COVID-19 Detection in Chest X-Ray Images

Isaac McPadden

In 2020 in the United States, COVID-19 (referred to as COVID from here on out) was rapidly spreading and there was a major shortage of test kits. There is significant overlap in COVID symptoms and other respiratory illnesses like influenza and pneumonia, so a COVID diagnosis cannot be made using just symptoms. A possible workaround for the lack of COVID test kits was to use x-ray images of a COVID victim’s lungs. With the potential for a high volume of COVID patients it would be convenient to develop a program that could diagnose COVID without requiring the input of a doctor. The program would need to be able to distinguish between lungs with COVID, healthy lungs, and lungs with a non-COVID illness. Using x-ray images (see Figure 1) from Shams et. al. from Kafrelsheikh University and Mansoura University (Shams, Elzeki, & Abd Elfattah, 2020), I used four different machine learning models to attempt to make such a program. In this paper I will analyze the performance of the four models to determine which model is best and if said model is good enough to use in a program for diagnosing COVID from lung x-rays.



Figure 1. Lung x-rays of COVID lungs, normal lungs, and pneumonia lungs going left to right.

To begin, it is important to understand what data is being fed to the machine learning algorithms to build the models. The image data consists of 221 chest x-rays of COVID patients, 234 chest x-rays of patients with normal lungs, and 148 chest x-rays of pneumonia patients totaling 603 x-ray images. Image data are two-dimensional grids of pixel data. In the simplest digital image forms, each pixel datum is a tuple containing numbers indicating the intensities of red, blue, and green light that combine to produce a color (Layton, Grisel, & Blondel, 2023). In grayscale images, the tuple contains just a single number in the range of 0-255 indicating that pixel’s brightness, with 0 being completely dark (black) and 255 being completely lit up (white). X-ray images are black and white so even if the image files contained color information, they were imported in grayscale to reduce memory usage. The images were imported as two-dimensional arrays from folders that were named for the x-ray diagnosis: “COVID-19 cases”, “Normal\_”, and “Pneumonia”. The images were each labeled with the folder they came from.

The images required extra preprocessing before they were usable by the computer. First, they were scaled because Scikit-Learn’s machine learning algorithms expect each row of data to be the same size (Thakur, 2022). Other options would be to add blackspace or whitespace to smaller images to increase their size and crop larger images to decrease their size but that is more destructive to the original image data (Thakur, 2022). In this case, the images were scaled using the Scikit Image Python library (Stéfan van der Walt, 2014) to images 256 pixels by 256 pixels. Going larger significantly increased processing time. Next, the pixel data were scaled from 0-255 to 0-1 by dividing the pixel values by 255 since many scikit learn functions require floats between 0 and 1 (Layton, Grisel, & Blondel, 2023). Finally, the pixel arrays were flattened to one-dimensional arrays because they require less memory storage than two-dimensional arrays (sagnikmukherjee, 2021). This is done by appending each row of the array to the end of the first row. Finally, the image labels were converted into numeric labels using Scikit Learn’s label encoder function for Scikit Learn functions that require numbers only. COVID lung x-rays were labeled with a zero, normal lung x-rays were labeled with a one, and pneumonia lung x-rays were labeled with a two.

The four machine learning algorithms used to generate x-ray classification models were a linear support vector classifier (LSVC), a stochastic gradient descent classifier (SGDC), a k-nearest neighbors classifier (KNNC), and a multi-layer perceptron classifier (MLPC). Each model was originally built using the default values included in the Sci-kit Learn library. The LSVC model and MLPC model required some modifications to default values to produce a valid model. For the LSVC, the tolerance was increased from 1e-4 to 1e-2 and the maximum iterations value was doubled from 1000 to 2000. For the MLPC, the maximum iterations value was increased from 200 to 500 and the hidden layer sizes was changed from (100) to (50, 50).

The primary scoring method was balanced accuracy which weights the accuracy to account for the dataset having different numbers of images for each classification. The balanced accuracy values are taken from the test sets. The LSVC and the SGDC had very similar results which makes sense because they are both types of support vector machine. The balanced accuracy of the LSVC was 0.88 with a precision of 1.00 for the COVID x-ray images. The balanced accuracy of the SGDC was 0.85 with a precision of 1.00 for the COVID x-ray images. The recall for COVID x-rays on both was lower at 0.88 for both. Both of these facts suggest the models are under-diagnosing COVID but it isn’t misdiagnosing COVID at all. In each, the recall on normal lung x-rays was higher than the recall for the pneumonia x-rays suggesting the models are more likely to misdiagnose COVID as pneumonia than normal lungs. The KNNC model performed worse than LSVC and SGDC with a balanced accuracy of 0.76. It had a precision of 0.98 for COVID x-rays with a recall of 0.86. For KNNC, the precision and recall for normal lung x-rays (0.71 and 0.70) and pneumonia x-rays (0.58 and 0.72) was considerably lower suggesting this model is mixing up normal lungs with pneumonia lungs and vice-versa. The MLPC had a balanced accuracy of 0.60 with a 1.00 precision and 0.76 recall for COVID x-rays and 0.52 precision and 1.00 recall for normal x-rays. In this case, the model is underdiagnosing COVID and is over-diagnosing normal lungs which is the opposite of the desired outcome. It also nearly completely misclassified pneumonia lungs as healthy lungs with pneumonia having a precision of 0.50 and a recall of only 0.03.

Having gotten working models of each type, the hyperparameters were tuned by performing a grid search. The grid search parameters, including the best results from the grid search can be found in the python notebook included with this paper. The balanced accuracy of the best LSVC model was 0.88, the same as the default model value of 0.88. The best estimator for the SGDC improved on balanced accuracy, increasing from 0.85 to 0.88. Of note, ‘constant’ and ‘invscaling’ learning rates failed to converge. The KNNC grid search improved the balanced accuracy from 0.76 to 0.79 and continued to mix up normal and pneumonia classifications. The MLPC balanced accuracy improved from 0.60 to 0.88, with the majority of misclassifications being COVID classified as pneumonia and pneumonia being classified as normal.

With a maximum balanced accuracy of 0.88 on three of the four models, classification of lung x-rays using a computer could be feasible someday, however PCR tests and antibody tests are far more accurate and getting the diagnosis wrong could cost lives. To improve the classification models, a larger dataset would be important, either with more images or by making minor transformations to the images in the current dataset. Additionally, using a distributed computing platform would allow for a much greater range of hyperparameter testing and optimization. It would also allow for higher resolution images to be used. Considering the SGDC and LSVC models showed little improvement after hyperparameter tuning, the MLPC would be the best model for further investigation.

# Works Cited

Layton, R., Grisel, O., & Blondel, M. (2023). *Color Quantization Using K-Means*. Retrieved from Scikit Learn: https://scikit-learn.org/stable/auto\_examples/cluster/plot\_color\_quantization.html

Pedregosa, F. V. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. Retrieved from scikit-learn.org

sagnikmukherjee. (2021, January 22). *Impact of Image Flattening*. Retrieved from Geeks for Geeks: https://www.geeksforgeeks.org/impact-of-image-flattening/

Shams, M., Elzeki, O., & Abd Elfattah, M. (2020, June 9). *Chest X-ray images with three classes: COVID-19, Normal, and Pneumonia*. Retrieved from Mendeley Data: https://data.mendeley.com/datasets/fvk7h5dg2p/1

Shubhtripathi. (2021, January 9). *KNN*. Retrieved from Medium: https://shubh-tripathi.medium.com/knn-7be936cea811#

Stéfan van der Walt, J. L.-I.-i. (2014, June 19). *scikit-image: image processing in Python*. Retrieved from PeerJ: https://peerj.com/articles/453/

Thakur, A. (2022, March 24). *How to Handle Images of Different Sizes in a Convolutional Neural Network*. Retrieved from Weights & Biases: https://wandb.ai/ayush-thakur/dl-question-bank/reports/How-to-Handle-Images-of-Different-Sizes-in-a-Convolutional-Neural-Network--VmlldzoyMDk3NzQ