

Using Deep Learning Neural Networks To Classify X-Ray Images

Jason Si

Southridge school, Vancouver, V4A9N8, Canada

Abstract. In early 2020, the virus known as COVID-19 emerged as a significant global public health concern. The urgent need to diagnose infected individuals quickly led to the development of diagnostic equipment. However, this situation also raised concerns regarding the effectiveness and reliability of such equipment. In this paper, we present an alternative approach using deep learning neural networks to diagnose patients. Our method uses a convolutional neural network (CNN) model capable of classifying COVID-19 using X-ray images of patients' lungs. We uploaded two sets of X-ray images categorized into subclasses of Normal, Covid, and Virus, and trained a MobileNet model. The trained MobileNet model showed capabilities in accurately classifying X-ray images among the three classes. This approach has the potential for further development and application in the future to benefit the medical field.

Keywords: COVID- 19; diagnostic equipment; CNN model; MobileNet model; X-ray.

1. Introduction

In early 2020, the global outbreak of Coronavirus Disease 2019 posed a significant challenge worldwide. The Centers for Disease Control and Prevention (CDC) developed COVID-19 test kits, but encountered issues and delays due to faulty kits [1,2]. These complications hindered the early response to the pandemic, while the number of cases rose rapidly worldwide. By May, testing improved, but with the death toll reaching 100,000 in the US, the situation remained far from ideal [2]. However, we believe a deep neural network (DNN) might have proven valuable in dealing with the diagnosis of diseases during the COVID-19 pandemic.

The use of Article Intelligence (AI) within healthcare has been rapidly growing [17]. Using neural network models such as DNN, or more specifically convolutional neural networks (CNN) [3], the use of deep learning neural networks to classify and identify images of large datasets such as our diagnostic images has quickly broken into the medicine field [17, 18].

Using CNN in healthcare offers numerous advantages. Firstly, it enables healthcare workers to diagnose diseases and conditions without trained eyes while still maintaining enhanced speed and accuracy [17]; Deep learning models excel at classifying medical images such as X-rays, enabling the detection of abnormalities that may be difficult to detect visually. Secondly, image classification can aid doctors and researchers in planning personalized treatment plans for patients. By analyzing medical images, image classification helps determine the most suitable treatment for each individual patient [18, 19]. Additionally, keep in mind that as the technology develops, expect it to grow and adopt new abilities.

In this paper, a ModelNetV1 model [12] was employed to diagnose and classify medical images, with a specific focus on X-rays images. The deep learning model was developed to identify three types of X-rays: normal, COVID-19 positive, or positive for another virus. The datasets used for training and evaluation were sourced from Kaggle, an online community for machine learning and data science. The first dataset, curated by Pranav Raikote [14] , comprised approximately 50 high-quality images categorized into COVID-19, normal, or virus. The second dataset, created by Manu Siddhartha [15], consisted of over 500 images, also categorized into COVID-19, normal, or virus.

2. Related Work

The classification of medical images using machine learning began in 2000's [20]. Through two decades, with the advancement and development of deep learning models and growth in medical datasets, deep neural networks models have been developed to do a similar task our experiment aims for.

CheXNet is a Dense Convolutional Network (DenseNet) trained on the ChestX-ray14 dataset [21], by Pranav Rajpurkar, Jeremy Irvin, and Kaylie Zhu. It holds 121 deep learning layers with the capabilities of detecting the probability of pneumonia and a heatmap able to locate areas that indicate pneumonia [4]. The DenseNet [5] used in the model uses a different connection pattern; each layer is densely and directly connected to each subsequent layer, hence the name DenseNet. Using DenseNet, CheXNet can classify pneumonia within X-Ray images with better gradient flow, parameter efficiency, and encourages feature reuse.

AlexNet (2012) is a eight-layered Convolutional Neural Network modeled by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. The overall neural network architecture contains five convolution layers and three connected layers.AlexNet in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 [23] had a top-5 error rate of 15.3% and a top- 1 error rate of 16.4% on its datasets. This dataset consists of over 1.2 million labeled images, belonging to 1,000 different classes. Using the ILSVRC 2010 dataset, the network managed to achieve a top- 1 test set error rate of 37.5% and top-5 test set error rate 17.0%, surpassing the best model in the ILSVRC 2010 competition [6].

VGG (2014) is a deep convolutional neural network architecture designed by the Visual Geometry Group at the University of Oxford. The VGG network achieved an impressive top-5 error rate of 7.3% in the ImageNet Large Scale Visual Recognition Challenge 2014. Compared to prior models, VGG uses a small convolution 3x3 filter which frees up space for the model to go deeper. The model contains 13 convolution layers and 3 fully connected layers [7]. VGG16 has also proven itself to be especially competent when it comes to medical diagnoses. Because of VGG16's smaller filter, using X-ray images of the lung, Yadav et al. were able to achieve a mean of 83.4% accuracy in classifying five individual models of normal and pneumonia infected lungs [8, 9].

GoogLeNet is a deep convolutional neural network architecture introduced by Christian Szegedy. At the ILSVRC 2014, GoogLeNet achieved a top-5 classification accuracy of 93.33%. The concept that GoogLeNet introduced is that of an Inception architecture. By concentrating in a single, localized region, and smaller convolutions, the model can increase its width and depth allowing higher accuracy [9] .

ResNet (2015), is a deep convolutional neural network architecture designed by Kaiming He, Xiangyu Zhang. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015, the ResNet architecture achieved a top-5 error rate of 3.57%. ResNet introduces the concept of residual connections, which allow the model to learn the residual mappings rather than directly learning the desired underlying mapping. By skipping certain layers and directly input information from earlier layers to later ones, ResNet helps the training of deeper networks and improves its performance [10] .

VIT(2020) or vision transformer is a deep learning architecture for image classification introduced by Alexey Dosovitskiy, Lucas Beyer. In the original paper introducing VIT, they reported a top-1 accuracy of 88.4% and a top-5 accuracy of 98.6% on the ImageNet training dataset. The major difference between VIT and other CNNs is its transformer architecture. Originally designed as a language model, using a transformer model, VIT treats images as sequences of localized patches and applies a series of self-attention mechanisms to capture global dependencies between the patches. This allows the model to accurately learn the spatial relationships between the patches and recognize complex patterns within the images [11] .

3. Method

3.1 Dataset Preparation

Collecting the necessary health information for model training can pose challenges, as it raises privacy concerns. Additionally, the consistency of each images' quality can prove to be difficult as the sizes of these datasets increase [13, 22]. Keeping these issues in mind, we sourced the data from Kaggle. Kaggle is an online community for data scientists, where users can share datasets, competitions, and models for public use. As mentioned, the two datasets we used were from Pranav Raikote [14] and Manu Siddhartha [15]. Combined, the datasets add up to around 2000 x-ray images, a decent number to train our model.

Table 1: Details of the medical X-Ray dataset.

	COVID- 19	Normal	Virus	Total
Training Set	359	594	495	1448
Testing Set	87	144	124	355
Total Set	446	738	619	1803

In total, 1803 images were uploaded onto the platform, out of which 1,448 images were used to train the model. The collective of images is separated into three different classes: 495 images in the virus class, 594 in the normal class, and 359 images in the COVID class. These images were used to train the model, while the rest of the images (355 images) were reserved for the testing set. In total, the model maintained an 80%/20% train and test split (Table 1, above).

All three sets of images (Fig. 1, below) are converted to grayscale, and we resized all of the images to 96x96 to ensure consistency amongst them. During the model's preparation, the model takes in raw data from the datasets and uses signal processing to extract features. Signal processing analyzes, modifies, and processes the input data. Digital image processing then examines the input through an algorithm, manipulating and transforming images to avoid distortion during the classification process.

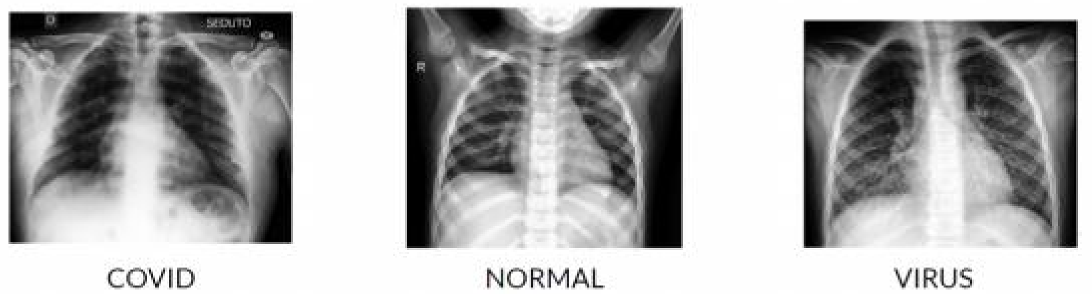

Figure 1: Examples from dataset: x-ray images of Covid, normal, and virus infected lungs.

3.2 Model Choice

In the experiment, we employed the MobileNetV1 0.25 model to classify images into their respective categories. The utilization of depthwise separable convolutions minimizes the number of parameters and computations necessary for the model, while still maintaining high precision. Essentially, the initial layer of the MobileNetV1 model involves two successive layers. The first layer conducts depthwise convolution by applying a set of small filters to each input channel separately. The second layer applies a pointwise convolution, which combines the output of the depthwise convolution across all channels by performing a 1x1 convolution. Using this method, MobileNetV1 has the ability to accurately classify and identify medical images [12, 16] .

However, as previously mentioned, many other models have the ability to classify the three datasets. For example, using models such as AlexNet, VGG, and ResNet, it is possible to achieve similar results. With the use of rectified linear unit (ReLU) activation functions, AlexNet will be able to function and achieve similar results [6]. Similarly, VGG's deep learning model that consists of multiple convolutional layers with a small filter size, followed by max-pooling layers with a stride of 2 pixels will have the ability to substitute for the model used in this experiment [7]. On the other hand, as mentioned above, ResNet will be able to classify the images using its residual connections, allowing the input to flow directly to the output. By doing so, the network can learn to identify residual or "shortcut" connections that capture the difference between the input and the output of the network[10]. Thus, there are multiple options for selecting an appropriate model architecture for classifying the three datasets.

3.3 Training

For the experiment, we trained the datasets using Edge Impulse. Edge Impulse is an online platform that enables developers to quickly build and conveniently deploy trained machine learning models. The platform provides a range of tools and services for collecting, processing, and analyzing data from sensors, as well as building and deploying machine learning models on embedded devices. Edge Impulse also provides a variety of pre-built machine learning models for building custom models, which can be trained and tested on the collected data. It supports a range of machine learning algorithms. Since the platform simplifies the process of collecting and labeling data from sensors, we can use it to simulate a model that can classify COVID and other viruses.

First, we inputted the dataset into Edge Impulse for training (Figure 2, below). It provides the ability to modify several parameters, including the training cycle, the learning rate, and a few advanced training settings. The number of training cycles is set to 150. The number of training cycles controls the amount of epochs the Neural Network (NN) will train on. Each epoch equals one cycle through the entire algorithm. Learning rate controls how fast the Neural Network (NN) will learn. It avoids the model from learning the data too fast and overfitting. Data augmentation will shift the training data randomly which allows a greater amount of training cycles while also minimizing overfitting. The validation set size represents how much of the training data will be stored to validate the model later on. By enabling a Profile int8 model, it will reduce the memory and computing requirements of my computer.

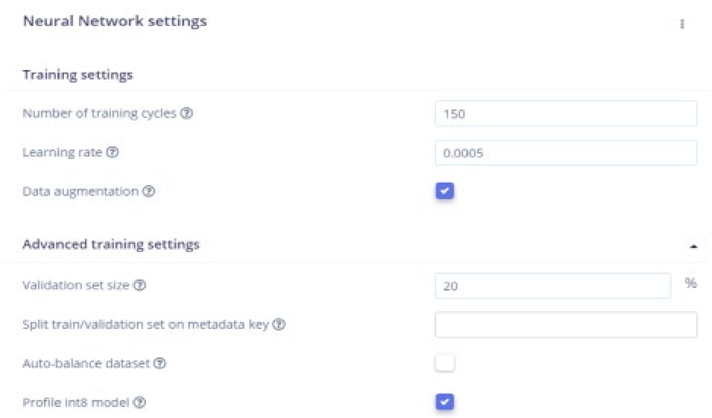

Figure 2: The details of the Neural Networks settings.

4. Experiments

The classification of medical images using machine learning began in 2000's [20]. Through two decades, with the advancement and development of deep learning models and growth in medical datasets, deep neural networks models have been developed to do a similar task our experiment aims for.

4.1 Experiment Result

In Edge Impulse, it is possible to modify the neural network to suit the requirements of the model. There are several neural network models available in Edge Pulse that can be customized. For this project, as previously mentioned, we chose the MobileNetV1 96x96 0.25 model by Edge Impulse, with 150 epochs, a learning rate of 0.0005, and a 25% validation set. The input layer had a total of 9,216 features. The model achieved an 85.5% accuracy rate over the training dataset while maintaining a loss of 0.40. It can be inferred that COVID was the most accurately classified category, but the model showed some confusion between Normal and Virus categories. The inference time was 150 ms, and the training time was approximately 20 minutes.

Table 2: Confusion matrix of the model.

	COVID- 19	Normal	Virus	COVID- 19
COVID- 19	98.5%	0%	1.5%	98.5%
Normal	4.4%	82.5%	13.2%	4.4%
Virus	9.9%	9.0%	81.1%	9.9%
F1 Score	0.88	0.86	0.83	0.88

4.2 Evaluation

The results (Table 2) shows that the model does not confuse COVID with Normal, but 1.5% of the COVID class was classified as Virus, which is great news as our goal is to determine and diagnose COVID- 19 patients. However, the Normal class only achieved an accuracy of 82.5%. The model tends to mix up Normal with Virus, as 13.2% of its model classify Normal as Virus. Although not a huge deal as being cautious with Normal patients leads to more good than harm, the Virus class encountered the most challenging issue. The Virus class had an accuracy of 81.1%, confusing Normal patients at 9.0%. This margin of error would be devastating as misdiagnosing infected patients might lead to long term repercussions.

For the virus datasets, of the 9.0% that is mixed with the normal set, in a lot of instances, the shape of the lung was very similar (Figure 3, below). The virus x-ray that was classified as normal shows a lung that is narrow on the top and wide at its base. At the same time, both lungs expand in the right direction confusing the model of their classification. Similarly, this is also the reason for why 13.2% of the normal dataset were classified as viruses. The similarity between the two x-rays exposes the danger when it comes to medical image classification. Tiny differences that could be only identified by trained eyes could become one of the issues when it comes to medical diagnoses.

Figure 3: Comparison between a virus and normal X-ray.

The model version used currently, is a quantized model (int8). However, when the model is switched to an unoptimized model, the overall accuracy seems to have gone up.

Table 3: Confusion matrix of the unoptimized model.

	COVID- 19	Normal	Virus
COVID- 19	98.5%	0%	1.5%
Normal	4.4%	82.5%	13.2%

Virus	9.9%	9.0%	81.1%
F1 Score	0.88	0.86	0.83

Compared with the other model, the unoptimized model's accuracy (Table 3) increased by 3.8% while the loss decreased by 11%. The COVID class's percentage of accuracy, however, has gone down, classifying some of its images as Normal. The Normal class increased in its accuracy, classifying lesser images as Viruses. Last, the Virus class's accuracy also increased, after classifying fewer of its images as COVID.

Even though an unoptimized model requires more memory and computation to run, leading to high run time. However, unoptimized neural network models provide a trade off between performance and resource usage. For example, a larger model with more parameters may achieve higher accuracy or generalization. In some cases, this trade-off may not be desirable, but in this case, the unoptimized model works better with classifying data [24].

5. Conclusion

In this experiment, we trained a deep learning Neural Network model to classify X-ray images from three different classes: normal, Covid, and other viruses. Using Edge Impulse, an online platform, we were able to simulate what a model like MobileNetV1 will perform on these datasets. By using the model we were able to achieve 89.3% accuracy which is slightly above average. Though there are still issues and limitations within the virus and normal classes, it is very possible medical image processing using deep neural networks will develop and medical diagnosis on patients will become simpler and faster. Additionally, the analysis of vast amounts of medical data such as medical images, genetic data, and patient history, will continue to aid healthcare professionals for the future.

References

[1] Centers for Disease Control and Prevention. (n.d.). COVID- 19 Timeline. Retrieved June 26, 2023, from https://www.cdc.gov/museum/timeline/covid19.html.

[2] AJMC. (n.d.). A Timeline of COVID- 19 Developments in 2020. The American Journal of Managed Care. Retrieved June 26, 2023, from https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020.

[3] IBM. (n.d.). Convolutional Neural Networks. Retrieved June 26, 2023, from https://www.ibm.com/topics/convolutional-neural-networks.

[4] P. Rajpurkar, J. Irvin, K. Zhu, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:1711.05225, 2017.

[5] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. arXiv preprint arXiv:1608.06993, 2018.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25,1097- 1105.

[7] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks For Large-Scale Image Recognition. arXiv preprint arXiv: 1409.1556, 2015.

[8] S. S. Yadav, and S. M. Jadhav. (2019). Data mining and big data: a selected review. Journal of Big Data, 6(1), 25.

[9] C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions. arXiv preprint arXiv: 1409.4842, 2014.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. arXiv preprint arXiv: 1512.03385, 2015.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al. An Image Is Worth 16X16 Words: Transformer For Image Recognition At Scale. arXiv preprint arXiv: 2010.11929, 2021.

[12] A. G. Howard, M. Zhu, B. Chen. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861, 2017.

[13] Dal-Ré, R., Dorsey, E. R., & Li, G. (2021). The ethics of clinical trials in the age of COVID- 19: Views from experts in leading research countries. BMC Medical Ethics, 22(1), 42.

[14] COVID-19 ImageDataset. (n.d.). Retrieved from https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset.

[15] COVID-CXR Image Dataset (Research). (n.d.). Retrieved from

[16] https://www.kaggle.com/datasets/sid321axn/covid-cxr-image-dataset-research.

[17] Towards Data Science. (n.d.). Review: MobileNetV1 — Depthwise Separable Convolution Light Weight Model. Retrieved June 26, 2023, from https://towardsdatascience.com/review-mobilenetv1-depthwise-separable-convolution-light-weight-m odel-a382df364b69.

[18] Southern Medical Association. (n.d.). AI in Medical Diagnosis. Retrieved June 26, 2023, from https://sma.org/ai-in-medical-diagnosis/.

[19] A. Gouy, J. Attia, and T. Montavon. (2021). COVID- 19 pandemic: Linking infection to inflammation and immune response. Frontiers in Genetics, 11, 607471.

[20] Drexel University. (n.d.). Artificial Intelligence in Medicine: Pros and Cons. Retrieved June 26, 2023, from https://drexel.edu/cci/stories/artificial-intelligence-in-medicine-pros-and-cons.

[21] V. Kaul, S. Enslin, and S. A. Gross. (2020). History of artificial intelligence in medicine. Gastrointestinal Endoscopy, 92(4), 807-812.

[22] X. Wang, Y. Peng, L. Lu. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. arXiv:1705.02315, 2017.

[23] Analytics Vidhya. (n.d.). Healthcare Data Problems. Medium. Retrieved June 26, 2023, from https://medium.com/analytics-vidhya/healthcare-data-problems-b2016aaee61d

[24] O. Russakovsky, J. Deng, H. Su. ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575, 2015.

[25] MathWorks. (n.d.). What Is int8 Quantization and Why Is It Popular for Deep Neural Networks? MathWorks Newsletters & Articles. Retrieved June 26, 2023, from https://www.mathworks.com/company/newsletters/articles/what-is-int8-quantization-and-why-is-it-po pular-for-deep-neural-networks.html