Original Article

Deep learning for detecting corona virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study

Shuyi Yang[1#], Longquan Jiang[2#], Zhuoqun Cao[2], Liya Wang[3], Jiawang Cao[4], Rui Feng[2], Zhiyong Zhang[1,5], Xiangyang Xue[2], Yuxin Shi[1], Fei Shan[1]

[1]Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China; [2]School of Computer Science, Fudan University, Shanghai 200433, China; [3]Department of Radiology, Affiliated Longhua People's Hospital, Southern Medical University, Shenzhen 518109, China; [4]Academy of Engineering & Technology, Fudan University, Shanghai 200433, China; [5]Fudan University, Shanghai 200433, China

Contributions: (I) Conception and design: S Yang, L Jiang, R Feng, Z Zhang, X Xue, Y Shi, F Shan; (II) Administrative support: Z Zhang, X Xue, Y Shi, F Shan; (III) Provision of study materials or patients: S Yang, L Wang, Z Zhang, Y Shi, F Shan; (IV) Collection and assembly of data: L Jiang, Z Cao, J Cao, R Feng; (V) Data analysis and interpretation: L Jiang, Z Cao, J Cao, R Feng; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Fei Shan, MD. Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, No. 2901 Caolang Road, Jinshan, Shanghai 201508, China. Email: shanfei_2901@163.com.

Background: To evaluate the diagnostic efficacy of Densely Connected Convolutional Networks (DenseNet) for detection of COVID-19 features on high resolution computed tomography (HRCT).

Methods: The Ethic Committee of our institution approved the protocol of this study and waived the requirement for patient informed consent. Two hundreds and ninety-five patients were enrolled in this study (healthy person: 149; COVID-19 patients: 146), which were divided into three separate non-overlapping cohorts (training set, n=135, healthy person, n=69, patients, n=66; validation set, n=20, healthy person, n=10, patients, n=10; test set, n=140, healthy person, n=70, patients, n=70). The DenseNet was trained and tested to classify the images as having manifestation of COVID-19 or as healthy. A radiologist also blindly evaluated all the test images and rechecked the misdiagnosed cases by DenseNet. Receiver operating characteristic curves (ROC) and areas under the curve (AUCs) were used to assess the model performance. The sensitivity, specificity and accuracy of DenseNet model and radiologist were also calculated.

Results: The DenseNet algorithm model yielded an AUC of 0.99 (95% CI: 0.958–1.0) in the validation set and 0.98 (95% CI: 0.972–0.995) in the test set. The threshold value was selected as 0.8, while for validation and test sets, the accuracies were 95% and 92%, the sensitivities were 100% and 97%, the specificities were 90% and 87%, and the F1 values were 95% and 93%, respectively. The sensitivity of radiologist was 94%, the specificity was 96%, while the accuracy was 95%.

Conclusions: Deep learning (DL) with DenseNet can accurately classify COVID-19 on HRCT with an AUC of 0.98, which can reduce the miss diagnosis rate (combined with radiologists' evaluation) and radiologists' workload.

Keywords: COVID-19; deep learning (DL); high resolution computed tomography (HRCT)

Page 2 of 7

Yang et al. DL for detecting COVID-19

Introduction

Since December 2019, some novel corona virus (COVID-19) patients emerged in Wuhan, Hubei, China, while early cases were related to the seafood market with wild animal trade (1). The new type of coronavirus was detected from infected airway epithelial cells (2), and named as sever acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) (Gorbalenya AE. Severe acute respiratory syndrome-related coronavirus – The species and its viruses, a statement of the Coronavirus Study Group [J]. BioRxiv 2020). SARS-CoV-2 are RNA viruses belonging to the family Coronaviridae and the order Nidovirales, as well as severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) (2-4), which have emerged as major global health threats since 2002 and 2012 that spread to 37 countries and 27 countries (3-5). On January 20, 2020, the National Health Commission of the People's Republic of China categorized the COVID-19 pneumonia into type B infectious disease and announced to take it as type A measures in preventing and controlling this disease. To mitigate the spread of SARS-CoV2, the national government has implemented traffic control in Wuhan and some nearby cities since Jan 23–24, 2020.

At present, COVID-19 patients are diagnosed mainly by real-time PCR (RT-PCR) to detect SARS-CoV-2 nucleic acid. Recently, due to the limited supply of RT-PCR kit and false-negative nucleic acid cases emerged, some experts have proposed to diagnose suspected cases using the time-saving chest computed tomography (CT) rather than RT-PCR (6). The typical clinical symptoms, epidemiological history and positive CT images are vital indicators to the identification of suspected patients. How to identify the positive CT images from massive number of CT images quickly and accurately, especially in the epidemic area, is a hot topic recently.

Deep learning (DL) has been proved to be an effective method on CT images to classify common imaging signs of lung diseases in recent years. Sun *et al.* proposed a Lung Image Database Consortium (LIDC) database and tested some DL methods include Convolutional Neural Network (CNN) (7). Kang *et al.* show that improved CNN perform better on LIDC classification problems (8). Besides all that, DL has been widely used in the automatic diagnosis (9), lung nodules detection (10) and classification (11). Densely Connected Convolutional Networks (DenseNet) (12), an improved CNN, which can extract the shallow features and inner representation of the image simultaneously, shows great performance in ImageNet (13) classification task.

Our hospital is the designated center for diagnosis and management of infectious diseases in Shanghai and is also WHO designated training organization for new emerging infectious diseases. The purpose of this study is to develop a DL model to evaluate the efficacy of DenseNet for detection of COVID-19 features on high resolution CT (HRCT).

Methods

Datasets

The Ethic Committee of Shanghai Public Health Clinical Center, Fudan University approved the protocol of this study and waived the requirement for patient informed consent (No. YJ-2020-S035-01).

From January 20 to Feb 1, 2020, a total of 156 confirmed cases of COVID-19 in Shanghai were admitted to our hospital. In this study, the inclusion criteria were followed as: (I) patients with a positive SARS-CoV-2 nucleic acid antibody detected by the Center for Disease Control (CDC), Shanghai, and re-checked by national CDC; (II) patients with chest thin-section HRCT scanning; (III) HRCT showed pulmonary lesions; (IV) healthy persons' HRCT images originated from the same CT scanner; the exclusion criteria was patients also with criteria of other pneumonia, such as bacterial or other virus pneumonia.

Two hundred and ninety-five patients (male: female, 154:141; median age, 37.5 years old; age range, 15–80 years) were enrolled in this study, in which, 149 were healthy person and 146 were patients with positive SARS-CoV-2 nucleic acid antibody (10 patients without pulmonary lesions on HRCT were excluded). Then we divided these CT images into three separate non-overlapping cohorts, one was used for algorithm development (training set, n=135, healthy person, n=69, patients, n=66), another was used for hyperparameter selection during development (validation set, n=20, healthy person, n=10, patients, n=10), and the rest was used for algorithm testing, (test set, n=140, healthy person, n=70, patients, n=70).

Image preprocessing

It was impossible to train the algorithms using full volumetric HRCT images, which may consist of 300–500 axial image slices per patients. Therefore, each
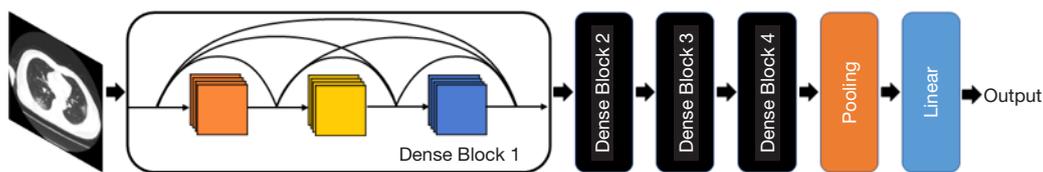
Figure 1 The workflow of DenseNet with 4 blocks. Dense Block 1 illustrate a three-layer block with dense connectivity. Pooling and Linear refer to global average pooling and fully connected layer.
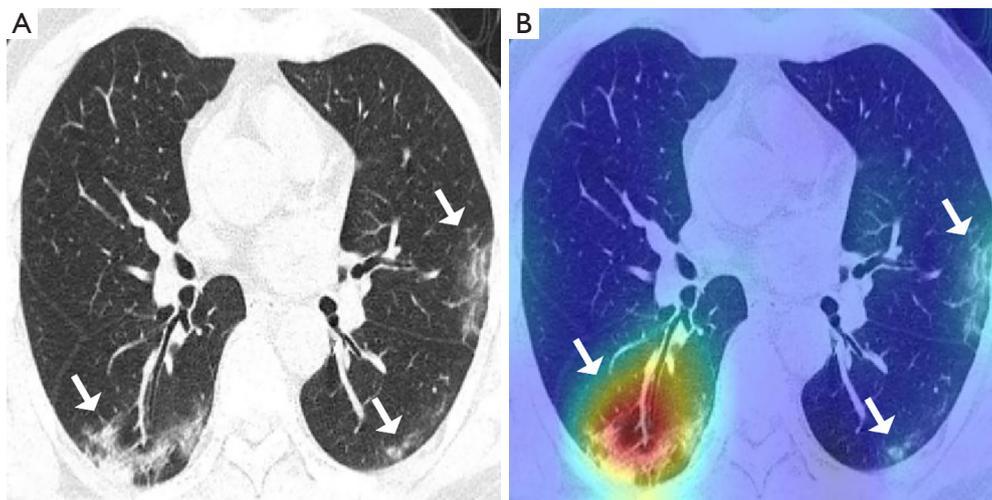

Figure 2 CAM calculated by output feature maps of the last convolutional layer. (A) HRCT shows GGOs with consolidation in the 3 segments of the lung (→); (B) The red and light blue regions represent areas activated by the DenseNet, while the dark purple background represents areas that are not activated. This shows that the DenseNet is focusing on parts of the image where the disease is present (→). HRCT, high resolution CT.

CT scan underwent a preprocessing step to select target slices before algorithm training. For patients' CT scans, images containing ground glass opacity (GGO), GGO with consolidation or consolidation were selected (14). While, for healthy control, we first identified the pulmonary parenchyma area by lung segmentation, and then images containing pulmonary parenchyma were selected out every 3 slices. Lung windowing (window width= 1,200 Hu, window level= −600 Hu) is performed over all image slices to increase the internal contrast of lungs CT scans.

Algorithm development

We implemented our method with *PyTorch1.1.0* (15). The workflow of the model building was demonstrated in *Figure 1*. The process of training was optimized by stochastic gradient descent (SGD) with step learning Rate, which initial values were: the momentum was 0.9, Step

Size was 5, and considering the features of the lung CT images should not be affected by the direction, we used Random Flip as a data argumentation method. Training for our model on CT images took 20 epochs with batch size 32 image with two Nvidia 1080Ti GPU. Test process took about 3 seconds for a batch of 32 images, and about 30 seconds for whole volumes of CT Scan for a patient.

Algorithm testing

The class activation maps (CAM) (16), calculated by output feature maps of the last convolutional layer, highlighted the importance of the image region to the prediction (*Figure 2*). The CAM classified as coronary pneumonia; red region indicates the discriminative region used by CNN to identify the category.

The performances of models were evaluated by assessing the accuracy on test set of 74 patients and 70 healthy
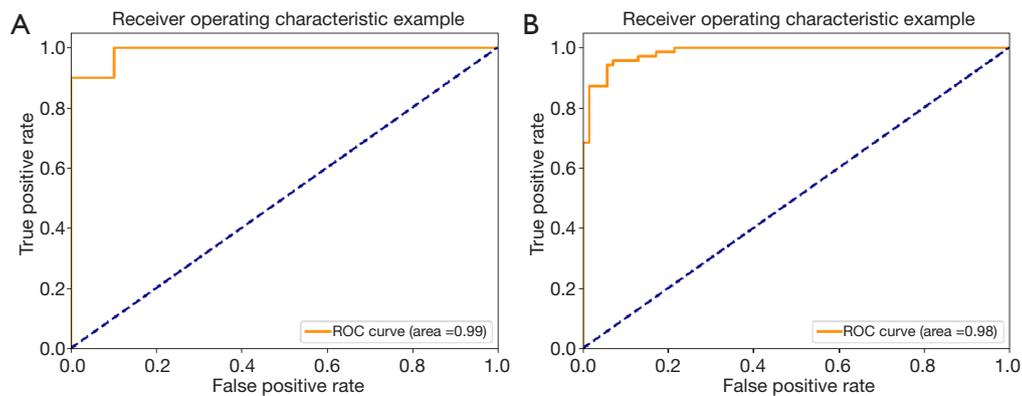
Page 4 of 7

Yang et al. DL for detecting COVID-19

Figure 3 Receiver operating characteristic plots for COVID-19 identification for the DenseNet algorithm (A: validation set; B: test set).
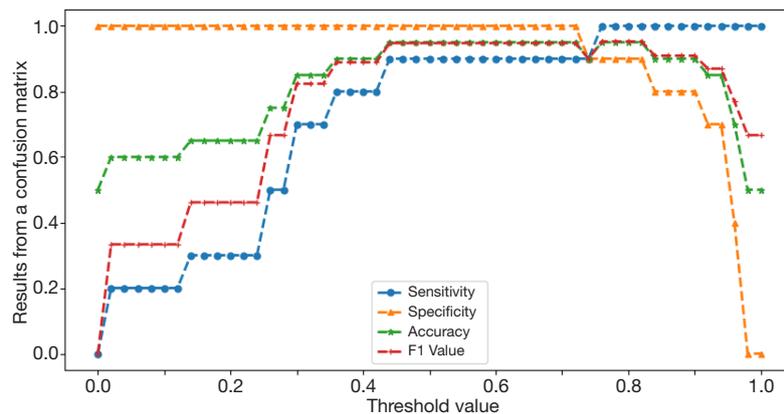

Figure 4 The sensitivity, specificity, accuracy and F1 value along with the change of threshold value (validation set).

persons. The output of the model represents the probability of the image classified into the two categories (patients or healthy person). Then, the classification result of one object is achieved by removing 10% of the images from the subject before and after and averaging the probability of the rest images. A radiologist with 6 years of chest imaging diagnosis experience was also blindly to evaluate the CT images in the test set.

Statistical analysis

All statistical analyses were performed by using *Python 3.7.6* and *sklearn 0.22.1*. On the validation and test sets, receiver operating characteristic curves (ROC) and area under the curves (AUCs) were determined. Sensitivity, also named true positive rate (TRR), was percentage of positive patients who were correctly discriminated. Specificity, also named true negative rate (TNR), was percentage of negative

persons who were correctly discriminated. Accuracy was the percentage of subjects with TRR and TNR. AUC was an index to measure the performance of the classifier. F1 score was a measure of the accuracy of a binary model. The COVID-19 patients were supposed to positive cases. We accessed the sensitivity, specificity, accuracy and F1-score, respectively, along with the change of threshold value. The radiologist diagnostic efficacy was also evaluated by sensitivity, specificity and accuracy (fourfold table analysis).

Results

The performance was evaluated on subject-level. The DenseNet algorithm model yielded an AUC of 0.99 (95% CI: 0.958–1.0) in the validation set and 0.98 (95% CI: 0.972–0.995) in the test set (*Figure 3*). Due to the purpose of our study, we expected a higher value of sensitivity, so the threshold value was identified as 0.8 in our study (*Figure 4*).

Table 1 DenseNet Algorithm performance with the threshold of 0.8

Parameter	Validation dataset	Test dataset	Radiologist
AUC (95% CI)	0.99 (0.958–1.0)	0.98 (0.972–0.995)	–
Accuracy (%)	95	92	95
Sensitivity (%)	100	97	94
Specificity (%)	90	87	96
F1 score (%)	95	93	–

AUC, area under the curve.

Table 2 Fourfold table of radiologist's diagnostic efficacy on the test set

	COVID-19 patients	Healthy person
Radiologist positive	66	3
Radiologist negative	4	67

Sensitivity =66/70 (94%); Specificity =67/70 (96%); Accuracy = (66+67)/140 (95%).

For validation and test sets, the accuracies were 95% and 92%, the sensitivities were 100% and 97%, the specificities were 90% and 87%, and the F1 values were 95% and 93%, respectively (*Table 1*). The sensitivity of radiologist was 94%, the specificity was 96%, while the accuracy was 95% (*Table 2*).

Discussion

As of Feb 21, 2020, nearly 76,769 cases of COVID-19 have been confirmed globally, while 75,569 are in China with at least 2239 deaths (17). The suspected patients should be confirmed by RT-PCR of COVID-19 nuclei acid test (5,6). For the epidemic source area, the limited amount of RT-PCR kit results in many suspected patients even the unknown potential patients, who are vital source of infection. Chest CT is widely used to detect pulmonary lesions and evaluate the therapeutic effect timely (7). Due to the overburdened medical work, especially for the radiologists in the infection specialist hospital, it is easy to miss the small pulmonary lesions, especially for the patients in an earlier stage (14), to make the misdiagnosis, which is not conducive to early treatment and isolation.

DL has demonstrated superior performance in lung nodule classification of CT image (11). Compared to other CNN model, DenseNet contain more shorter connections between layers which can strengthen feature propagation and encourage feature reuse. Such a deep supervision (18) mechanism has proven to be beneficial for medical imaging, such as UNet ++ (19). In this study, we trained a DL algorithm to detect COVID-19, which were confirmed by positive nuclei acid test results. In the test set, the sensitivity of DenseNet classifier was 97%, while the accuracy was 92%, which is similar to a young radiologist with a relative mild burdened medical work situation. But, compared to a radiologist who needs 5–10 minutes to diagnose a CT scan, DenseNet classifier only takes about 30 seconds to give the final result.

We also rechecked the false positive and false negative cases misclassified by DenseNet classifier. We found that, to the false positive cases in the healthy persons group, the misidentification areas focused on the like-GGO generated by the partial volume effect of pulmonary vessels through the axial section obliquely, the respiratory motion artifacts mainly located in the basal segment of the lung or the subpleural nonspecific inflammation results from the cardiovascular physical motion. To the false negative cases in COVID-19 group, the pulmonary lesions were all solitary and focal, manifested as pure GGOs (size: 1.5 and 0.89 cm, CT value: –759.55 and –588.42 Hu). We also found that some slight lesions were misdiagnosed by radiologist but detected by DenseNet classifier accurately. Those indicated that our model reflects high sensitivity and specificity, although with some misdiagnosed cases. The radiologists could achieve more accurate results, combined with our model. Such an improvement was of great clinical significance. It could save a lot of doctors' diagnosis time, devoted more valuable time to other professional work, and greatly improved clinical care for patients.

There are some limitations in our study. Firstly, it is a pilot study about the DL application in a new emerging pneumonia detection with limited number of COVID-19 patients. Secondly, in order to evaluate the efficacy of DenseNet for detection of COVID-19 on CT, we enrolled the healthy and COVID-19 patients without other pathogen pneumonia. The DL applied in differentiating COVID-19 from other pathogen pneumonia is under research in our team. Thirdly, the impact of equipment differences is not currently considered, and subsequent research needs to include data from different sources to verify the generalization ability of the model.

Conclusions

In conclusion, we developed a DL model with human-level performance for detecting COVID-19 on high-

Page 6 of 7

Yang et al. DL for detecting COVID-19

resolution CT. The accuracy detection with an AUC of 0.98, which is valuable for reducing the miss diagnosis rate and radiologists' workload.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm.2020.03.132). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The Ethic Committee of Shanghai Public Health Clinical Center, Fudan University approved the protocol of this study and waived the requirement for patient informed consent (No. YJ-2020-S035-01).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. WHO. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. (accessed Feb 20, 2020). Available online: https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020
2. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265-9.
3. Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003;348:1953-66.
4. de Groot RJ, Baker SC, Baric RS, et al. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J Virol 2013;87:7790-2.
5. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;395:689-97.
6. Xinhua. Virus-hit Wuhan speeds up diagnosis of patients. (accessed Feb 6, 2020). Available online: https://enapp.chinadaily.com.cn/a/202002/06/AP5e3be074a3103a24b1106147.html
7. Sun W, Zheng B, Qian W. Computer aided lung cancer diagnosis with deep learning algorithms. SPIE Medical Imaging. Medical Imaging 2016: Computer-Aided Diagnosis, 2016.
8. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med Phys 2017;44:e360-75.
9. Peng Z, Xinnan X, Hongwei W, et al. Computer-Aided Lung Cancer Diagnosis Approaches Based on Deep Learning. J Comput Aided Design Comput Graph 2018;30:90.
10. Li Z, Li L. A novel method for lung masses detection and location based on deep learning, 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2017.
11. Song Q, Zhao L, Luo X, et al. Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. J Healthc Eng 2017;2017:8314740.
12. Huang G, Liu Z, Weinberger KQ, et al. Densely connected convolutional networks. 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2016.
13. Deng J, Dong W, Socher R, et al. ImageNet: a Large-Scale Hierarchical Image Database. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2009.
14. Song F, Shi N, Shan F, et al. Emerging Coronavirus 2019-nCoV Pneumonia. Radiology 2020. [Epub ahead of print].
15. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2017.
16. Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:2921-9.

17. WHO. Coronavirus disease 2019 (COVID-19) situation report - 32. (accessed Feb 22, 2020). Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

18. Lee CY, Xie S, Gallagher P, et al. Deeply-supervised nets. In: Artificial intelligence and statistics. 2015:562-70.

19. Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, 2018:3-11.