75 other viral)	73%

Table 10: Ablation Table

5 FUTURE WORK

Not only image-based datasets, but also text-based datasets, we are inspired by the related field clinical NLP, a specialization of NLP that allows computers to understand the rich meaning that lies behind a doctor's written analysis of a patient, and through tracking the patient's health history, researchers can achieve a comprehensive view on the background of one patient sample, thus the work can be fulfilled with more details. This is also another way of data augmentation, since the COVID-19 CT scans are few available to the public now, and its sensitive feature inspired our move towards the text.

It is also important to notice that one reason lies behind this pandemic is its speed of infection and since the coronavirus is spread by respiratory droplets and sneezes, humans get infected easily by breathing in the air. We can imagine in the future that when we walk into a hospital waiting room, instead of encountering a human we encounter a robot who's able to help you.

From several following aspects, the robots can assist this outbreak in a significant way. Humans are easily infected but robots will never. From transmitting life essentials to goods delivery, zero contact can be easily accomplished by using medical robots. Furthermore, since everyone is advised social distancing, teleoperating would invoke a brand new meeting with doctors, without worrying about the possible infection from both sides. Last but not the least, thousands of research confirmed a stronger connection and reliability of human robot interaction. In the next step, research can involve the presence of medical robots, interaction between the patients for the sake of the ideal experience of health care.

6 CONCLUSION

In this work, we present COVID-Neural Networks for classification of COVID-19, other lung infection pneumonia (bacterial and viral) and healthy people. These three models were trained on a publicly available dataset and have shown excellent classification capability (recall score of COVID-19). We also performed data augmentation in order to increase the training set size and improve generalization.

We would like to emphasize that even though using neural networks on classification tasks is very promising and accurate, it is not meant to be directly employed for clinical diagnosis at current stage. The goal of this work was to show that using different training techniques enable us to train models that are computationally efficient and accurate. In order to make this work clinically useful requires training with a larger dataset and testing in the wild with a larger cohort.

REFERENCES

[1] Deng, J., Dong, W., Socher, R., Li, L., Li, K., amp; Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvpr.2009.5206848.

[2] Howard, J., amp; Gugger, S. (2020). Fastai: A Layered API for Deep Learning. Information, 11(2), 108. doi:10.3390/info11020108.

[3] Ng, M., Lee, E. Y., Yang, J., Yang, F., Li, X., Wang, H., . . . Kuo, M. D. (2020). Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. Radiology: Cardiothoracic Imaging, 2(1). doi:10.1148/ryct.2020200034.

[4] Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., amp; Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. Jama. doi:10.1001/jama.2020.3786.

[5] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., . . . Xia, L. (2020). Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology, 200642.

doi:10.1148/radiol.2020200642.
[6] Lee, J. (2020). COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image (Preprint). doi:10.2196/preprints.19407.
[7] Abbas, A., Abdelsamea, M., amp; Gaber, M. (2020). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. doi:10.1101/2020.03.30.20047456.
[8] Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). doi:10.1109/wacv.2017.58.
[9] Lu, T., amp; Pu, H. (2020). Computed Tomography Manifestations of 5 Cases of the Novel Coronavirus Disease 2019 (COVID-19) Pneumonia From Patients Outside Wuhan. Journal of Thoracic Imaging, 35(3). doi:10.1097/rti.0000000000000508.