COVID-19 Detection from Chest X-ray images using Convolutional Neural Network

Hieu T. Le and Voratham Tiabrat

Columbian College of Arts & Sciences, George Washington University

DATS6501: Capstone Project

Dr. Abdi Awl

May 6, 2021

Table of Contents

Glossary of Terms

Artificial neural network: a computing system designed to simulate the structure and function of the brain that analyzes and processes information and signals.

Chest computed tomography (CT) scan: an imaging test using X-rays and computer technology to capture pictures of the organs and structures inside the chest.

Chest X-ray (CXR): an imaging test using X-rays to provide a back-and-white picture of the organs and structures inside the chest.

Computer-aided diagnosis (CADx): is a system utilizing artificial intelligence methods to assist doctors in the interpretation of medical images.

Convolutional neural network (CNN): is a class of deep neural network designed for processing pixel data such as images.

Cost-sensitive learning: a type of machine learning that takes the costs of misclassification into account when training a machine learning model.

COVID-19: the acronym for the full name coronavirus disease 2019 which is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Data augmentation: a technique used to increase the diversity of data by adding slightly modified copies or newly created synthetic data from existing data without collecting new samples.

RT-PCR: the acronym for the full name reverse transcriptase polymerase chain reaction which is a laboratory technique combining reverse transcription of RNA into DNA and amplification of specific DNA targets using polymerase chain reaction.

Introduction

Background

Coronavirus disease 2019 or COVID-19 is an infectious disease caused by a newly discovered strain of coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This novel coronavirus was discovered in Wuhan, Hubei province, China by the end of 2019, and spreads across the world in 2020.

Since the outbreak of the disease and its severe impacts on human life worldwide, the World Health Organization (WHO) announced COVID-19 as pandemic in March 2020. According to the most recent reports from WHO, 223 countries have been facing with COVID-19, and there are approximately 141 million infected patients and more than 3 million confirmed deaths globally.

A major step in fighting the pandemic is to detect infected individuals at an early stage and put them under special treatment and careful supervision. The most popular and effective testing method is the laboratory test with the use of RT-PCR protocols. Another useful method is computed tomography (CT) scan which is an imaging technique that focuses on finding the abnormalities of infected chests and uses them for screening infected chests.

Problem statement

Despite the importance of COVID-19 detection, there remains difficulties in providing people with an easy access to current testing methods. The consequence is that governments cannot keep track of infected individuals and control the spread of the pandemic, leading to further detrimental impacts on health systems, economies and environmental domains globally.

Problem elaboration

One of the major challenges is the insufficiency of testing resources. In developing countries and rural areas, people are not able to get access to COVID-19 testing due to poor infrastructure, unaffordability of expensive RT-PCR based protocols and inadequate lab capacity. On the other hand, developed countries have to deal with the increasing demand for coronavirus tests, which is the primary reason for the shortages of testing kits, testing laboratories and well-trained doctors. Furthermore, viral mutations of SARS-CoV-2 can lead to the emergence and spread of new coronavirus variants that are different from the predominant one. Consequently, more false negative test results may be concluded due to changes in viral genome and protein structures of these genetic variants.

Motivation

Inspired by the success of computer-aided diagnosis (CADx) in medical research and the advance of deep learning in computer vision tasks, this capstone project aims to build a CADx system using convolutional neural network architectures to screen COVID-19 patients from Chest X-ray (CXR) images. The benefits of this method are the cost efficiency of CXR imaging technique and the high performance of CNN in image classification.

Project scope

In this project, CXR images are collected from public sources such as Kaggle and GitHub to generate a coronavirus data including 3 categories: healthy or normal chest, viral pneumonia and COVID-19. The purpose is to develop a CNN model with high sensitivity and specificity when classifying COVID-19 examples.

Literature Review

As the preparation for this project, relevant research projects are reviewed to develop background of coronavirus detection using medical images. In “COVID-19 image data collection”, Cohen et al. (2020) state the importance of developing a pneumonia database in the context of the pandemic and take the initiative to collect open images of COVID-19, MERS, SARS and ARDS. The authors assemble medical images from publications and online resources to create an image data including chest X-rays and CT scans which can be used for computational analysis. Another article published by Nishio et al. (2020) discusses benefits and drawbacks of real time RT-PCR laboratory tests and CT scans and explains the potential of CADx with CNN models for screening COVID-19 patients from CXR images. Furthermore, relevant GitHub projects are referred to improve knowledge about methodologies, challenges and creative solutions in classifying COVID-19 CXR images.

In general, most of studies about detecting COVID-19 have to face the class imbalance due to the limitation of open COVID-19 medical images. There are only few hundred COVID-19 cases but thousands of non-COVID-19 samples available for training, which prevents the models from learning diverse patterns of coronavirus variants. Some projects handle this issue by downsampling the number of non-COVID-19 data. However, this method does not reflect the reality that there would be always much more non-COVID-19 individuals than infected ones in a population.

Methodology

Dataset description

The dataset includes approximately 26,000 medical images of COVID-19 cases, viral pneumonia and normal chests collected from multiple online resources. There are two types of medical images including chest X-rays and chest CT scans with a variety range of resolutions and formats (png, jpg, gif, dcm …). The CXR images are captured in various projections: posteroanterior (PA), anteroposterior (AP), and lateral views.

Data collection

This project assembles medical images from popular public data sources such as GitHub, Kaggle, …, some online websites and social media platforms. The source links for these datasets are listed below.

Public data sources:

- https://github.com/ieee8023/covid-chestxray-dataset
- https://github.com/ml-workgroup/covid-19-image-repository
- https://github.com/armiro/COVID-CXNet
- https://github.com/agchung/Figure1-COVID-chestxray-dataset
- https://github.com/agchung/Actualmed-COVID-chestxray-dataset
- https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
- https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
- https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

Online websites:

- https://radiopaedia.org/cases
- https://aimi.stanford.edu/resources/covid19
- https://www.sirm.org/category/senza-categoria/covid-19/
- https://www.eurorad.org/

Social media:

- https://twitter.com/
- https://www.reddit.com/r/Radiology/
- https://www.reddit.com/r/science/

Feature engineering

In this project, only CXR images with PA and AP views are collected for classification because these projections sufficiently show the damage of the lung due to the virus and are useful for pattern recognition.

Datasets from GitHub repositories are classified using the label and the image path stored in the metadata. On the other hand, the images from Kaggle, social media and online websites have been already pre-labelled.

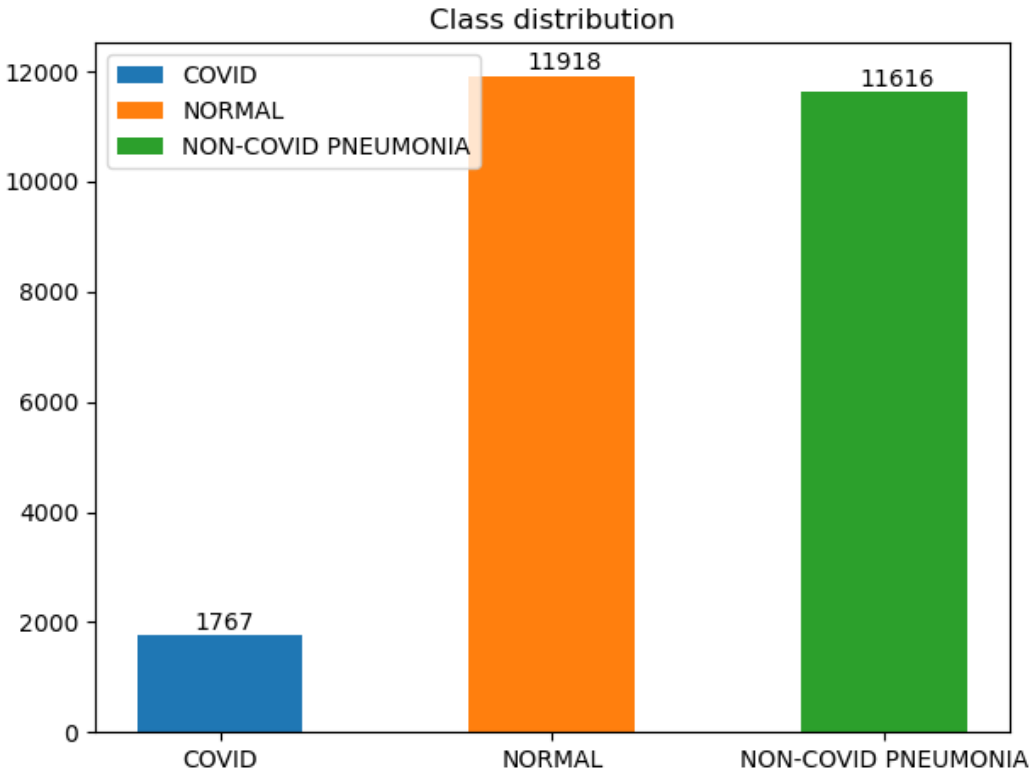
After the images are correctly classified into correct categories, they are all converted to png format and resized to 128×128 pixel.

As the data come from many resources, it is possible that there are similar images between different sources, therefore, removing duplicated images is necessary to avoid data leak.

The actual coronavirus dataset contains 25,301 different CXR images of 3 categories: COVID-19, viral pneumonia and normal chests.

Figure 1

Class distribution of the coronavirus dataset



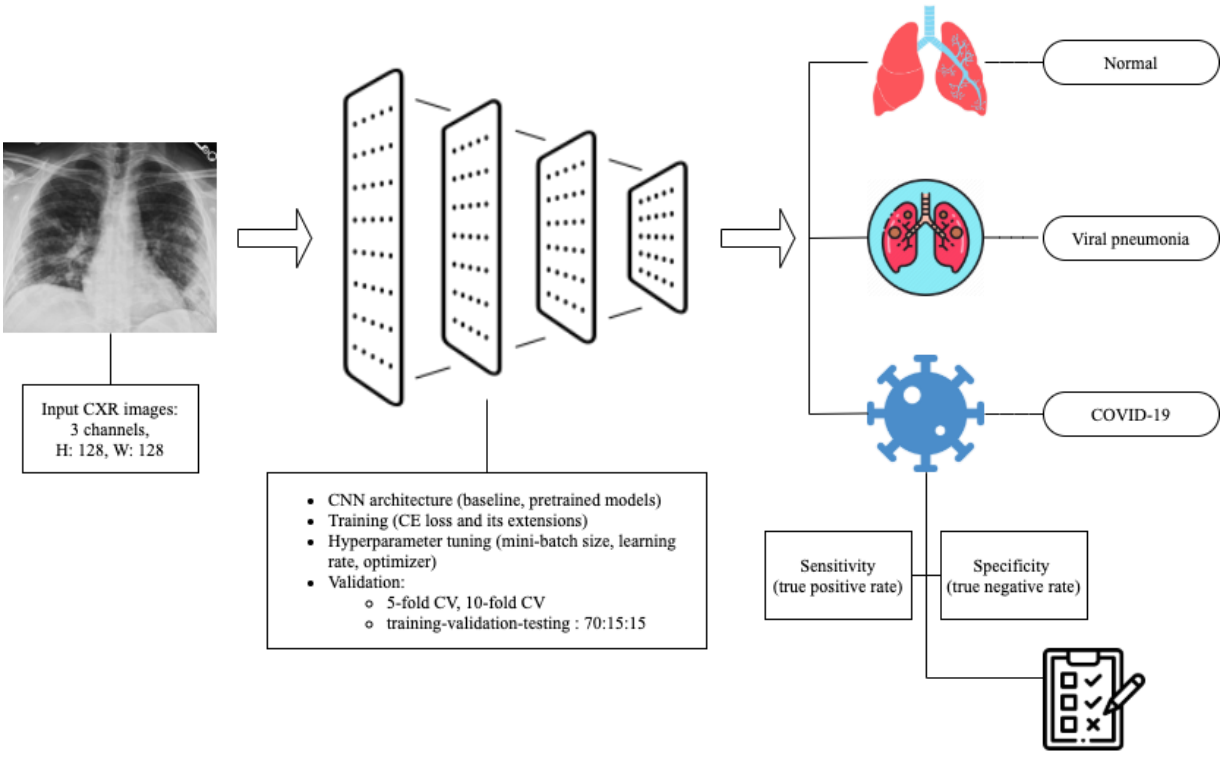
Finally, the data is split into training, validation and testing sets (ratio: 70:15:15) which are used for training, hyperparameter tuning and model evaluation, respectively.

It is worthwhile to note that cross-validation is a useful technique to assess the hyperparameters of a model and mitigate the randomness when selecting the validation set. However, this method is computationally expensive for complex convolutional neural networks, and the most minority class in the dataset has more than 1,700 samples which are sufficient for learning. Therefore, cross-validation is not used in this project.

Data modeling

Figure 2

The workflow of modeling process



- The purpose is to build CNN models classify the input CXR images into 3 categories (normal, viral pneumonia and COVID-19), and use the sensitivity and specificity of COVID-19 cases for evaluation.
- A simple architecture with 2 convolutional layers is used as the baseline model. Pretrained models such as ResNet, DenseNet and VGG are used as backbone architectures to implement transfer learning since they are effective in pattern recognition studies.
- The models are trained on the training set, fine-tuned on the validation set and the final evaluation is made on the testing set.

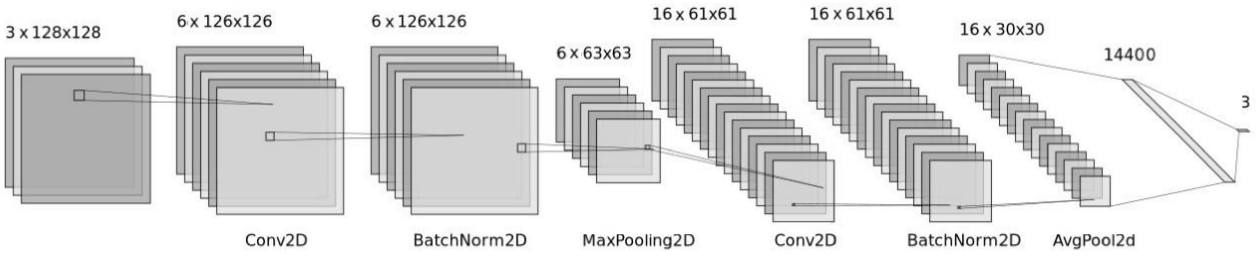
- Cross-entropy (CE loss) is the loss function.
- Hyperparameters for fine tuning are on optimizer, mini-batch size and learning rate. Adam, SGD and RMSprop are optimization algorithms. The mini-batch size is selected from 32, 64, 128, 256 and 512. The learning rate is scheduled to increase exponentially from 10^{-7} to 10^{-2} and the most optimal value is found using the cyclical learning rate method introduced by Smith (2017) (see Appendix B for an example).
- Machine learning techniques, tricks and innovative ideas such as alternative loss functions, data augmentation and differential learning rates are utilized to find the best modeling solution.

Results and Analysis

Baseline model

Figure 3

Architecture of the baseline convolutional neural network



Note. The activation function for the flatten layer is ReLu.

Figure 4

The predictions of the baseline model on the testing data

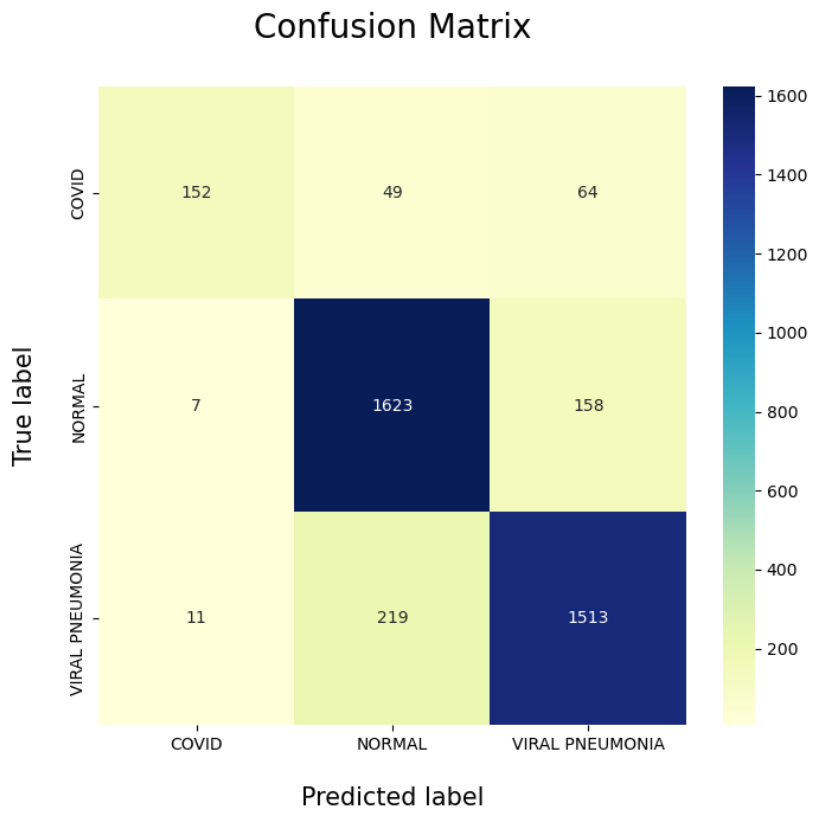


Table 1

Hyperparameters and evaluation metrics of the baseline model

Hyperparameters	Sensitivity		Specificity	
	validation	testing	validation	testing
optimizer = Adam learning rate = 3×10^{-4} mini-batch = 128	0.6113	0.5736	0.9943	0.9949

According to the confusion matrix in Figure 4, the majority of normal and viral pneumonia cases are correctly classified, but there are a lot of false negatives when predicting COVID-19. As can be seen from the calculation in Table 1, the sensitivity for classifying the testing data is about 57.36% which indicates a high false negative rate at 42.64%. This issue is also commonly known for current testing methods such as RT-PCR test, antigen test and antibody test (Shmerling, 2020).

A major reason for high false negative rate is the shortage of information about coronavirus and its variants, whereas there are more details and knowledge about the causes of viral pneumonia as well as the structures inside a health chest. This leads to the dominance of non-COVID-19 cases (both normal and viral pneumonia) and makes testing samples tend to be classified into these majorities. Therefore, it is critical to implement methods that handle the class imbalance to reduce the false negative rate and improve the model's sensitivity to coronavirus.

Transfer learning

ResNet34 is selected as the backbone architecture for transfer learning. The last fully connected (FC) layer is modified to match the output of the project, whereas the convolutional base is kept in its pretrained form.

The pretrained model is trained in two different ways. The first ResNet34 model has its convolutional base frozen while only the top FC layer is trained. In the second Resnet34, all the layers are unfrozen and trained from their pretrained weights.

For both models, the most optimal optimizer is Adam and the mini-batch size of the training set is 128.

Table 2

Learning rate and evaluation metrics of ResNet34 models

Model	Learning rate	Sensitivity	Specificity
Only train FC layer (1^{st} ResNet34)	10^{-4}	0.6642	0.9844
Train all layers (2^{nd} ResNet34)	10^{-5}	0.8264	0.9932

As can be seen from Table 2, pretrained models have much better performance than the baseline when predicting the positives of COVID-19. In the second model whose all layers are trained, the true positive rate is improved by approximately 25% compared to the baseline.

It is worthwhile to note that the second ResNet34 gives higher scores than the first one. This means that there is a difference between the ImageNet and our dataset, thus, pretrained weights need to be adjusted. However, since the learning rate to train all layers is only one-tenth of the learning rate to train only the FC layer, these adjustments to the convolutional base are relatively small compared to the change of the top layer. This finding suggests the model be trained with differential learning rates to optimize the learning ability.

Methods, tricks and ideas

Cost-sensitive learning and focal loss are implemented to address the main challenge in this project – the class imbalance. By adding class weights to the loss function based on the data distribution, cost-sensitive learning helps all classes to be treated equally during the training. On the other hand, focal loss extends CE loss by a factor featured by a focusing parameter gamma

that down-weights easy positives and focuses on hard negatives (see Appendix A). Moreover, the well-known data augmentation technique is included to increase variability and diversity of the data and reduce overfitting. Last but not least, different modeling strategies are experimented to find the best way for models to learn the training samples.

Findings

After building many models and analyzing their results, we come up with two most critical findings. The first one is that cost-sensitive learning has better performance than focal loss in terms of addressing the class imbalance. The reason would be that the outputs are 3 categories, while focal loss is shown to be more successful in binary classification and multi-label classification tasks.

Secondly, the fully connected layer and the convolutional base should be trained separately to avoid the corruption of pretrained weights due to the randomness of initial weights in the fully connected.

Final solution

The strategy to build the best models is explained below (more details about other models can be found in Appendix C):

- The weighted cross-entropy is the loss function.
- Transfer learning is implemented to build the backbone CNN architecture.
- Two different training sets are generated using data augmentation (random crop, random rotation, random flip and normalization). The first set contains only

augmented images, whereas the second one contains both augmented and original images.

- The model is trained on the first training set with its convolutional base frozen.
- After that, all the layers are unfrozen, and the model is trained entirely on the second training set with a smaller learning rate than the value used to train the first set.
- Fine tuning for both training processes is done on the same validation set.

Table 3

The performance of 3 different CNNs using the best solution

Model	1st learning rate	2nd learning rate	sensitivity	specificity
ResNet34	10^{-4}	10^{-5}	0.9547	0.9955
DenseNet121	3×10^{-4}	3×10^{-5}	0.9396	0.9949
VGG16	10^{-4}	10^{-5}	0.9321	0.9963

Note. All models have the same optimal optimizer (Adam) and mini-batch size (256).

According to Table 3, all the models have better scores than the baseline. The prediction of true positives is significantly improved, whereas the true negative rate remains above 99%. The best architecture is ResNet34 with a sensitivity of 95.47% and a specificity of 99.55%. It seems that the complexity of ResNet34 is more suitable for the input sizes. However, the performances of pretrained models would vary if the images have higher resolutions and the number of augmented images is increased.

Conclusion

Project conclusion

This capstone project studies a novel method for COVID-19 detection using computer-aided diagnosis with deep convolutional neural network as the backbone architecture. Assembling open CXR images from online articles, public data resources and social media platforms, we create a coronavirus dataset of three categories: COVID-19, viral pneumonia and normal chests. After experimenting many techniques and training strategies, we combine cost-sensitive learning, transfer learning, data augmentation and differential learning rates to build a ResNet34 model that gives a sensitivity of 95.47% and a specificity of 99.55%. The result is a great improvement compared to our baseline convolutional architecture whose sensitivity is only 57.36%.

However, despite a promising performance, the method is not recommended as a replacement for real-world coronavirus detection. More empirical research and careful examinations need be conducted to evaluate its practicability and effectiveness.

Project limitation

The project remains some limitations that can be improved for future studies. The first issue is the availability of open COVID-19 medical data. There are only few sources to collect COVID-19 CXR images, but there is still no official releasement of a coronavirus database for computational analytics. This reduces the diversity of training samples and prevents our models from learning more unseen patterns and abnormalities that differentiate coronavirus from viral pneumonia.

Another challenge is the restricted computing power of our virtual machine's GPU that prevents the models from training high resolution images and optimizing computationally expensive tasks such as cross-validation and data augmentation.

Future research

There are several objectives to be implemented in future research to expand the project scope and mitigate existing limitations. The diversity of our coronavirus dataset can be improved by spending more time on collecting medical images or seeking assistance from coronavirus testing laboratories and hospitals. The CXR images would be resized to a higher resolution to provide clearer pictures inside the chest and help the convolutional neural network to retrieve more crucial information. An extension of this study is to include bounding box annotation to locate special patterns and abnormalities of infected chests which are useful for developing vaccines and medical treatments. Finally, there are other pretrained models such as SqueezeNet, MobileNet, EfficientNet and Inception to be tried as the backbone architecture for our CADx system.

References

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., & Xia, L. (2020). Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China. *A Report of 1014 Cases. Radiology, 296*(2). https://doi:10.1148/radiol.2020200642

Ahmed, S., Yap, M. H., Tan, M., & Hasan, M. K. (2020). ReCoNet: Multi-level Preprocessing of Chest X-rays for COVID-19 Detection Using Convolutional Neural Networks. https://doi.org/10.1101/2020.07.11.20149112

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481-2495. https://doi:10.1109/tpami.2016.2644615

Binnicker, M. J. (2020). Challenges and Controversies to Testing for COVID-19. *Journal of clinical microbiology*, *58*(11), e01695-20. https://doi.org/10.1128/JCM.01695-20

Chowdhury, M. E. H., Rahman, T., Khandakar, A., & Mazhar, R. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia?. *IEEE Access, vol. 8*, pp. 132665-132676, 2020, doi:10.1109/ACCESS.2020.3010287

Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 Image Data Collection. *arXiv.org*. https://arxiv.org/abs/2003.11597

Jacobi, A., Chung, M., Bernheim, A., & Eber, C. (2020). Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clinical imaging, 64*, 35–42. https://doi.org/10.1016/j.clinimag.2020.04.001

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). Focal Loss for Dense Object Detection. *arXiv.org*. https://arxiv.org/abs/1708.02002

Mahdy, L. N., Ezzat, K. A., Elmousalami, H. H., Ella, H. A., & Hassanien, A. E. (2020). Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine. https://doi.org/10.1101/2020.03.30.20047787

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, J. G. (2020). Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical image analysis*, *65*, 101794. https://doi.org/10.1016/j.media.2020.101794

Nishio, M., Noguchi, S., Matsuo, H., & Murakami, T. (2020). Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Scientific reports*, *10*(1), 17532. https://doi.org/10.1038/s41598-020-74539-2

Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. *arXiv.org*. https://arxiv.org/abs/1506.01186

Watson, J., Whiting, P. F., & Brush, J. E. (2020). Interpreting a covid-19 test result. *The BMJ*. https://doi.org/10.1136/bmj.m1808

Winther, H. B., Laser, H., Gerbel, S., Maschke, S. K., B. Hinrichs, J., Vogel-Claussen, J., Wacker, F. K., Höper, M. M., & Meyer, B. C. (2020). COVID-19 Image Repository (Version 1). *figshare*. https://doi.org/10.6084/m9.figshare.12275009.v1

Appendices

Appendix A

Focal loss function

Cross-entropy loss: $CE = -\sum_{c=1}^{n} y_c \log(p_c)$

Where:

n is the number of classes.

y_c is the binary indicator if the class c is the correct prediction

p_c is the predicted probability of class c.

Log is natural logarithm

Focal loss: $FL = (1 - e^{-CE})^{\gamma} \times CE$

Where γ is the focusing parameter (if γ = 0, FL = CE).

The factor with focusing parameter gamma will down-weight easy positives but focus on hard negatives.

For example: γ = 2.

If the correct class is predicted with probability p = 0.99 which indicates an easy positive, the focusing factor is: $(1 - 0.99)^2 = 0.0001$. This makes the loss function down-weighted by 10, 100 times.

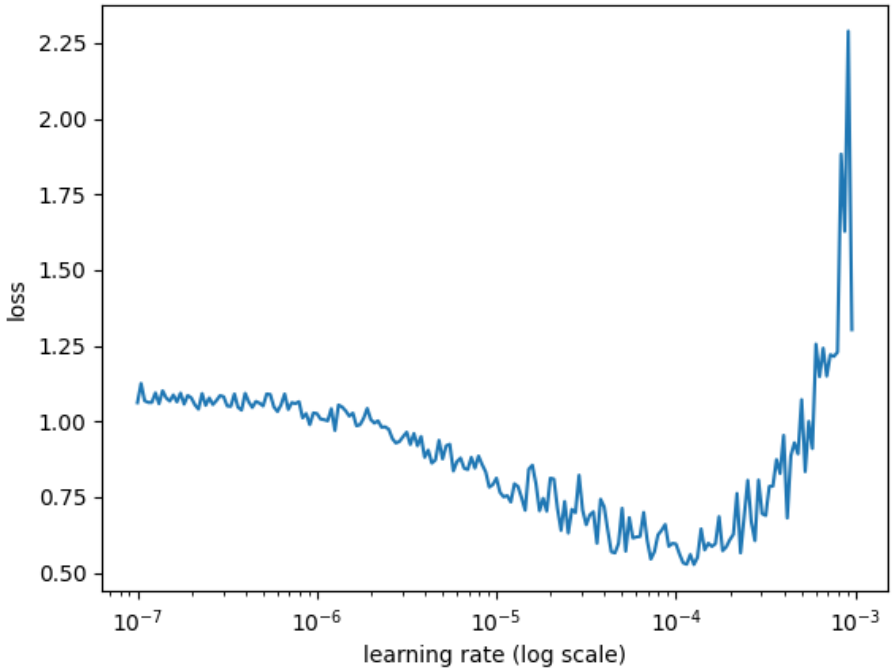
On the other hand, if the correct class is predicted with probability p = 0.01 which indicates a hard negative, the focusing factor is: $(1 - 0.01)^2 = 0.9801$. The decrease is much smaller than the reduction in easy positive case.

Appendix B

Cyclical learning rate method

Figure B

Change of the loss when exponentially increasing the learning rate



An example of cyclical learning rate method is shown in Figure B1. The learning rate is scheduled to increase exponentially from 10^{-7} to 10^{-3}. The loss slowly decreases at the beginning, then accelerates its reduction until reaching 10^{-4}. After that, the loss starts going up and does not converge.

As can be seen from the figure, the loss has the fastest decreasing rate when the learning rate is between 10^{-5} and 10^{-4}. This suggests that the most optimal learning rate would be around this range.

Appendix C

Information and results of the models built in the project

Model	Training set	Trained	Loss	Sensitivity	Specificity
Baseline	original	all layers	CE loss	0.5736	0.9949
Baseline	original	all layers	weighted CE	0.8792	0.9496
Baseline	original	all layers	Focal loss	0.7434	0.9864
Baseline	oversampled	all layers	CE loss	0.7358	0.9705
ResNet34	original	FC layer	CE loss	0.6642	0.9844
ResNet34	original	all layers	CE loss	0.8264	0.9932
ResNet34	augmented	FC layer	weighted CE	0.8868	0.9425
ResNet34	original & augmented	all layers	weighted CE	0.9547	0.9955
ResNet34	augmented	FC layer	focal loss	0.6943	0.9875
ResNet34	original & augmented	all layers	focal loss	0.8981	0.9972
VGG16	augmented	FC layer	weighted CE	0.8377	0.9487
VGG16	original & augmented	all layers	weighted CE	0.9321	0.9963
DenseNet121	augmented	FC layer	weighted CE	0.8604	0.9731
DenseNet121	original & augmented	all layers	weighted CE	0.9396	0.9949

Note. Original indicates the real CXR training images. Augmented means the augmented images creating by transforming the original images.