Detecting COVID-19 from Chest X-Ray Images

Abdullah AlOthman
Felipe Buchbinder
Joseph Krinke
JingJing Shi

Duke MIDS

April 2020

Abstract

Millions of people have been infected with the novel coronavirus COVID-19. The current technology used to detect COVID-19 is a PCR-based test, which many countries have had difficulty producing, administering, and analyzing. Examining lung images has been proposed as a method to identify patients suffering from the disease, but it is time-consuming for radiologists to analyze the images manually. Artificial intelligence has been proposed as a way to efficiently identify COVID-19 cases using lung images. In this paper we create a novel dataset and use it to fine-tune an InceptionV3 and an Xception model and use them in an ensemble. Our results show our model can distinguish between disease classes, achieving 89% overall accuracy with 97% precision and 93% recall on the COVID-19 class. This report, however, is merely a feasibility study, as existing datasets are not yet reliable enough to be used in earnest.

Introduction

Viruses are infectious agents that can only replicate within the cells of living hosts [1]. Disease outbreaks caused by viruses often pose a risk to public health [2]. In the past two decades, lower respiratory tract infections caused by viruses have become the most deadly infectious diseases [3]. There have been several outbreaks caused by influenza and coronaviruses. These include severe acute respiratory syndrome (SARS) in 2003 [4] and Middle East respiratory syndrome (MERS) in 2012 [5].

Like SARS and MERS, the COVID-19 is a highly contagious pathogen, and has spread to 210 countries with a total of 1,852,356 confirmed cases in four months as of April 12th, 2020[6]. Lessons from past diseases have catalyzed the sharing of genetic information of COVID-19 globally, and this information has enabled the rapid development of diagnostic tests. The COVID-19 reverse-transcription-polymerase-chain-reaction (RT-PCR) test is the current diagnostic test for COVID-19, and it is used for the in-vitro qualitative detection of COVID-19 RNA in upper and lower respiratory specimens taken from individuals [7]. However, there are several limitations to the COVID-19 RT-PCR test, which lead to difficulties when attempting to implement mass testing. The test produces large numbers of false negatives such that the CDC recommends that healthcare providers combine results with symptom, exposure, and geographic information to make a diagnosis[8]. Additionally, many countries do not have the necessary laboratory space allocated to process tests; only about 140,000 tests have been done daily in the US, which is not sufficient given the speed of the spread of COVID-19. Consequently, a different diagnostic method may prove useful in helping governments track and fight the disease.

This has motivated the scientific community to look for alternative ways to detect the virus. Computer Tomography (CT) scans and X-rays of the patient's lungs (collectively called chest radiographs) are established methods of diagnosing the presence of COVID-19, with some

even arguing that a chest CT is indispensable for the proper diagnosis of this disease[9]. Indeed, Li & Xia[10] found that doctors failed to correctly identify COVID-19 in CT images in only 3.9% of cases. A large study in China[11] showed that chest CT-scans have a high recall for COVID-19 diagnosis, and might be considered for the COVID-19 screening, especially in high epidemical areas. Since CT and X-ray machines are much more widespread than COVID-19 tests, detecting COVID-19 from these images could be an important weapon in confronting this disease.

CT-scans and X-rays are diagnostic tests that are already available throughout the world. In addition, most medical personnel are already trained in operating X-ray machines, so no additional training is necessary. Moreover, X-ray diagnostics results are ready quickly, reducing the time lapse between examination and intervention. This speed could be further increased if a faster method of examining chest radiographs was developed. This paper aims to use machine-learning techniques to quickly identify the presence of COVID-19 in a chest radiograph, thereby demonstrating a potential automatic diagnostic technique.

Background

One of the major challenges the world is facing with the COVID-19 pandemic is the shortage of testing. Many countries, including the US, face a shortage of tests, with demands for testing greatly exceeding the available supply[12, 13]. As a result, a number of scholars have attempted to develop machine learning approaches to detect COVID-19 from CT or X-ray images.This follows developments applied in the contexts of other diseases, such as SARS[14] and pneumonia[15, 16]. Narin and others[17] used a ResNet50 with transfer learning and obtained 98% accuracy on a dataset of 50 X-rays of COVID-19 patients and 50 X-rays of healthy patients. Not only is this a very small sample, it is also perfectly balanced, which is unrealistic in practice. A similar result is attained by Salman[18], who correctly classified all 260 images of a balanced sample of COVID-19 and healthy patients. Yet, this study is subject to the same caveats as Narin[17].

Zhang et al.[19] use a much larger dataset, with 100 COVID-19 patients and 1431 healthy patients. They propose a convolutional neural network (CNN) built on top of a pre-trained ImageNet backbone. With it, the authors managed to accurately identify 96% of diseased patients, but only 76% of the healthy ones. Feng et al.[20] tackled the somewhat different problem of distinguishing between images of people who have COVID-19 and those of people who have viral pneumonia. This is arguably a more difficult problem than separating between diseased and healthy patients. Still, they achieve a recall of 90.70% and a specificity of 83.30% on a dataset with 1658 COVID-19 patients and 1028 viral pneumonia patients. Unlike

Narin[17] or Zhang et al.[19], however, Feng et al. base their analysis in CT, rather than X-ray imaging.

To put these model's performances into perspective, we may recall Li and Xia`s (2020) finding that doctors failed to identify COVID-19 in only 3.9% of CT images, meaning doctors had a recall of 96.1%. This is much higher than Feng *et al.*'s [20] 90.7%, which has also used CT images, but is at par with Zhang *et al*`s [19] model constructed with X-ray imaging. The benchmark of medical error rate is more clearly outperformed by Apostolopoulos and Mpesiana[21], who use transfer learning on a series of state-of-the-art CNNs. The authors use a dataset with 224 images with confirmed COVID-19, 700 images with confirmed common bacterial pneumonia, and 504 images of healthy patients. They find that a VGG-19 network achieves a recall of 92.85% and a specificity of 98.75%. The authors argue, however, that recall is more important than specificity when dealing with health issues, and therefore preferred the MobileNet v2 architecture, which achieved a recall of 99.10% and a specificity of 97.09%. Note that Apostolopoulos and Mpesiana [21] do not distinguish between viral and bacterial pneumonia, a critique they acknowledge themselves. They therefore proceed to test the MobileNet v2 architecture on a new dataset which allows for such distinction. But only the MobileNet v2 architecture is run on this new dataset, and it produced a recall of 98.66% and a specificity of 96.46%.

These analyses show the potential of applying machine learning techniques to the problem of detecting COVID-19 using chest radiographs. However, they are limited by their small sample sizes. In this paper we address these limitations by combining several lung-disease datasets to produce an image collection containing images of normal lungs, those with pneumonia, and COVID-19.

Data

COVID-19 Radiography Database [22][23]

The COVID-19 dataset was collected by a research team from Qatar University, Doha, Qatar and the University of Dhaka, Bangladesh with their collaborators from Pakistan and Malaysia. The dataset contains 219 COVID-19 positive cases, 1,341 normal cases and 1,345 viral pneumonia cases. All the images are in PNG file format and their dimensions are 1024-by-1024 pixels.

NIH Chest X-ray Data [24]

The NIH X-ray dataset consists of 112,120 labeled X-ray images from 30805 unique patients. The labels were constructed using natural language processing but are expected to be over 90% accurate. The images all have dimensions of 1024x1024. We used 644 images from

this dataset; 322 of adult normal lungs and 322 of adult lungs with pneumonia.We used this number as it was all the pneumonia images available in the dataset.

Table 1: Distribution of Classes Across Datasets

	COVID-19 Database	NIH	Total
Normal	1341	322	2069
Pneumonia	1345	322	1667
COVID-19	219	0	219

In the COVID-19 dataset, 21 duplicate images were removed from the analysis. This is a developing situation, so the images available for analysis are limited and have imperfections. It is also important to note that the files from the COVID-19 database had a mixture of adult and child images, with all of the pneumonia pictures being of children. To address this we added additional data from the NIH Chest X-ray Dataset [24] containing adult pneumonia images to ensure that the model was not inadvertently learning to identify X-rays of children.

Methods

We split the dataset into training, validation, and test subsets with sizes of 70%, 10%, and 20%, respectively. The split is stratified based on the class distribution, meaning that each subset has the same distribution of classes. We augment the data with random horizontal flipping and rotations, we then scale the images to 299x299 to feed into the model. We used an InceptionV3 model and an Xception model (see Figure 1) both pre-trained on the ImageNet dataset and replaced the classification layers with our own. Since the dataset is imbalanced, we assigned different weights to the 3 classes based on their relative proportions in the training set. After exploring the search space by monitoring the performance on the validation set, we decided to use the Adam optimizer with a learning rate of 1e-4 and a decay of 1e-6. We use categorical cross-entropy for our loss function. We then weigh the outputs of the models based on their performance on the validation set and then average those weighted outputs to determine the class of the image. Our methodology is summarized in Figure 2.

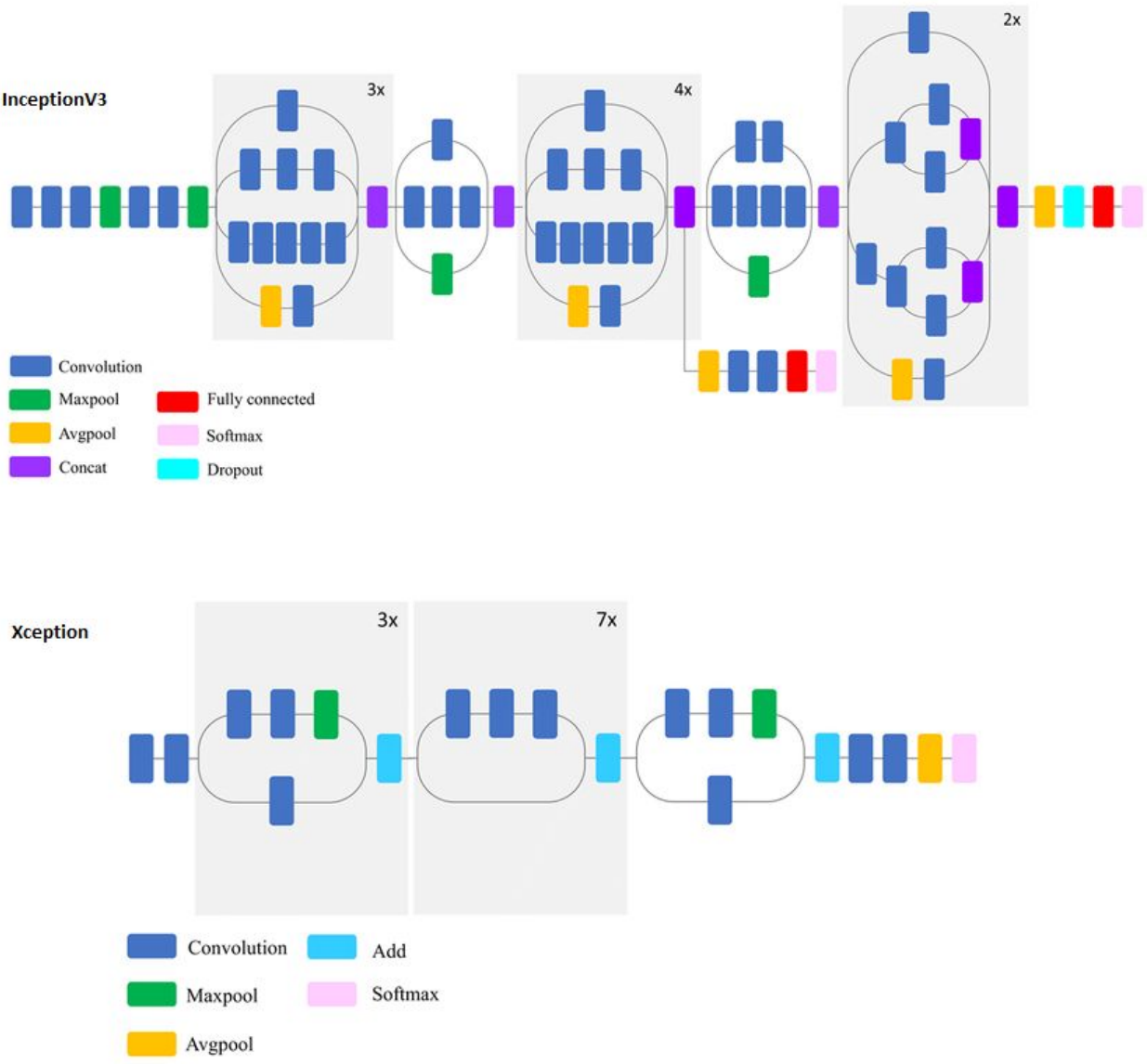

Figure 1. Schematic diagrams of the InceptionV3 and the Xception models (compressed view) [25].

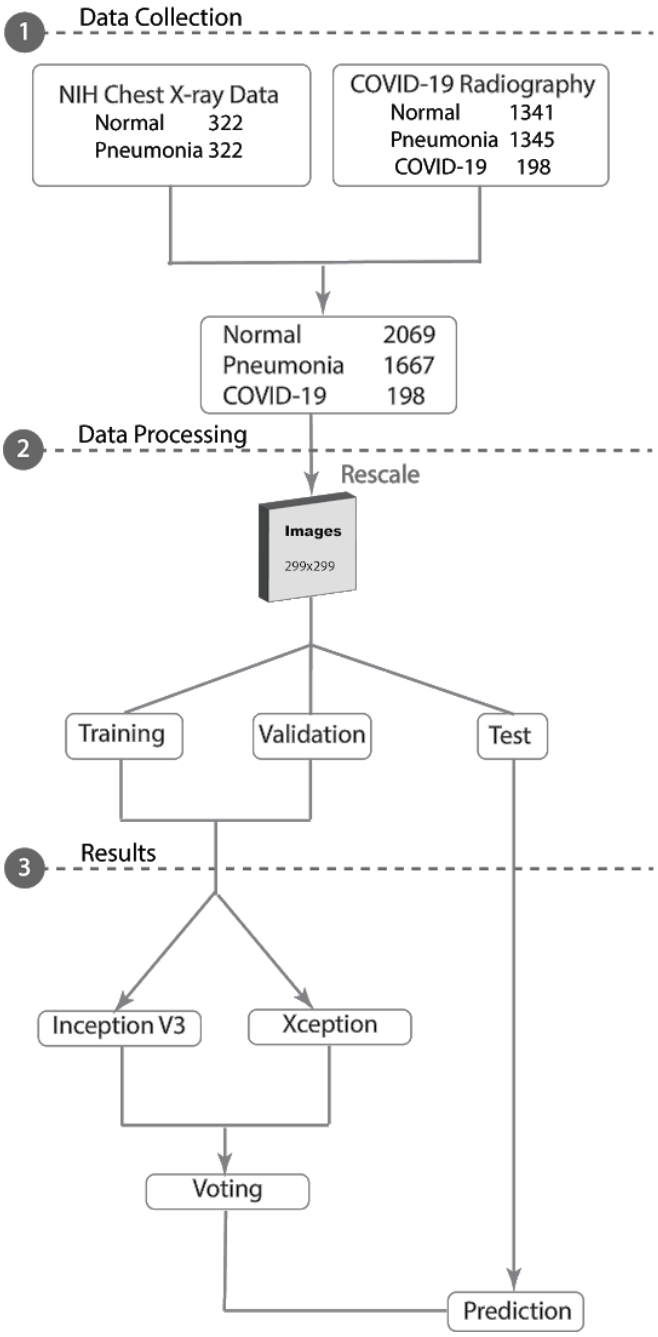

Figure 2. Diagram of Data Collection and Modeling Process

Results

The test set is composed of 706 images: 40 COVID-19 cases, 333 normal cases and 333 other/pneumonia cases. Figure 3 displays the performance achieved on this test set for the

Inception V3 and for the Xception model, as well as for an *ensemble* model that computes the weighted average of the votes of these two models. The InceptionV3 has better recall, whereas the Xception has better precision. The weighted model achieves the best result, with an 89% overall accuracy score, a 93% recall and a 97% precision for the COVID-19 class. Figure 3 displays the confusion matrix and the ROC curves for each category.

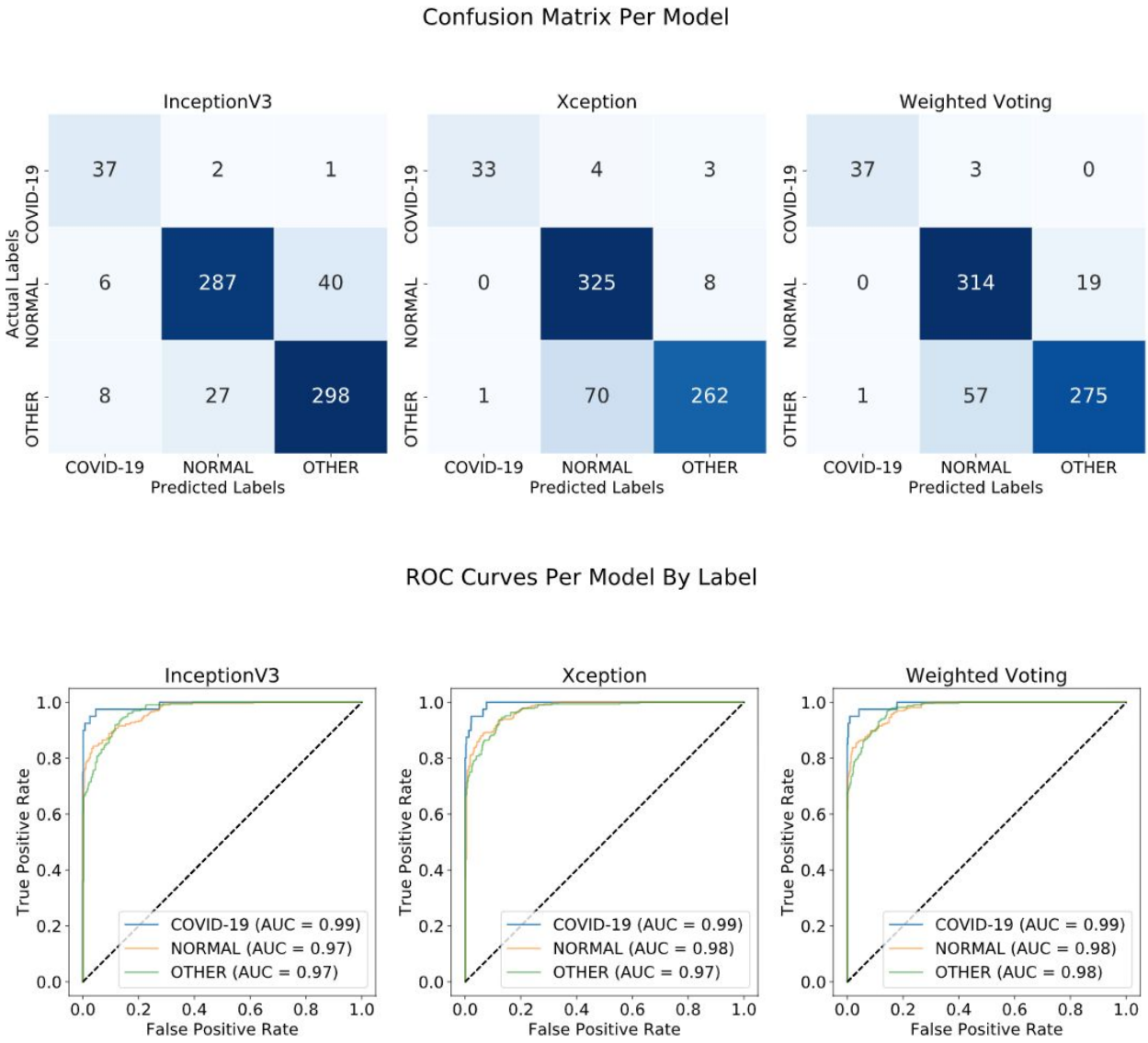

Figure 3. Model Performance on Test Set

[other indicates pneumonia]

Examining misclassified images may give some insight into where and why our model fails, but we must acknowledge the limited number of samples and our lack of medical expertise when making any hypothesis. It may be that the variable sources of the images had a confounding effect on the classifications. For example, the COVID-19 images are collected from

multiple sources and publications which means that they have higher ‘noise’ signals than the other classes (See Figure 4).

It is also important to consider the limitations of the images themselves. It is possible that by the time that a COVID-19-positive individual receives a chest scan their disease has progressed significantly. It may be that our model would not be effective in early detection of the disease, as those patients may not have sufficient lung damage to be detected. Additionally, the datasets used have not been peer-reviewed or collected in a standardized fashion. Images were taken on people of varying ages and sizes, on different machines, and those images were processed differently. There are also images marked by physicians which could lead to possible information leakage.

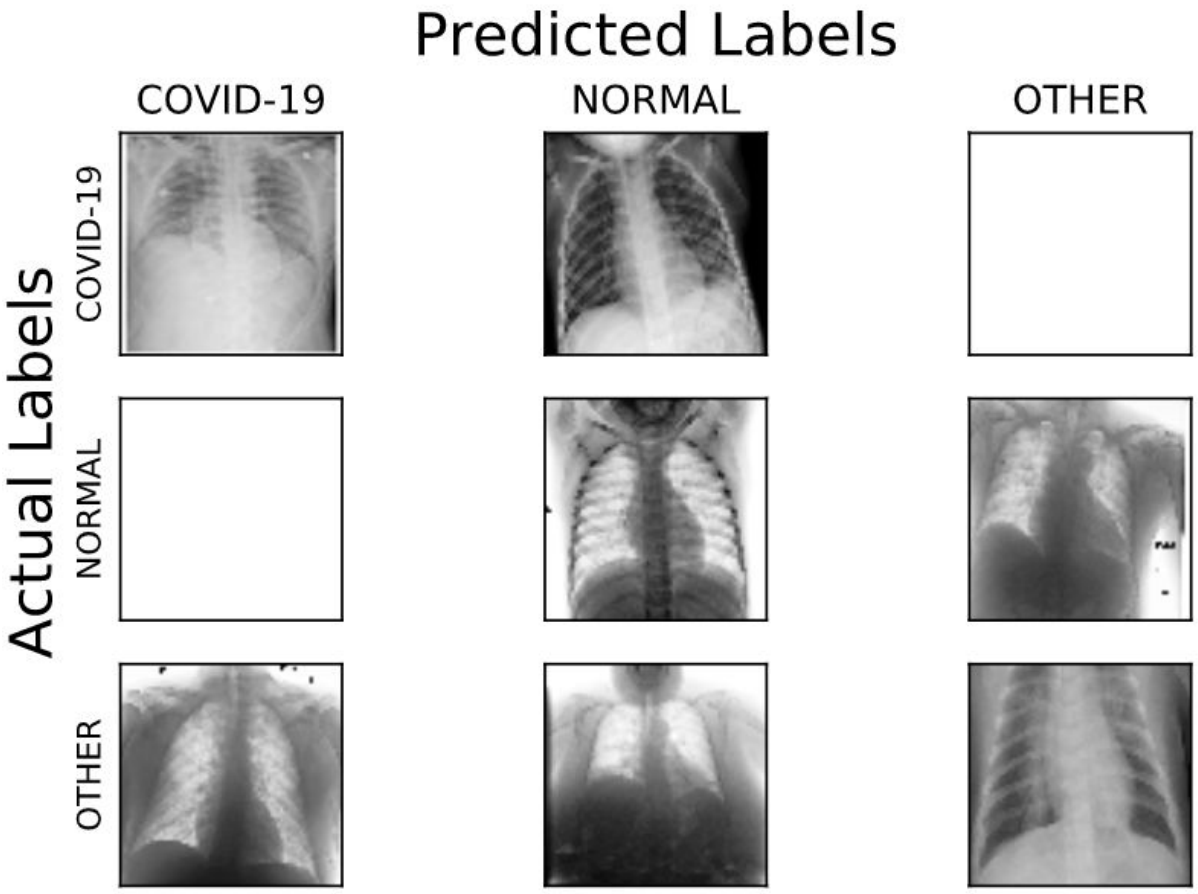

Figure 4. Samples of Classified Images

[other indicates pneumonia, blanks indicate that images with that actual and predicted label combination do not exist in the test set]

We are particularly interested in distinguishing COVID-19 cases from non-COVID-19 cases. When performing this distinction, our model achieves an AUC of 0.99. While this distinction is the most important, our model also achieves a 0.98 AUC in identifying both healthy patients ("Normal") and patients with other forms of pneumonia ("Other").

In the literature, the model which is most closely comparable to ours is that of Apostolopoulos and Mpesiana [21], since they also classify patients as having COVID-19, being normal, or having pneumonia. Their model is also the best so far, so it serves as a competitive benchmark. Their model achieves a recall of 98.66% and a specificity of 96.46%. Our model has a recall of 93% and a specificity of 99.8%. Apostolopoulos and Mpesiana's [21] model is superior to ours in terms of sensitivity, but has a lower specificity.

Conclusion

Our algorithm consisted of a weighted voting of two CNNs (InceptionV3 and Xception) that were trained to classify a patient as having COVID-19, other disease, or being normal. This model achieved an AUC of 0.99, with 93% recall and 97% precision on the COVID-19 class. In contexts where testing with RT-PCR is unfeasible, algorithms such as the one developed in this paper could help screen patients. However, any algorithm must still prove its reliability under proper clinical trials. Nonetheless, this study demonstrates that neural networks may be a valuable tool in fighting COVID-19.

References

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Introduction to pathogens," in Molecular Biology of the Cell. 4th edition: Garland Science, 2002.

[2] D. E. Bloom and D. Cadarette, "Infectious Disease Threats in the 21st Century: Strengthening the Global Response," Frontiers in immunology, vol. 10, p. 549, 2019.

[3] A. T. Pavia, "Viral infections of the lower respiratory tract: old viruses, new viruses, and the role of diagnosis," Clinical Infectious Diseases, vol. 52, no. suppl_4, pp. S284-S289, 2011.

[4] Y. Huang, "The SARS epidemic and its aftermath in China: a political perspective," Learning from SARS: Preparing for the next disease outbreak, pp. 116-36, 2004.

[5] C. f. D. C. a. Prevention. "Middle East Respiratory Syndrome." https://www.cdc.gov/coronavirus/mers/index.html (accessed.

[6] W. H. Organization, "Coronavirus disease (COVID-19) Pandemic," 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[7] C. f. D. C. a. Prevention, "Coronavirus Disease 2019 (COVID-19) Information for Laboratories," 2020. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/lab/index.html.

[8] C. f. D. C. a. Prevention, "CDC - 2019-nCoV Real-Time RT-PCR Diagnostic Panel," 2020. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/downloads/Factsheet-for-Patients-2019-nCoV.pdf.

[9] J. Liu, H. Yu, and S. Zhang, "The indispensable role of chest CT in the detection of coronavirus disease 2019 (COVID-19)," European Journal of Nuclear Medicine and Molecular Imaging, p. 1, 2020.

[10] Y. Li and L. Xia, "Coronavirus Disease 2019 (COVID-19): Role of chest CT in diagnosis and management," American Journal of Roentgenology, pp. 1-7, 2020.

[11] T. Ai et al., "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," Radiology, p. 200642, 2020.

[12] T. W. Post. (2020) How U.S. coronavirus testing stalled: Flawed tests, red tape and resistance to using the millions of tests produced by the WHO
. Available: https://www.washingtonpost.com/business/2020/03/16/cdc-who-coronavirus-tests/

[13] E. J. Emanuel et al., "Fair allocation of scarce medical resources in the time of COVID-19," ed: Mass Medical Soc, 2020.

[14] X. Xuanyang, G. Yuchang, W. Shouhong, and L. Xi, "Computer aided detection of SARS based on radiographs data mining," in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2006: IEEE, pp. 7459-7462.
[15] P. Rajpurkar et al., "Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
[16] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, "Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database," Computers & Electrical Engineering, vol. 78, pp. 388-399, 2019.
[17] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," arXiv preprint arXiv:2003.10849, 2020.
[18] F. M. Salman, S. S. Abu-Naser, E. Alajrami, B. S. Abu-Nasser, and B. A. Alashqar, "COVID-19 Detection using Artificial Intelligence," 2020.
[19] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection," arXiv preprint arXiv:2003.12338, 2020.
[20] F. Shi et al., "Large-Scale Screening of COVID-19 from Community Acquired Pneumonia using Infection Size-Aware Classification," arXiv preprint arXiv:2003.09860, 2020.
[21] I. D. Apostolopoulos and T. A. Mpesiana, "COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," Physical and Engineering Sciences in Medicine, p. 1, 2020.
[22] Chowdhury, M.E., et al., Can AI help in screening Viral and COVID-19 pneumonia? arXiv preprint arXiv:2003.13145, 2020.
[23] M.E.H. Chowdhury, T.R., A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, COVID-19 RADIOGRAPHY DATABASE. 2020.
[24] Wang, X., et al. Hospital-scale Chest x-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. in IEEE CVPR. 2017.
[25] Mahdianpari, Masoud & Salehi, Bahram & Rezaee, Mohammad & Mohammadimanesh, Fariba & Zhang, Yun. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. Remote Sensing. 10. 1119. 10.3390/rs10071119.
[26] Ai, T., et al., Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, 2020: p. 200642.