Data Augmentation for Covid-19 Classification

Allen Chang
Dornsife College of Letters Arts and Sciences
University of Southern California
Los Angeles, California
achang30@usc.edu

Abstract

In December 2019, a series of acute atypical respiratory disease, which would later be known as Covid-19, was observed among a small group of people in China. However, the disease spread rapidly, infecting many people in a short period of time [1]. At the time, there was little known about the virus and no known treatment and vaccines. At present, more research has been conducted, and there is more abundant data for disease analysis. Here, I create a robust deep learning model architecture to classify whether a patient has coronavirus based on chest X-ray scans. I also experiment with scaling and blurring the training data and observe that both augmentation methods improve certain performance metrics. Lastly, I observe that differences in the air content of lungs is the primary marker for Covid-19 diagnosis through chest X-rays.

Introduction

The novel SARS-CoV-2 virus is thought to have originated from a seafood market in Wuhan, China. The disease caused by this virus is referred to as Coronavirus disease 19 (Covid-19). The disease spread rapidly and was later declared a pandemic by the World Health Organization. Covid-19 mainly affects the respiratory system with a wide variety of related symptoms. Symptoms can be severe, such as hypoxia and acute respiratory distress syndrome [1].

According to the WHO, as of July 2021, there have been approximately 190,833,853 confirmed Covid-19 cases and 4,100,087 confirmed Covid-19 related deaths [2]. With Covid-19 being a widespread pandemic, having a quick and accurate method to determine whether a patient is infected is crucial. Typically, deep learning models benefit from increased training data. Thus, I will experiment with augmentation methods including scaling and blurring on chest X-ray images to increase training sample size and examine whether they result in better model performance. Furthermore, I will characterize differences between the X-ray images of healthy and infected individuals.

Materials and Methods

My cohort includes 4500 individuals in total, sampled from the original dataset (here) to accommodate memory and runtime constraints [3], [4]. **Table 1** shows specific sampling metrics.

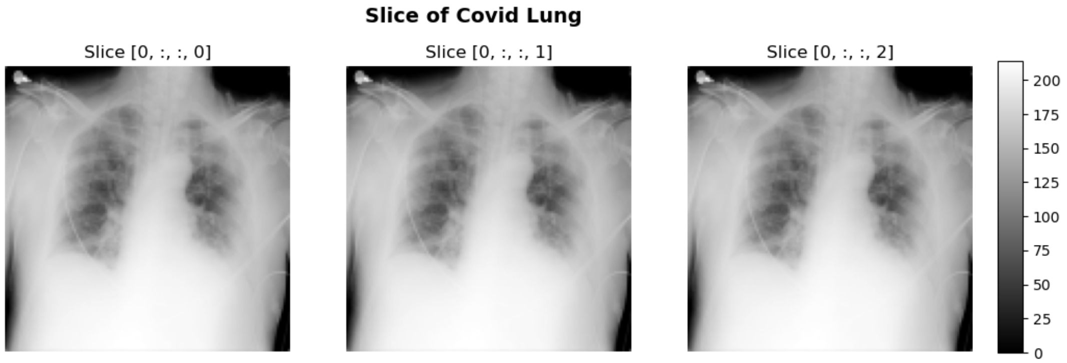
Table 1 *Data Distribution across Training, Validation, and Test Sets*

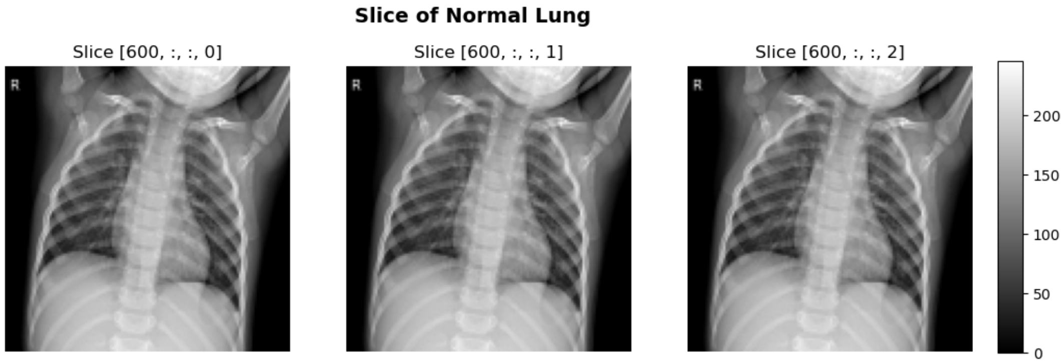
	Training	Validation	Test
Covid-19 Positive	500	500	500
Covid-19 Negative	1500	1000	500

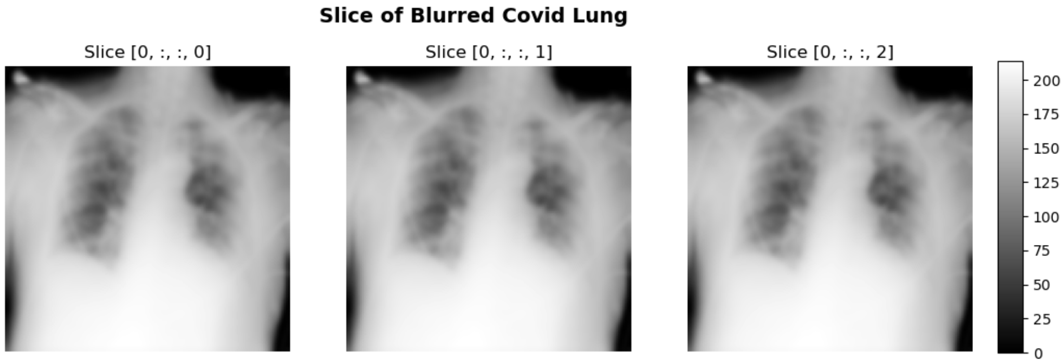
My model architecture comprises of an initial input layer followed by 4 2D convolutional blocks. Each convolutional block consists of a convolutional layer with increasing units (8, 16, 32, and 64) using a 3x3 kernel and rectified linear unit (ReLU) activation. Ridge (L2) kernel regularizer with a lambda value of 0.01 was used for each convolutional layer in the model. Max pooling (2x2) and batch normalization enhance feature extraction. Dropout layers with rates of 0.5 are used on the last three convolutional blocks. These blocks are followed by a flattening layer followed by another block consisting of a 128-unit dense layer with ReLU activation and a dropout layer with a 0.5 dropout rate. Finally, results are output through a single unit dense layer using sigmoid activation. The Adam optimizer minimizes the binary cross-entropy loss function over 100 epochs with a batch size of 16. The epoch with peak validation accuracy is saved.

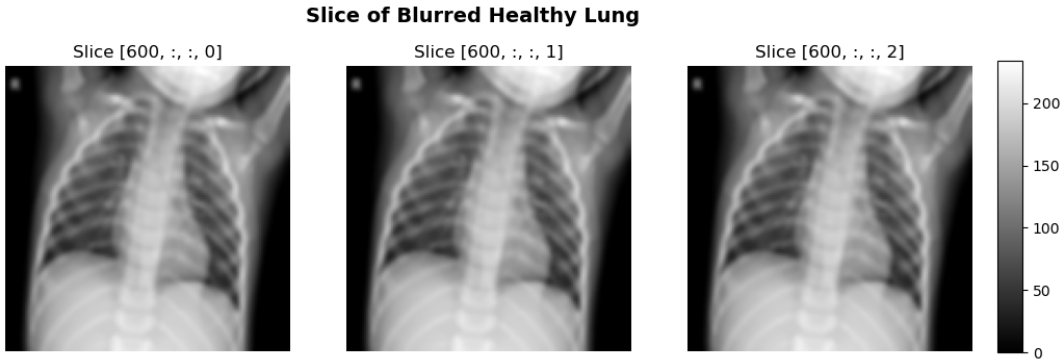
My control model was trained on duplicated unaltered training data. My second model was trained on unaltered data concatenated with Gaussian blurred data, where each image was blurred using a random standard deviation between 1 and 1.05 and a kernel size of 5x5. My third model was trained on unaltered data concatenated with scaled data where each image was randomly scaled anywhere between 0% and 5%. Lastly, my fourth model was trained on unaltered data concatenated with data that was first blurred then scaled using the parameters above. An example of augmented lungs is shown in **Figure 1**. Area under the precision-recall curve (AUPRC), f1-score, precision, recall, and accuracy were reported for each model.

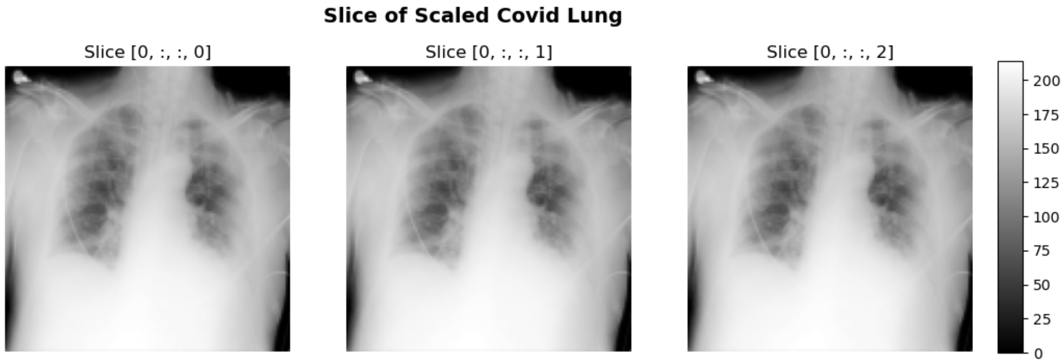
The test set was analyzed to visually identify differences between Covid-19 positive and healthy lungs. Two images (**Figure 5**) were generated for infected and healthy lungs by taking the average voxel intensities over all infected and healthy lungs in the test set.




Slice of Normal Lung
Slice [600, :, :, 0]
Slice [600, :, :, 1]
Slice [600, :, :, 2]
R
200
150
100
50
0


Slice of Blurred Covid Lung
Slice [0, :, :, 0]
Slice [0, :, :, 1]
Slice [0, :, :, 2]
200
175
150
125
100
75
50
25
0


Slice of Blurred Healthy Lung
Slice [600, :, :, 0]
Slice [600, :, :, 1]
Slice [600, :, :, 2]
200
150
100
50
0


Slice of Scaled Covid Lung
Slice [0, :, :, 0]
Slice [0, :, :, 1]
Slice [0, :, :, 2]
200
175
150
125
100
75
50
25
0

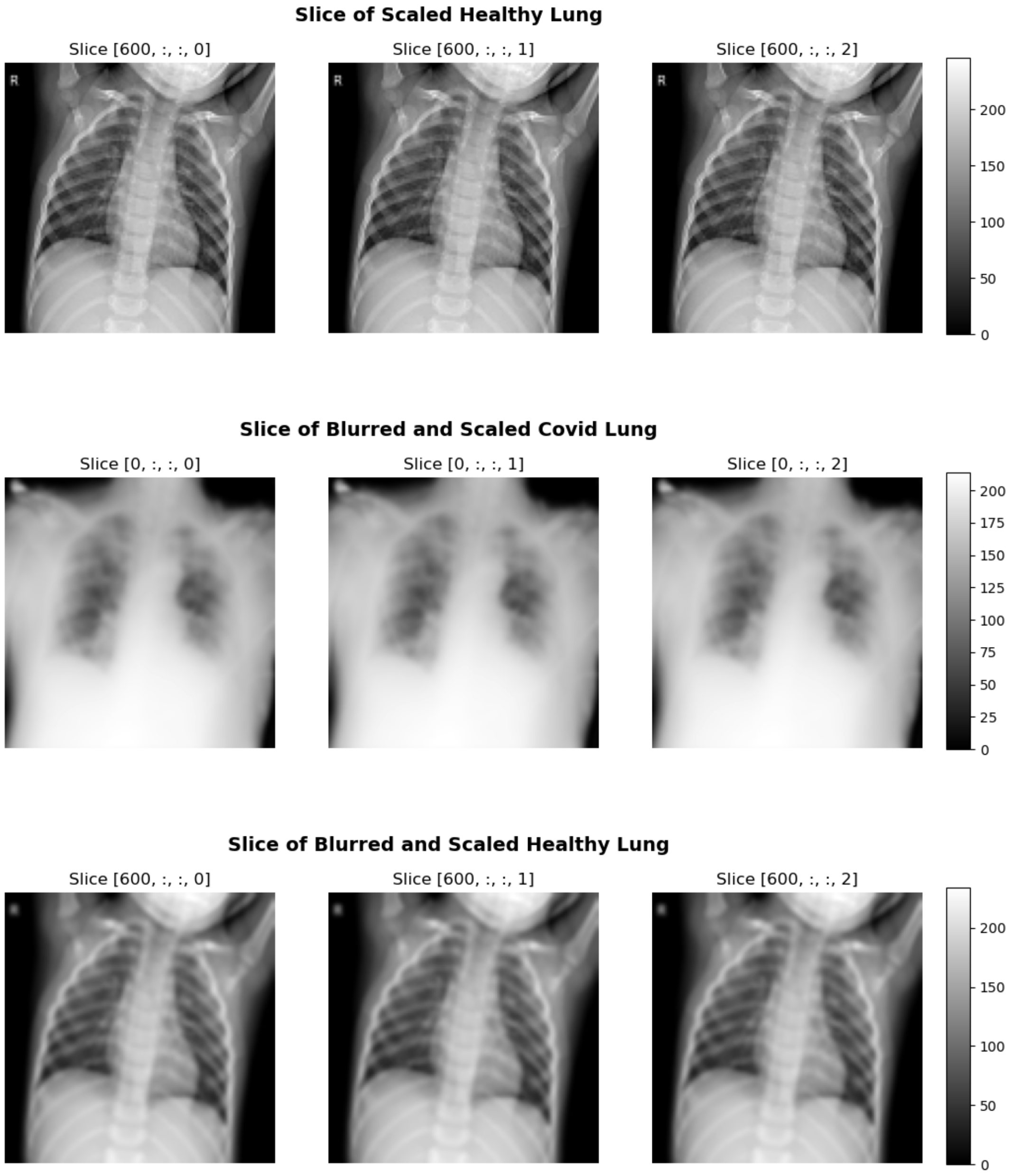
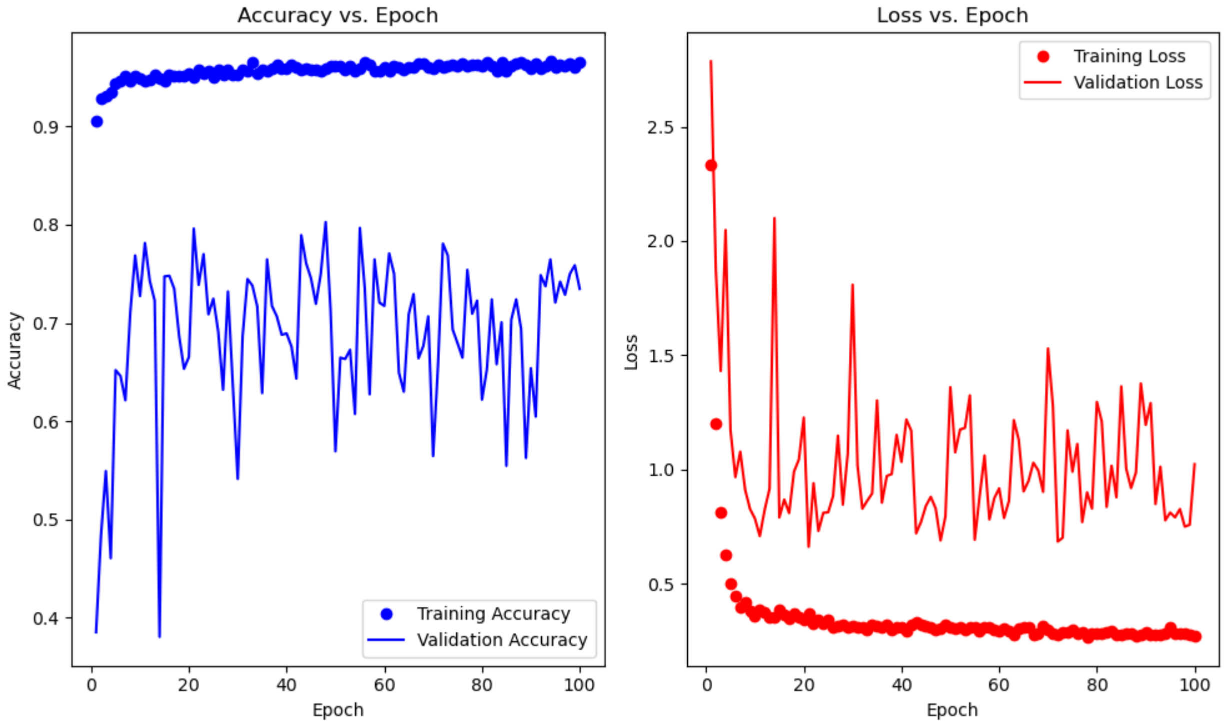
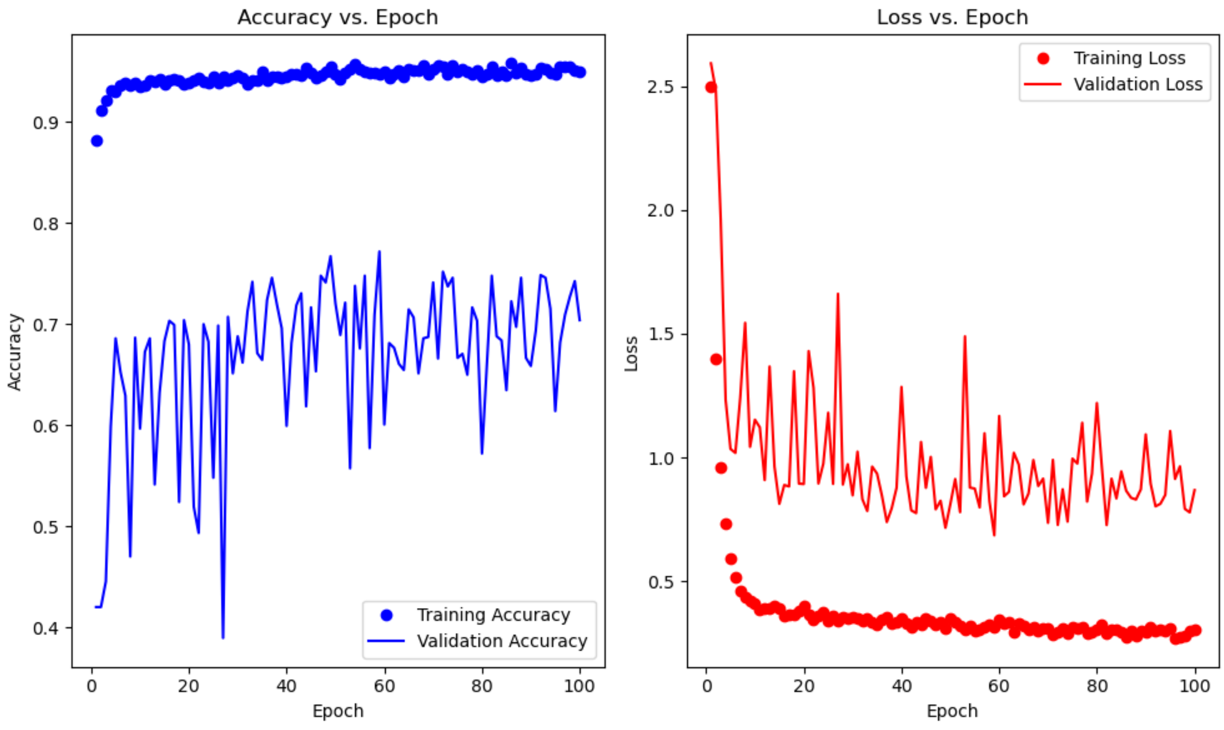


Figure 1 *The first dimension represents the selected image. Here, the first 500 X-rays are Covid positive lungs, and the next 1500 images are Covid negative lungs. Thus, I have selected the 0th and 600th slice as example images to demonstrate augmentation. The next two dimensions represent a coronal view of the lung. The last dimension allows for the three coronal slices to be shown.*

Results and Discission

In **Figure 2**, the validation loss begins to converge around the 20th to 40th epoch for all models. Additionally, the model trained on blurred data appears to have least fluctuation of validation loss throughout training. Both models that were trained on scaled data appear to be overfitting beyond the 80th epoch as a slight upwards trend in validation loss begins to appear.


Accuracy vs. Epoch
Accuracy
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100
Epoch
Training Accuracy
Validation Accuracy
Loss vs. Epoch
Loss
0.5
1.0
1.5
2.0
2.5
0
20
40
60
80
100
Epoch
Training Loss
Validation Loss


Accuracy vs. Epoch
Accuracy
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100
Epoch
Training Accuracy
Validation Accuracy
Loss vs. Epoch
Loss
0.5
1.0
1.5
2.0
2.5
0
20
40
60
80
100
Epoch
Training Loss
Validation Loss

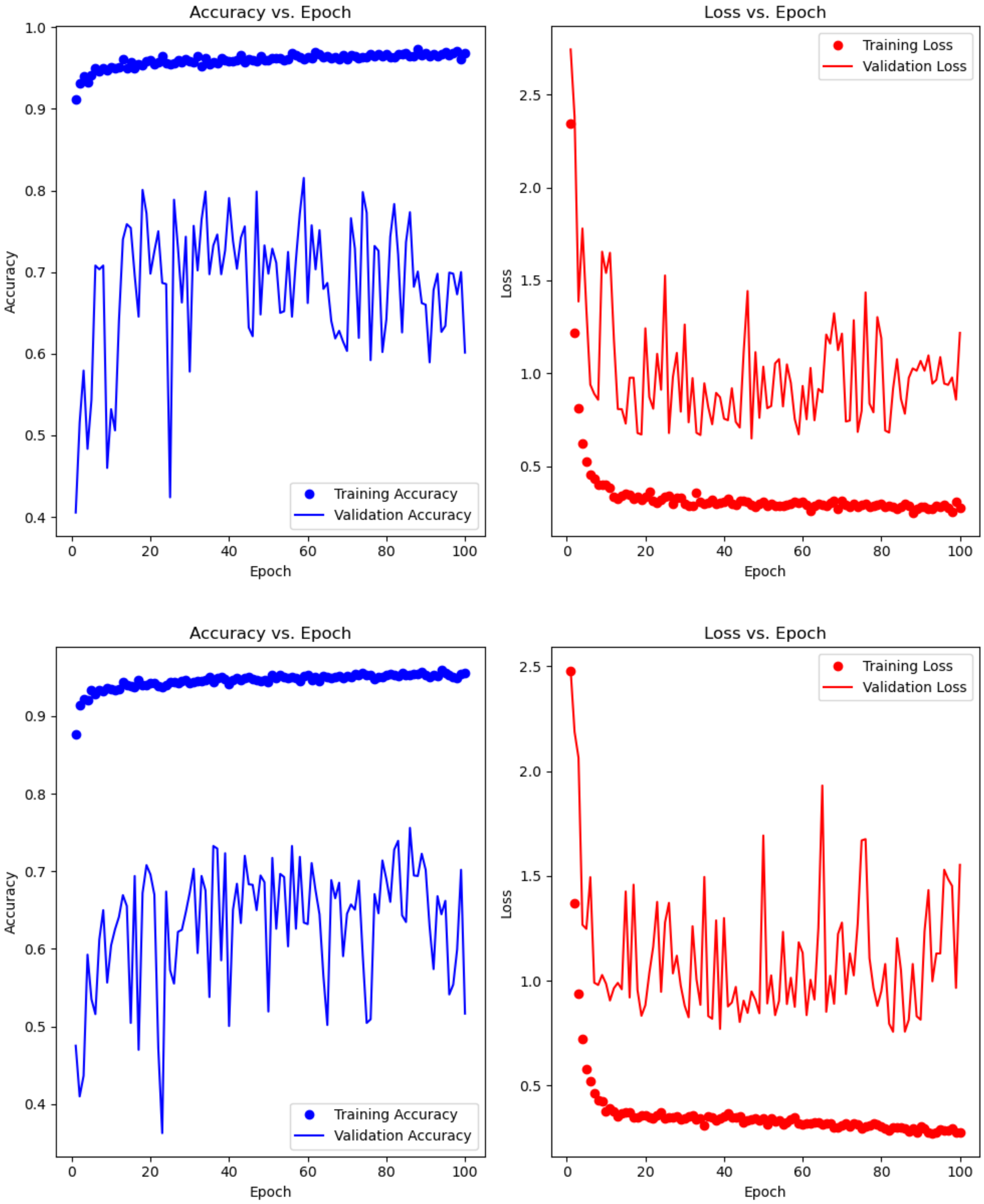


Figure 2 *Validation and accuracy for train and test sets for the control model, the model trained on blurred data, the model trained on scaled data, and the model trained on blurred and scaled data (top to bottom) are shown.*

As seen in **Table 2** and **Figure 3**, the model trained on blurred and scaled data has the highest AUPRC, followed by the model trained on scaled data, the model trained on blurred data, and the control model. Additionally, **Table 2** shows that the model trained on blurred data has the highest F1-score followed by the model trained on scaled data, the model trained on blurred and scaled data, and the control model. The same ranking is observed for accuracy. The model trained on blurred data has highest precision, followed by the model trained on scaled data, the control model, and the model trained on blurred and scaled data. Lastly, the model trained on blurred and scaled data has the highest recall, followed by the control model, the model trained on scaled data, and the model trained on blurred data.

In **Figure 4**, the model trained on blurred and scaled data has the highest false positive rate followed by the control model, the model trained on scaled data, and the model trained on blurred data. As expected, the ranking for false negative rates is the opposite.

Table 2 *Model Performances*

	Control Model	Model Trained on Blurred Data	Model Trained on Scaled Data	Model Trained on Blurred and Scaled Data
AUPRC	0.910	0.917	0.918	0.950
F1-score	0.869	0.884	0.874	0.872
Precision	0.812	0.878	0.827	0.791
Recall	0.934	0.890	0.926	0.970
Accuracy	0.859	0.883	0.866	0.857

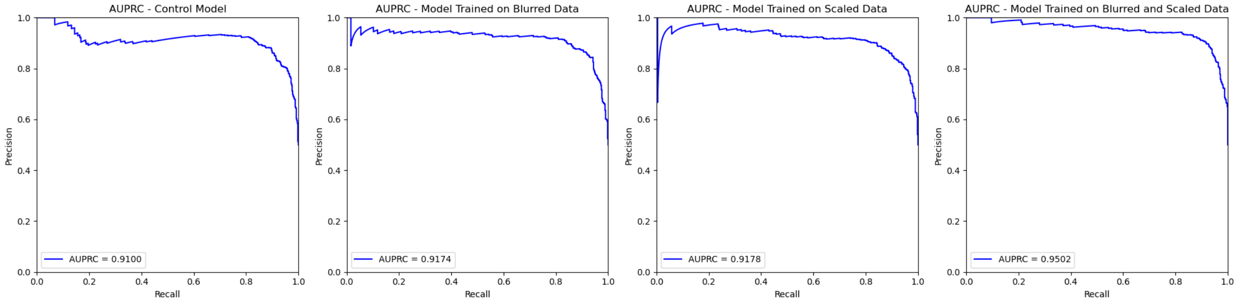


Figure 3 *AUPRC curves for the models are shown.*

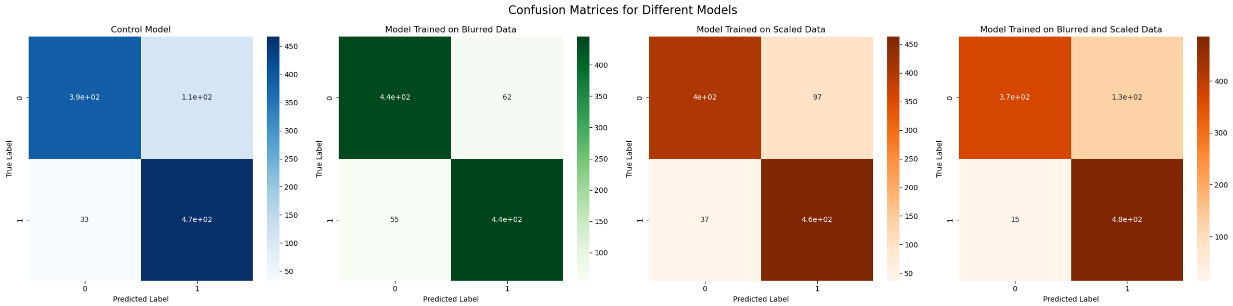


Figure 4 *Confusion matrices for the models are shown.*

While it is difficult to conclusively determine the superior model, I have shown that blurring and scaling are viable augmentation methods to tailor a model to cater to specific scenarios. For instance, in the case of Covid-19 diagnosis, a false negative can have serious implications. On the other hand, a false positive can lead to further clinical examination by a healthcare professional. In this case, the model trained on blurred and scaled data is the most suitable choice, given its high recall and minimal false negative rate.

An effective augmentation method should realistically simulate potential variations in unaltered chest X-rays. This enables the model to identify features indicative of real-world scenarios, which is important for accurate predictions. Motion-induced blurring is common during X-ray imaging [5]. Thus, incorporating blurring as an augmentation method is a realistic method to enhance the training data. Similarly, scaling is also an effective method since lung volume varies from individual to individual. For instance, the lung volume of an adult female is typically 10-12% smaller than that of a male counterpart with equivalent height and age [6].

I also examine differences between healthy and Covid positive lungs and found that the most notable difference between the X-rays is where the lungs are located. This region appears to have lower voxel intensities with a noticeable smaller black area in **Figure 5**. Since the lungs are predominantly air, they should appear black on the X-ray. In fact, Covid-19 pneumonia is most noticeable in radiographs when there is a loss of a normal black appearance in the lung [7]. An increased whiteness, typically referred to as a "ground glass appearance," due to increased density is typically seen in Covid-19 infected lungs [7]. Nevertheless, typical markings, such as blood vessels, within the lungs are still noticeable in infected patients, which can be visualized in **Figure 1** [7].

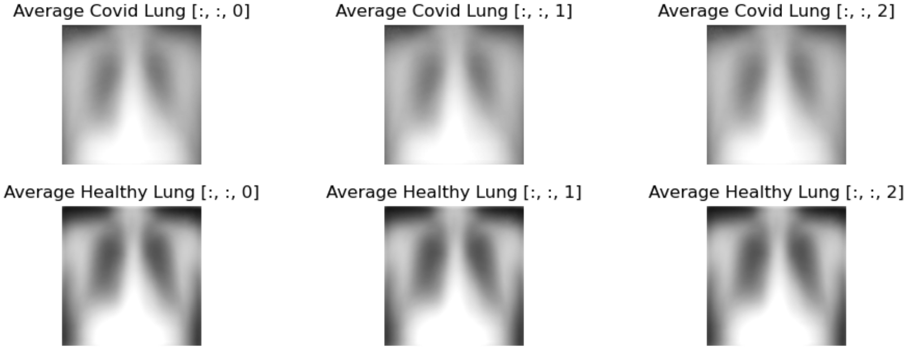

Average Covid Lung [:, :, 0]
Average Covid Lung [:, :, 1]
Average Covid Lung [:, :, 2]
Average Healthy Lung [:, :, 0]
Average Healthy Lung [:, :, 1]
Average Healthy Lung [:, :, 2]

Figure 5 *Averages of healthy and infected lungs are shown.*

Conclusion

My results have shown that Gaussian blurring and scaling are viable augmentation methods. Similarly, combining the augmentation methods also results in an increase in certain performance metrics. Nevertheless, more research needs to be conducted on a wider selection of augmentation parameters and model architectures. Furthermore, a limitation is that it is not feasible to have a test set that represents the true ever evolving distribution of Covid-19 positive individuals in a population.

References

[1] K. Yuki, M. Fujiogi, and S. Koutsogiannaki, “COVID-19 pathophysiology: A review,” *Clin. Immunol.*, vol. 215, p. 108427, Jun. 2020, doi: 10.1016/j.clim.2020.108427.

[2] S. U. Rehman, S. U. Rehman, and H. H. Yoo, “COVID-19 challenges and its therapeutics,” *Biomed. Pharmacother.*, vol. 142, p. 112015, Oct. 2021, doi: 10.1016/j.biopha.2021.112015.

[3] M. E. H. Chowdhury *et al.*, “Can AI Help in Screening Viral and COVID-19 Pneumonia?,” *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.

[4] T. Rahman *et al.*, “Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images,” *Comput. Biol. Med.*, vol. 132, p. 104319, May 2021, doi: 10.1016/j.compbiomed.2021.104319.

[5] W. Huda and R. B. Abrahams, “X-Ray-Based Medical Imaging and Resolution,” *Am. J. Roentgenol.*, vol. 204, no. 4, pp. W393–W397, Apr. 2015, doi: 10.2214/AJR.14.13126.

[6] F. Bellemare, A. Jeanneret, and J. Couture, “Sex Differences in Thoracic Dimensions and Configuration,” *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 3, pp. 305–312, Aug. 2003, doi: 10.1164/rccm.200208-876OC.

[7] J. Cleverley, J. Piper, and M. M. Jones, “The role of chest radiography in confirming covid-19 pneumonia,” *BMJ*, p. m2426, Jul. 2020, doi: 10.1136/bmj.m2426.

https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/data