Grand Valley State University
ScholarWorks@GVSU

Technical Library School of Computing and Information Systems

2020

Automatic Detection from Chest X-Rays using Deep Learning

Evan Lang
Grand Valley State University

Follow this and additional works at: https://scholarworks.gvsu.edu/cistechlib

ScholarWorks Citation

Lang, Evan, "Automatic Detection from Chest X-Rays using Deep Learning" (2020). *Technical Library*. 367.
https://scholarworks.gvsu.edu/cistechlib/367

This Project is brought to you for free and open access by the School of Computing and Information Systems at ScholarWorks@GVSU. It has been accepted for inclusion in Technical Library by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

Automatic Detection of COVID-19 from Chest X-Rays using Deep Learning

Evan Daniel Lang

A Project Submitted to

GRAND VALLEY STATE UNIVERSITY

In

Partial Fulfillment of the Requirements

For the Degree of

Master of Science in Applied Computer Science

School of Computing and Information Systems

December 2020

GRAND VALLEY STATE UNIVERSITY

The signatures of the individuals below indicate that they have read and approved the project of Evan Daniel Lang in partial fulfillment of the requirements for the degree of Master of Science in Applied Computer Science.

Dr. Greg Wolffe, Project Advisor Date

Graduate Program Director Date

Unit head Date

Contents

Abstract

Over the past year, COVID-19 has affected countries world-wide. COVID-19 detection tests have allowed us to control the spread of the disease; however, COVID-19 testing kits are highly specialized and difficult to procure in quantity. X-rays, on the other hand, have broad clinical usage and therefore tend to be readily available. As a result, radiologists have begun using chest X-rays to diagnose COVID-19 in patients with respiratory distress. The goal of this research project was to demonstrate that deep learning can be used to automatically detect COVID-19 from chest X-rays. Automated detection with deep learning models could help make X-ray diagnosis even more efficient, relieving the burden on radiologists, especially those not specifically trained in detecting COVID-19. However, training deep learning models for computer vision tasks requires a large amount of labeled data, which does not yet exist for COVID-19 chest X-rays. Therefore, this project utilized transfer learning to adapt a pre-trained ResNet50 model for the COVID-19 classification task. Results demonstrated that this approach can successfully classify chest X-rays (normal versus viral pneumonia versus COVID-19) with 94% overall accuracy. In the binary classification problem (COVID-19 versus other), this approach could classify chest X-rays with an even higher accuracy of 98%. In this project, I also implemented gradient class activation maps to visualize different convolutional layers of the model. These visualizations highlight those areas within each X-ray the model found most significant, providing a degree of interpretability to the classifications. In summary, this research project established the viability of using deep learning for detecting respiratory conditions via computer vision, despite the current limited availability of data.

Introduction

The Coronavirus (COVID-19) pandemic has dramatically affected every country, infecting millions of people and causing over one million deaths (John Hopkins CSSE). The most widely used testing procedure involves taking a sample from the respiratory system via nasal swabbing, processing the sample in a lab, and receiving a test result at least 1 to 2 days later (Xie et al.). This lag in receiving test results has made it especially difficult to control the spread of COVID-19 (Wang et al.). Because the virus is typically associated with respiratory distress (cough, trouble breathing), chest X-rays may provide another more immediate route for detecting COVID-19 in patients experiencing respiratory issues (Xie et al.). Chest X-rays are a widely available technology and have been used to detect similar respiratory conditions, such as SARS and MERS (Narin et al.). The goal of my project was to train a computer vision model to analyze chest X-rays and differentiate between COVID-19, other forms of viral pneumonia, and a healthy respiratory system. In addition to evaluating the model's accuracy, I explored visualization methods to understand how the model made its classifications.

Computer vision is a growing area of computer science in which researchers have demonstrated that convolutional neural networks (CNNs), a type of neural network for processing multi-dimensional inputs, can perform very well in image classification. Several researchers applied CNNs to different medical images, including chest X-rays (Hashmi et al.). During the course of this project, other researchers have published work focused on detecting COVID-19 from chest X-rays, illustrating the growing appeal of research in this area.

However, due to the recent appearance of COVID-19, large datasets of COVID-19 chest X-rays are not yet available. I gathered data from publicly available sources (Kaggle and the IEEE data port). These datasets have 3,384 images consisting of 3 classes: 1,341 healthy chest

X-rays, 1,344 viral pneumonia X-rays, and 699 COVID-19 X-rays. Due to the small size of this dataset, it would be difficult to train a computer vision model from scratch. Instead, I used a pre-trained computer vision model. This pre-trained model was trained for a different image classification task, and I used transfer learning to analyze the chest X-ray classification problem.

Transfer learning refers to training a model on one data source, and then applying the trained model to a new dataset. Transfer learning often requires fine-tuning the weights of the pre-trained model and it allows deep networks to be used on new research problems with smaller datasets. For my project I used ResNet50, a pre-trained model developed by Microsoft researchers (He et al.). These researchers trained ResNet50 on ImageNet, a publicly available dataset with over one million labeled images. I fine-tuned ResNet50 with the gathered chest X-ray dataset and evaluated the model's ability to differentiate between COVID-19, viral pneumonia, and healthy X-rays. I performed a similar experiment to differentiate between COVID-19 and other X-rays, where other included viral pneumonia and healthy combined.

In addition to accuracy evaluation, I looked at computer vision model visualization methods. For computer vision models to be used in a medical setting, medical professionals will need to understand how the model made a classification (Vellido et al.). Neural networks have traditionally been "black box," but recent research has developed visualization methods to gain insight into how models make classifications. I applied gradient class activation maps (Grad-CAM), which use gradient information from the final fully connected layer, backpropagated to the convolutional layers (Selvaraju et al.). This generates a heat map highlighting which areas of the image the model used in making its classification. Clinicians may review the highlighted areas in the heat maps, providing an additional source of validation and confidence for patients, and potentially furthering the clinician's understanding of the disease itself.

Related Work

The application of deep learning to different fields, including computer vision, has become a promising research area. Previous work has used deep learning computer vision models to differentiate between healthy and pneumonia chest X-rays with 98% accuracy (Hashmi et al.). COVID-19 is one type of viral pneumonia, and researchers have applied the same methods to differentiate between healthy and COVID-19 chest X-rays with 99% accuracy (Narin et al.). Multi-class problems, where models differentiate between more than two classes, are likely more representative of a real-world use case. Researchers have used pre-trained computer vision models to differentiate between healthy, viral pneumonia, and COVID-19 chest X-rays with 98% accuracy (Wang et al**.,** Selvaraju et al.). This work showed that COVID-19 chest X-rays have features that allow it to be differentiated from viral pneumonia chest X-rays.

However, achieving high accuracy with COVID-19 chest X-ray detection models is not the only challenge in this space. Models within the medical field need to be interpretable by the clinicians that are using them. Thus, we need to incorporate visualization methods to understand how the model made its classification. Researchers have used Grad-CAM to understand which areas of the X-rays were used for classification. This additional interpretability would make it more acceptable to use a computer vision model in a clinical setting (Selvaraju et al.).

Methods

For my project, I performed classification tasks similar to those in related work. In the first task, I performed multi-class classification, differentiating between COVID-19, viral pneumonia, and healthy X-rays. In the next task, I performed binary classification, differentiating between COVID-19 and other X-rays. I applied deep learning to these classification problems using a pre-trained model, and I explored visualization methods for interpreting these models. In this paper, I present the Grad-CAM visualization results, as these were the easiest to decipher.

Data

I analyzed COVID-19 chest X-rays from two publicly available datasets: one from researchers at Qatar University and University of Dhaka (on Kaggle) (Rahman et al.), and the second from IEEE (Cohen et al.). Combined, these datasets have 1,341 normal, 1,344 viral pneumonia and 699 COVID-19 labeled chest X-rays. I applied standard image transformations where I resized images to 256x256. I also randomly cropped or flipped training images to prevent the model from learning based on image quality or image position. Finally, I normalized the images which allows the CNN to converge more quickly.

CNN Background

A convoluted neural network (CNN) is a type of network designed to handle multi-dimensional inputs, making these networks ideal for images. Images are passed through several types of layers within CNNs: convolutional layers, pooling layers, and fully connected layers. This architecture is illustrated in Figure 1. The initial convolutional layers use filters to detect different features, such as outlines, within each image. When building the CNN, the model architect chooses the number and size of the filters in each convolutional layer. These filters then generate a feature map, and the next step in a CNN is typically to apply an activation function to that feature map. The activation function helps combat vanishing gradients, a problem that prevents successful model training. The rectified linear unit (ReLU) function is commonly used as an activation function in deep learning networks. After applying the activation function, the feature maps are passed through a pooling layer. These pooling layers downsample the image by aggregating the values within in a sliding window. This slowly reduce the spatial size of the feature maps, in turn decreasing the number of parameters to learn. There are two main types of pooling: maximum and average pooling, which take the maximum or average of the values in the sliding windows. These operations repeat based on the number of layers in the CNN. Eventually, the layers should reduce the input image to a single dimensional representation. This new representation is an input to fully connected layers, which behave like a traditional neural network and can be used for classification. Next, a softmax function is applied to the output of the fully connected layers to create a membership probability distribution for each class. The final classification is the class with the highest membership probability.

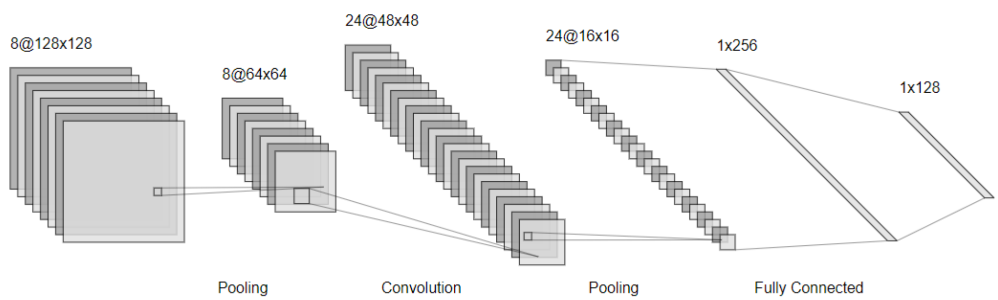

Figure 1. Example of Convolutional Neural Network (CNN) Architecture (Lenail et al.)

Figure 1 shows how the alternating convolutional and pooling layers slowly reduce the dimensionality of the image into a form that can be input into the fully connected layers. The size and stride of the filters and windows, the type of pooling, as well as the number and order of layers, represent design parameters which require tuning.

Model Training

The size and number of layers also results in many parameters which need to be learned. Given the many parameters, CNNs require a large amount of data for training. Because of the recency of COVID-19, chest X-ray datasets with labeled COVID-19 images are not widely available. However, we can use smaller datasets and employ transfer learning. More specifically, we can train a model on a large dataset, and then retrain or fine-tune the final layers of the network with a smaller dataset (Oh et al.). The earlier layers will identify general patterns, but the later layers will then learn more specific features for the new problem.

In this project, I fine-tuned ResNet50 for a multi-class (COVID-19 versus viral pneumonia versus healthy) and a binary (COVID-19 versus other) classification task. ResNet50 was pre-trained on the ImageNet database. In my preliminary analysis, I experimented with different versions of ResNet (ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152) and found that ResNet50 best balanced classification accuracy with runtime. ResNet50 is composed of 50 layers and over 23 million trainable parameters. I implemented the model fine-tuning in Python using Pytorch and I performed the training using a NVIDIA GeForce 1050ti GPU.

I used a five-fold cross validation framework, in which I assigned each image to one of five folds. I trained the model to minimize cross-entropy loss, and this was repeated five times, each time training on three folds, validating on another fold, and testing on the held-out fold. During training, I used a random weighted sampler to balance the number of samples per class. I used the default ResNet50 hyperparameters, as well as some hyperparameters that were used in related work (Narin et al.). Specifically, I used a training batch size of 8, a learning rate of 1e-5, and a momentum of 0.9. I used a maximum of 50 training epochs, but used my validation set to implement early stopping and prevent overfitting.

Model Evaluation

I used four different metrics to evaluate the model. I used cross-entropy loss to evaluate the model during training. Cross-entropy loss measures the difference between the probability the model classified an image in each class, and the ground truth probability that the image was in that class. After training, I measured the model accuracy, precision, and recall. Accuracy is the ratio of correctly classified samples to total samples, and it provides an overall assessment of the model's performance. Precision and recall describe how the model performs on each class. Precision is the ratio of true positives that have been identified to all positives that have been identified. Recall is the ratio of true positives that have been identified to all positives.

Visualization method

I used CNN visualization methods to gain insights into the model's classifications. Traditionally, neural networks, especially deep networks such as CNNs, have been thought of as "black boxes," and it has been difficult to understand how models have made classifications. In the last couple years, researchers have made progress in visualizing the weights in CNNs, which can lend some insight into why a model behaved in a certain way (Zeiler et al.). I explored the use of these methods on the COVID-19 chest X-ray dataset as this interpretability would make it more feasible to use this technology in a clinical setting.

My preliminary work demonstrated that gradient class activation maps (**Grad-CAM**) was a promising visualization method, due to the ease of using Grad-CAM compared to other methods that had many parameters to tune. Grad-CAM uses the gradients from the feature maps of the final convolutional layer to project a heat map onto the image. I adapted a GitHub implementation to generate the Grad-CAM visualizations for this project (Nakashima et al.).

Results/Discussion

Model Results

Figure 2 shows the loss during training for each fold in each experiment. For the binary classification experiment, loss dropped quickly and stopped decreasing within 15 epochs. For the multi-class classification experiment, loss dropped quickly and stopped decreasing within 30 epochs. This illustrates the effectiveness of transfer learning with a pre-trained network, as the fine-tuning of the CNNs converged rather quickly.

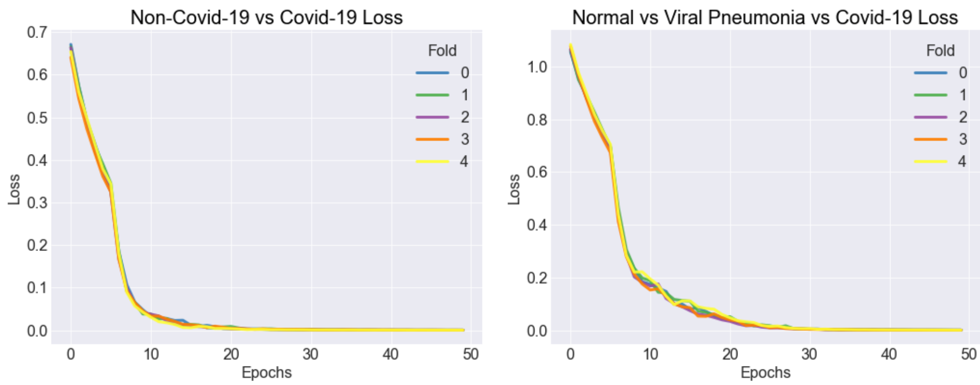

Figure 2. Training Loss

Table 1 shows the accuracy, precision and recall for the binary classification experiment. The model achieved 98% accuracy, with overall precision and recall of 98%. This shows that the model successfully differentiated between the two classes.

Table 1. Test Results

Class	Accuracy	Precision	Recall
Overall	0.98	0.98	0.98
Non-Covid-19	--	0.98	0.99
COVID-19	--	0.99	0.98

Figure 3 shows the confusion matrix for the binary classification experiment. The diagonal line shows that the model had 99% accuracy for the non-COVID-19 class and 98% accuracy for the COVID-19 class. I trained the model on the multi-class problem to understand which non-COVID-19 X-rays were most likely to be misclassified, and vice versa.

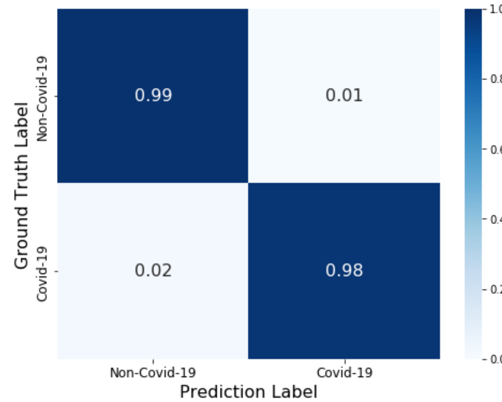

Figure 3. Confusion Matrix

Table 2 lists the model accuracy, precision, and recall across all folds and classes for the multi-class classification problem. Overall, the model performed with 94% accuracy, and with precision and recall of 94% as well. The normal and viral pneumonia classes had relatively low precision and recall, ranging from 89% to 95%. This shows that the model struggled to differentiate between the normal and viral pneumonia chest X-rays. This may be because viral pneumonia is a relatively general term, making it more difficult to classify. The precision and recall for the COVID-19 class was 97% and 99%, respectively. A high recall here is especially valuable, as misclassifying a positive case of COVID-19 is likely more damaging than misclassifying a negative case.

Table 2. Test Results

Class	**Accuracy**	**Precision**	**Recall**
Overall	0.94	0.94	0.94
Normal	--	0.91	0.95
Viral Pneumonia	--	0.95	0.89
COVID-19	--	0.97	0.99

The confusion matrix in Figure 4 provides additional insight into these results. It shows that COVID-19 X-rays are classified correctly with 99% accuracy, and those that are classified incorrectly are classified as viral pneumonia. Viral pneumonia X-rays, when misclassified, tend to be misclassified as normal. Normal X-rays, when misclassified, tend to be misclassified as viral pneumonia. This suggests that the difference between normal and viral pneumonia X-rays may be less clear than the difference between those categories and COVID-19, again highlighting the potential of these tools in diagnosing COVID-19.

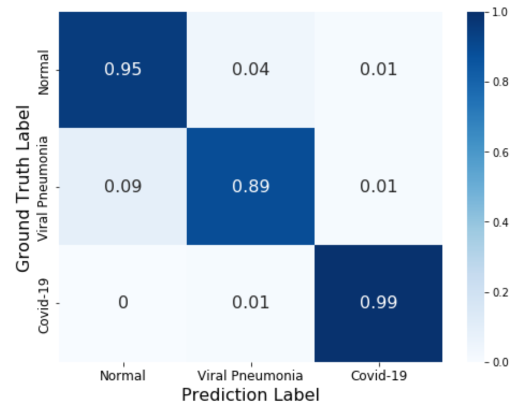

Figure 4. Confusion Matrix

Visualization Results

Using the multi-class fine-tuned CNN, Figure 5 illustrates the grad-CAM visualizations. These images highlight X-ray areas that each CNN layer found to be important. Highlighting these areas improves the interpretability of the model's process. This could potentially also help clinicians within the diagnostic process to minimize errors. These visualization results show a progression of features over the layers: the earlier layers (1 and 2) learn to differentiate between edges of the bones, the mid layers (3 and 4) start to detect different objects within the lung area, and the final layer (5) uses the middle area in one of the lungs to make the final classification.

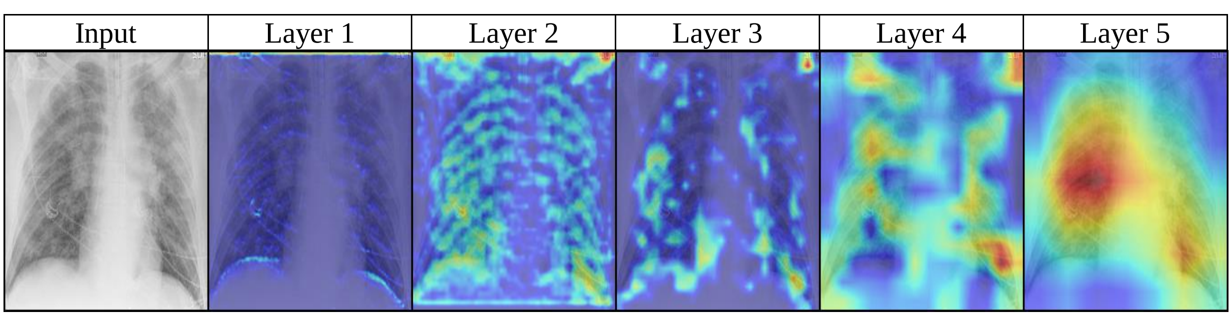

Figure 5. Example of grad-CAM across 5 layers

Figure 6 shows additional examples from the final convolutional layers of COVID-19 positive X-rays. The goal of the visualizations is to gain insight into how the model behaves and provide some understanding of the classification process. This provides additional validation that the model is making classifications for similar reasons that a clinician might. The Grad-CAM images show that the model is identifying features within the lungs and not focusing on different aspects of image markings or quality.

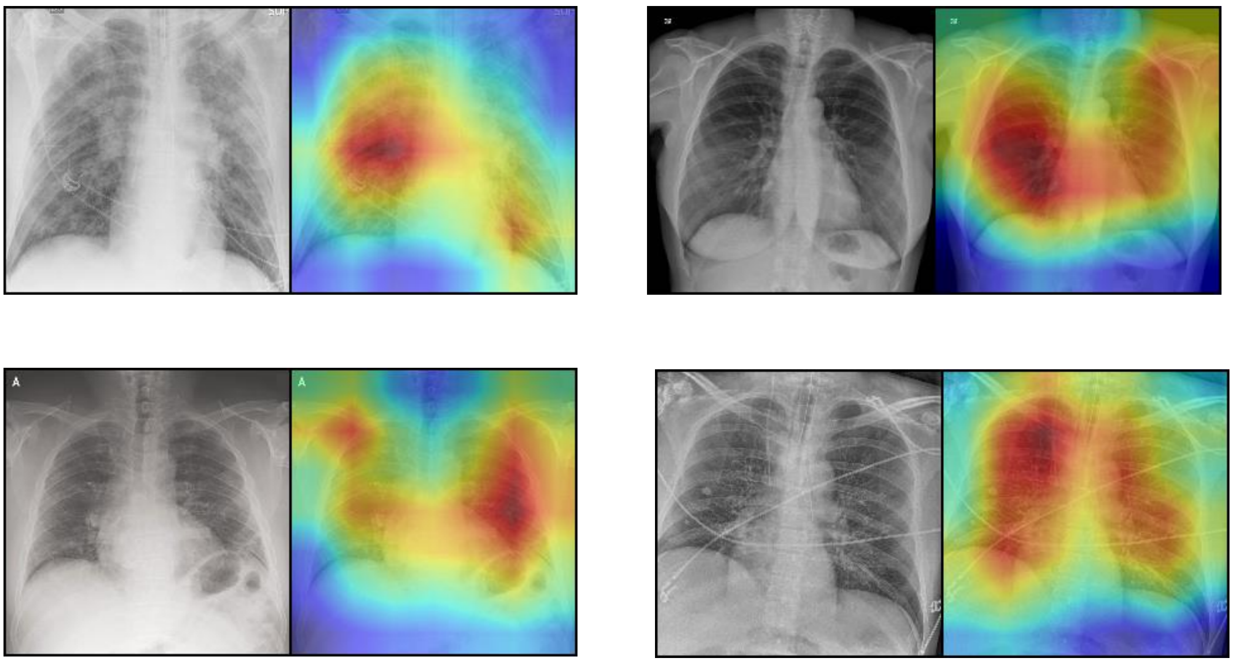

Figure 6. Examples of grad-CAM Final Layers from COVID-19 Samples

The goal of the model is to provide a diagnostic aid for patients who have respiratory issues. However, this model may be limited to patients who experience COVID-19 respiratory symptoms. Even with that limitation, these tools may help clinicians confirm severe cases of COVID-19 more quickly than laboratory tests. This may be especially important in countries where the chest X-ray infrastructure already exists, but COVID-19 tests are not widely available.

Conclusions

The COVID-19 pandemic has had a devastating impact on the world. In addition to laboratory testing, chest X-rays are an effective and globally available diagnostic tool that can allow clinicians to differentiate between COVID-19 and other respiratory diseases. In this project, I demonstrated that these classifications could be made automatically by fine-tuning pre-trained computer vision models. My results show that we can accurately classify COVID-19, viral pneumonia, and healthy chest X-rays with 94% accuracy. This accuracy increases to 98% when we only classify COVID-19 versus other chest X-rays. In addition, I show that grad-CAM can be used to visualize which areas of the image the model used to make its classifications.

In future work, I am interested in exploring the use of Generative Adversarial Networks (GANs) with small datasets. GANs learn image features from a training set and generate additional image training samples to improve overall classification. It would be interesting to see how much data is required to train an effective GAN, and how the size of the GAN-augmented dataset affects the classification performance. This may inform, for future diseases, on how much new data would be sufficient to develop similar diagnostic tools.

Bibliography

Center for Systems and Engineering at John Hopkins University. (n.d.). COVID-19 Dashboard. Retrieved October 12, 2020, from *https://coronavirus.jhu.edu/map.html*

Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., . . . Islam, M. T. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access, 8*, 132665-132676. doi:10.1109/access.2020.3010287

Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988, https://github.com/ieee8023/covid-chestxray-dataset, 2020*.

Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. *Diagnostics, 10*(6), 417. doi:10.3390/diagnostics10060417

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2016.90

Lenail, A. (n.d.). NN-SVG. Retrieved November 20, 2020, from http://alexlenail.me/NN-SVG/LeNet.html

Nakashima, K. grad-cam-pytorch, GitHub repository, https://github.com/kazuto1011/grad-cam-pytorch

Narin, A., Kaya, C., & Pamuk, Z. (2020). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*

Oh, Yujin, Sangjoon Park, and Jong Chul Ye. "Deep learning covid-19 features on cxr using limited training data sets." *IEEE Transactions on Medical Imaging* (2020).

Rahman, T. (April, 2020) COVID-19 Radiography Database, Version 2. Retrieved August 13, 2020 from *https://www.kaggle.com/tawsifurrahman/covid19-radiography-database*

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

Vellido, A. (2019). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, 1-15.

Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports, 10*(1). doi:10.1038/s41598-020-76550-z.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in different types of clinical specimens. *Jama*, *323*(18), 1843-1844.

World Health Organization. (2001). *Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children* (No. WHO/V&B/01.35). World Health Organization.

Xie, X., Zhong, Z., Zhao, W., Zheng, C., Wang, F., & Liu, J. (2020). Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology*, 200343.

Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.