Diagnosis of COVID-19 with a Deep Learning Approach on Chest CT Slices

Fatma Muberra Yener
Biological Data Science
Institute of Graduate Studies
Istanbul Medeniyet University
Istanbul, Turkey

Ayse Betul Oktay
Department of Computer Science
Istanbul Medeniyet University
Istanbul, Turkey
abetul.oktay@medeniyet.edu.tr

***Abstract*— Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) first broke out in Wuhan, China and COVID-19 disease spread throughout the world by its highly contagious nature. High death numbers have caused a massive panic across the globe. Fast and early diagnosis is the key for preventing the virus from spreading. Besides PCR test, computed tomography (CT) of lungs is also used for diagnosis of COVID-19. Since the amount of testing kits for the diagnosis is insufficient and the conventional diagnosis methods are slow, developing AI-based fast diagnosis tools is not only an alternative way but also an urgent requirement for such alarming situations as those people faced with today. In this study, we employed three popular CNN models, VGG16, VGG19, and Xception, to classify CT scans of suspected patient cases as COVID-19 infected and non-COVID-19. VGG16 achieved 93% accuracy with the best parameters on the test set.**

Keywords—deep learning; convolutional neural networks; transfer learning; COVID-19; computed tomography; chest

I. INTRODUCTION

A recent epidemic of a new Coronavirus Disease (COVID-19) has had a great impact on the nations worldwide. One of the most important features of the virus which belongs to the large subfamily of Coronaviruses varying from common cold to Middle East Respiratory Syndrome Coronaviruses (MERS-CoV) is the speed of spreading among people. That's why the main precaution taken against the disease is quarantine. The virus, officially named as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is identified by the common symptoms such as cough, shortness of breath, fatigue, myalgia and abnormalities (glass-like opacities) on chest computed tomography images (CT) apart from genetic analysis. Although some patients may develop more severe complications like pulmonary and organ injuries, advanced viral pneumonia, acute respiratory distress syndrome, and some patients may be asymptomatic [1]. In order to prevent the disease from rapid spreading and to follow the quarantine measures, the diagnosis by screening great number of potential cases using both viral genetic testing and CT scan is vastly crucial.

Since the availability and accuracy of nucleic acid testing have some limitations along with the accessibility for the test kits in the infected zone is low, medical radiological imaging is an alternative way for COVID-19 diagnosis. CT images of COVID-19 patients indicate common features such as basal-predominant ground-glass opacities with round or oval morphology and consolidation, holding higher sensitivity to those features over X-ray imaging [2]. It is challenging for radiologists to distinguish between chest CT images of patients with COVID-19 and those with other viral pneumonia due to the non-specificity of the symptoms [3]. Hence, an extensive need for automated classification systems of COVID-19 CT images arises to diagnose using the power of artificial intelligence. On the other hand, deep learning-based CT evaluation has a capability to respond fast to the crowd gathered in the hospitals and waiting for their results, which is critical to reduce cross-contamination in the hospitals [4].

Deep learning, one of the subtopics of machine learning, has proven to perform efficiently on CT images regarding feature extraction, localization, segmentation and classification [5,6,7,8,9]. Convolutional neural network (CNN) based models were developed for the detection of lymph nodes on Mediastinum and Abdomen CT scans by *Roth et al.* [10] and pulmonary nodule by *Van Ginneken et al.* [11] on chest CT scans. *Gao et al.* [12] built a CNN classifier for interstitial lung diseases using chest 2-D CT slices.

For developing robust CNN-based models, the huge amount of data for training is a major necessity [13]. Therefore, standardized and massive databases are available for public use such as ImageNet, Pascal [14,15]. Transfer learning involves pre-trained models on large datasets as a framework to the tasks using other datasets. ImageNet dataset including over 1000 classes with 1.2 million natural images is often used for pre-training the framework models because there are no other datasets in similar extent and quality. ImageNet pre-trained models adapt successfully to the medical image tasks, not only saving time and computational power but also raising the performance metrics [11,16].

There are various popular CNN architectures available for transfer learning such as ResNet, VGG16, Inception, Xception, etc. [17,18,19,20]. In this study, VGG16, VGG19 and Xception network architectures pre-trained on ImageNet is used for detection of the COVID-19. It is stated that two open datasets, independently including 746 and 3227 chest CT images from patients with COVID-19, healthy people as well as those with other non-relevant chest diseases, are used for the validation of the CNNs.

© IEEE 2021. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

II. Dataset and Methods

A. Dataset

One of the datasets used in this study is built by *Zhao et al.* [21]. They downloaded preprints from medRxiv and bioRxiv sent between 19th January and 25th March, 2020. These papers are about COVID-19 patient cases and some of them contains CT images of the patients. Figures and captions were extracted in good quality. CT images of the patients were separated from the rest by hand. According to the caption of each image, whether each one represents positive or negative was decided. Finally, 349 CT images are positively labeled and 397 negatively labeled.

Alternative sources were also used to retrieve images. We retrieved a second reliable dataset, collected voluntarily for public use from multiple sources, from Kaggle database [24].

In total, A balanced dataset with 3227 CT images were used, we split it into 2 subsets as training (85%) and test (15%) sets respectively. Our dataset is a mixture of images from the patients with COVID-19, other chest infection cases and healthy people.

In our workflow, grayscale CT images were preprocessed: Every image was resized to 224x224 pixels using "Image.resize" function from Python Imaging Library (Pillow) and normalized dividing each pixel by 255.

B. Model Architecture

Transfer learning approach leverages our system in many ways. Its main purpose is to extract knowledge from one or more tasks and then to adapt it to a different task. After some architectural experiments, we preferred VGG16 network architecture filled with the stacks of convolutional layers whose receptive fields are very small; either 3X3 or 1x1. ReLU nonlinearity is carried out after each convolutional layer in order to enhance the discriminative properties of decision function [18]. Its pre-trained version is ready for usage in Keras Application module. Although 3-channel data shape is compulsory for the pre-trained model, our data have only one channel. To fit our data to input shape, we stacked a single-channel image three times and made it look like a 3-channel image. Same procedure was applied to the whole dataset.

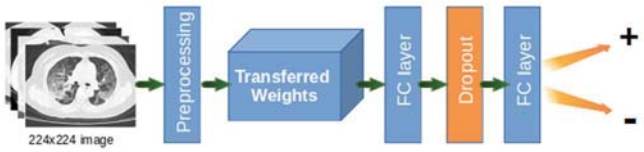


Fig. 1. Transfer learning model for COVID-19

In the model we proposed, VGG16 architecture with pretrained weights on ImageNet are transferred. The last fully-connected layers are trimmed and new two are added. Only last two fully connected layers set as trainable, the rest of the weights are frozen. Our implementation based on Keras library.

C. Training

In this section, the model is trained during approximately 35 epochs, although the number of epochs was hard coded as 100 since some checkpoints were put in use for early stopping. Such checkpoints were EarlyStopping and ReduceLROnPlateau. The arguments of EarlyStopping; Monitor, mode, minimum delta and patience were set to validation loss, min, 10^{-7} and 15 respectively. Those for ReduceLROnPlateau; Monitor, factor, patience, minimum delta and mode were set to validation loss, 0.1, 10, 10^{-8} and min in respective order.

Mini-batch learning with batch size of 30 was applied to proposed CNN model. Adam optimizer as optimizing method and binary cross entropy as loss function were used. Dropout with the rate of 0.5 was chosen to carry out regularization in order to get rid of overfitting. In the preliminary designs of the network, hyperparameters were fine-tuned and learning rate (LR) was adopted as 10^{-3} and 10^{-4}.

D. Performance Metrics

Performance measures are accuracy, precision, sensitivity, F1 score, ROC-AUC score, confusion matrix. Let TN be the correctly classified COVID-19 negative images, FN be the incorrectly classified COVID-19 positive images, TP be the correctly classified COVID-19 positive images and FP be the incorrectly classified COVID-19 negative images.

Confusion matrix is a tabulated representation of true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP), so called error matrix. The performance metrics used in this study are given below:

$$\text{Sensitivity (recall)} = \frac{TP}{TP+FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{F1 score} = 2.\frac{Precision.Recall}{Precision+Recall} \tag{3}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

ROC, Receiver operating characteristic curve is a curve where X-axis is False Positive Rate and Y-axis is True Positive Rate. It is used for deciding which threshold is better for a specific binary classification task. Area under the ROC (ROC-AUC) score is used for deciding which binary classification method is better for specific data [22].

III. Experimental Results

A. Architectural Comparisons

Three different models were experimented to observe the impact of the architectural design preferred: VGG16, VGG19 and Xception. All the other options, hyperparameters and pipeline are exactly the same as those mentioned in the training section.

TABLE 1 RESULTS OF DIFFERENT CNN ARCHITECTURES (LR=10^{-3}, FOR + CLASS)

	Acc	Precision	Recall	F1 measure	ROC-AUC
VGG16	.91	.90	.94	.92	.91
VGG19	.91	.90	.93	.91	.90
Xception	.65	.65	.74	.69	.64

One can build a model with unique architecture from scratch but we used well-known architectures. The decision which architecture is suitable for the dataset is made by simply choosing the one with the best performance metrics, especially ROC-AUC. The metrics were screened during training and testing, VGG architectures perform better than Xception at chest CT COVID-19 dataset (Table 1). In Table 1, VGG16 and VGG19 has 0.91 accuracy, however recall of VGG16 is 0.01% higher than VGG19. VGG has small size filters and ReLU activation where Xception has modified depth wise separable convolution. In most of the studies in the literature, Xception outperforms many architectures, however for the chest CT image slices, it has 0.65 accuracy.

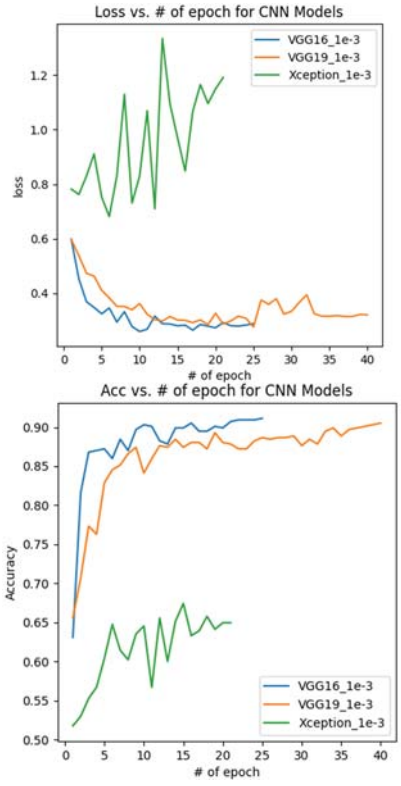


Fig. 2. Loss and Accuracy Values for Models (LR=10^{-3})

The convergence of VGG19 model is slower than that of VGG16 model as seen in Fig 2. Xception model seems like unstable and unconverged. We applied the same hyperparameters (10^{-3} and 10^{-4}) and none of them yielded a significant outcome. VGG16 is the best model regarding both minimum loss and maximum accuracy. The fluctuations and zigzags in the plots stem most probably from mini-batches. Since we shuffled all samples and divided them into mini-batches, it is unlikely for them to have evenly distributed features. The emphasis should be made on Xception plot having very deep zigzags which cannot be accepted as normal.

B. *The Effect of Learning Rate*

Learning rate is a hyperparameter that needs to be fine-tuned during training or validation. The strategies carried out for hyperparameter optimization are grid search and random search. Studies showed that random trials are more efficient than trials on a grid. Allocating the same cost for both, random search can come up with superior models [23]. In our models, we used manual random search for learning rate hyperparameter. VGG19 achieved the best results when learning rate is 10^{-3} and 10^{-4}.

TABLE 2 RESULTS OF MODELS WITH DIFFERENT LEARNING RATES (FOR + CLASS)

	Acc	Precision	Recall	F1	ROC-AUC	LR
VGG16	.91	.90	.94	.92	.91	10^{-3}
VGG16	.93	.91	.96	.94	.93	10^{-4}

In the final setup, we kept model architecture unchanged and decreased learning rate to 10^{-4}. It outperformed the previous VGG16 model with learning rate of 10^{-3} according to all metrics. Our dataset is multifarious, thereby it is inferred with ease that VGG16 model with LR 10^{-4} generalizes well.

If we compare Loss/Accuracy vs epoch graphs in Fig. 3 again, the model with LR 10^{-4} is smoother and fluctuated less in terms of loss and accuracy. Imbalanced up and downs are seen more in the VGG16_10^{-3} model's graph. This may due to the non-uniform distribution of samples in each batch. When accuracy was saturated with the increasing epoch, graphs of both models become plateau-like shape. The first model (VGG16_10^{-3}) shows lower ROC-AUC score than the second one (VGG16 _10^{-4}).

IV. CONCLUSION

In this study, we proposed plain and useful CNN models for the purpose of diagnosing COVID-19. In our model, VGG16 architecture with pretrained weights on ImageNet was used for the binary classification of CT images of patients with COVID-19 and non-COVID ones. CT images we used are publicly available. Despite high accuracy, sensitivity and F1 scores were achieved, The network could have generalized more over diverse data if larger number of images existed for training.

Providing radiologists with an effective AI-based computer aided system in the scope of fast and early diagnosis of recent ongoing epidemic is aimed in this study. Much more effort is needed to deploy a promising system for the diagnosis of the disease to the currently used technological devices but this work

represents a hope-inspiring step towards it. Designing such early diagnosing machinery can be the future of healthcare system. Availability of them is vital particularly during emergencies like COVID-19 outbreak case.

As future work, collecting more data and building deeper and more robust models are intended.

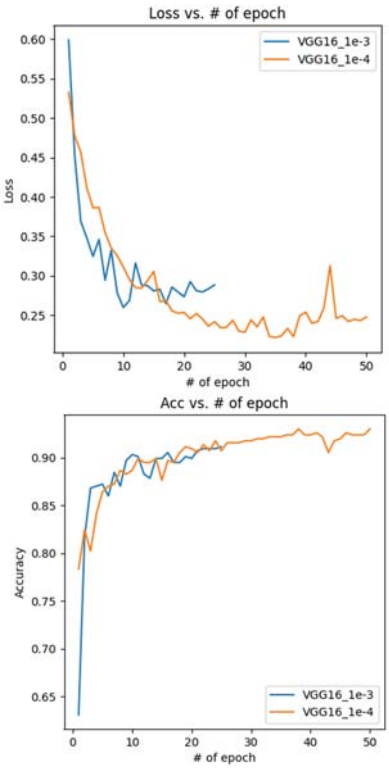


Fig. 3 Loss and Accuracy Values for Models with different LR

ACKNOWLEDGEMENT

Dataset used for training and test is available at https://github.com/UCSD-AI4H/COVID-CT and the second dataset is from https://www.kaggle.com/covidct

REFERENCES

[1] Hu, Z., Song, C., Xu, C., Jin, G., et al (2020). Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Science China Life Sciences, 63(5), 706-711.

[2] Schmitt, W., & Marchiori, E. (2020). Covid-19: round and oval areas of ground-glass opacity. Pulmonology.

[3] Zu, Z. Y., Jiang, M. D., Xu, P. P., Chen, W., Ni, Q. Q., Lu, G. M., & Zhang, L. J. (2020). Coronavirus disease 2019 (COVID-19): a perspective from China. Radiology, 200490.

[4] Ye, Guangming, Hualiang Lin, Liangjun Chen, Shichan Wang, Zhikun Zeng, Wei Wang, Shiyu Zhang et al. "Environmental contamination of SARS-CoV-2 in healthcare premises." Journal of Infection (2020).

[5] Ciompi, F., de Hoop, B., van Riel, S. J., Chung, K., Scholten, E. T., Oudkerk, M., ... & van Ginneken, B. (2015). Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Medical image analysis, 26(1), 195-202.

[6] Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2015, June). Multi-scale convolutional neural networks for lung nodule classification. In International Conference on Information Processing in Medical Imaging (pp. 588-599). Springer, Cham.

[7] Kuruvilla, J., & Gunavathi, K. (2014). Lung cancer classification using neural networks for CT images. Computer methods and programs in biomedicine, 113(1), 202-209.

[8] Lu, F., Wu, F., Hu, P., Peng, Z., & Kong, D. (2017). Automatic 3D liver location and segmentation via convolutional neural network and graph cut. International journal of computer assisted radiology and surgery, 12(2), 171-182.

[9] Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013, September). Mitosis detection in breast cancer histology images with deep neural networks. In International conference on medical image computing and computer-assisted intervention (pp. 411-418). Springer, Berlin, Heidelberg.

[10] Roth, H. R., Lu, L., Seff, A., Cherry, K. M., Hoffman, J., Wang, S., ... & Summers, R. M. (2014, September). A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations. In International conference on medical image computing and computer-assisted intervention (pp. 520-527). Springer, Cham.

[11] Van Ginneken, B., Setio, A. A., Jacobs, C., & Ciompi, F. (2015, April). Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In 2015 IEEE 12th International symposium on biomedical imaging (ISBI) (pp. 286-289).

[12] Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H. C., ... & Xu, Z. (2018). Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(1), 1-6.

[13] Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.

[14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[15] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303-338.

[16] Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. (2015, April). Chest pathology detection using deep learning with non-medical training. In 2015 IEEE 12th international symposium on biomedical imaging (ISBI) (pp. 294-297). IEEE.

[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[18] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[19] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[20] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D, et al. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[21] Zhao, J., Zhang, Y., He, X., & Xie, P. (2020). COVID-CT-Dataset: a CT scan dataset about COVID-19. arXiv preprint arXiv:2003.13865. In press.

[22] Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine learning, 77(1), 103-123.

[23] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. The Journal of Machine Learning Research, 13(1), 281-305.

[24] Open dataset, Second dataset from https://www.kaggle.com/