Predicting COVID-19 From Chest X-Ray Images Using Transfer Learning

Shirin Yamani [1], **Mahsa Malek** [2]
[1]*Department of Software Engineering, University of Calgary, Calgary, Alberta, Canada*
[2]*Department of Software Engineering, University of Calgary, Calgary, Alberta, Canada*

Abstract: For the past almost two years COVID-19 pandemic is causing a major outbreak around the globe by having a severe impact on the health and life of many people. During this chaotic period, one effective step to combat with COVID-19 is the ability to detect the infected patients at the very early stage, and put them under special care. Relatively, clinical studies have shown that COVID-19 infected patient has specific abnormalities in his/her chest radiograms. In this research, inspired by earlier works, we study the application of deep learning models to detect COVID-19 patients from the abnormalities in their chest X-ray images. We used a publicly available dataset of 5000 Chest X-rays as our input followed by a Transfer Learning based approach on a subset of 2,000 radiograms of this dataset to train ResNet18 model, to identify COVID-19 disease in the chest X-ray images. Afterwards, we evaluated this model on the remaining 3,000 test set images, then examined sensitivity and specificity rates on different thresholds to evaluate the performance of our model.

Keywords: COVID-19 Detection, X-ray Image Processing, Transfer Learning, ResNet18 Model

1 Introduction

The recent outbreak of the novel coronavirus disease (COVID-19) started in Wuhan at the end of 2019 and its spread across the world, caused a major effect on global health. Due to the unavailability for vaccine at the early 2020, early diagnosis of the virus has been very crucial to combat the spreading of the virus in society. As known, one of the fastest solutions for pneumonia diagnosis, is Chest radiography imaging (e.g., X-ray). Chest X-rays are valuable resources since they show visual indexes correlated with COVID-19 very accurately [1]. During the early course of COVID-19, ground glass pattern is seen in areas that edges the pulmonary vessels and may be difficult to appreciate visually [2]. Such subtle abnormalities can only be interpreted by expert radiologists. However, considering huge rate of suspected people and limited number of trained radiologists, automatic methods for identification of such obvious abnormalities can assist the diagnosis at early stages of this disease. Therefore, we believe that Machine Learning (ML) solutions are potentially powerful tools for solving such problems.

So far, a small dataset of COVID-19 X-ray images was collected, which made it possible for AI researchers to train ML models to perform automatic COVID-19 diagnostics from X-ray images [3]. These images were extracted from academic publications reporting the results on COVID-19 X-ray and CT images. We used these images as our positive samples for COVID-19. We then used a subset of images from ChexPert dataset [4], as the negative samples for COVID-19 detection. The combined dataset has around 5,000 Chest X-ray images (called COVID-Xray-5k), which is divided into 2,000 training, and 3,000 testing samples.

Afterwards, An ML framework was employed to predict COVID-19 from Chest X-ray images. Unlike the classical approaches for medical image classification which follow a two step procedure (hand-crafted feature extraction + recognition), we use an end-to-end deep learning framework which directly predicts the COVID-19 disease from raw images without any need of feature extraction.

In this framework, we train a popular CNN, namely; ResNet18 [5] which has achieved promising results in several tasks in recent years. We trained ResNet18 on COVID-Xray-5k dataset, and analyze its performance for COVID-19 detection. Since so far there is a limited number of X-ray images publicly available for the COVID-19 class, we cannot simply train these models from scratch. Therefore, Transfer Learning based strategy was adopted to address the COVID-19 image scarcity issue in this work:

- We first used data augmentation to create transformed version of COVID-19 images (such as flipping, small rotation, adding small amount of distortions), to increase the number of samples; Then instead of training the model from scratch, we fine-tune the last layer of the pre-trained version of our selected model (ResNet18) on ImageNet [6]. This way, the model can be trained with less labeled samples from each class.

The above strategy helped to train the ResNet18 network with the available images, and achieve reasonable performance on the test set of 3,000 images.

2 Related Works

Since the start of COVID-19, researchers quickly divided their effort on combating it by focusing on developing a vaccine in one hand [7] and detecting COVID-19 using PCR and imaging systems on the other hand. Here, we review studies devoted to the use of radiography images to aid and complement PCR in diagnosing COVID-19 cases. [8] built a deep CNN on ResNet50, InceptionV3 and Inception-ResNetV2 models for the classification of COVID-19 Chest X-ray images to normal and COVID-19 classes.

They reported a significant correlation between CT image results and PCR approach. [9] proposed a method to detect COVID-19 using X-ray images based on deep feature and support vector machines (SVM). They collected X-ray images from GitHub, Kaggle and Open-I repository. They extracted the deep feature maps of a number of CNN models and conclude that ResNet50 is performing better despite the small number of images used in their investigation. [10] proposed a simple CNN of 16 layers only to detect COVID-19 using both X-ray and CT scans and reported good performance but the dataset used is small. [11] focused on segmenting COVID-19 CT scans using a deep learning approach known as VB-Net and reported dice similarity of 91%±10%. [12] used CT images to predict COVID-19 cases where they deployed Inception Transfer Learning model to establish an accuracy of 89.5% with specificity of 88.0% and sensitivity of 87.0%. In [13] a number of CNN architectures that are already used for other medical image classifications, evaluated over a dataset of X-ray images to distinguish the coronavirus cases from pneumonia and normal cases. CNN's adopted on a dataset of 224 images of COVID-19, 700 of non- COVID19 pneumonia, and 504 normal where they report overall accuracy of 97.82. In [14] authors experimented several CNN architectures classify normal X-ray images with COVID-19 X-rays and they report excellent classification accuracy, sensitivity and specificity. However, they trained their CNNs based on 50 images from each of the normal and COVID-19 classes which may result in some sort of bias in the training phase.

In all the works discussed here, to the best of our knowledge, we did not notice either of the aformentioned studies, have taken ResNet18 into consideration as their pretrained model. Therefore, in the following we aim to train ResNet18 model on a dataset of 5,000 images with binary labels from Chest X-ray images, and evaluated its performance on a test set of 3,000 images for the purpose of COVID-19 detection.

3 Dataset

In this work, we used **COVID-XRay-5K dataset** as an input to our model. This dataset includes overall number of 5,000 images which has been created from two sources; 1) **Covid-Chestxray-Dataset** (https://github.com/ieee8023/covid-chestxray-dataset), for COVID-19 X-ray samples and 2) **ChexPert dataset** [4] for Non-COVID samples. COVID-19 samples from Covid-Chestxray-Dataset are extracted from a several publications, which all has been verified to have a clear sign of COVID-19 by the researchers [15]. The Non-COVID samples of this dataset has been uniformly selected from sample images of ChexPert. As far as known, for each classification problem in deep learning, we need to have two categories of data; 1) Training set (to train the model on) and 2) Test set (to evaluate the proposed model). Considering that, from this dataset, 450 COVID-19 images has allocated to training phase along with 100 to the test set; Also 2,000 Non-COVID-19 images has allocated to training phase along with 3,000 to the test set. It is worth mentioning that those 450 images were achieved after several data-augmentation techniques (e.g. randomly resized crop, random rotation, random horizontal/vertical fip.) via Python Augmentor library. (https://github.com/mdbloice/Augmentor). **Fig. 1** shows a sample images from COVID-Xray-5k dataset. The images in the first row show 4 COVID-19 images. The images in the second row are 4 sample images of a sub-category in Non-COVID images from ChexPert, and the images in the remaining rows denote non-COVID cases.

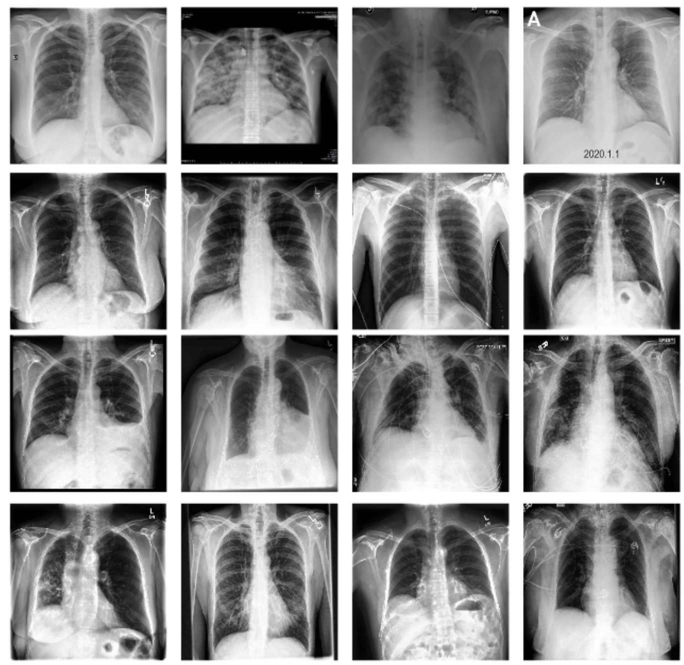

Figure 1: *Sample images from COVID-Xray-5k dataset*

4 Proposed Framework

To overcome the limited data sizes, Transfer Learning was used to fine-tune a popular pre-trained deep neural networks; namely, ResNet18 [16], on the images from COVID-Xray-5k dataset.

4.1 Transfer Learning Approach

In practice, very few people train an entire Convolutional Network from scratch (with random initialization), because it is relatively rare to have a dataset of sufficient size. Instead, it is common to pre-train a model on a very large dataset (e.g. ImageNet, which contains 1.2 million images with 1,000 categories), and then use it either as an initialization or a fixed feature extractor for the task of interest. And that is exactly the definition of Transfer Learning. The two major Transfer Learning scenarios look as follows:
1) Pre-trained model as fixed feature extractor in which we only remove the last fully-connected layer, then treat the rest of the model as a fixed feature extractor for the new dataset (i.e. the internal weights of the pre-trained model are not adapted to the new task, only a classifier is trained

on top of it to perform classification).
2) Fine-tuning the whole network in which we not only replace and retrain the classifier on top of the network on the new dataset, but to also fine-tune the weights of the pre-trained network by continuing the backpropagation. It is possible to fine-tune all the layers of the network, or it's possible to keep some of the earlier layers fixed (due to overfitting concerns) and only fine-tune some higher-level portion of the network.
In our case, since the number of images in the COVID-19 category is very limited, we only fine-tune the last layer of the CNN, and essentially use the pre-trained models as a feature extractor. We then evaluate the performance of ResNet18 on the test set. In the following section (4.2), we provide a quick overview of the ResNet18 architecture for the purpose of COVID-19 detection.

4.2 Residual ConvNet (i.e. ResNet18) for Covid-19 Recognition

As discussed in the former section, in this work we use the pre-trained ResNet18, trained on ImageNet dataset as our feature extractor network. ResNet18 is one of the most popular CNN architecture, which provides easier gradient flow for more efficient training by implementing the definition of learning residual functions with reference to the layer inputs, instead of learning un-referenced functions. The core innovation of ResNet is the use of skip connections that skips one or more layers aiming to solve the problem of vanishing gradient that may result in stopping the weights in the network to further update/change. In other words, this would help the network to provide a direct path to the very early layers in the network, making the gradient updates for those layers much easier. The overall architecture of ResNet18 model, and how it is used for COVID-19 detection is shown in **Fig 2**.

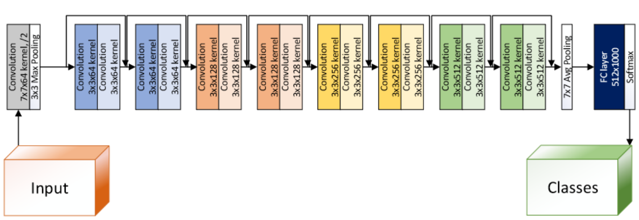

Figure 2: *High-level architecture of ResNet18 model*

4.3 Model Training

The employed model is trained with a cross-entropy loss function, which tries to minimize the distance between the predicted probability scores, and the ground truth probabilities (derived from labels), and is defined as:

$$LOSS_{CE} = -\sum_{i=1}^{N} p_i \log q_i \tag{1}$$

where pi and qi denote the ground-truth, and predicted probabilities for each image, respectively. We can then minimize this loss function using stochastic gradient descent algorithm (SGD).

5 Experimental Setup and Results

5.1 Model Hyper-parameters

We fine-tuned the ResNet18 model for 100 epochs. The batch size is set to 20, and ADAM optimizer is used for loss function optimization, with a learning rate of 0.0001, implemented in Pytorch (https://pytorch.org/). All images are down-sampled to 224x224 before being fed to the neural network (as these pre-trained models are usually trained with a specific image resolution).

5.2 Evaluation Metrics

There are different metrics which can be used for evaluating the performance of classification models, such as classification accuracy, sensitivity, specificity, precision, and F1-score. Since the current test dataset is highly imbalanced (100 COVID-19 images, 3000 Non-COVID image), sensitivity and specificity are two proper metrics which can be used for reporting the model performance. The equations for calculating them are summarized bellow (2), (3).

$$Sensitivity = \frac{\text{correctly predicted covid-19 images}}{\text{total covid-19 images}} \tag{2}$$

$$Specificity = \frac{\text{correctly predicted non-covid-19 images}}{\text{total non-covid-19 images}} \tag{3}$$

5.3 Model Predicted Scores

As mentioned before, we have picked ResNet18 as our pre-trained network. The model predicts a probability score for each fed image, which shows the likelihood of the image being detected as COVID-19. By comparing this probability with a cut-off threshold, we can get a binary label showing if the image is COVID-19 or not. An ideal model in this case should predict the probability of all COVID-19 samples close to 1, and non-COVID samples close to 0. In **Fig. 3** we shown the ResNet18 result of the predicted probability scores for the images in the test set.

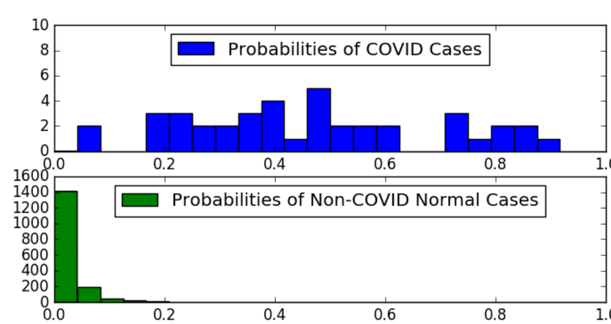

Figure 3: *ResNet18 probability scores prediction on test set*

As shown in the **Fig.3.** ResNet18 predicts a probability score showing the chance of the image being COVID-19. However, we need to set different thresholds with which we can compare these scores aiming to infer if the image is COVID-19 or not. Therefore, the predicted labels are used to estimate the sensitivity and specificity of each model. Depending on the value of the cut-off threshold, we can get different sensitivity and specificity rates for each model. Tables 1 summarizes the RestNet18 sensitivity and specificity rates for different thresholds.

Table 1: *Sensitivity and Specificity rates of ResNet18 model, for different threshold values.*

Threshold	Sensitivity	Specificity
0.10	100%	72.5%
0.15	97.6%	89.9%
0.20	94.3%	91.9%
0.25	90.0%	95.6%

5.4 Confusion Matrix

A confusion matrix is a technique for summarizing the performance of a classification algorithm. The core reason beyond using confusion matrix is because classification accuracy alone can be misleading if we have an unequal number of observations in each class or if we have more than two classes in our dataset. Therefore, the confusion matrix calculation can give us a better idea of what our classification model is getting right and what types of errors it is making. Hence, in **Fig. 4.** in order to see the exact number of correctly samples as COVID-19 and Non-COVID, the confusion matrix of the ResNet18 on 3,100 test set has been calculated.

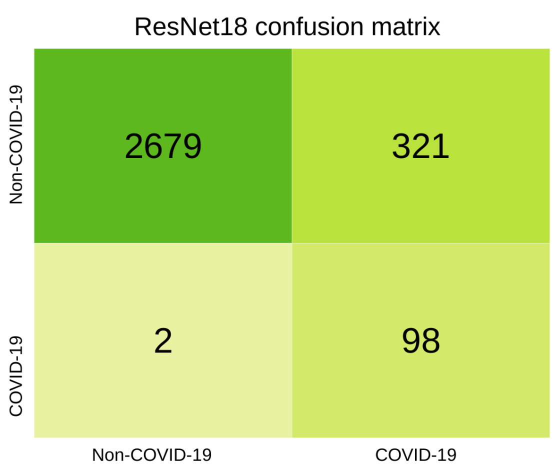

Figure 4: *ResNet18 confusion matrix*

6 Conclusion

This research presented a Transfer Learning based approach for COVID-19 detection from Chest X-ray images, by fine-tuning a pre-trained convolutional models; namely, ResNet18 on our training set. We used COVID-Xray-5k (composing of two image dataset sources). We evaluated the performance of our models on the test set of the our data, in terms of sensitivity, and specificity. We also provided confusion matrix of our model as a tool for summarizing the performance of our model. In future works, we will investigate on how to use GAN jointly with deep Transfer Learning for COVID-19 detection in chest X-ray images. This is because, the lack of COVID-19 chest X-rays datasets is one of the major problem in scientific community. Our main future focus will be on collecting all the possible images for COVID-19 that exists until the writing of this report and use the GAN network to generate more images to help in the detection of this virus along with its potential variant from the available X-rays images with the highest accuracy possible.

Acknowledgment

We would like to wholeheartedly thank Prof. Roberto Medeiros de Souza for his excellent teaching, material and for always providing insightful comments that help us to improve our work.

References

[1] Kanne, Jeffrey P., Brent P. Little, Jonathan H. Chung, BrettM. Elicker, and Loren H. Ketai. ”Essentials for radiologists on COVID-19: an updateradiology scientific expert panel.” Radiology (2020): 200527.

[2] Hansell, David M., Alexander A. Bankier, Heber MacMahon,Theresa C. McLoud, Nestor L. Muller, and Jacques Remy. ”Fleischner Society: glossary of terms for thoracic imaging.” Radiology 246, no. 3 (2008): 697-722.

[3] Cohen, Joseph Paul, Paul Morrison, and Lan Dao. ”COVID-19 image data collection.” arXiv preprint arXiv:2003.11597, 2020.

[4] Irvin, Jeremy, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund et al. ”Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison.” In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 590-597. 2019.

[5] He, Kaiming, et al. ”Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[6] Krizhevsky, Alex, et al. “ImageNet Classification with Deep Convolutional Neural Networks.” Communications of the ACM, vol. 60, no. 6, May 2017, pp. 84–90. DOI.org (Crossref), https://doi.org/10.1145/3065386

[7] Oxford University, “COVID-19 vaccine development Oxford Vaccine Group.” [Online]. Available: https://www.ovg.ox.ac.uk/news/covid-19-vaccine

[8] T. Ai et al., “Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease (COVID-19) in China:A Report of 1014 Cases,” Radiology, p. 200642, Feb. 2020

[9] P. Kumar and S. Kumari, “Detection of coronavirus Disease (COVID-19) based on Deep Features,” preprints.org, no. March, p. 9, Mar. 2020.

[10] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan, “Diagnosing COVID-19 Pneumonia from X-Ray and CT Images using Deep Learning and Transfer Learning Algorithms,” arXiv, Mar. 2020.

[11] F. Shan et al., “Lung Infection Quantification of COVID-19 in CT Images with Deep Learning,” arXiv, Mar. 2020.

[12] S. Wang et al., “A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID19),” medRxiv, p. 2020.02.14.20023028, Apr. 2020.

[13] I. D. Apostolopoulos and T. Bessiana, “Covid-19:Automatic detection from X-Ray images utilizing Transfer Learning with Convolutional Neural Networks,” Phys. Eng. Sci. Med., Mar. 2020.

[14] A. Narin, C. Kaya, and Z. Pamuk, “Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks.

[15] Minaee, Shervin, et al. “Deep-COVID: Predicting COVID-19 from Chest X-Ray Images Using Deep Transfer Learning.” Medical Image Analysis, vol. 65, Oct. 2020, p. 101794. DOI.org (Crossref), https://doi.org/10.1016/j.media.2020.101794.

[16] He, Kaiming, et al. ”Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.