Small-Net: A COVID-19 Classifier for Underrepresented Demographics

Name: Brendan Wallace-Nash
Student Number: n9993304
Name: Bella Qian
Student Number: n10661450
Assignment: Assignment 1 COVID-19 Chest X-ray Analysis
Due Date: 14/ 11/2021

Abstract

The COVID-19 pandemic has had a drastic effect on the medical industry, no-more so than the frontline medical staff. Currently molecular swab testing is the gold standard for testing people for COVID-19, but this method leaves medical staff vulnerable as they must come in close contact with potentially COVID-19 positive patients to conduct the test. Not only does medical swab testing put the medical staff at risk but it also takes a significant amount of time between initial testing and results. These problems with molecular swab testing have demonstrated an opening for an artificial intelligence (AI) driven testing method that can mitigate against the problems of molecular swab testing. COVID-NET is an open-source machine learning driven classifier that can identify COVID-19 positive patients from X-ray images at an accuracy of 93.3%. While the results of the COVID-NET model are very impressive, it is our understanding that often the data used to train models like COVID-NET typically have biased datasets that are overly represented by men. The aim of our research is to establish if convolutional neural network (CNN) can be trained off a very small data set (2000 images) to allow for medical centres performing COVID-19 testing on potentially underrepresented demographics to build a CNN prediction model suited to their needs. Our CNN for predicting COVID-19 in patients is called Small-NET and is trained off a fraction of the 13,975 images used by COVID-NET. While the resulting test accuracy of Small-Net was not as high as that of COVID-NET, Small-Net still produced a sensitivity (97.5%) higher than that of molecular swab testing. Further testing has still to be done on Small-Net to ensure that it is effective when presented with datasets of unique demographics and alternative medical imaging techniques.

Introduction

In 2019 the World Health Organisation classified the outbreak of SARS-CoV-2 (COVID-19) as a pandemic. COVID-19 has since reshaped how many industries operate, but no industry has had to adapt as much as the medical industry. One of the key factors in curbing the spread of COVID-19 has been mass testing, which can be used to identify infected people in the population and identify areas with higher rates of COVID-19. Current testing methods can range largely in how long it takes for patients to get their results back with molecular tests taking 48-72 hours and antibody testing taking 1 to 3 weeks (Sullivan, 2021). Molecular testing has become the gold standard in COVID-19 testing and while it is an effective method there are limitations and risks to this method. Centre for Health Security (2021) explained that molecular tests amplify bits of viral RNA so that viral infection can be detected using a specialized test. The procedure begins by taking a sample from a potentially infected person's nose or mouth (saliva), where virus might be found. Nasal swab is the most common molecular testing method and involves medical staff taking off the patient's facial mask and performing a nasal swab test. While the medical staff is wearing personal protective equipment (PPE) the mass number of times they conduct the procedure daily leaves them vulnerable to the exposure of an active case of COVID-19. Molecular testing also requires staffed monitoring of each test to see if the positive indicator is present on the testing device (Hutch, 2021). The coupling of risk, time, and cost of molecular testing makes it less than an effective method to be performing the COVID-19 test on such a large scale.

There has been an increasing discussion about alternative methods that can be used to test for COVID-19, with a school of thought trending towards the use of medical imaging devices such as computed tomography (CT-scans) and chest X-rays being used to test for COVID-19. The use of CT scans of lungs from COVID-19 patients in Wuhan found that there are clinical features that define the lungs of a COVID-19 infected patient (Huang Da, 2020). Being able to use medical imagery to test for COVID-19 will allow for real-time results to be obtained on patients, which will significantly cut down on the delay between testing and results. The suggestion of using medical imagery to diagnose COVID-19 is not an attempt to phase out molecular testing but rather to use the two in tandem. Having an alternative testing method at medical hubs can allow for triaging of patients with more severe cases to get real-time medical imagery testing and for asymptomatic patients to be given the molecular testing method. Medical imagery also provides a greater level of protection from COVID-19 transmission to the medical staff as they can conduct screening without any contact or interaction with the patient (Dennie, 2020). While CT scans provide more detailed imagery of the clinical indicators of COVID-19 in infected patients' lungs, they are a less abundant form of

equipment, especially in developing countries, leading research to focus on X-ray technology as it is more accessible (Silverstein, 2016).

While medical imagery can produce a real-time diagnosis of COVD-19 it is limited by the fact that a radiologist will have to examine the medical imagery. The mass scale of COVID-19 testing that is currently being conducted would not be manageable by the number of radiologists in a hospital, which is typically around 2 per hospital (Garland, 1963). This is where adopting machine learning classifiers to diagnose images can allow for an efficient diagnosis rate of X-ray images. While the ethics of having artificial intelligence (AI) and machine learning (ML) techniques in the medical industry has long been questioned, there is a growing consensus an AI-driven medical industry could revolutionise the sector, and its use in this situation could drastically increase the efficiency of testing for COVID-19 (Sara Gerke, 2020). There is a vast array of deep neural networks, but the convolutional neural network (CNN) is the most prominent in image classification and has already been successfully used to identify pneumonia from chest X-rays (Yadav, 2019).

While there is already an open-source tailored CNN that can diagnose lung X-ray images with high efficacy (COVID-NET), their model is trained on 13,975 high-quality X-ray pictures (Linda Wang, 2020). While their efforts in the development of the open-source COVID-NET classifier are commendable, the researchers would be the first to say that the model is still not production-ready (Linda Wang, 2020). The identified problem in COVID-NET this paper aims to address is the potential bias that could be caused by the sample data used. While the data set used in the COVID-NET was well-sourced from a diverse range of locations there was no reference to percentages of gender and age demographics that made up the dataset. This could become problematic as there has been evidence to suggest that women are often underrepresented in AI training datasets, and as a result, the AI models typically perform worse on women (Robbins, 2020). The X-ray dataset gathered for the COVID-NET training also only featured one method of X-ray imagery. While the COVID-NET dataset did use imagery obtained from using the standard method of X-ray imagery there has been growing research into the use of dark-field X-ray imaging (Morton, 2021). Dark-field X-ray imagery produces a more prominent black image with higher contrast and is being touted as the most significant advancement in X-ray since its inception (Morton, 2021). Building a model that does not incorporate this technique could reduce the integrability of the COVID-NET model if dark-field X-ray is to become the norm method of medical imagery.

The research conducted in this paper aims to build a model that can identify COVID-19 infected patients from X-rays but be able to have a model that is built off a much smaller dataset than that of COVID-NET. This will allow for individual hospitals to perform their own training on a

dataset relevant to their environment, whether that is training the model to be suited to a particular gender or age demographic, or X-ray imagery type, assuring a greater level of transparency and interoperability than that of COVID-NET.

Methods

The approach to this study was to develop a CNN to identify COVID-19 from a small dataset of X-rays from both COVID-19 positive patients and negative patients, the CNN from this research will be referred to as Small-Net. The purpose of the small data set is to replicate a dataset of X-rays a medical centre could generate within a few days of using X-ray testing to train the Small-Net to suit their patient demographic and X-ray methodology. The dataset consists of patients with COVID-19, which were labelled positive (binary label is 1), while patients with either no respiratory problems or pneumonia were labelled negative (binary label is 0). The collection, processing and subsequent Small-Net architecture the data was used to train will be explained below. All data and code used in this report will be provided on GitHub here.

COVID-19 dataset was comprised of X-ray images from the COVIDx dataset. The COVIDx dataset was developed in the making of COVID-NET and consists of data from five different datasets.

1. COVID-19 Image Data Collection (Cohen, 2020).
2. COVID-19 Chest X-ray Dataset Initiative (Chung, 2020).
3. Actualmed COVID-19 Chest X-ray dataset initiative (Wang, 2020).
4. RSNA Pneumonia Detection Challenge Dataset (Rahman, 2021).
5. COVID-19 Radiography database (Radiological Society of North America, 2019).

A text file was provided by the COVIDx dataset to split the data into training and testing datasets. The original dataset provided by COVIDx is represented in *figure 1* and demonstrates a large disparity between the number of positive and negative patients in the training data, while the split between training and testing data is imbalanced at 97.5% train and 2.5% testing.

Figure 1: COVIDx Data Distribution

Type	COVID-19 Positive	COVID-19 Negative	Total
Train	2158	13794	15952
Test	200	200	400

Of the original COVIDx dataset 900 (5.5%) files that originated from the COVID-19 Radiology database were removed because of formatting problems, it was deemed acceptable to remove

these as they made up such a small percentage of the data. The dataset from the COVIDx dataset was then randomised before reducing it to a training sample of 2,430 and a test sample of 400. The randomising of the training data was to ensure that there was a variety of data from all the sources that make up the COVIDx data set. As demonstrated in *figure 2* there is an even distribution of positive and negative data in both the training and the testing.

Figure 2: Small-Net Dataset

Type	COVID-19 Positive	COVID-19 Negative	Total
Train	1215	1215	2430
Test	200	200	400

All images from the COVIDx have had the top 8% of the image cropped to control against text embedded in the photo, although there were still images with text embedded in them as seen in *figure 3*. The images were then resized to 250x250 to ensure all photos were the same size and that there are fewer parameters and thus making the Small-Net less computationally heavy. The process of image resizing was done by trial and error, with the aim to make the image size smaller to be computationally efficient while still maintaining clarity of the details in the image. *Figure 4* shows an original image compared to its resized image. As you can see in *figure 4* there are light grey vein like features in the centre of the lungs, these features were found to be unique to COVID-19 positive patients, so the aim was to make these features still definable even in the resized image. The resized images were set to grayscale and then transformed to an array of pixel values. The corresponding label for each picture was changed to binary, with 1 representing positive and 0 representing negative. The resulting X and y datasets for test and training were converted to a pickle file and can be found on our GitHub.

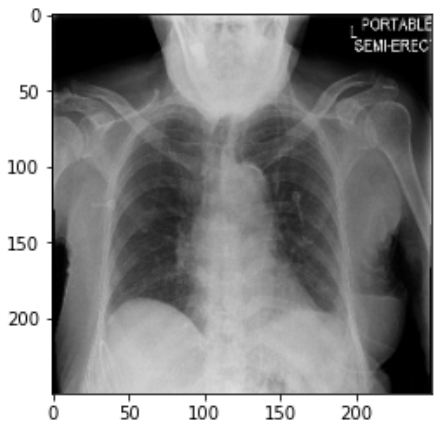

Figure 3: Text in Image

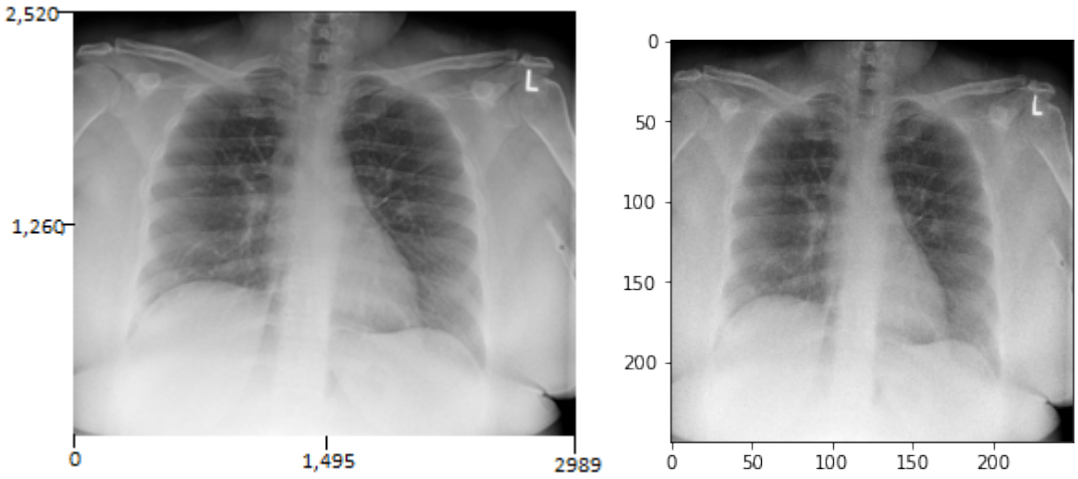

Figure 4: Original Image (left) Transformed for CNN (right) (metric: Pixels)

Small-Net architecture was developed using *TensorFlow* and *Keras.* The reasoning for choosing these packages to develop the Small-Net architecture is because they are open source so they will be available to be implemented by medical centres and the simplicity of the tools in these packages make the architecture of the Small-Net model succinct and readable. The architecture of the model was derived from trial and error of modifying parameters, with the focus being on making a fast model that could be trained on computers available in medical centres. While the aim is always to have a model with high test accuracy, interoperability with the computing power at medical centres was also needed to be considered as it would be likely that the training of the model would be run-on desktops without advanced GPU's and would have to be trained in the smaller time frame than that of architectures that take hours. The resulting architecture for the Small-Net can be seen in *Table 1*. The architecture of Small-Net consists of 6 filter layers, 1 hidden layer, and an activation layer. Kernel size was kept at 3x3 for all convolutional layers as it is obtaining more detail from the images while being computationally efficient. The architecture of Small-Net follows a pattern of increasing filter sizes, with the 2nd and 3rd, and 4th and 5th filters having the same value. The increase in filter size is to allow for Small-Net to identify more complex patterns as the filters get larger. The image data was normalised before being entered into the CNN by dividing each image by 255. Normalising images by dividing the array by 255 is done as the colour scale of grey-scale images in between the values of 0-255 making all images values in the ray between 0-1.

Table 1: Small-Net Architecture

Layer	Filters	Kernel/Pool Size	Activation

1st Convolutional Layer	32	3x3	Relu
1st Max pool 2D		2x2	
2nd Convolutional Layer	64	3x3	Relu
2nd Max Pool 2D		2x2	
3rd Convolutional Layer	64	3x3	Relu
3rd Max Pool 2D		2x2	
4th Convolutional Layer	128	3x3	Relu
4th Max Pool 2D		2x2	
5th Convolutional Layer	128	3x3	Relu
5th Max Pool 2D		2x2	
6th Convolutional Layer	256	3x3	Relu
Max Pool 2D		2x2	
Hidden Layer (Dense)	512		Relu
Activation Layer (Dense)	1		Sigmoid

The method for compiling Small-Net can be seen in *table 2*. A call-back function was also used in the training of the Small-Net model to allow for early stopping at an epoch where the validation accuracy was over 95%. While using early stopping can possibly stop the training prematurely and risk missing out on a higher validation accuracy, it was conducted as the validation accuracy did not increase stably and it was not possible to run the model for a set number of epochs and guarantee validation accuracy would be above 90%. Epochs for the training were set to 30 and batch size set to 80, these parameters were chosen as they will be optimal for the types of computers available at medical centres. Training data used to train the model was split into subsets of 80% training and 20% validation.

Table 2: Small-Net Compiler

Loss Function	Optimizer	Metrics
Binary Cross-Entropy	Adam	Accuracy

Results

To ensure the interoperability of Small-Net to computers in a medical centre, all training and testing of Small-Net were conducted on a 2020 MacBook Air M1 with 8 gigabits (a computer with

similar specifications to typical workstation computers). The total training and testing time of the Small-Net model took 14 minutes in total.

The results of Small-Net can be seen in *figure 5*. Small-Net only has marginal underfitting with training accuracy being 93.78% and validation accuracy being only 95.52%. The difference in cross-entropy loss is again marginal between the training and test data, with training cross-entropy loss at 0.14 and validation cross-entropy loss at 0.11. *Figure 5* displays that validation loss did not drop as consistently as training loss, but the validation loss of 0.11 is reprehensive of a good model as it remains close to that of a cross-entropy loss of 0 that would indicate a perfect model.

Figure 5: Accuracy and Loss of Test and Train Data

A comparison of how Small-Net's results in accuracy and computational complexity can be found *in table 3*. Note that model computational complexity and accuracy data for VGG-19, ReNet-50 and CVID-NET are taken from COVID-NET report written by Wang (2020) and the same testing data was used for all models. While Small-Net does have lower accuracy than COVID-Net, ResNet-50 and VGG-19 It should be noted that while Small-Net is trained and tested on COVIDx data, it is only run on a fraction of the data that COVID-Net, VGG-19, and ResNet-50 were run on. Small-Net also has fewer parameters than all three other architectures, representing the smaller depth of the architecture and its lower computational complexity, making it more suitable for workstation computers.

Table 3: Models Computational Complexity and Accuracy

Architecture	Params (unit of millions)	Accuracy (%)
VGG-19	20.37	83.0
ResNet-50	24.97	90.6
COVID-Net	11.75	93.3
Small-Net	00.70	76.5

Figure 6 shows a confusion matrix for predicted positive and negative diagnoses and true diagnoses. There is a good level of model sensitivity at 97.5% and positive predictive value at 68.66% and negative predictive value at 95.69%. This indicates that of the 200 active cases tested only 2.5% were wrongly diagnosed as not active. A low false negative rate is one of the most critical metrics in medical testing of highly contagious diseases as it means that these patients will receive the right treatment and in the case of an infectious disease the proper quarantine methods can be taken. We acknowledge that Small-Net does have a rather high false positive rate. In a medical testing context, this means that some patients are wrongly categorised as COVID-19 patients and some extra unnecessary quarantine restrictions could be applied to them. It should also be noted that the COVIDx dataset had pneumonia patients included in the non-COVID-19 patient group, so we are not able to tell exactly know how many of the false positives and false negatives are a result of pneumonia patients being misdiagnosed as COVID-19 patients and if that was a significant factor in the sensitivity of the model. The misdiagnosing of pneumonia patients as COVID-19 could be problematic if the treatment needed differs substantially between the two diseases, causing pneumonia patients to go untreated.

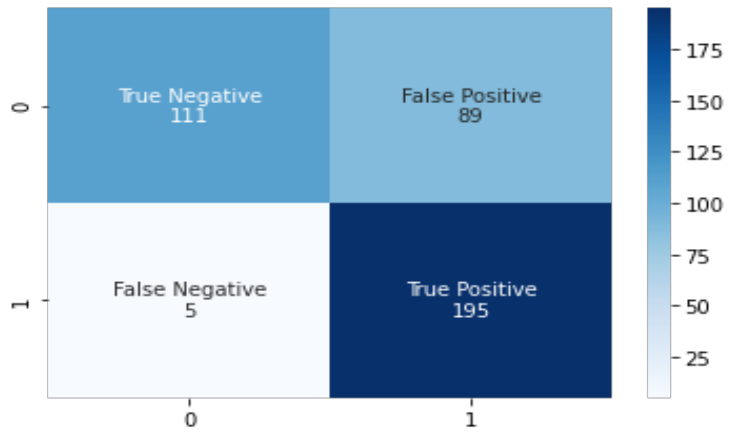

Figure 6: Small-Net Sensitivity

Conclusion

This study looked at implementing Small-Net's on a fraction of the COVIDx dataset used by COVID-Net, while maintaining a computational complexity that could be implemented by medical centres. Small-Net demonstrated to be able to produce a high accuracy in correct diagnosis of COVID-19 positive patients from the small amount of data used in training while being accomplished on a basic specification computer in a very accessible time. Small-Net also demonstrated a sensitivity higher than that of standard methods with a sensitivity of 97.5%, marginally above that of the sensitivity of molecular swab testing (90.7%) (Kanji, 2021). Comparatively, it still falls short of the level of accuracy that COVID-Net can produce but the portability of Small-Net means medical centres can apply Small-Nets architecture to a sample that is representative to their population demographic through the use of a small number of X-rays that could be acquired responsively by medical centres. The ability for medical centres to be able to build a represented dataset of their clinic allows for a greater layer of transparency in the potential biases involved in AI-driven clinical techniques.

The limitations and ethical considerations should be understood when considering Small-Net. As seen in the results, while Small-Net has a good level of sensitivity to predicting true positive cases, it does however report a high number of false positives. This could both lead to health and economic implications for patients as they could be treated for the wrong disease and be put into quarantine wrongly, effectively taking them out of the workforce for a period of time. While false negatives are less than that of molecular swab testing it still presents the risk of misdiagnosing an active patient and allowing for potential community outbreaks.

While the efficiency of a classifier such as Small-Net could greatly aid medical staff in diagnosing patients with COVID-19 through faster testing results than standard molecular swab tests and providing greater protection to medical staff, it is still not ready for production use. There still needs to be qualitative testing on Small-net to ensure that the model is using relative information to identify COVID-19 in patients and establish if it is wrongly diagnosing pneumonia patients. Furthermore Small-Net needs to be trained and tested on the datasets it was intended for. Training and testing Small-Net on small datasets of demographics such as women or children will prove the efficacy of Small-Net. There will also need to be a procedure in place on how a technical programming procedure such as Small-Net could be implemented in hospitals where there might not be a great understanding of the domain. Further testing and planning for Small-Net could make it the go to method for safely testing COVID-19 in underrepresented demographics and for less common medical imaging techniques.

References

Audrey Chung, L. W. (2020, May 9). *Figure1-COVID-chestxray-dataset*. Retrieved from GitHub: https://github.com/agchung/Figure1-COVID-chestxray-dataset

Center for Health Security. (2021). *Antigen and Molecular Tests.* https://www.centerforhealthsecurity.org/covid-19TestingToolkit/testing-basics/types-of-COVID-19-tests/antigen-and-molecular-tests.html

Chaolin Huang Da, Y. W. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 497-506, URL: https://www.sciencedirect.com/science/article/pii/S0140673620301835.

Cohen, J. P. (2020). COVID-19 image data collection. *Electrical Engineering and Systems Science*, https://arxiv.org/abs/2003.11597.

Dennie, C. H. (2020). Canadian Society of Thoracic Radiology/Canadian Association of Radiologists Consensus Statement Regarding Chest Imaging in Suspected and Confirmed COVID-19. *Canadian Association of Radiologist Journal*, 470-481, DOI: https://doi.org/10.1177/0846537120924606.

Fred Hutch. (2021, September 21). *How Scientists Test for COVID-19*. Retrieved from Fred Hutch: https://www.fredhutch.org/en/research/diseases/coronavirus/serology-testing.html

Garland, L. H. (1963). The Radiological Service—How Many Radiologists Are Advisable? *Radiology*, DOI: https://doi.org/10.1148/80.4.686.

Jamil N. Kanji, N. Z. (2021). False negative rate of COVID-19 PCR testing: a discordant testing analysis. *Open Access*, URL: https://virologyj.biomedcentral.com/articles/10.1186/s12985-021-01489-0#:~:text=By%20extension%2C%20the%20sensitivity%20(1,%25%20CI%2082.6%E2%80%939 8.9%25).

Linda Wang, Z. Q. (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Report*, doi: 10.1038/s41598-020-76550-z.

Morton, W. (2021). X-ray dark-field imaging may help diagnose lung disease. *diagnostic imaging*, URL: https://physicsworld.com/a/x-ray-dark-field-imaging-may-help-diagnose-lung-disease/#:~:text=X-ray%20dark-field%20chest%20imaging%20%E2%80%93%20a%20new%20technique,prototype%20and%20tested%20the%20system%20in%20healthy%20patients.

Radiological Society of North America. (2019). *RSNA Pneumonia Detection Challenge*. Retrieved from Kaggle: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data

Robbins, R. (2020). AI systems are worse at diagnosing disease when training data is skewed by sex. *HEALTH TECH*, url: https://www.statnews.com/2020/05/25/ai-systems-training-data-sex-bias/.

Samir S. Yadav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Open Access*, URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0276-2#:~:text=Notably%2C%20the%20convolutional%20neural%20network,professional%20expertise%20to%20label%20them.

Sara Gerke, T. M. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Elsevier Public Health Emergency Collection*, 295–336, doi: 10.1016/B978-0-12-818438-7.00012-5.

Silverstein, J. (2016, September 2016). *Most of the World Doesn't Have Access to X-Rays*. Retrieved from The Alantic: https://www.theatlantic.com/health/archive/2016/09/radiology-gap/501803/

Sullivan, D. (2021, March 9). *How Long Does It Take to Receive COVID-19 Test Results?* Retrieved from Health Line: https://www.healthline.com/health/how-long-does-it-take-to-get-covid-results

Tawsifur Rahman, D. M. (2021, March). *COVID-19 Radiography Database*. Retrieved from Kaggle: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

Wang, A. W. (2020, May 6). *Actualmed COVID-19 Chest X-ray Dataset Initiative*. Retrieved from Github: https://github.com/agchung/Actualmed-COVID-chestxray-dataset