COVID-19 and Pneumonia Detection using Computed Tomography Scans with YOLO3 Darknet-53

Youngho Kim[1], Daekyung Kim[2], Jey Kang[3], Byungkon Kang[4], and Deepak Pradhan[5]

I. INTRODUCTION

66,418,437, which is more than South Korea's population, is the total number of worldwide coronavirus cases as of December 5^{th}, 2020. 1,530,642 people have died, and 46,065,730 have recovered after being infected by the virus. The number of active coronavirus patients is still 19,001,886, and 106,102 have progressed to be at a severe state [10].

The coronavirus disease (COVID-19) originated from Wuhan, China in December 2019. Previous cases of coronavirus outbreaks include Middle East respiratory syndrome (MERS-CoV) and severe acute respiratory syndrome (SARS-CoV). Past examples of automated chest radiography (X-ray) diagnoses are not as easily accessible as current cases due to the comparatively little attention the outbreaks received, and the subsequent lack of interest in the radiography data.

Recently, many hospitals have begun to realize the necessity of automated COVID-19 detection due to the rapid growth in the number of COVID-19 patients. Radiography physicians and doctors were able to differentiate between pneumonia-infected and healthy lungs through patients' chest X-ray images and computed tomography (CT) scans. Now, however, doctors have to recognize the differences between COVID-19 and regular pneumonia from X-ray images and CT scans. The problem here is that the symptoms shown in the X-ray and CT scans of COVID-19 and pneumonia are similar enough that a regular doctor cannot easily differentiate the two. Specialized radiologists can diagnose COVID-19 more easily, but there are far fewer specialists than regular doctors [1].

There are two solutions to this problem. The first solution is providing a machine learning classification model to do the differentiation. With less work and involvement from the doctors, the process of accurately diagnosing the patient will become faster. In the cases where the machine learning model fails to provide high confidence in diagnosing, a possible secondary solution would be to use CT scans instead of X-ray images. Using CT scans may provide higher accuracy in classifying COVID-19 compared to X-ray images due to the higher resolution of the produced scans [4]. It is also important to note that doctors find it difficult to recognize COVID-19 from visual inspection of CT scans, which is why machine learning is needed to corroborate the diagnosis.

II. BACKGROUND

A. Chest Radiography (X-ray) and Computed Tomography (CT) datasets

Chest X-ray image represents whether the patient is pneumonia-infected or normal by looking at the haziness and impurities around the lung area. Chest X-ray images, however, depend on the X-ray machines, environments, and distance between the machines and a patient. It often gives haziness even if the patient seems normal. A computed tomography (CT) scan of a human chest represents better features and results than those of an X-ray image. A CT scan can produce a tomographic image of a human body by taking multiple X-ray measurements from different angles.

B. Convolutional Neural Network (CNN)

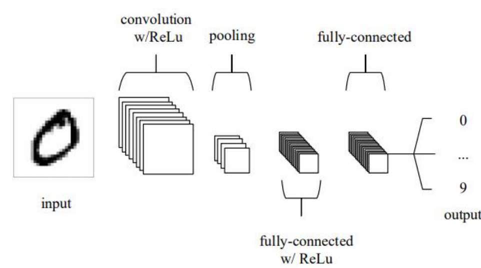

Fig. 1. An simple CNN architecture, comprised of just five layers [9]

A convolutional neural network (CNN) is one of the machine learning techniques used in image recognition and detection problems [7]. CNN works as follows; the images provided as input will be converted into a format that can be used in the neural network. To calculate these images, in the neural network, the images must be converted into matrix format. By reading the matrix, the system will determine which image belongs to which class by differentiating the images. Each layer in the network learns small differences of the labeled images during the training session, and the model makes predictions on the new images.

C. Transfer Learning

One of the biggest difficulties faced by software developers is the limited number of datasets. Deep learning models need a lot of data to be trained, but the data of the novel COVID-19 are limited because it suddenly happened last year. To train a new model without any pre-trained models, labeling this data is costly and time-consuming. By transfer learning, a model can effectively learn much more data faster and accurately from a pre-trained model [7].

III. EXISTING MODELS

A. *Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multi-centre study*

A total of 1485 CT scan images from 495 patients (one coronal, axial, and sagittal slice from each patient) were used. Out of these, 395 patients (294 COVID-19, 101 non-COVID pneumonia) were used for training, and the remaining 100 patients were divided in half for validation and testing. There were no regular patients (non-COVID-19 or pneumonia) used in the dataset. The CT scan slices were chosen by finding the slice with the maximum lung area through segmentation. Images (originally 512x512px) were resized to 256 x 256px [2].

The network used was based on the Resnet50 architecture, with each type of CT scan (coronal/axial/sagittal) used as input instead of the separated R, G, and B channels of an image, which are typically used for training Resnet-based networks. Each image was run through a Res block, then concatenated into a Dense layer.

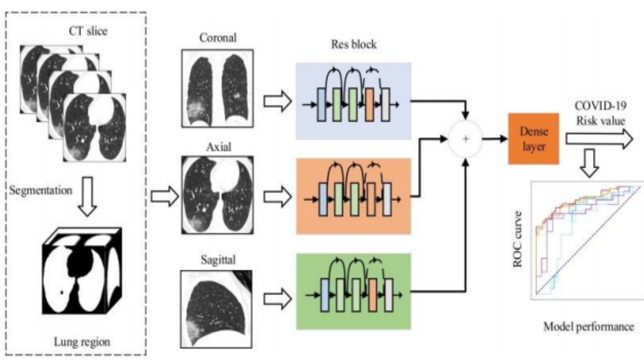

Fig. 2. Proposed structure and input for the multi-view neural network

The main flaw with this network was its inability to generate results at a reasonable accuracy. The authors tested both a model trained with regular axial slices of CT scans and the special tri-image network; the former achieved an accuracy of 0.62 on the testing set and the latter achieved 0.76. Interestingly, the triple-input model (dubbed "multi-view" in the paper) was designed to output risk scores rather than a classification. The authors set an arbitrary value (0.653) as the cut-off for the risk score [2]. This may be of more use than a simple classification for use as a supplementary aid during a visual inspection of CT scans by a doctor.

B. *Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods*

This study performs automatic coronavirus detection in two similar stages. The first stage did not use a feature extractor to train the model. The four subsets were just transformed into vectors and were classified by the support vector machine (SVM). In the second stage, five different feature extraction processes were used. They are Grey Level Co-occurrence Matrix (GLCM), Local Directional Patterns (LDP), Grey Level Run Length Matrix (GLRLM), Grey Level Size Zone Matrix (GLSZM), and Discrete Wavelet Transform (DWT). These methods extract the features from the image to improve the accuracy of the classification. During the classification process, cross-validation methods were used (2-fold, 5-fold, and 10-fold) [8].

The dataset they used contains 150 computed tomography (CT) images from the Societa Italiana di Radiologia Medica e Interventistica. The images in the dataset were from different CT tools. This situation can lead to difficulties in the classification process because grey-levels in CT images sometimes represent an infected area in one CT tool, and grey-levels can also represent non-infected areas in other CT tools. To solve this problem, they divided it into four different subsets and divided the patch dimensions into 16x16, 32x32, 48x48, and 64x64. Patch regions represent the lung area from the CT scans [8]. Subset 1 contains 5912 non-coronavirus patches and 6940 coronavirus patches, subset 2 contains 942 non-coronavirus patches, and 1122 coronavirus patches, subset 3 contains 255 non-coronavirus patches and 306 coronavirus patches, and subset 4 contains 76 non-coronavirus patches and 107 coronavirus patches.

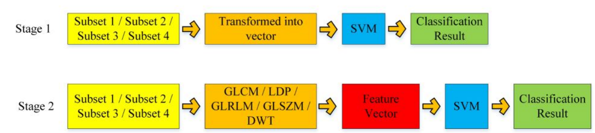

Fig. 3. The classification process for Stage 1 and Stage 2 [8]

As a result, Five different evaluation methods were used: Sensitivity, Specificity, Accuracy, Precision, and F1-score. Subset 1 had 5912 non-infected and 6940 infected patches, and they were classified by Stage 1 and Stage 2. The best result was obtained as 99.68 percent in Stage 2 with a 10-fold cross-validation method and with Gray Level Size Zone Matrix extraction process [8]. Subset 2 had 942 non-infected and 1122 infected patches, and the best classification result was obtained as 99.37 percent in Stage 2 with 10-fold cross-validation and with the Discrete Wavelet Transform extraction process. Subset 3 had 255 non-infected and 306 infected patches, and the best accuracy was 99.67 percent in Stage 2 with 10-fold cross-validation and Discrete Wavelet Transform extraction process. Finally, with 76 non-infected and 107 infected patches, subset 4 had the best accuracy of 97.28 percent in Stage 2 with 10-fold cross-validation and with the Discrete Wavelet Transform extraction process.

IV. DATASET

A. *"CT Scans for COVID-19 Classification"*

For the binary classification model and the second model of workflow, a dataset from 2 hospitals, Union Hospital (HUST-UH) and Liyuan hospital (HUST-LH), was used to train the model. There are 5705 non-informative CT images, 4001 Positive COVID-19 CT images, and 9979 Negative COVID-19 CT images. The author of this dataset also uploaded the pre-processed version of each image, using threshold and binary image processing.

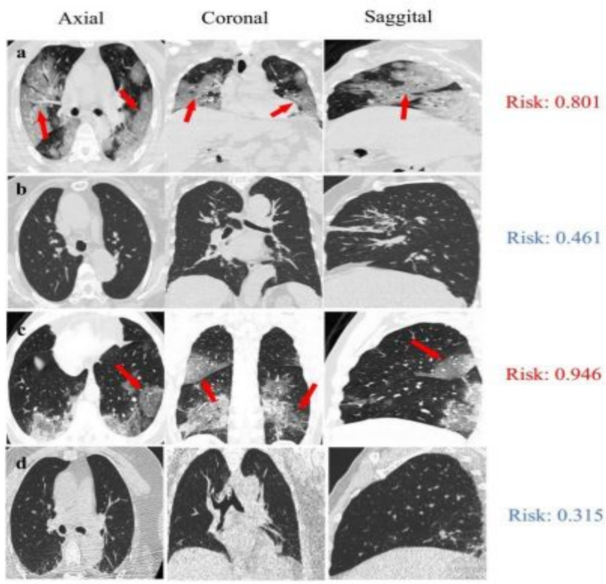

Fig. 4. Example of risk values output by the multi-view model. Risk values higher than the cutoff are labeled red.

Dataset URL: https://www.kaggle.com/azaemon

B. "CT Images for COVID NORMAL PNEUMONIA Mendeley"

In order to train enough Pneumonia CT images to the pre-trained model, we used 2105 pneumonia CT images and additionally used transfer learning method to train more COVID and Normal dataset: 1681 COVID CT images and 2103 Normal CT images.

Dataset URL: https://www.kaggle.com/anaselmasry

V. MODEL CHOICE

From four different types of network types of YOLO: Darknet-19, ResNet-101, ResNet-152, and Darknet-53, Darknet-53 was chosen because of its high confidence and fast speed. Darknet-53 also achieves the highest measured floating-point operations per second, meaning that the GPU is better utilized by the network structure. This means that it made the structure more efficient and faster than ResNet-101 and ResNet-152, which have just way too many layers and are not very efficient [3]. Additionally, Darknet-53 is based on CNN.

Backbone	Top-1	Top-5	Bn Ops	BFLOP/s	FPS
Darknet-19 [15]	74.1	91.8	7.29	1246	**171**
ResNet-101[5]	77.1	93.7	19.7	1039	53
ResNet-152 [5]	**77.6**	**93.8**	29.4	1090	37
Darknet-53	77.2	**93.8**	18.7	**1457**	78

Fig. 5. Comparison of backbones. Accuracy, billions of operations, billion floating point operations per second, and FPS for various networks [3].

Interestingly, when CT-scan pre-trained models were used for transfer learning in this model, they all failed to have higher accuracy than using COCO 80 class dataset for the transfer learning. Hence, we realized that there were not enough COVID-19 CT dataset to train an empty model. By deciding to use the pre-trained coco 80 class dataset model, it had better confidence and F1-score when detecting the three classes (pneumonia, COVID-19, and normal) from a CT-scan than other pre-trained models.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
1×	Convolutional	64	3 × 3	
1×	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
2×	Convolutional	128	3 × 3	
2×	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
8×	Convolutional	256	3 × 3	
8×	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
8×	Convolutional	512	3 × 3	
8×	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
4×	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 6. Darknet-53

VI. DATA PRE-PROCESSING

A. LabelImg

CT-scans are not always as clear as expected. Various hospitals have different settings, CT-scanning machines, and environments. To reduce the complexity of the training images, the LabelImg tool was used to specify the clarity of an input image that the model shall learn from.

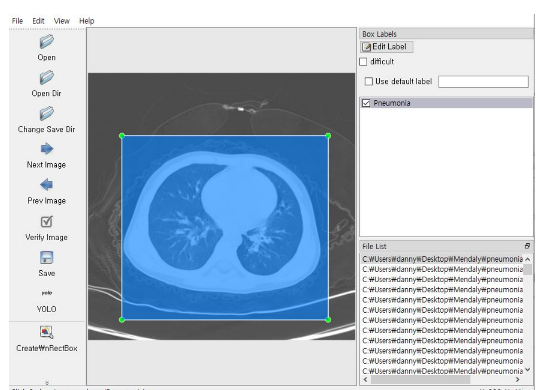

Fig. 7. LabelImg usage

When a set of CT-scans are well labeled as shown in Fig. 5, a .txt file with the position of the specific rectangle box drawn on that image will be created. Both txt and jpg (or png) files are used to train Darknet-53 model. When Darknet-53 is used in YOLO framework, then these txt files are required to specify the labels of each input image.

Please refer to: https://github.com/tzutalin/labelImg.

B. Segmentation

Thanks to Wanshan Ning, the developer of HUST-19, CT-scan segmentation was successfully done with his help.

In a CT-scan of a chest area, we not only see the lungs but also see the diaphragm as well. Some patients may have cracks or impurities on their diaphragm, causing the model to learn also about the diaphragm. To prevent the model from learning some false information of CT-scans, segmentation was added to the pre-processing part to extract only the lung area from a CT-scan.

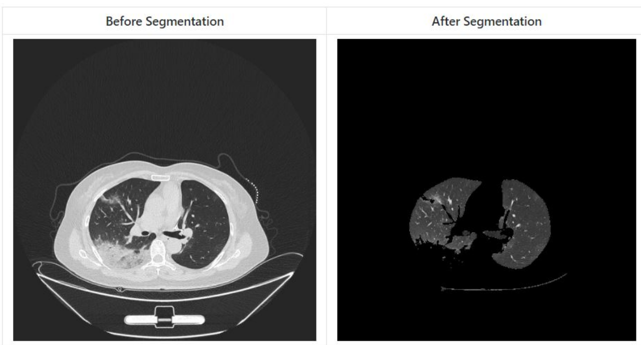

Fig. 8. Lung parenchyma extraction using segmentation

For further information about the segmentation and its code, please refer to: https://github.com/rxYoungho.

VII. BINARY CLASSIFICATION MODEL

Binary classification models often provide better accuracy and F1-score than multiple classification models with more than two classes. To test if Darknet-53 is the right network model to classify the COVID-19 from a CT-scan, we first trained the model with only two classes: COVID-19 and Normal patients.

Fig. 9. Detection result of binary classification model

The model was trained with 100 representative chest CT-scans for each COVID-19 and Normal class. When tested using cross-validation of 100 images out of 1000 images, the F-1 score was about 0.95 when detecting the Normal CT-scan. The F-1 score of classifying the COVID-19 case was about 0.99. One of the reasons for their high accuracy is the segmentation method used both in input data and trained data.

COVID-19 Detection

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	1.00	0.91	0.95	97
accuracy			0.91	97
macro avg	0.50	0.45	0.48	97
weighted avg	1.00	0.91	0.95	97

Normal Detection

	precision	recall	f1-score	support
0	1.00	0.98	0.99	99
1	0.00	0.00	0.00	0
accuracy			0.98	99
macro avg	0.50	0.49	0.49	99
weighted avg	1.00	0.98	0.99	99

Fig. 10. Binary Classification F1-score

Even though it had great accuracy in cross-validation, the model had a critically low F1-score when detecting a set of different chest CT-scans from different hospitals. The model was tested by 12939 CT-scans.

Unseen Data [F1-Score of both Normal and COVID-19 class]

	precision	recall	f1-score	support
0	0.75	0.80	0.77	9305
1	0.40	0.35	0.37	3674
accuracy			0.67	12979
macro avg	0.58	0.57	0.57	12979
weighted avg	0.65	0.67	0.66	12979

Fig. 11. Binary Classification F1-score of unseen CT-scan data

To improve the accuracy of this model, various chest CT-scans from different hospitals are needed.

VIII. MULTIPLE CLASSIFICATION MODEL

The most challenging part of this model is differentiating pneumonia and COVID-19 cases. Both have cracks in the lung area and contain water in the lung area. To solve this problem, we have created a workflow to test at most twice for one image. Firstly, when the model gets one raw input image (without any segmentation method), it classifies what the image is. In a low probability, the model will fail to differentiate or give critically low confidence (less than 0.70) in both COVID-19 and pneumonia cases.

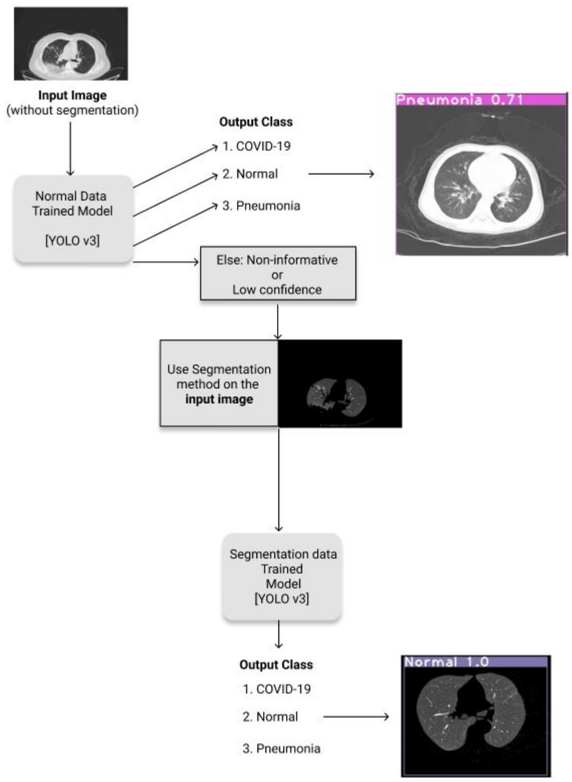

Fig. 12. COVID-19 CT Scan Workflow using two models

Then, the model will send the non-informative or low-confidence image to a different model that is trained with segmented training images. From the non-informative image that is sent from the previous model, the lung parenchyma will be extracted by using the segmentation method used in the latter model.

The final F1-score of this workflow was about 0.98 and an accuracy of 0.96 when detecting the COVID-19 case.

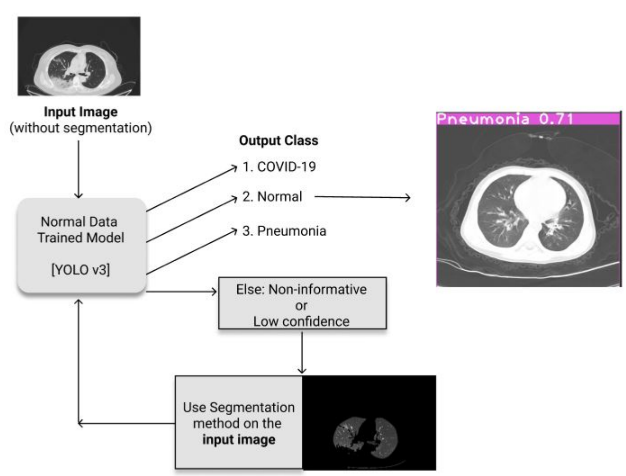

Fig. 13. COVID-19 CT Scan Workflow using one model

As shown in Fig. 13. we also made a workflow using one model, but it failed to give appropriate accuracy and detection confidence because the first model that gets the raw input data was only trained by the segmented train dataset. There was no difference in using one model workflow and using just one model.

COVID-19 detection F1-Score using Multiple Classification model

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	1.00	0.96	0.98	1666
accuracy			0.96	1666
macro avg	0.50	0.48	0.49	1666
weighted avg	1.00	0.96	0.98	1666

Fig. 14. F1-Score of COVID-19 detection case using two models workflow

IX. CONCLUSION

The addition of the segmentation method caused a better detecting accuracy than the other existing models mentioned above. As more CT-scans and COVID-19 datasets are collected, we hope to reinforce this model so that it really can be used as a supporter for the doctors. The coronavirus confirmed case in one day in South Korea is about 600 on average. If this model can help doctors and radiologists to detect both the COVID-19 and Pneumonia, we hope this model to be used in medical usage as soon as possible.

REFERENCES

[1] Al-Karawi, Dhurgham, et al. "Machine Learning Analysis of Chest CT Scan Images as a Complementary Digital Test of Coronavirus (COVID-19) Patients." medRxiv (2020).
[2] Wu, Xiangjun et al. “Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study.” European journal of radiology vol. 128 (2020): 109041. doi:10.1016/j.ejrad.2020.109041.
[3] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).
[4] Che Azemin, Mohd Zulfaezal, et al. “COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings.” International Journal of Biomedical Imaging, 18 Aug. 2020, doi:https://doi.org/10.1155/2020/8828855.
[5] Wang H, Wang L, Lee EH, Zheng J, Zhang W, Halabi S, Liu C, Deng K, Song J, Yeom KW. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. Eur J Nucl Med Mol Imaging. 2020 Oct 23:1–9. doi: 10.1007/s00259-020-05075-4. Epub ahead of print. PMID: 33094432; PMCID: PMC7581467.
[6] Ioannis D. Apostolopoulos1, Tzani A. Mpesiana, “Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks”, Physical and Engineering Sciences in Medicine, 3 April 2020
[7] Ali Narin, Ceren Kaya, Ziynet Pamuk, “Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks”, Version 3, arXiv ,20 Mar 2020
[8] Mucahid Barstugan, Umuy Ozkaya, Saban Ozturk, “Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods”, Version 1, arXiv, 20 Mar 2020
[9] O’Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." arXiv preprint arXiv:1511.08458 (2015).
[10] “Coronavirus Cases:” Worldometer, www.worldometers.info/coronavirus/.