Deep Learning Diagnosis of SARS-CoV-2 Through X-Ray Images

Zhengru (James) Fang, Upper Saint Clair High School, 21 March 2020

Pittsburgh, Pennsylvania

fangz@uscstudents.org

Introduction:

COronaVIrus Disease (COVID-19) is currently a global pandemic that has reached 160 countries and regions by March 19, 2020, with more than 240,000 total cases (1). This pneumonic viral infection originated in Wuhan, Hubei, China at the end of 2019 (2). Unfortunately, the 240,000 cases only include the confirmed cases. An average of 5.1 days is elapsed before patients show symptoms after infection (3), and a further lag will occur between when the patient shows symptoms and is confirmed with COVID-19 (4). In both of these phases, the patient is infected and may be infecting others (5). Thus, the number of confirmed cases would be marked lower than that of the total infected number of people.

COVID-19 currently has an average mortality rate of around 2-3% (7), but this figure varies widely around the world. In Wuhan, the mortality rate is around 4% (8). However, Italy currently records a mortality rate of 7.7% (9), as its healthcare resources are overstrained and this effect is further compounded with infection of hospital workers, which reduces the meager amount of healthcare resources available to patients. This strain on healthcare resources can be observed in other regions as the number of COVID-19 cases rise quickly and has been observed in Wuhan (10, 11) .

The rate of infection and deaths in South Korea has been markedly slower than that of many other countries, such as Italy, Spain and the United States, even though strict movement restrictions have been in place in all three latter countries (1,12). This can be correlated with a high per-capita testing where screening has found many asymptomatic COVID-19 patients - the per capita testing is more than thirty times that of the United States (13).

Thus, the director of the World Health Organization advised countries to "test, test, test", in order to screen out the unconfirmed, infected patients and prevent these unconfirmed cases from further spreading, thus controlling the rate of infection (14). Unfortunately, testing kits require a chemical component, which impacts the speed that it can be produced and distributed to hospitals around the world. Not all testing kits are effective. Recent evidence shows that the RT-PCR method of detecting COVID-19 has a low positive rate in the early stages (15).

In this investigation, the use of a convolutional neural network on detecting COVID-19 from X-ray images is investigated. Since the infrastructure for X-ray scans are in place in most countries, as compared to the infrastructure for biochemical testing for COVID-19, this will mean that patients have faster access to testing for COVID-19. There are key distinctions between pneumonia and coronavirus pneumonia, as evidenced by X-ray images of SARS (16). As COVID-19 infects by entering the same receptor protein (ACE2) as SARS does (17), it is reasonable to assume COVID-19 infection patterns will be similar to that of SARS.

Methods

    a. Dataset
    1. *COVID-19 Algorithm Dataset*

A total of 146 pneumonia lung samples were collected from a Github repository compiled by Joseph Paul Cohen of University of Montreal for pneumonia images (18). The lung samples represented X-ray and CT scans. The syndromes represented include COVID-19, SARS, streptococcus, ARDS, and pneumocystis. The pneumonia lung samples represent a wide geographic diversity, from China, United States, Canada, Italy, Taiwan, Israel, South Korea, Vietnam and Australia. The 146 images represented 74 different patients, some of whom had multiple X-ray scans over time. The number of images on the Github repository will continually increase as new X-ray scans of COVID-19 patients are constantly uploaded.

The samples from the Github repository are filtered in order to only include the relevant COVID-19 images:
1) The 146 images were then filtered out to include only X-ray scans (-23 scans, leaving 123 images).
2) The 123 remaining images are filtered out for positions other than posterior-anterior (PA) or anterior-posterior (AP) view (-11 scans, leaving 112 images).
3) The 112 images are then finally filtered to only include SARS and COVID-19 images (-12 scans, leaving 100 images).

The remaining images only represent the COVID-19 and SARS pneumonia examples. For the other two categories that will be studied in this project (Normal, healthy patients and other pneumoniae, including both bacterial and viral pneumoniae), the data would be collected from a dataset published in the Kermany et al. 2018 paper in *Cell* (19). The training and validation and label datasets for each category would be listed below. The exclusion of lateral images can be justified due that lateral images only exist in the COVID-19 dataset, but not in the Cell dataset, thus would lead to the dataset algorithm overfitting and predicting all lateral images as patients with COVID-19.

Distribution of X-ray scans in each category (Total: 5,956)	Normal	Non-COVID Pneumonia	COVID-19 Pneumonia
Training	1341	3875	86 (75 COVID-19, 11 SARS)
Validation	242	398	14 (all COVID-19)

Care was taken to place all X-rays of a certain patient all within either the training or validation datasets, but not spread through the two different phases (training and validation). The fourteen COVID-19 scans in the validation datasets represented eight different patients. Since the data is published on Github and on cell, the data has already been deidentified. A possible problem is the skewing in the COVID-19/Cell dataset, which will be addressed in section b), on training the algorithm.

Figure 1. Clockwise from top left: 1) Normal, Healthy lung, 2) Bacterial Pneumonia, 3) Viral Pneumonia, and 4) COVID-19 Pneumonia
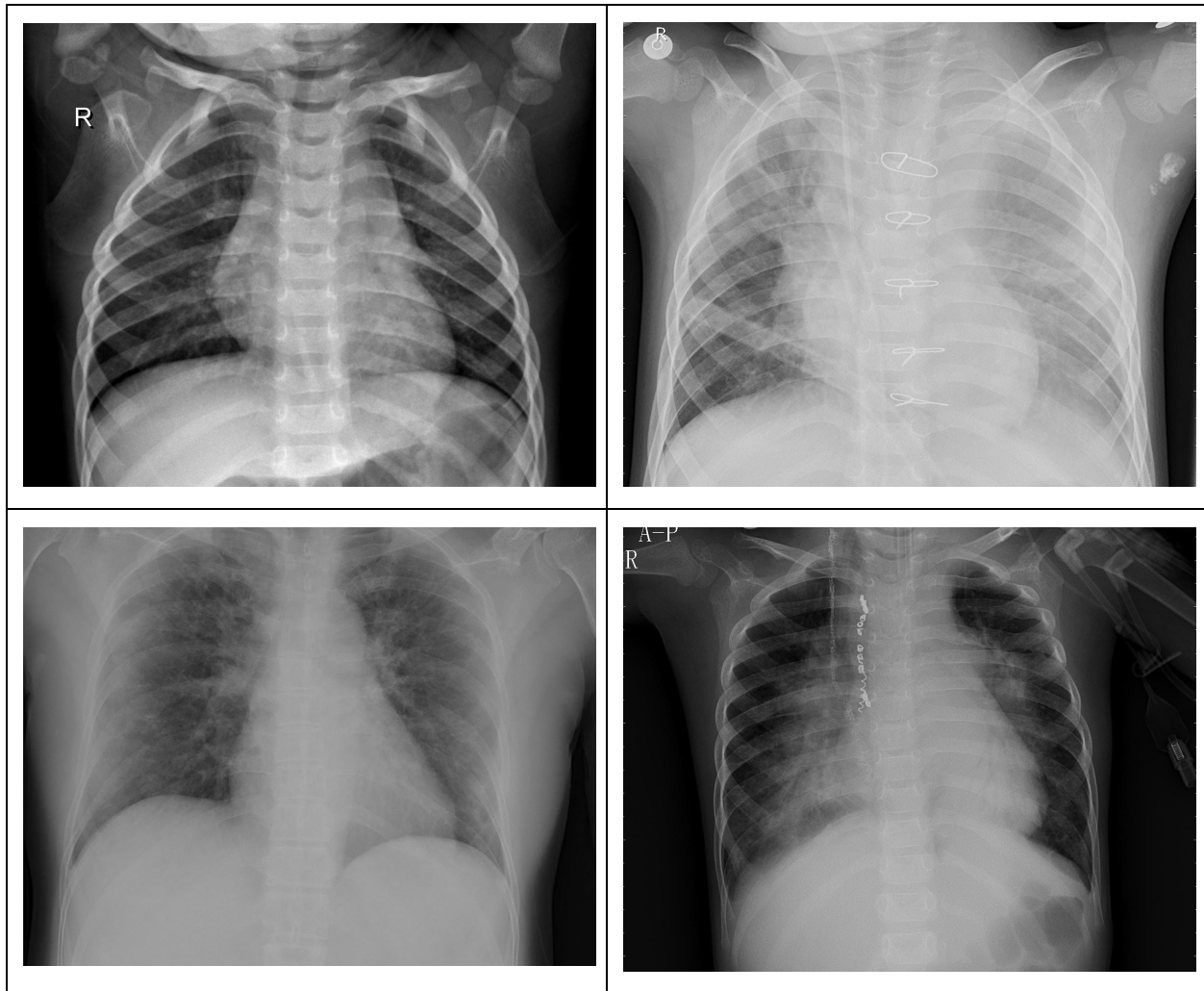
2. *Transfer Learning Dataset*

To improve performance, this algorithm employs transfer learning. Transfer learning will be discussed in greater detail in section b), algorithm development. The general principle is the use

of weights and pattern-recognition capabilities from previous models (20). This algorithm will use transfer learning twice.

The first transfer learning dataset is the ImageNet dataset (21). The ImageNet dataset is a dataset with more than 14 million images divided into 21841 diverse categories. The advantages of using weights from the ImageNet dataset is that the earlier layers would contain very fine pattern recognition areas, such as for detecting small lines and small patterns (20). The time required to train the model would therefore also be reduced, as many of the fine weights and patterns are already established when the neural network has been trained on this large corpus of 14 million images. The weights from the pretrained network were downloaded for use on the second transfer learning dataset. The last layer will be changed to three output neurons.

The second transfer learning dataset is the 2018 RSNA Pneumonia Detection Challenge dataset (22). In this dataset, there are 26,684 total images. The three categories are Normal, No Lung Opacity / Not Normal (indication of other lung diseases), and Lung Opacity, with 8,851, 11,821 and 6,012 examples respectively. This second transfer learning dataset is designed to increase the number of lung images that the algorithm is trained on, therefore improving its pattern recognition capabilities. The neural network's exposure to 26,684 images would allow the neural network to be trained on 5 times the number of images than had the neural network been trained on the 5,956 images from the *Cell* and the Cohen datasets. All of the images are in either PA or in AP positions, again warranting the exclusion of lateral (L) images from the COVID-19 dataset. This dataset of 26,684 total images is divided randomly into 24,184 training images and 2,500 validation images.

b. Algorithm Development

This program is developed using the Fast.AI library, which allows researchers to easily train neural networks and to use preset tools in order to preprocess and to evaluate models (23). Its advantages include flexibility while allowing researchers to quickly train models within a few lines of code. This program is written in Python.

1. *Architecture*

The model architecture for this algorithm is the EfficientNet-B4 released by Google in 2019 (24). The EfficientNet is an improved series of image recognition architectures through scaling of depth, width and resolution of images. The EfficientNet-B4 features a balance between complexity and performance, as the EfficientNet series after B4 have many more parameters (increased complexity of training) for fewer returns in increases in performance.

Figure 2. An EfficientNet model. This is depicting EfficientNet-B0 (24).

Figure 3. Chart of accuracy vs. number of parameters. Notice that gains in accuracy quickly levels off after EfficientNet-B4 (24).

	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**78.8%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.1%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.6%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**
†Not plotted		

2. Transfer Learning Intuition

Transfer learning was discussed briefly in section a). The intuition behind transfer learning is that it allows the fine weights learned from previous tasks to be further tuned to suit a particular task, thus reducing the amount of training required, because the basic pattern recognition capabilities are already in place (20). It is especially useful for smaller datasets such as the

COVID-19 dataset that was compiled using a small Github database and the Cell dataset. The difference between the RSNA dataset and the COVID-19 & Cell dataset, such as the different classes, makes it less desirable to train the two datasets together, compounded with the relative class imbalance between the two datasets.

3. Image Preprocessing

For the RSNA dataset: in the label file for RSNA, since the original intention of the dataset is to predict bounding boxes for lung opacities, there were duplicates within the dataset to indicate multiple bounding boxes in a lung patient. For this task, however, the bounding boxes task was removed, and the algorithm was only to predict whether the lung X-ray image was Normal, No Opacity / Not Normal, or Lung Opacity. The DICOM images, originally with size 1024 x 1024, were sized down to 3 x 256 x 256 pixel JPEGs to save and converted from grayscale to RGB (without affecting the colour of the RGB images), thus making training faster as loading from DICOM was inefficient. The data was inputted using size 3 x 224 x 224 pixel into EfficientNet.

For the COVID 19/Cell dataset: Since all of the images are of different sizes, all images were sized down into size 3 x 224 x 224, before training the neural network for a second time.

The amount of image preprocessing in the deep learning algorithm is minimal compared to other machine learning methods, where features had to be artificially extracted from the image. Artificial extraction of features is often inefficient and would lead to waste as much of the data is thrown away (25). It would also lead to lower performance because there is less data for the algorithm to base its decision on and errors in labelling features would have greater impacts as the relatively few features will make each error more significant.

4. Image Augmentation

The dataset is augmented in multiple image transforms. The image transforms are horizontal flip, rotation of up to 10 degrees, zoom of 1.1x, lighting and contrast change, and random warp. Image augmentation serves to increase the robustness of the algorithm since it increases the amount of data given to the algorithm and the different angles that the algorithm looks at the images, thus improving its performance on different datasets and angles. Data augmentation reduces overfitting as the deep learning algorithm is forced to predict on a diverse number of different perspectives. Given the limited number of public pneumonia datasets available, image augmentation will be an attractive option to build more robust models. Image augmentation has shown to increase performance on the ImageNet Dataset, from AlexNet, and has been extended to other deep learning models, such as image segmentation with R-CNN (26).

5. Training

This algorithm is trained using cosine learning rate annealing. Learning rate increases until further increases in learning rate hurts training, in which the weights are not updated and the algorithm starts from a small learning rate again. This makes sure that, in the end, the algorithm ends up with "stable" weights resistant to a difference in the distribution between the training, validation, and real-life data as a large learning rate is needed to "destabilize" the algorithm, thus the algorithm is stable over a greater distribution of data. Learning rate annealing also allows for faster training of neural networks as this means fewer updates have to be done with the optimized "large steps" towards the minimum error (27).

Overfitting is reduced with weight decay. Weight decay refers to multiplying the weights at the end of each batch by a number just slightly less than 1. It serves as regularization because the compounded effects would make the outputs tend towards zero, cutting out the overfitting patterns. Another technique is dropout - where a number of neurons are turned off during training, forcing the algorithm to not just depend on a certain set of neurons in performing its task, thus the algorithm becomes more robust (28).

In each of the two algorithms, training passes through two different phases. The first phase of training is just training the top layer, to just train the classifier to recognize the largest patterns indicating the presence or absence of lung disease. The intuition is, again, that ImageNet weights are already well fine-tuned to recognize the earlier level features, thus, only the more abstract and complex features should be adjusted to fit to the task (29). The second level is fine-tuning the model. The earlier layers are unfroze, however they will pass through discriminative learning rates. The learning rates, from smallest to largest, are for early, middle and later layers in this order. The reasoning is that the earliest layers have very fine-tuned layers that need to be only slightly adjusted, while the later layers need more fine-tuning in order to serve the task (30).

The problem of skewed classes was raised in section a). Skewed classes could mean that the deep learning algorithm will just gloss over the images in the minority classes, in favor of predicting the majority classes accurately (31). The number of COVID-19 images are especially lacking as it only represented 100/5958 images (86/5304 in training and 14/654 in validation), but is the most important class since this algorithm is designed to predict whether patients have COVID-19 pneumonia or not. To resolve the skewed class problem, the minority classes (COVID-19/SARS and normal images) would have the training error increased if the algorithm misses these two classes. The ratio of loss for those three classes {COVID-19/SARS, pneumonia, normal} are {11, 1, 2} for the first phase (top-level training), and {11, 1, 3} for the second phase training (fine-tuning).

6. Visualization: Grad-CAM

Deep learning algorithms have often been criticized as black boxes. While deep learning is strong at classification, they often cannot justify their decisions. Despite this, they are strongly preferable to other machine learning methods, such as random forests and logistic regression, as they allow for less flexibility, thus their performance is lower than that of deep learning. In many ways, deep learning is the only way to construe the fine relationships between the pixels. One of the ways to understand the decision-making process is through visualization by Grad-CAM. Since deep learning algorithms do not work in all situations, as they rely on data given by humans and can fail when it is prompted to perform tasks outside of its domain), it is imperative to understand conditions that would make the algorithm fail by understanding its decision making process. Given the possibility that the algorithm fails and the difficulty to see when the algorithm will fail, that is one of the main reasons that medical professionals may be hesitant to start using DLAs in diagnosis (32, 33).

Visualization of algorithms is the use of heatmaps in order to explain the algorithm's focus on the different sections of the lung images. It serves to explain which areas of an image is more important to the deep learning algorithm, thus pulmonologists can understand the basis for the algorithm's decision. While this does not explain the ways that patterns are extracted or the overall weighting of the patterns, visualization still provides a strong start to understanding the areas that the algorithm considers important. The technique used to visualize the deep-learning algorithm will be Grad-CAM. To perform this visualization, the model would take in an image, pass through every layer, and then resize every layer of the neural network, add the activations up, before finally outputting a feature map of the regions that the algorithm has decided to be important. That way, the entire decision process is documented (32).

Figure 4. From Selvaraju et al. 2016 - shows the process of Grad-CAM (32)
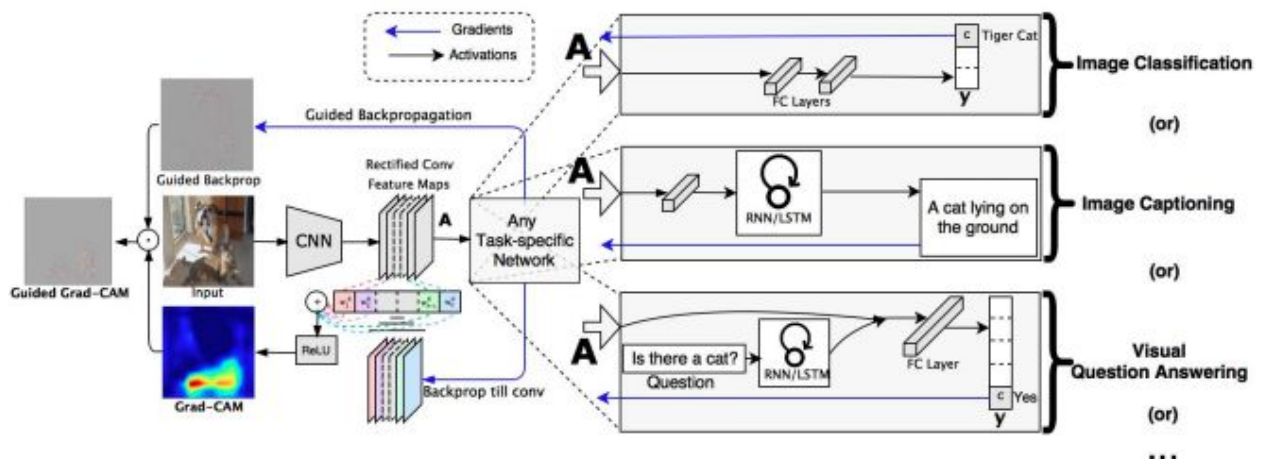
Results and Discussion

    a. Evaluation

The model will be evaluated on its overall accuracy and its precision and recall for each category. Other evaluation metrics include the speed of the predictions. Accuracy refers to the overall percent correct within the validation dataset. Precision refers to the percent correct for *all predictions* made towards a class. Recall refers to the percent of samples correctly classified or discovered *within a class*. All evaluation is made with respect to the validation dataset. The F1 score will also be used, as F1 combines both precision and recall (34).

Precision is important because a high precision will mean fewer false positives, thus healthcare resources would not be used on healthy patients and healthy patients would not instead become infected with COVID-19 while in hospitals, while recall is important because false negatives mean that some patients do not receive the care they need and will still spread the coronavirus to other people in a community.

    b. Result data

The overall accuracy on the validation dataset is at 92.97% (608 out of 654 samples). Below are the confusion matrix and the table illustrating precision, recall and F1 for each class.

Figure 5. Confusion matrix of the results. Note that all 14 images in COVIDnSARS category contains just COVID-19 images.

Table for precision, recall and F1 on each class.
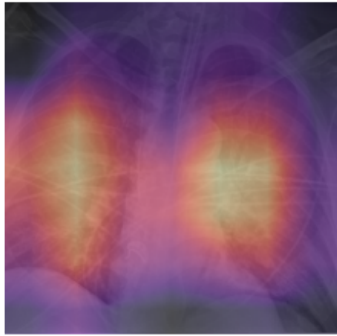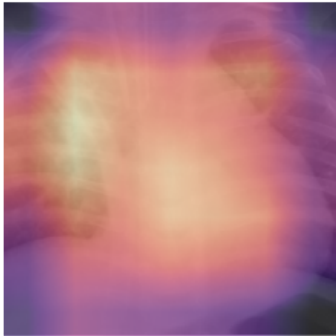
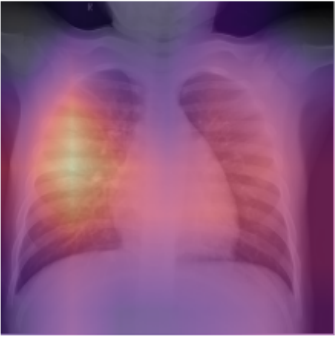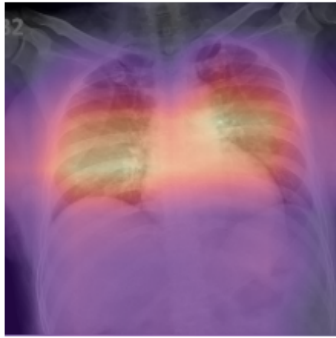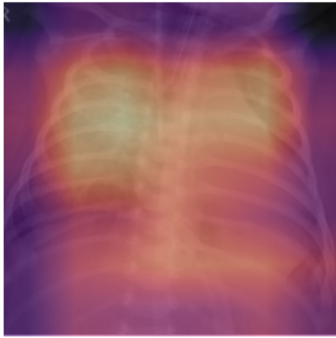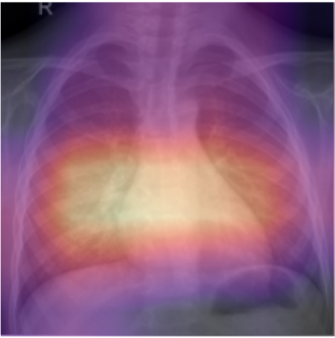|	COVID-19	Other Pneumonia	Normal
Precision	1	0.9055	0.9757
Recall	1	0.9874	0.8305
F1	1	0.9446	0.8973
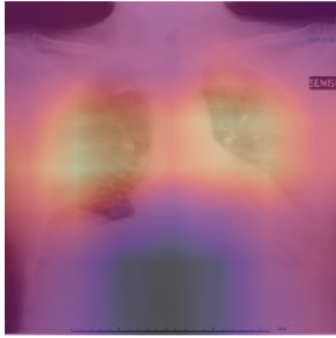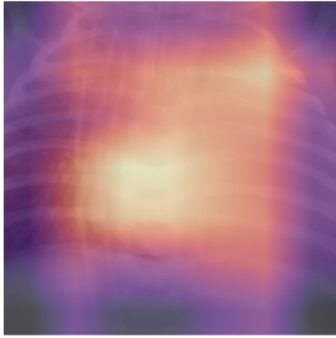
The algorithm was able to loop through all 654 validation examples in the validation dataset within 10 seconds using Nvidia RTX 2070, indicating near-instantaneous output of results.

c. Evaluation of Grad-CAM data

From the Grad-CAM results, the algorithm was able to locate the areas which indicated whether there was pneumonia or not, and if there is, whether the pneumonia was of COVID-19 or not. However, the resolution can still be improved as the algorithm did not indicate which specific areas in the lungs indicate pneumonia.

Figure 6. Visualization on the validation dataset. Three examples for each class is provided here.

	COVID-19	Other Pneumonia	Normal
1			

| 2 | |
| 3 | |

d. Discussion

As COVID-19 continues to spread and with vaccines estimated at more than a year away (35), the most efficient solution would therefore be to test large numbers of people in order to screen out all symptomatic and all asymptomatic patients. COVID-19's impact can be demonstrated as even advanced health systems, such as that of Lombardy's, became overwhelmed by the sheer number of patients (36). Mass testing has found success in containing the COVID-19 outbreak in numerous countries, such as in South Korea (1,12). Local regions have also found success with mass testing, such as in Vo Village in Northern Italy (37). Thus, the World Health Organization advises countries to test as much as possible in order to screen out for cases of COVID-19.

While RT-PCR testing can chemically confirm the presence of COVID-19 virus, there are important drawbacks with RT-PCR testing. RT-PCR requires the production of chemical kits - and distribution and production both require time. The RT-PCR may miss many cases in the early phases of infection. Finally, the testing itself requires time to obtain laboratory results. However, the infrastructure for X-ray scans of lungs are already in place, and this deep learning detector can therefore be used to detect COVID-19 with a high precision and recall. The patient can then be immediately placed into isolation when detected positive, therefore not infecting other healthy people.

The deep learning algorithm was able to process the images, filter for the most useful areas in the lungs to indicate whether the patient is healthy, has COVID-19 pneumonia or has other pneumonia infection, by spotting for the features that indicate COVID-19 pneumonia. They can thus assist physicians in making decisions more accurately.

The speed that the deep learning algorithm worked at indicated that it can quickly infer on images (654 in less than 10 seconds), however these are predictions on an Nvidia RTX 2070. Thus, further investigation should be conducted to understand how quickly prediction may proceed on a slower CPU.

The result of full precision and recall should be heeded with caution. Even though the heatmap can classify the areas that were most important to the algorithm in making its decision, there are only 14 validation images and 86 training images. Thus, it is necessary to increase the number of X-ray images in order to produce a reliable model. While the images represented many geographic regions, 86 images are still not diverse enough to represent all COVID-19 patients. The full precision and recall can possibly be attributed to the algorithm basing decisions on the features *not present* within the *Cell* dataset that was used to obtain data for the other two categories, rendering the classification a matter of distinguishing between different datasets.

Future improvements would be to include more data, preferably hundreds or thousands of images for both training and validation. The new dataset would encompass an even greater geographic diversity, thus increasing its applicability to be used throughout the world.

Conclusion

In this study, the deep learning algorithm was able to detect COVID-19 with full precision and recall and then justify its decisions. An automated system would allow doctors to screen for large numbers of COVID-19 patients quickly using an x-ray. However, development on larger datasets is advised in order to increase the robustness of this model.

References:

1. Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. doi: 10.1016/s1473-3099(20)30120-1
2. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 395(10223), 497–506. doi: 10.1016/s0140-6736(20)30183-5
3. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., … Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From

Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine. doi: 10.7326/m20-0504

4. Wu, Z., & Mcgoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. JAMA. doi: 10.1001/jama.2020.2648

5. Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., … Hoelscher, M. (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. New England Journal of Medicine, 382(10), 970–971. doi: 10.1056/nejmc2001468

6. Jung, S.-M., Akhmetzhanov, A. R., Hayashi, K., Linton, N. M., Yang, Y., Yuan, B., … Nishiura, H. (2020). Real time estimation of the risk of death from novel coronavirus (2019-nCoV) infection: Inference using exported cases. MedRxiv. doi: 10.1101/2020.01.29.20019547

7. Bialek, S., Boundy, E., Bowen, V., Chow, N., Cohn, A., Dowling, N., … Sauber-Schatz, E. (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. MMWR. Morbidity and Mortality Weekly Report, 69(12). doi: 10.15585/mmwr.mm6912e2

8. Mizumoto, K., Kagaya, K., & Chowell, G. (2020). Early epidemiological assessment of the transmission potential and virulence of coronavirus disease 2019 (COVID-19) in Wuhan City: China, January-February, 2020. MedRxiv. doi: 10.1101/2020.02.12.20022434

9. Lazzerini, M., & Putoto, G. (2020). COVID-19 in Italy: momentous decisions and many uncertainties. The Lancet Global Health. doi: 10.1016/s2214-109x(20)30110-8

10. Li, R., Rivers, C., Tan, Q., Murray, M. B., Toner, E., & Lipsitch, M. (2020). The demand for inpatient and ICU beds for COVID-19 in the US: lessons from Chinese cities. MedRxiv. doi: 10.1101/2020.03.09.20033241

11. Peltier, E., & Minder, R. (2020, March 20). Europe Struggles to Combat Coronavirus as Spain Passes Grim Milestone. Retrieved from https://www.nytimes.com/2020/03/20/world/europe/coronavirus-spain.html

12. Tilford, C. (2020, March 21). Coronavirus tracked: the latest figures as the pandemic spreads. Retrieved from https://www.ft.com/coronavirus-latest

13. White, E. (2020, March 11). Coronavirus testing: how are the hardest-hit countries responding? Retrieved from https://www.ft.com/content/dd416102-5d20-11ea-b0ab-339c2307bcd4

14. WHO head: 'Our key message is: test, test, test'. (2020, March 16). Retrieved from https://www.bbc.com/news/av/world-51916707/who-head-our-key-message-is-test-test-test

15. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., … Xu, B. (2020). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). doi: 10.1101/2020.02.14.20023028

16. Müller, N. L., Ooi, G. C., Khong, P. L., & Nicolaou, S. (2003). Severe Acute Respiratory Syndrome: Radiographic and CT Findings. American Journal of Roentgenology, 181(1), 3–8. doi: 10.2214/ajr.181.1.1810003

17. Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. Journal of Virology, 94(7). doi: 10.1128/jvi.00127-20

18. Cohen, J. P. (2020, February 14). ieee8023/covid-chestxray-dataset. Retrieved from https://github.com/ieee8023/covid-chestxray-dataset

19. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., … Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, 172(5). doi: 10.1016/j.cell.2018.02.010

20. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? ArXiv:1411.1792 [Cs]. http://arxiv.org/abs/1411.1792

21. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & 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

22. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3462–3471. https://doi.org/10.1109/CVPR.2017.369

23. Howard, J., & Gugger, S. (2020). Fastai: A layered api for deep learning. Information, 11(2), 108. https://doi.org/10.3390/info11020108

24. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. ArXiv:1905.11946 [Cs, Stat]. http://arxiv.org/abs/1905.11946

25. Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4–21. doi: 10.1109/jbhi.2016.2636665

26. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0

27. 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

28. Loshchilov, I., & Hutter, F. (2017). Decoupled Weight Decay Regularization. ICLR.

29. Brock, A., Lim, T., Ritchie, J. M., & Weston, N. J. (2017). FreezeOut: Accelerate Training by Progressively Freezing Layers. Paper presented at NIPS 2017 Workshop on Optimization, Long Beach, United States.

30. Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. ArXiv:1801.06146 [Cs, Stat]. http://arxiv.org/abs/1801.06146

31. Longadge, R., & Dongre, S. (2013). Class Imbalance Problem in Data Mining Review. ArXiv, abs/1305.1707.

32. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2019). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision, 128(2), 336–359. doi: 10.1007/s11263-019-01228-7

33. Topol, E. J. (2019). Deep medicine: how artificial intelligence can make healthcare human again. New York: Basic Books.

34. Sasaki, Yutaka. (2007). The truth of the F-measure. Teach Tutor Mater.

35. Why will it take so long to make a coronavirus vaccine that can prevent COVID-19? (2020, March 13). Los Angeles Times. https://www.latimes.com/science/story/2020-03-12/why-does-it-take-so-long-to-make-a-coronavirus-vaccine

36. Nacoti Mirco, Ciocca Andrea, Giupponi Angelo, Brambillasca Pietro, Lussana Federico, Pisano Michele, Goisis Giuseppe, Bonacina Daniele, Fazzi Francesco, Naspro Richard, Longhi Luca, Cereda Maurizio, & Montaguti Carlo. (n.d.). At the epicenter of the covid-19 pandemic and humanitarian crises in italy: Changing perspectives on preparation and mitigation. Catalyst Non-Issue Content, 1(10). https://doi.org/10.1056/CAT.20.0080

37. Armitage, R. (2020, March 21). A town held a coronavirus experiment and now it's 'the healthiest place in Italy'. Retrieved from https://www.abc.net.au/news/2020-03-21/one-italian-town-is-bucking-the-countrys-coronavirus-curve/12075048