OOD detection of COVID-19 from Chest X-ray images

Ziyi Xiang **Shuai Wu** **Mahan Tourkaman** **Fang Stålmarck**

Abstract

A new variant of the coronavirus, COVID-19, has plagued the world since December 2019. The virus is particularly dangerous due to its long incubation period and being highly contagious both during and after this period. Due to the shortage of testing kits, developing new testing methods became an urgent task. There exist multiple research about using deep learning methods on radioactive images, X-rays[7][13] and CT[6] to determine whether the patient has the Corona or not. Although there are many mysteries surrounding COVID-19, it is highly probable that outlier could occur in the data set, therefore we propose to train the image recognition system with out-of-distribution(OOD) detection, in order to handle outliers. The usefulness of the proposed OOD method was deemed to be inconclusive, due to both vanilla and OOD classifiers have achieved high accuracy (98%), the reason behind it is thought to be because of the lack of training and testing samples, which outliers were not commonly occurred.

1 Introduction

1.1 Background

COVID-19, the disease caused by the Coronavirus (SARS-CoV-2), originated from Wuhan, China, in December 2019, has since then spread across the entire whole world. The virus is transmitted between humans via droplets and has an incubation period of 2-14 days [1]. During the incubation time, the patient will infect others despite having mild or no symptoms, making this virus particularly dangerous. On March 12th, WHO officially declared the outbreak of COVID-19 a global pandemic [14]. At the time of this writing, there are more than 4 million confirmed cases, and about 280 thousand deaths across the world [10].

The Swedish government took a passive approach against the coronavirus and relies on herd immunity[12], which requires a large portion of the population to get infected by the virus, in order to build up immunity against it. The number of corona patients will thereby be in millions, it is highly improbable that everyone with as little as mild symptoms gets tested with a test kit. Developing alternative testing methods becomes an urgent task.

Although some researchers have deemed X-ray images to give inconclusive result[3], they also show that the number of normal chest X-rays(58.3%) overlap with the patient with mild or medium symptoms(80.3%)[5], meaning X-ray images contains enough information to determine whether a severe patient have COVID-19 or not. Due to the fact that severe patients are considerably more contagious than patients with mild symptoms[5], on top of that, most of the more severe patients already have taken an X-ray image, running a test on the image will cut the cost of doing other redundant tests. Hereby, we propose a testing method that relies on the patient's chest X-ray scan and recommends it to be used on a patient with severe respiratory symptoms.

1.2 Goal

The goal of this project is to build a functioning COVID-19 testing method based on the patient's chest X-ray, by building an image recognition neural network with out-of-distribution detection on training to handle outlier samples. The accuracy need not be compared with state of the art, but rather against the same neural network without outlier handling, determine whether the proposed outlier handling method is useful on COVID-19 detection.

2 Related work

There are multiple papers about distinguishing COVID-19 patients from either healthy individual or patient with other diseases with chest x-ray images using convolution neural networks[7][13], which have shown promising results. The proposed outlier handling method can easily be implemented on other neural network-based detection methods in order to increase their performance if the proposed method was decided to be useful.

Related research was also conducted on chest ct-scans, using deep learning methods, achieving both high sensitivity and high accuracy on COVID-19 classification[6]. Although CT-scan gives better contrast than normal chest X-ray images, making it easier for doctors to identify sickness, it is not necessary for deep learning methods, where similar accuracy is achieved using both X-ray images and CT-scans. On top of that, CT-scans usually have a higher cost, slower imaging time, and gives a higher radiation dose, therefore testing COVID-19 on chest X-ray images is actually more beneficial compared with chest CT scans.

3 Data

3.1 Data-set

In this research, we combined two different data-sets and used 128x128 pixels for training. For COVID data we used an open-source COVID-19 image data collection approved by the University of Montreal's Ethics Committee on the Github[2]. It is the only open sourced X-ray dataset for COVID-19 at the knowledge of the authors. For non-COVID data, we used the Chest X-Ray data-set from NIH[8], which contains X-ray images for 14 diseases and images for healthy bodies.

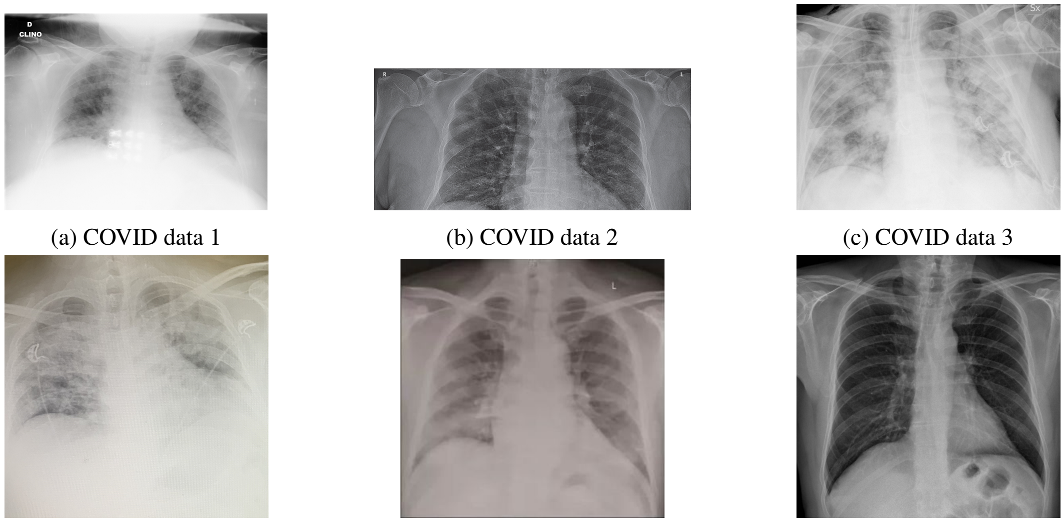
A problem when training our classification emerges, which is the lack of sufficient COVID samples. These data in the medical industry are often heavily protected due to privacy reasons, in our case, very few COVID chest x-ray could be used. In the Github repository mentioned above, the data set only contains around a hundred COVID x-ray images and multiple images may belong to the same patients.

Another problem for this COVID data set is the images are received from different sources, making format, and quality inconsistent. Pictures from different sources may have different sizes, brightness, scaling or even filming angle.1

3.2 Data augmentation

To address these problems, we apply data augmentation on what little data we have. Data augmentation is a technique that generates new samples from the original ones. Studies have shown the effectiveness of extending data set using this approach in image classification through simple techniques like cropping, flipping, scaling, rotating, and color-changing [11].

The technique addressing the problem of lacking sufficient data can also be used to improve the imbalanced population between classes. In our case, we have a data set of three classes: COVID, healthy and other diseases, which we lack sufficient COVID samples. After applying data augmentation, we extend our data set to 3000 images, where COVID data extends to around 800 samples. The samples for testing are completely new and weren't used in training, thus the test accuracy will not affect by data augmentation. The color-changing, flipping, and all sorts of brightness adjustments were applied to all samples, in order to produce new samples, while COVID samples have created some extra data using techniques like cropping, sliding, and rotating which changes the values of image array without modifying the content.

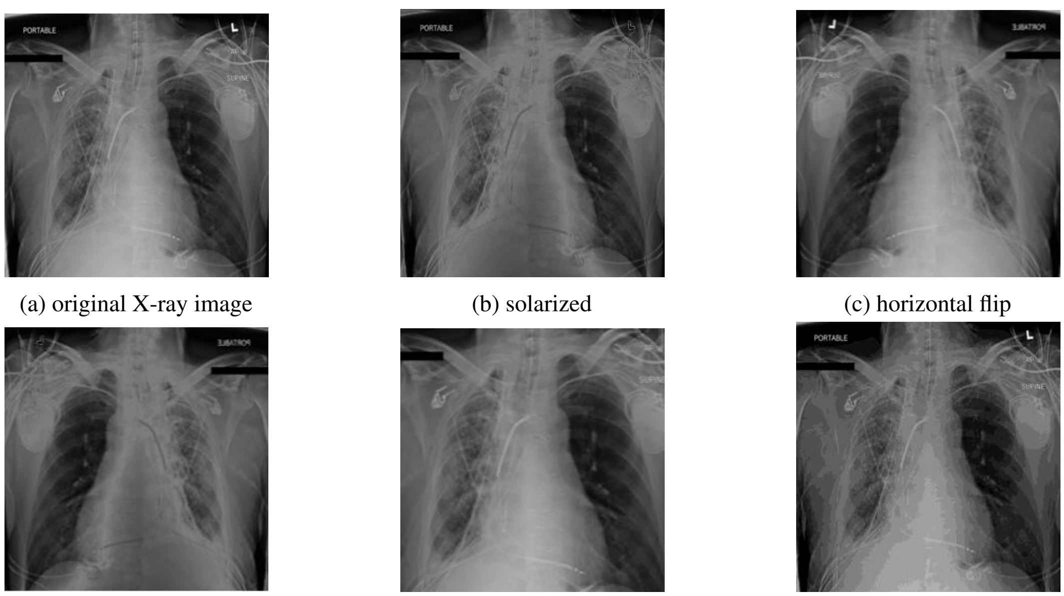


(a) COVID data 1 (b) COVID data 2 (c) COVID data 3

(d) COVID data 4 (e) COVID data 5 (f) COVID data 6

Figure 1: six samples from GitHub repository

Some examples of data augmentation on COVID data2.



(a) original X-ray image (b) solarized (c) horizontal flip

(d) solarized flip (e) cropping (f) posterized

Figure 2: six simples of data augmentation from COVID data-set

4 Method

In this project, we used ResNet as our backbone. ResNet is a convolutional neural network architecture, that has shown promising results on many computer vision tasks.

4.1 Convolutional neural networks architectures

The convolutional neural networks are popularly used for image classification and have specialization for being able to recognize patterns like shapes, edges, and objects[9]. This property of pattern detection makes it efficient in image analysis. The Convolutional layer receives input from the previous layer, transforms it using convolution operation, and forwards to the next layer.

Each of the convolutional layers can have 'filters' which are small square matrices initialized with random numbers. The filter works like a sliding window slide across the input matrix to calculate the dot product to create a new filtered matrix. From a high-level view, each filter may capture a different pattern in the image, such as shape, edges, and object. The deeper layers create more sophisticated 'filters'.

4.1.1 ResNet

In theory, if we stacking deeper layers on a neural network it should performs better but it is not case in practice and we cannot just simply stacking layers together[4]. There is a limit for stacking layers on plain feed-forward networks and researchers found that the performance may decrease after a certain point. The idea of stacking deeper layers to get better performance is like adding more identity layers which the networks should perform at least as well as the smaller one, while in practice all layers learn simultaneously which lead to a decrease of performance.

To address this problem, ResNet utilizes skip connection that jumps over some layers which make the network easier to learn the identity function. Skip connection is an addition of original input data to the output data before feeding into the next layer. By utilizing this feature, the layer can learn at least the identity function, but can also learn more complex relation under the original network architectures. This shortcut concept makes us being able to learn a deeper network in practice.

4.2 OOD detection

The classification generalized well for test samples that are from a similar distribution(i.e., in-distribution). However, there are many unknown or unseen (out-of-distribution) samples in the real world that classifiers are not confident enough to give a high accuracy estimation. The medical branch often faces the problem of anomalous data with unjustified causes. Being able to handle outliers is, therefore, an important trait a classifier should have when doing medical-related classifications.

OOD detection is used to solve the problem of outliers. There exist multiple ways to do out of distribution detection to account for data anomaly, this project adopted a method which utilizes probability decomposition. Unlike the traditional approach, this approach does not require out of distribution samples on training, which solves the main problem out of distribution detection methods face: difficult to obtain out of distribution samples.

By utilize Bayes theorem:

$$p(y_{in}|d_{in}, x) = \frac{p(y_{in}, d_{in}|x)}{p(d_{in}, x)}$$

Yen-Chang Hsu et.al claims that, one could learn the underlying distribution without the use of OOD samples. The details of the algorithm refer to their paper[15].

5 Experiment

This section will show how the experiment is done, as well as what the results are. Two models were used within the experiment: The vanilla ResNet18 and the ResNet18 with OOD detection as outlier handler.

Both classifiers used a constant learning rate of $2 * 10^{-3}$, trained from scratch for 10 epochs . A data set separated from the training set was used as a test set. The test set has 30 samples from each class, normal, nonCOVID-19 disease, COVID-19. The performance(Accuracy) of both classifiers on the test set was evaluated, deciding whether the proposed method is useful or not.

5.1 Result

98% Accuracy on validation set was achieved by both the vanilla version(98.22%) and the modified OOD version(98.37%) of ResNet18.3 We also get 93% and 94.8% on an independent testing set for respective methods. Therefore the effectiveness of the proposed OOD method is deemed to be inconclusive, more about this will be discussed in the next section.

6 Conclusion and discussion

Both of our vanilla ResNet and OOD ResNet converge and archive 98% accuracy on the validation set. The different initialization may explain the variations between accuracy since the training is resource demanding, we could not train the model until converge. Since we do not have access to a larger data-set for testing our result, we are not confident enough to draw any simple conclusion on top of the current data.

The main problem in our project is the lacking of a sufficient amount of data. We understand the difficulty to collect these images due to the privacy policy for sensitive information in hospitals, it is not too mind-boggling that the data set of a newfound virus is small. On top of that, X-ray images in the Git repository have a variety of formats, which can sometimes lead to misclassification. Another potential problem caused by the dataset is, samples in the training set can be biased, either due to human factors in which the doctors misclassified the disease.

The data augmentation part performs surprisingly well which increases the accuracy by 6% on ResNet. Even without the augmentation, we can still get accuracy for over 90% in just 10 epochs which shows the effectiveness of classifying image data using CNN.

As to future researches, one could try to relate data from blood pressure, vital capacity, etc and decide whether the patient has COVID-19 based on it since it demands fewer medical resources and could show more insight about the virus.

References

[1] *Clinical Questions about COVID-19: Questions and Answers*. `https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html#Transmission`. Accessed: 2020-05-12.

[2] *COVID-19 image data collection*. `https://github.com/ieee8023/covid-chestxray-dataset`. Accessed: 2020-05-17.

[3] Ana Echenique. “Chest X-Ray Findings in 636 Ambulatory Patients with COVID-19 Presenting to an Urgent Care Center: A Normal Chest X-Ray Is no Guarantee”. In: (Apr. 2020).

[4] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: `1512.03385 [cs.CV]`.

[5] Yuefei Jin et al. “Virology, Epidemiology, Pathogenesis, and Control of COVID-19”. In: *Viruses* 12 (Mar. 2020), p. 372. DOI: `10.3390/v12040372`.

[6] Lin Li et al. “Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT”. In: *Radiology* (Mar. 2020), p. 200905. DOI: `10.1148/radiol.2020200905`.

[7] Ali Narin, Ceren Kaya, and Ziynet Pamuk. “Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks”. In: (Mar. 2020).

[8] *NIH chest X-rays*. `https://www.kaggle.com/nih-chest-xrays/data`. Accessed: 2020-05-17.

[9] Keiron O’Shea and Ryan Nash. “An Introduction to Convolutional Neural Networks”. In: *ArXiv e-prints* (Nov. 2015).

[10] World Health Organization et al. “Coronavirus disease 2019 (COVID-19): situation report, 112”. In: (2020).

[11] Luis Perez and Jason Wang. “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”. In: *arXiv e-prints*, arXiv:1712.04621 (Dec. 2017), arXiv:1712.04621. arXiv: `1712.04621 [cs.CV]`.

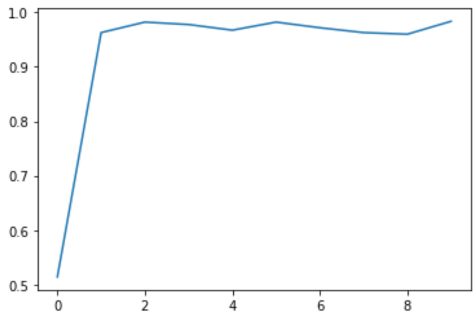
[12] *Tegnell: ”Vi kommer med besked”*. `https://www.svt.se/nyheter/tegnell-vi-kommer-med-besked`. Accessed: 2020-05-12.

[13] Linda Wang and Alexander Wong. “COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images”. In: (Mar. 2020).

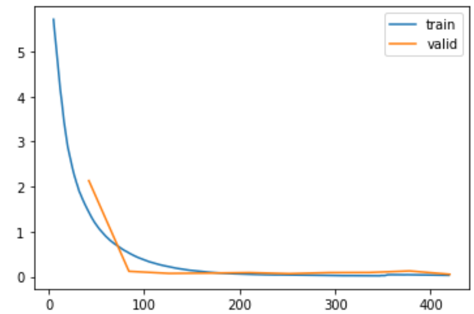
[14] *WHO announces COVID-19 outbreak a pandemic*. `http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic`. Accessed: 2020-05-12.

[15] Hongxia Jin Yen Chang Hsu Yilin Shen and Zsolt Kira. *Learning Out-of-distribution Detection without Out-of-distribution Data*. Sept. 2019.

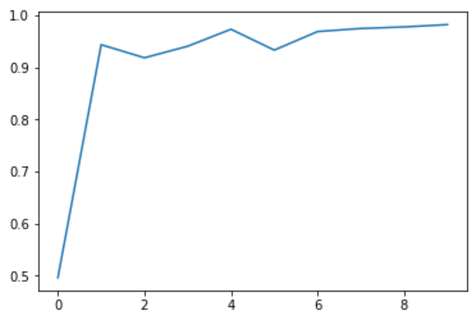
7 Appendix



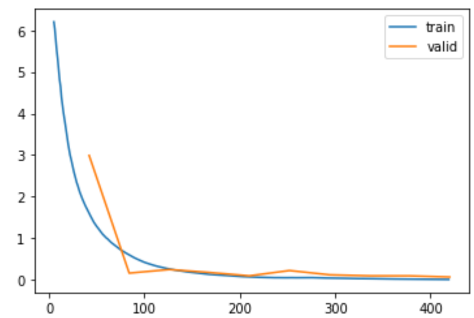
(a) Accuracy for vanilla ResNet18



(b) Loss for vanilla ResNet18



(c) Accuracy for OOD ResNet18



(d) Loss for OOD ResNet18

Figure 3: Loss and validation accuracy during training