Designing a Physical X-Ray Filter for Optimal Classification of COVID-19 Pneumonia using Convolutional Neural Networks

Mitchell W. Hutmacher*
Department of Biomedical Engineering
Duke University
Durham, NC 27708
mitchell.hutmacher@duke.edu

Abstract

The overall purpose of this project was twofold. The first task was to determine the efficacy of convolutional neural networks (CNN) as means of classifying X-ray images of lungs. Two types of lungs were assessed, one group displayed pneumonia as a result of a COVID-19 infection, while the other set of images displayed healthy lungs. In addition to this classification, a physical layer representing a thin layer of material meant to block a certain proportion of the X-rays was trained using the dataset. The total size of the dataset that was used was 1,750 images, with half of the images being healthy, and the other half being positive for COVID-19. The CNN used was InceptionV3 (cite source paper), using a stochastic gradient descent optimizer with a learning rate of 0.75 and trained over the course of 50 epochs. The area under the Receiver-Operator Characteristic (ROC) curve (AUC) was found to be 0.996, which indicates excellent classification. In addition, when tested, only five false positives and four false negatives were found out of a total of 352 test images classified. Using this information, I can conclude soundly that the InceptionV3 architecture with the addition of a physical layer representation can reliably predict the status of a patient when given a chest X-ray.

1 Introduction

Since the first outbreaks started occurring worldwide early in the year 2020, the novel coronavirus and how to minimize its spread has been the main focus of every government on the planet. One key part of minimizing the spread of this infectious disease is to accurately identify individuals who have it, so that they can self-isolate and prevent further spread of the disease. Unfortunately, many of the rapid tests used to detect COVID-19 are ineffective, with some studies placing the sensitivity of pharyngeal swabs at approximately 60-70%, meaning that only 60-70% of the patients who are positive for COVID-19 are told that they are positive [1]. In light of this, this project aims to evaluate the efficacy of CNNs to classify chest X-rays of patients who are either negative for COVID-19, or are experiencing pneumonia induced by a COVID-19 infection.
Chest X-rays are a very common imaging modality [1]. A chest X-ray machine is present in most hospitals and, more importantly, safe for the patient, with one chest X-ray being "about the same amount of radiation people are exposed to naturally over the course of about 10 days" [2]. The lungs of COVID-19 patients who have pneumonia show signs of the illness in the form of build-up in the lower lungs [3].

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

1.1 Physical Layer

In addition to measuring the viability of X-ray imaging as a means of diagnosing COVID-19 infections, this project also aims to design a novel "physical layer" that would optimally augment the X-rays transmittance into the X-ray detector in order to best diagnose the infections. In the physical world, this simulated physical layer would be a thin layer of a material that can block X-rays, such as lead glass. The layer would be of variable thickness, designed by the convolutional neural network, with thicker regions blocking larger percentages of the X-rays. A thin layer of material was chosen that augments onto the exterior of the machine in order to maximize the usefulness of the attachment. During the pandemic, hospitals could make a request for one such filter, which would ideally make their X-rays more discernable to machine learning algorithms, allowing for more accurate rapid diagnoses of potential COVID infections.
In an X-ray image, the brightness, or numerical value, of each pixel represents the amount of X-rays that reach that area of the X-ray detector. This physical layer is represented by a matrix of equal dimensions to the X-ray, with values ranging from zero to one. The values in this physical layer correspond to the percentage of X-rays that would be blocked by the physical layer. A value of one would mean that 0% of the X-rays that struck that pixel on the detector were blocked, while a pixel value of zero would correspond to 100% of the X-rays that struck that pixel on the detector were blocked.

The end goal of this model is to not only be able to accurately classify the X-ray images that correspond to X-rays of patients with or without pneumonia induced by COVID-19, but also to design a novel X-ray filtration layer that is able to increase the ability of the model to classify images.

2 Related Work

Due to the ability of CNNs to classify sets of data more accurately than humans, medical images have been a frequent target of neural network research. In particular, multiple groups have tried to use medical imaging classification to classify images from multiple imaging modalities in an attempt to diagnose COVID-19 infections quickly and accurately.
The authors of one study used InceptionV3, and compared it to other networks, to classify healthy lungs and lungs with pneumonia, though not pneumonia specifically induced by COVID-19 [8]. The author of that paper found that the best classification accuracy came from ResNet18, followed closely by Inception. In the confusion matrices, the authors found that after training the Inception network, the sensitivity was approximately 96% and the specificity was approximately 95%. In addition, the author found that their trained Inception network had an area under the ROC curve (AUC) of 0.99, indicating excellent performance on the classification of pneumonia versus healthy lungs.
One study investigated the use of CNNs to classify COVID-19 lung CT scans, as compared to both healthy lungs and the lungs of patients who had pneumonia that was not induced by COVID-19 [9]. Another study used CNNs with transfer learning to classify multiple medical imaging modalities, including healthy lungs versus viral pneumonia versus bacterial pneumonia [10]. They found an AUC of 96.8% for detecting the difference between healthy lungs and all lungs with pneumonia. They also found an accuracy, sensitivity, and specificity of 92.8%, 93.2%, and 90.1%, respectively.
The research in this field has shown that imaging the lungs of a patient with pneumonia is possible, and that InceptionV3 is an excellent network to accomplish this, given its relatively low computational cost and its excellent performance. This project builds on these previous findings by attempting to design a physical layer that can optimally remove portions of the X-rays that reach every pixel in order to make an easier image for my network to classify.

3 Methods

3.1 InceptionV3

InceptionV3 is a neural network developed in 2015 that was based off of the GoogleNet neural network [4]. The network was designed to be less computationally intensive than its competitors, while still maintaining high levels of classification accuracy. This architecture is made up of a series of convolutional and pooling layers, along with multiple concatenations and other techniques to maximize performance, as shown in 1. The network also has demonstrated good performance on low

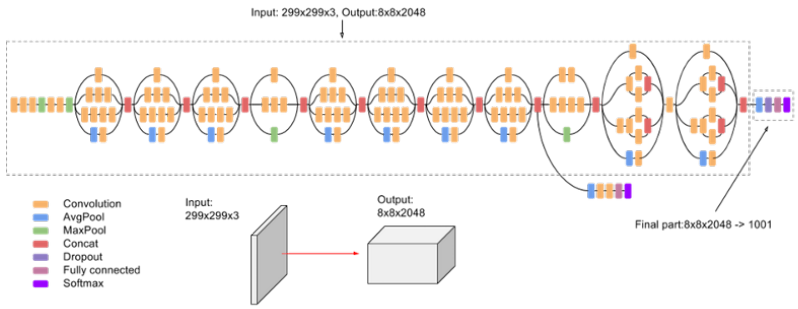

Figure 1: High level overview of the InceptionV3 model [7].

Table 1: Sensitivity, Specificity, and Accuracy

Sensitivity	Specificity	Accuracy
0.9898	0.9795	0.9847

resolution images, which is an important factor for this project, given that the memory limitations on Google Colab require the images to be downsampled to a lower resolution.

3.2 Data Modification

The data used for this project was retrieved from Kaggle, a website used to host data for machine learning [5][6]. In total, there were 1,932 images of healthy lungs and 875 images of lungs that had pneumonia induced by COVID-19. One decision that was made was to remove a significant amount of the COVID-19 negative images in order to attain a 1:1 ratio of positive to negative samples. This resulted in the removal of 856 healthy lung X-rays being removed from the training set and 201 healthy lung X-rays being removed from the testing set. The images that were removed were randomly selected from all of the indices in the data sets that corresponded to healthy lungs.
The images from the datasets chosen had a large range of possible values, with some images having axis lengths of less than 200 to some images having axis lengths of more than 2000. In addition, the ratios of axis lengths were not consistent across the different images. In order to remedy this, I made the decision to crop evenly from the sides of the image if the columns were the long axis, but if the rows were the long axis then the data would be cropped from the top. This decision was made because studies have suggested that the lower areas of the lungs were more likely to demonstrate symptoms of pneumonia from COVID-19 [3].

3.3 Implementation of the Physical Layer

For the physical layer in this project, the matrix was initialized to all 1s. This was a conceptual choice, meant to simulate the layer originally having a thickness of 0 all over. Due to InceptionV3 requiring 3 channel images, the 1 channel physical layer was concatenated 3 times to form a $n \times n \times 3$ sized matrix, with each $n \times n$ matrix being the same. Only one $n \times n$ matrix was trained, resulting in 65536 trainable elements in the matrix, when the images were resized to 256×256. In addition, the values of each element in the physical layer was restricted to a value between 0 and 1. This helped to model the percentage of X-rays that were allowed to pass, where a value of 0 would represent all of the X-rays being blocked by the physical layer. The physical layer, like the rest of the model, was trained by a stochastic gradient descent optimizer over the course of 50 epochs at a learning rate of 0.85.
One significant alteration that had to be made to the physical layer in order for it to function properly with the InceptionV3 architecture that I chose was that the physical layer, along with the images, had to be duplicated three times along an axis. This is because the InceptionV3 architecture will only accept images of size $n \times n \times 3$. The physical layer was only trained as a single layer and was simply concatenated before the multiplication with the input image.

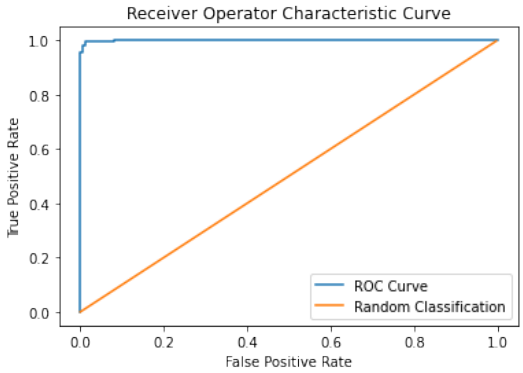

Figure 2: ROC curve of the figure after being trained for 50 epochs with a learning rate of 0.85.

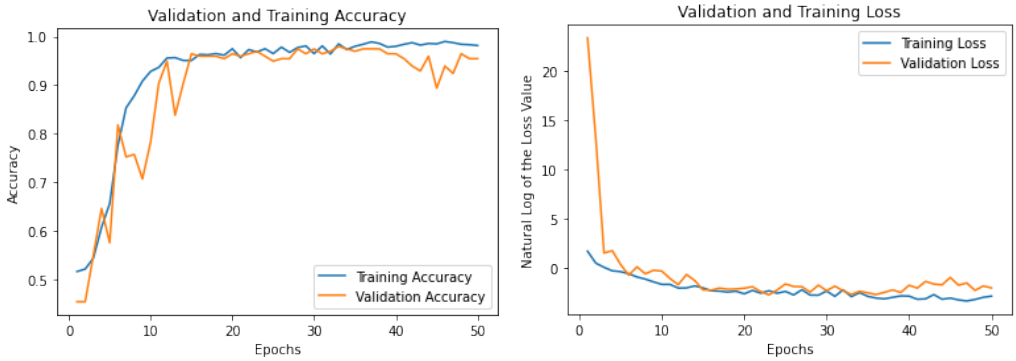

Figure 3: Training and validation accuracy measurements over the course of 50 epochs. Natural logarithm of training and validation loss measurements over the course of 50 epochs.

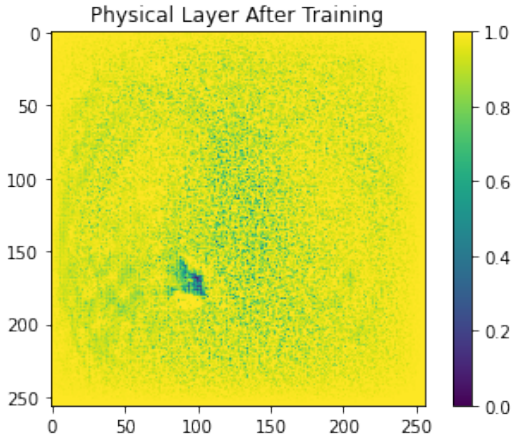

Figure 4: Physical layer after training for 50 epochs. The values are equivalent to the portion of X-rays that are allowed through the layer.

Table 2: Confusion Matrix

	Predict Negative	Predict Positive
Is Negative	172.4	3.6
Is Positive	1.8	174.2

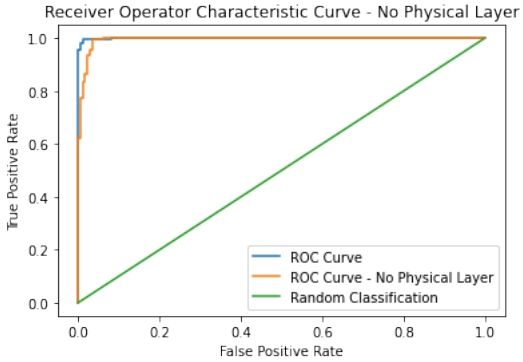

Figure 5: ROC curves of the model when trained with and without the physical layer.

4 Results

4.1 Sensitivity, Specificity, and Accuracy

1 displays the average sensitivity, specificity, and accuracy of this trained model after it was trained 5 different times on the same set of data with the same settings. The sensitivity describes the number of truly positive images that were classified correctly over the total number of images that were classified as positives. After my model had been trained, it classified 175 images as positive correctly and 1 image as positive when it had been negative, resulting in a sensitivity of 99.43%.
The specificity calculates the total number of images that were correctly classified as negative out of all of the images that were classified as negative. Similarly, after the model was trained it calculated 175 images as negative correctly and 1 image as negative incorrectly, resulting in a specificity of 99.43%.
The accuracy of the model quantifies the total number of images successfully classified out of the total number of images that were tested. Over 5 iterations of the model, the average accuracy was 98.47%.

4.2 Confusion Matrix

2 shows the average values for true negative, false positive, false negative, and true negative image classifications over the course of 5 iterations of the model. The model was found to accurately classify both the negative and positive images with a high probability.

4.3 ROC Curve

The Receiver-Operator Characteristic (ROC) curve, shown in 2 compares the true positive rate as you alter the threshold for considering an image a positive instance to allow more images to classified as positive falsely. The network outputs probabilities for each option, in this case for COVID-19 positive or COVID-19 negative. Increasing the false positive rate means that the network would classify images as positive with a lower probability than it would originally.
In my ROC curve, the curve is nearly a right angle and the AUC of this ROC curve was 0.999, as compared to the average AUC over 5 iterations of 0.999. This implies that even at high thresholds for positive classification, the network is able to reliably classify images correctly.

4.4 Loss and Accuracy

3 shows that over the course of the 50 epochs, the accuracy rose from around 50% for both the training and validation accuracies to mid to high 90s. This suggests that not only was the model able to develop weights that fit the training algorithm, but it was also able to generate weights that would fit data that it was not trained on. The natural logarithm of the loss function was taken because the early values of the loss function for the validation data were extremely large, so taking the natural logarithm of the data was done to see the general trends.

4.5 Physical Layer

4 shows the physical layer after being trained for 50 epochs at a learning rate of 0.85. The model, when trained with the physical layer, was very effective at classifying the true negative and true positive images in the test dataset. The areas where the most X-rays were blocked was the center of the chest, where the spine and sternum make up the majority of the signal. In addition, there are areas of lighter X-ray stoppage in the centers of the lungs, to the right and the left of the central heavier region. This indicates that the optimal way to filter these X-rays was to leave more data in these regions. In addition, there is an aberration in the patient's right lung, which is likely a result of limited training. 5 demonstrates that the physical layer has an impact on the performance of the model, albeit a small, positive one.

5 Discussion

5.1 Effectiveness of Classification

There are a number of metrics that can be used to quantify the success of my model at classifying the COVID-19 positive and negative images. The ROC curve shown in 2 displays that the network was able to reliably identify positive examples with low levels of false positive classification. In addition to this, the AUC of that curve, 0.999, and the AUC after 5 iterations of my model, 0.9992, are also indicators of the success of the model. The measures shown in 1 also are indicative that the model is successfully classifying the data. The majority of the model's negative and positive inputs are classified as such.
In the future, the ability of this network to accurately classify images could be improved by pre-training the model using the ImageNet database. Past work in the field has shown that pre-training the InceptionV3 architecture with the ImageNet dataset improves the performance of the model [11]. Using these pre-trained weights was my original plan, but issues with some of the parameters required for the model to use the pre-trained weights conflicting with my own data resulted in me training the model from scratch.

5.2 Effectiveness of Physical Layer

5 shows that while the physical layer was beneficial to the success of the model, it also only resulted in meager increases in the model's ability to accurately classify COVID-19 positive and negative lung X-rays. The AUC for the model with the physical layer was 0.999, which means that it classified the test data nearly perfectly. The confusion matrix (2) shows that, on average, there were 3.6 false positives and 1.8 false negatives out of a total of 352 images classified.
In future experiments, a number of changes could be made in order to train a more effective physical layer. One reason for the meager gains associated with the physical layer is the low portion of X-rays blocked by it. This is largely due to training at low learning rates and for small numbers of epochs, resulting in small changes to the initialized values values, which is all a uniform value of one. Letting the model train for more epochs or training it with a higher learning rate. While I was testing my model, I noticed that as the epochs increased and as the learning rate increased, the physical would display more drastic filtering of the X-rays. Further research with this physical layer could be done to physically make it. In the future, I could determine the necessary path lengths for each pixel in the physical layer such that the proper amount of X-rays from prevented from reaching their target.
In addition, the physical layer could also be given a trainable blurring component, which may allow it to have a greater effect on the model. This trainable blurring component could conceptually represent a reduced aperture size on the X-ray itself, which would result in a blur on the final image.

References

[1] A. Nair, J.C.L Rodrigues, S. Hare, A. Edey, A. Devaraj, J. Jacob, A. Johnstone, R. McStay. E. Denton, and G. Robinson, "A British Society of Thoracic Imaging statement: considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic," Clinical Radiology **75.5**: 329-334 (2020).

[2] “Understanding Radiation Risk from Imaging Tests," (2018). Retrieved November 23, 2020 from https://www.cancer.org/treatment/understanding-your-diagnosis/tests/understanding-radiation-risk-from-imaging-tests.html.

[3] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. Hesketh, L. Yang, and C. Zheng, “Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia," Radiological Society Public Health Emergency Colletion (2020).

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision," Proceedings of the IEEE conference on computer vision and pattern recognition, 2818-2826 (2016).

[5] “COVID-2019 Dataset with Chest X-Ray Images," (2020). Retrieved November 24, 2020 from https://www.kaggle.com/zainaali/covid2019-dataset-with-chest-xray-images.

[6] “Chest X-ray (Covid-19 Pneumonia)," (2020). Retrieved November 24, 2020 from https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia.

[7] “Advanced Guide to Inception v3 on Cloud TPU | Google Cloud," (2020). Retrieved November 24, 2020, from https://cloud.google.com/tpu/docs/inception-v3-advanced.

[8] V. P. Vianna, “Study and development of a Computer-Aided Diagnosis system for classification of chest x-ray images using convolutional neural networks pre-trained for ImageNet and data augmentation," arXiv preprint **arXiv:1806.00839**, (2018).

[9] S. Hassantabar, M. Ahmadi, A. Sharifi, “Diagnosis and detection fo infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches," Chaos, Solitons Fractals **140** (2020).

[10] D.S. Kermany, M. Goldbaum, W. Cai, C.C.S Valentim, H. Lian, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, Ju. Dong, M.K. Prasadha, J. Pei, M.Y.L. Ting, J. Zhu, C. Li, S. Hewett, Ja. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V.A.N. Huu, C. Wen, E.D. Zhang, C.L. Zhang, O. Li, X. Wang, M.A. Singer, X. Sun, J. Xu, A. Tafreshi, M.A. Lewis, H. Zia, and K. Zhang, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," Cell **172.5**, 1122-1131 (2018).

[11] S.S. Yadav and S.M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis," Journal of Big Data **6.113** (2019).