

COVID-19 CHEST X-RAY IMAGES: LUNG SEGMENTATION AND DIAGNOSIS USING NEURAL NETWORKS

Alan Zhang

Chinese International School, Hong Kong SAR, China

ABSTRACT

COVID-19 has caused world-wide disturbances and the machine learning community has been finding ways to combat the disease. Applications of neural networks in image processing tasks allow COVID-19 Chest X-ray images to be meaningfully processed. In this study, the V7 Darwin COVID-19 Chest X-ray Dataset is used to train a U-Net based network that performs lung-region segmentation and a convolutional neural network that performs diagnosis on Chest X-ray images. This dataset is larger than most of the datasets used to develop existing COVID-19 related neural networks. The lung segmentation network achieved an accuracy of 0.9697 on the training set and an accuracy of 0.9575, an Intersection-over-union of 0.8666, and a dice coefficient of 0.9273 on the validation set. The diagnosis network achieved an accuracy of 0.9620 on the training set and an accuracy of 0.9666 and AUC of 0.985 on the validation set.

KEYWORDS

COVID-19, Chest X-ray Images, Diagnosis, Lung Segmentation, Neural Networks

1. INTRODUCTION

Coronavirus Disease 2019 (COVID-19) has impacted the world drastically over the past months. Some common symptoms of the disease that have been observed include fever, dry cough, and tiredness [11]. The World Health Organization declared it a pandemic on March 11, 2020 [14]. Presently, there have been more than 22 million confirmed cases of the disease and 780,000 deaths related to the disease around the world [10]. COVID-19 has undoubtedly caused worldwide social and economic disturbances, and more importantly tragic responses among citizens of the world.

During the pandemic, the machine learning community has been finding ways to combat the disease. The machine learning community has progressed rapidly in image processing tasks like segmentation and classification over recent years, including in the field of biomedical imaging. Although COVID-19 medical images are still not available in large quantities, experiments can be done on the available datasets. One method for initial screenings of COVID-19 is using Chest X-rays. There are several benefits of using Chest X-rays for COVID-19 screening [9]. First, Chest X-rays are widely available in healthcare centers because they are common tools that have a wide variety of uses. They are more common than CT scans, especially in less developed areas. Second, Chest X-rays can be performed rapidly on the side with viral tests to determine which patients need the most urgent care. Finally, there are portable Chest X-ray systems that reduce the risk of transmission compared to fixed systems like CT scanners. So far, there have been numerous neural networks that perform segmentation on COVID-19 CT images. However, there

are still relatively few that are designed to perform segmentation on COVID-19 Chest X-ray images.

The author would like to thank V7 Darwin for their efforts in building a large COVID-19 Chest X-ray Dataset that is effectively annotated [1]. This dataset is larger than most of the datasets used previously to develop COVID-19 related neural networks. Using this dataset, a U-Net based network that performs lung-region segmentation and a convolutional neural network that performs diagnosis on Chest X-ray images were trained.

U-Net [2] is the state-of-the-art network for image processing tasks like semantic segmentation. The network contains a contracting path followed by an expanding path. The expanding path is roughly symmetric to the contracting path. This unique U-shaped structure is effective for tasks like semantic segmentation because it first extracts the locations of the key features through the contracting path, and then expands the image back to the original size while keeping those features in place.

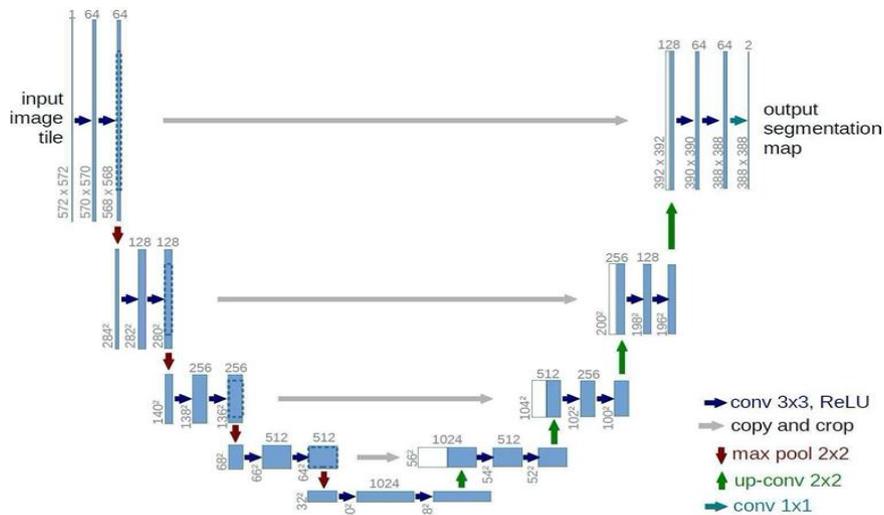


Figure 1. U-Net [2]

Convolutional Neural Networks are some of the most commonly used neural networks for image processing tasks. Their general structure consists of convolutional and pooling layers. The convolutional layers extract key features by performing convolution operations with kernels designed to detect relevant shapes. These kernels are learned during the training process. The pooling layers reduce the size of the image and increase computational efficiency [12]. For classification models, there are usually fully connected layers at the end of the network that categorize the image based on the higher-level features from the final convolutional layer.

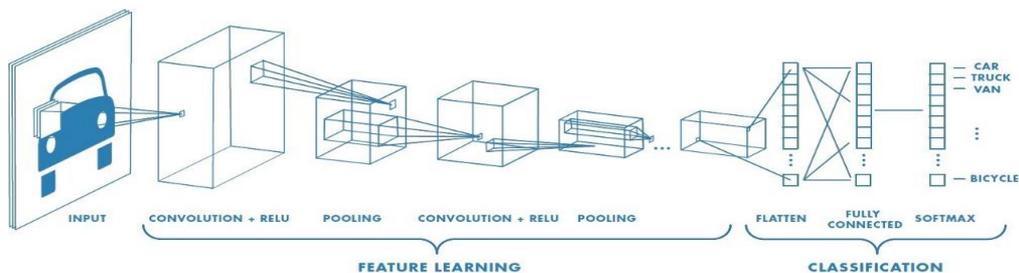


Figure 2. Convolutional Neural Network [12]

The paper is organized in the following structure. Section 2 presents related work and compares them with this study Section 3 presents and explains the dataset used in this study in more detail. Section 4 presents the lung segmentation network by showing the network architecture, explaining the training process, presenting the experimental results, and providing further discussions. Section 5 presents the diagnosis network in the same structure. Finally, section 6 concludes the paper.

2. RELATED WORK

A published review of the existing artificial intelligence techniques to process COVID-19 related medical images [17] is used in this paragraph to compare the performance of some existing classification networks on COVID-19 X-ray images. A total of three studies [18,20,21] that perform classification on X-ray images in the review will be discussed. The studies each use a total of 50 to 70 COVID-19 X-ray images, and those quantities are much fewer than the 439 COVID-19 X-ray images used in this study. Among these studies, two [20,21] achieved lower performance results than the network presented in this study, and the reason may be because more COVID-19 images are available in the V7 Darwin COVID-19 dataset [1]. One of the studies [18] achieved a high accuracy of 98% with a ResNet50 [24] network, which outperforms the network in this study. However, that study only relied on 50 COVID-19 and 50 normal images. The main advantage of the diagnosis network in this study is that it uses more COVID-19 images for both training and validation.

Currently, there are relatively few segmentation networks on COVID-19 X-ray images, and the majority of existing COVID-19 segmentation models are designed to work on CT-images instead. A related study called COVIDX-Net [19] relies on 25 COVID-19 X-ray images and 25 normal images, and achieved f1 scores of 0.89 and 0.91 for their VGG19 [22] and DenseNet [23] models respectively. Again, the segmentation network in this study has an advantage because it uses the newly published V7 Darwin COVID-19 dataset [1] that contains more COVID-19 X-ray images. Since COVID-19 medical images are not widely available at all, the existing neural networks could really benefit from having additional images during training and validation.

3. DATA

The V7 Darwin COVID-19 Chest X-ray Dataset [1] is used in this study. This dataset was chosen because it is one of the largest and most well-organized COVID-19 Chest X-ray datasets that the author found. It contains a total of 6504 RGB images. The images are labeled with tags that indicate whether COVID-19 is present. Among the 6504 images, 521 are labeled with COVID-19 and 5983 are not. The non-COVID images are labeled with other diseases like bacterial pneumonia. Out of the 6504 images, semantic lung masks are provided for 6396 frontal Chest X-ray Images, and among those 439 are labeled with COVID-19 and 5957 are not. The remaining 108 images are either CT scans or lateral Chest X-ray Images that do not contain lung masks.

Status	Number of Images
Total	6504
COVID	521
Non-COVID	5983
Semantic Lung Mask Provided	6396
Semantic Lung Mask Provided & COVID	439
Semantic Lung Mask Provided & Non-COVID	5957
Semantic Lung Mask Missing	108

Table 1. Data Description

The images have varying sizes. In this study, all of the images had their height and width resized initially to 256 pixels to increase computational efficiency.

4. LUNG SEGMENTATION ON COVID-19 CHEST X-RAY IMAGES

In this section, a U-Net [2,8] based network is used to perform lung segmentation on frontal Chest X-ray images. The network takes in an RGB X-ray image of size (256, 256, 3) and outputs a predicted pixel-wise lung mask of size (256, 256, 1). Each pixel has a predicted number between 0 and 1, where 0 represents a pixel without lungs and 1 represents a pixel with lungs. The network achieved an accuracy of 0.9697 on the training set and an accuracy of 0.9575, an Intersection-over-union of 0.8666, and a dice coefficient of 0.9273 on the validation set.

4.1. Network Architecture

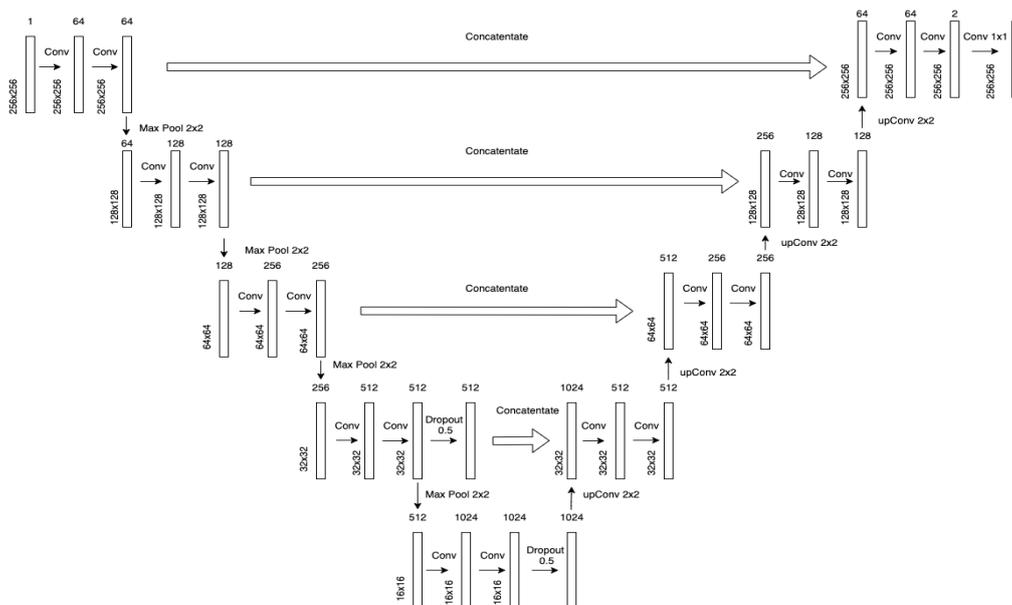


Figure 3. Lung Segmentation Network

The U-Net structure [2,8] is used as the foundation of this lung segmentation network. The input to the network is an RGB X-ray image of size (256, 256, 3) and the output is an image of size (256, 256, 1) that contains the predicted lung mask. Each layer in the contracting path typically follows the structure of 2 convolution operations with a kernel size of (3, 3) and a max pooling operation with a kernel size of (2, 2). Each layer in the expanding path typically follows the pattern of 2 convolution operations with a kernel size of (3, 3), an upsampling operation with a kernel size of (2, 2), and a convolution operation with a kernel size (2, 2). A ReLu activation function follows each of these convolution operations. Note that the padding is set so that the height and width of the image remain constant in each layer. The final convolution operation has a kernel size of (1, 1) and is followed by a sigmoid function that outputs a number between 0 and 1, indicating the probability of that pixel containing lungs.

4.2. Training

First, all the images in the dataset that do not have lung annotations are excluded from this study. In this specific dataset, these include CT images and lateral Chest X-ray Images. As a result, this study uses a total of 6396 Chest X-ray images. Among those images, there are a total of 439 COVID X-ray Images and 5957 non-COVID X-ray Images. The label of each of these images is a pixel-wise semantic mask that contains the lungs. Figure 4 shows an example of a Chest X-ray image and its semantic lung mask from the V7 Darwin COVID-19 dataset [1].

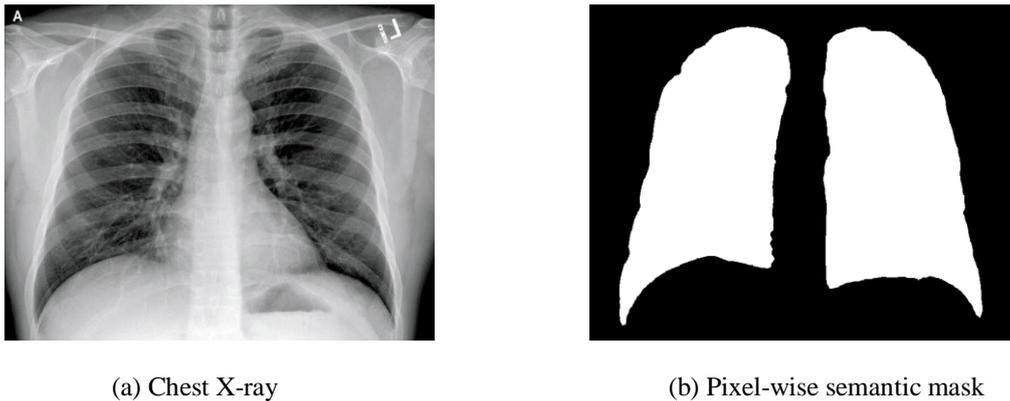


Figure 4. Example image in the dataset [1]

After the images have been reshaped to size (256, 256, 3), some masks were blurred and contained pixels with values between 0 and 1. In order to address this issue, a threshold of 0.5 is adapted so that all pixels that have a value of 0.5 or below are set to 0 and the rest are set to 1.

The training set consists of 70% of the 5957 non-COVID X-ray images and 70% of the 439 COVID X-ray images. The validation set consists of the remaining 30% of COVID X-ray images. Therefore, there are a total of 4477 images in the training set and 132 images in the validation set.

Status	Number of Images
Non-COVID	4170
COVID	307

Table 2. Training set of Lung Segmentation Network

Status	Number of Images
Non-COVID	0
COVID	132

Table 3. Validation set of Lung Segmentation Network

The network is trained with TensorFlow 2.3.0 [6] and Keras 2.4.3 [7]. It is compiled with the adam optimizer and the binary cross-entropy loss function. The network was trained for 15 epochs.

4.3. Experimental Results

The network received an accuracy of 0.9697 on the training set and 0.9575 on the validation set.

Figure 5 shows examples of the segmentation results of COVID Chest X-ray image from the V7 Darwin COVID-19 Dataset [1]. Since the predicted mask has a pixel-wise prediction between 0 and 1, a threshold of 0.2 is applied to the predicted mask so that all pixels less than 0.2 are classified as 0 and the rest as 1.

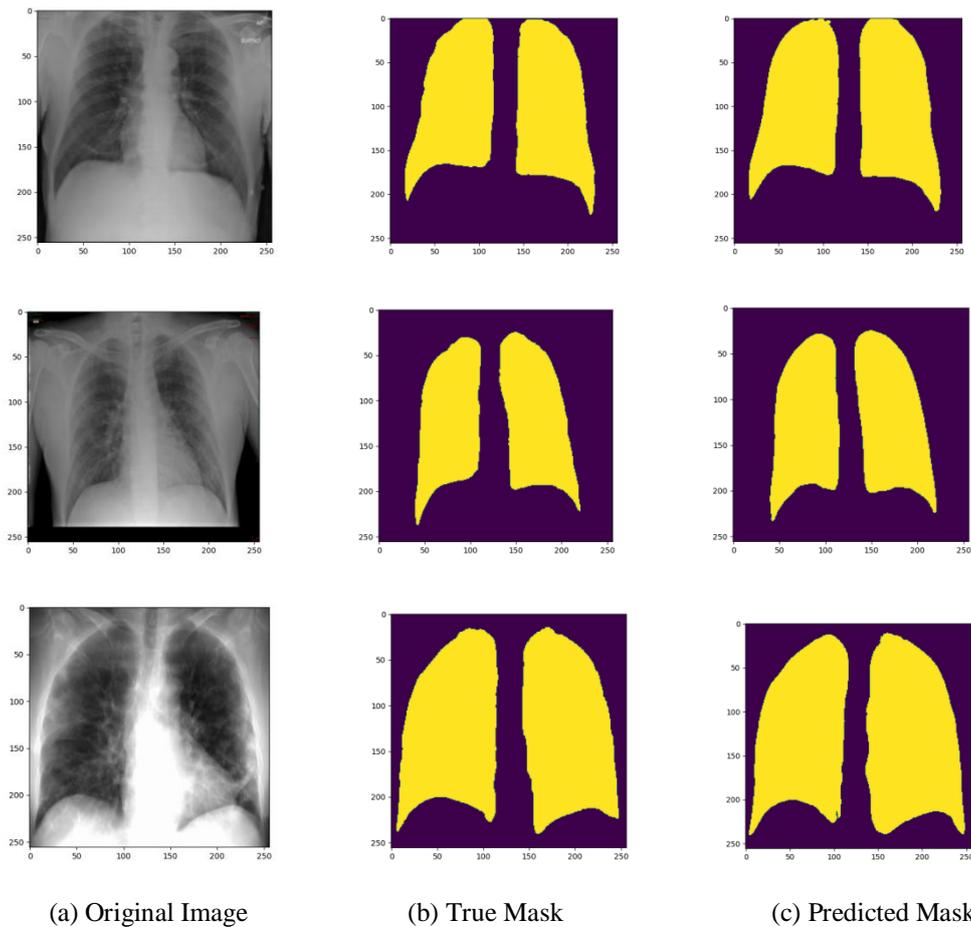


Figure 5. Segmentation results of COVID Chest X-ray images [1]

When the threshold of 0.2 is applied to the predicted mask, an Intersection-over-union of 0.8666 and a dice coefficient of 0.9273 are achieved on the validation set [3]. These statistics are generated by the procedure developed in [3].

4.4. Discussion

Overall, the U-Net performs well for the lung segmentation task. The final threshold of 0.2 is chosen because it yields the best results upon observations, and different thresholds can be experimented with further.

A training session prior to the current one used a training set that consists of 500 non-COVID X-ray images and 70% of the 439 COVID X-ray images and a validation set of the remaining 30% of COVID X-ray images. Note that the images were randomly shuffled before each training session. With that data, the network actually achieved a decent accuracy of 0.9538 on the training set and 0.9114 on the validation set. This result means that increasing the number of non-COVID X-ray images during the training process actually does not improve the performance of the network by a large margin.

Finally, the network discussed in this section is definitely not ready for clinical usages and would largely benefit from having more COVID Chest X-ray Images during training.

5. DIAGNOSIS ON COVID-19 CHEST X-RAY IMAGES

In this section, a convolutional neural network is developed to determine whether COVID-19 is present from a frontal Chest X-ray image. The network takes in an RGB X-ray image of size (246, 236, 3) and outputs a number between 0 and 1, where 1 represents the presence of COVID-19. The network achieved an accuracy of 0.9620 on the training set and an accuracy of 0.9666 and AUC of 0.985 on the validation set.

5.1. Network Architecture

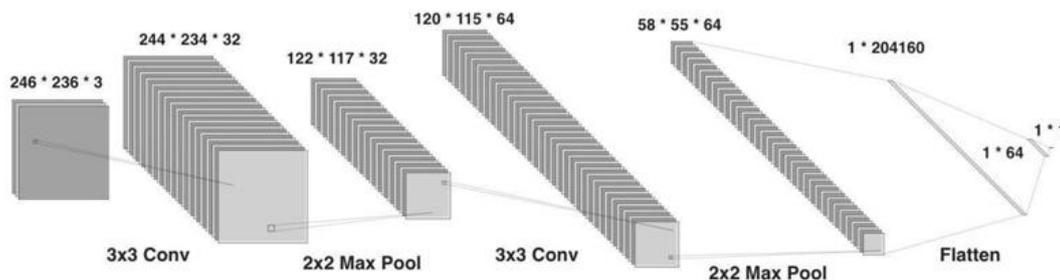


Figure 6. Diagram of the CNN

A classic convolutional neural network is used for the binary classification task. Figure 6 is generated with an online tool developed by Alex Lenail [4]. The input to the network is an RGB X-ray image of size (246, 236, 3) and the output is a number between 0 and 1. The network contains 2 convolution layers with convolutional operations of kernel size (3,3), 2 max pool layers with a kernel size of (2,2), 1 flattening layer, and 1 dense layer of size 64 and the final dense layer of size 1. The convolutional layers and the second to last dense layer are all followed by a ReLu activation function. The final dense layer is followed by a sigmoid function that outputs a number between 0 and 1.

5.2. Training

The same 6393 Chest X-ray Images are used to train this network as the previous network. The labels are no longer the semantic masks, but simply the number 0 or 1, where 1 represents the presence of COVID-19.

Before training or validation, all of the images are cropped so that the topmost 10 pixels, the leftmost 10 pixels, and the rightmost 10 pixels are removed. The reason for doing this is because a large number of non-COVID X-ray images in this dataset have the letter ‘R’ near the top or left region while some COVID X-ray images in this dataset tend to have the letter ‘L’ near the top or right region. If the letters are kept, the network would detect them as a highly relevant feature during the classification process, which would be unfavorable. Following the cropping procedure, the shapes of the images change from (256, 256, 3) to (246, 236, 3). Figure 7 shows an example of the cropping procedure on an X-ray image from the dataset [1].

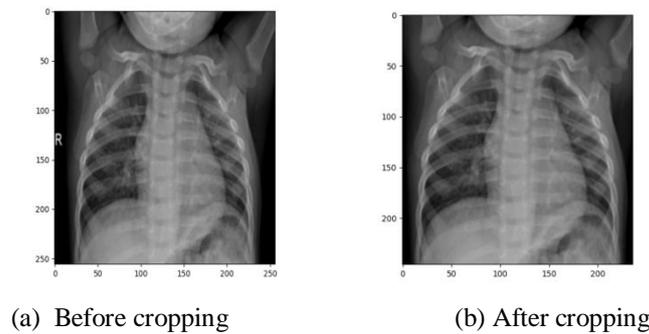


Figure 7. Example of image cropping [1]

The training set consists of 70% of the 5957 non-COVID X-ray images and 70% of the 439 COVID X-ray images. The validation set consists of the remaining 30% from both classes. Therefore, there are a total of 4477 images in the training set and 1919 images in the validation set.

Status	Number of Images
Non-COVID	4170
COVID	307

Table 4. Training set of Diagnosis Network

Status	Number of Images
Non-COVID	1787
COVID	132

Table 5. Validation set of Diagnosis Network

The network is trained with TensorFlow 2.3.0 [6] and Keras 2.4.3 [7]. It is compiled with the adam optimizer and the binary cross-entropy loss function. The network was trained for 2 epochs

and stopped due to the early stopping condition of an increase in validation loss over the patience of 1 epoch.

5.3. Experimental Results

The network received an accuracy of 0.9620 on the training set and 0.9666 on the validation set. The Receiver Operating Characteristic (ROC) curve is plotted for the validation set and the Area under the ROC curve (AUC) is 0.985 [5]. The X-Axis of the ROC curve is the False Positive Rate (FPR) and the Y-Axis of the ROC curve is the True Positive Rate (TPR) [16].

The formula to calculate FPR is using the number of true positives to divide the total number of true positives and false negatives [16]. In the context of this study, true positive is the case when a COVID image is correctly diagnosed and false negative is the case when a non-COVID image is falsely diagnosed with COVID [15].

$$\frac{TP}{TP + FN}$$

The formula to calculate TPR is using the number of false positives to divide the total number of false positives and true negatives [16]. In the context of this study, false positive is the case when a COVID image is falsely predicted to be non-COVID and true negative is the case when a non-COVID image is correctly predicted to be non-COVID [15].

$$\frac{FP}{FP + FN}$$

The ROC curve spans over different classification thresholds and plots the values of FPR and TPR [16]. Figure 8 shows the ROC curve as well as its corresponding AUC, both generated by the procedure developed in [5]. The network is successful because it achieved a satisfactory result of an AUC of 0.985.

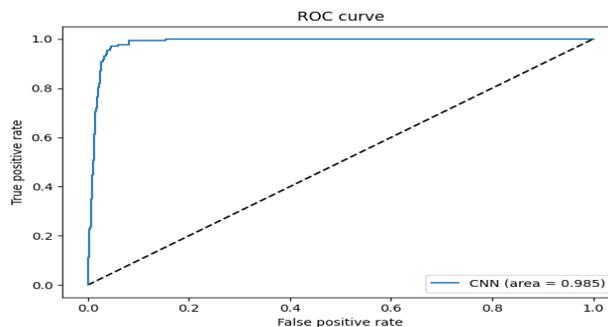


Figure 8. ROC of the convolutional neural network

5.4. Discussion

While the experiment results show excellent performance, there are many limitations to the network discussed in this section.

Although the images are cropped to remove some of the letters or labels, there are still many images in the dataset that have different kinds of markings. These markings should not be features that the network detects during the classification process. Figure 9 shows examples of X-ray images from the dataset [1] that contains abnormal markings.

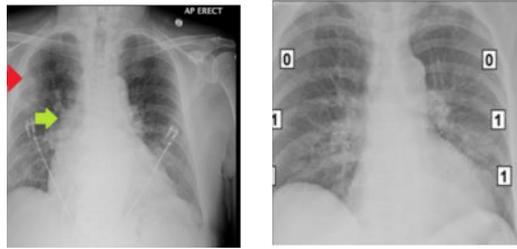


Figure 9. Examples of images that have markings [1]

A training session prior to the current one used a training set that consists of X-ray Images that have been cropped so that the leftmost and topmost 20 pixels were removed. With that data, the network achieved an accuracy of 0.9723 on the training set and 0.9666 on the validation set. However, the higher training set accuracy might be attributed to the network recognizing the letter 'L' that appears on the right side of many COVID Chest X-ray images. More methods of removing these bias sources should be experimented with, one of which is mentioned in the conclusion.

Finally, although this dataset is already one of the largest and most well organized, it still contains a limited number of COVID Chest X-ray Images. Therefore, the network discussed in this section is definitely not ready for clinical usages yet. The machine learning community needs to acquire more labeled images to develop better COVID-19 related networks.

6. CONCLUSION

In this study, the V7 Darwin COVID-19 Chest X-ray Dataset [1] is used to train a U-Net based network that performs lung-region segmentation and a convolutional neural network that performs diagnosis on Chest X-ray images. This dataset is larger than most of the datasets used to develop existing COVID-19 related neural networks. The U-Net achieved an accuracy of 0.9575, an Intersection-over-union of 0.8666, and a dice coefficient of 0.9273 on the validation set. The convolutional neural network achieved an accuracy of 0.9666 and AUC of 0.985 on the validation set.

There is one future extension that could be built upon this study. The predicted lung masks from the lung segmentation network could actually be applied to the X-ray images before they are then diagnosed. This way, only the lung regions would be considered during diagnosis. This would be very beneficial because other parts of the X-ray images such as markings and letters would not be considered. This method is mentioned in a study [13] which explains how to remove bias sources from available COVID-19 X-ray images.

The networks discussed in this study demonstrate satisfactory performances. They can be useful for making comparisons with future studies. As more COVID-19 data becomes available, the machine learning community can continue to develop better and more reliable networks to combat the pandemic. Both of the trained networks in this study can be downloaded at <https://www.dropbox.com/sh/561b9fmgxyw5iae/AABGiPgs8nwXn7TETbliNCiqa?dl=0>. It should be noted again that these networks were trained with TensorFlow 2.3.0 [6] and Keras 2.4.3 [7].

ACKNOWLEDGMENT

The author would like to thank Dr. Chong Chen for providing the hardware to train the lung segmentation network in this study.

REFERENCES

- [1] [Online]. Available: <https://darwin.v7labs.com/v7-labs/covid-19-chest-x-ray-dataset>.
- [2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, pp. 234–241, 2015.
- [3] E. Tiu, "Metrics to Evaluate your Semantic Segmentation Model," *Medium*, 10-Aug-2019. [Online]. Available: <https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2>.
- [4] "NN-SVG," NN SVG. [Online]. Available: <http://alexlenail.me/NN-SVG/LeNet.html>.
- [5] Chengwei, "Simple guide on how to generate ROC plot for Keras classifier," *DLology*. [Online]. Available: <https://www.dlology.com/blog/simple-guide-on-how-to-generate-roc-plot-for-keras-classifier/>.
- [6] TensorFlow. [Online]. Available: <https://www.tensorflow.org/>.
- [7] Keras. [Online]. Available: <https://keras.io/>.
- [8] Zhixuhao, GitHub, 27-Nov-2018. [Online]. Available: <https://github.com/zhixuhao/unet/blob/master/model.py>.
- [9] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," *arXiv preprint arXiv:2003.09871*, Mar. 2020.
- [10] "COVID-19 Map," Johns Hopkins Coronavirus Resource Center. [Online]. Available: <https://coronavirus.jhu.edu/map.html>.
- [11] "Q&A on coronaviruses (COVID-19)," World Health Organization. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses>.
- [12] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks-the ELI5 way," *Medium*, 16-Dec-2018. [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- [13] E. Tartaglione, C. A. Barbano, C. Berzovini, M. Calandri, and M. Grangetto, "Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data," *arXiv:2004.05405*.
- [14] "WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020," World Health Organization, 11-Mar-2020. [Online]. Available: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
- [15] "Classification: True vs. False and Positive vs. Negative." [Online]. Available: <https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative>.
- [16] "Classification: ROC Curve and AUC | Machine Learning Crash Course," Google. [Online]. Available: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>.
- [17] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19," *IEEE Reviews in Biomedical Engineering*, pp. 1–1, Apr. 2020.
- [18] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *arXiv:2003.10849*, 2020.
- [19] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," *arXiv:2003.11055*, 2020.
- [20] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection," *arXiv:2003.10769*, 2020.
- [21] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "COVID-19 screening on Chest X-ray images using deep learning based anomaly detection," *arXiv:2003.12338*, 2020.
- [22] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556*, 2014.
- [23] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.