medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)

Shuai Wang[1,*], Bo Kang[2,3,*], Jinlu Ma[4,*], Xianjun Zeng[5,*], Mingming Xiao[1,*], Jia Guo[3], Mengjiao Cai[4], Jingyi Yang[4], Yaodong Li[6], Xiangfei Meng[2,#], Bo Xu[1,#]

[1] Department of Molecular Radiation Oncology, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Key Laboratory of Breast Cancer Prevention and Therapy, Ministry of Education, Tianjin Clinical Research Center for Cancer, Tianjin Medical University Cancer Hospital and Institute, Tianjin 300006, China

[2] College of Intelligence and Computing,Tianjin University, Tianjin 300350, China

[3] National Supercomputer Center inTianjin, Tianjin 300457, China

[4] Department of Radiation Oncology, First Affiliated Hospital, Xian Jiaotong University, Xi'an, China

[5] Department of Radiology, Nanchang University First Hospital, Nanchang, China

[6] Department of Radiology, No.8 Hospital, Xi'an Medical College, Xi'an, China

* Equal contribution

corresponding authors:

Bo Xu, MD, PhD
Tianjin Medical University Cancer Institute and Hospital
Tianjin 300006, China
Email: xubo@tmu.edu.cn

Xiangfei Meng, PhD
National Supercomputer Center in Tianjin
Tianjin 300457, China
Email: mengxf@nscc-tj.cn

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Competing Interest Statement

The authors have declared no competing interest.

Funding Statement

None

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Abstract

The outbreak of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) has caused approximately 64,000 cases of Corona Virus Disease (COVID-19) in China so far, with that number continuing to grow. To control the spread of the disease, screening large numbers of suspected cases for appropriate quarantine and treatment measures is a priority. Viral nucleic acid testing based on specimens from the lower respiratory tract is the diagnostic gold standard. However, the availability and quality of laboratory testing in the infected region presents a challenge, so alternative diagnostic methods are urgently needed to combat the disease. Based on COVID-19 radiographical changes in CT images, we hypothesized that Artificial Intelligence's deep learning methods might be able to extract COVID-19's specific graphical features and provide a clinical diagnosis ahead of the pathogenic test, thus saving critical time for disease control. To test this possibility, we collected 453 CT images of pathogen-confirmed COVID-19 cases along with previously diagnosed with typical viral pneumonia. 217 images were used as the training set and the inception migration-learning model was used to establish the algorithm. The internal validation achieved a total accuracy of 82.9% with specificity of 80.5% and sensitivity of 84%. The external testing dataset showed a total accuracy of 73.1% with specificity of 67% and sensitivity of 74%. These results indicate the great value of using the deep learning method to extract radiological graphical features for COVID-19 diagnosis.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Introduction

The outbreak of atypical and person-to-person transmissible pneumonia caused by the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2, also known as 2019-nCov) has caused a global alarm. There have been nearly 64,000 confirmed cases of the Corona Virus Disease (COVID-19) in China, as of February 14, 2020. In addition to these China also has, more than 14,000 other suspected cases..

According to the WHO, 16-21% of people with the virus in China have become severely ill with a 2-3% mortality rate. With the most recent estimated viral reproduction number (R0), the average number of other people that an infected individual will transmit the virus to in a completely non-immune population, stands at about 3.77 [1], indicating that a rapid spread of the disease is imminent. Therefore, it is crucial to identify infected individuals as early as possible for quarantine and treatment procedures. The diagnosis of COVID-19 relies on the following criteria: clinical symptoms, epidemiological history and positive CT images as well as positive pathogenic testing. The clinical characteristics of COVID-19 include respiratory symptoms, fever, cough, dyspna, and viral pneumonia [2-5], however, these symptoms are nonspecific, as there are isolated cases where, for example, in an asymptomatic infected family a chest CT scan revealed pneumonia and the pathogenic test for the virus came back positive.

Once someone is identified as a PUI (person under investigation), lower

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

respiratory specimens, such as bronchoalveolar lavage, tracheal aspirate or sputum, will be collected for pathogenic testing. This laboratory technology is based on real-time RT-PCR and sequencing of nucleic acid from the virus [6, 7]. Since the beginning of the outbreak, the efficiency of nucleic acid testing has been dependent on several rate-limiting factors, including availability and quantity of the testing kits in the affected area. More importantly, the quality, stability and reproducibility of the detection kits are questionable. The impact of methodology, disease development stages, specimen collection methods, nucleic acid extraction methods and the amplification system are all determinant factors for the accuracy of test results. Conservative estimates of the detection rate of nucleic acid are low (between 30-50%), and tests need to be repeated several times in many cases before they can be confirmed.

Radiological imaging is also a major diagnostic tool for COVID-19. The majority of COVID-19 cases have similar features on CT images including ground-glass opacities in the early stage and pulmonary consolidation in the late stage. There is also sometimes a rounded morphology and a peripheral lung distribution [5, 8]. Although typical CT images may help early screening of suspected cases, the images of various viral pneumonias are similar and they overlap with other infectious and inflammatory lung diseases. Therefore, it is difficult for radiologists to distinguish COVID-19 from other viral pneumonias.

Artificial Intelligence involving medical imaging deep-learning systems has been developed in image feature extraction, including shape and spatial

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

relation features. Specifically, Convolutional Neural Network (CNN) has been proven in feature extraction and learning. CNN was used to enhance low-light images from high-speed video endoscopy with the limited training data being just 55 videos [9]. Also, CNN has been applied to identify the nature of pulmonary nodules via CT images, the diagnosis of pediatric pneumonia via chest X-ray images, automated precising and labeling of polyps during colonoscopic videos, cystoscopic image recognition extraction from videos [10-13].

There are a number of features for identifying viral pathogens on the basis of imaging patterns, which are associated with their specific pathogenesis [14]. The hallmarks of COVID-19 are bilateral distribution of patchy shadows and ground glass opacity [2]. Based on this, we believed that CNN might help us identify unique features that might be difficult for visual recognition. To test this notion, we retrospectively enrolled 453 CT images of pathogen-confirmed COVID-19 cases along with previously diagnosed typical viral pneumonia. We trained 217 images using the inception migration-learning model in order to establish the algorithm. We achieved a total accuracy of 83% with specificity of 80.5% and sensitivity of 84% for validation. The external testing showed a total accuracy of 73% with specificity of 67% and sensitivity of 74%. These observations demonstrate the proof-of-principle using the deep learning method to extract radiological graphical features for COVID-19 diagnosis.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Methods and Materials

Retrospective collection of datasets.

We retrospectively collected CT images from 99 patients, in which the cohort includes 55 cases of typical viral pneumonia and the other 44 cases from three different hospitals with confirmed nucleic acid testing of SARS-COV-2. The hospitals providing the images were Xi'an Jiaotong University First Affiliated Hospital, Nanchang University First Hospital and Xi'An No.8 Hospital of Xi'An Medical College. All CT images were de-identified before sending for analysis. This study is in compliance with the Institutional Review Board of each participating institutes. Informed consent was exempted by the IRB because of the retrospective nature of this study.

Overview of the proposed architecture.

Our systematic pipeline for the prediction architecture is depicted in **Figure 1**. The architecture consists of three main processes. 1) Randomly selection of ROIs; 2) Training of the CNN model to extract features; 3) Classification model training of fully connected network and prediction of multiple classifiers.

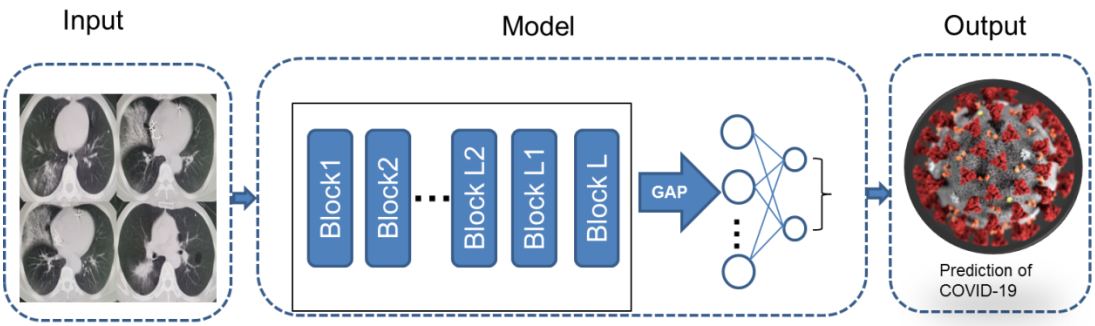


Figure 1 Deep learning algorithm framework. For each patient computed tomography scan, we randomly selected ROIs and used the Inception network to extract features and then make a prediction.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Intercept ROIs from CT images

CT images were recorded by video camera, which approximately 1*1 in-plane resolution. In order to reduce the computation complexity, and based on the signs characteristic of pneumonia, regions of interest (ROI) were extracted from CT images. For a ROI, it is sized approximately from 395*223 to 636*533 pixels. We picked 195 ROIs from 44 COVID-19 positive pneumonia patients and 258 ROIs from 50 COVID-19 negative patients. We built a transfer learning neural network based on Inception network. The entire neural network can be roughly divided into two parts: the first part uses a pre-trained inception network to convert image data into one-dimensional feature vectors, and the second part uses a fully connected network and the main role is for classification prediction. We randomly take 2-3 pictures from each case to form a training dataset. The number of various types of pictures in the training set is approximately equal, with a total number of 236. The remaining CT pictures of each case were used for validation. The model training is iterated 15,000 times with a step size of 0.01. In total, 236 ROIs were used to train the model and 217 ROI were extracted for validation.

Feature extraction using transfer learning

We modified the typical Inception network, and fine-tuned the modified Inception (M-Inception) model with pre-trained weights. During the training phase, the original Inception part was not trained, and we only trained the

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

modified part. The architecture of M-Inception is shown in **Table 1**. The difference between Inception and M-Inception in classification lies in the last fully-connected layers. We reduced the dimension of features before it was sent to the final classification layer. The training dataset made up of all those patches aforementioned. The Inception network is shown in **Table1**.

Table 1. The architecture of M-Inception

	Layer	Patch size/stride or remarks
Inception part	conv	3×3/2
	conv	3×3/1
	conv padded	3×3/1
	pool	3×3/2
	onv	3×3/1
	conv	3×3/2
	conv	3×3/1
	Inception	3x, 5x, 2x
	pool	8x8
	linear	logits
	softmax	classifier
Modified part	Fc1	$\begin{bmatrix} \text{batchnorm} \\ \text{dropout}(0.5) \\ \text{512d Linear} \end{bmatrix}$
	Fc2	$\begin{bmatrix} \text{batchnorm} \\ \text{dropout}(0.5) \\ \text{2d Linear} \end{bmatrix}$

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Prediction

After generating the features, the final step is to classify the pneumonia based on those features. Ensembling of classifiers was used to improve the classification accuracy. In this study, we combined Decision tree and Adaboost to produce the performance.

Performance evaluation metrics

We compared the classification performance using Accuracy, Sensitivity, Specificity, Area Under Curve (AUC), Positive predictive value (PPV), Negative predictive value (NPV), F1 score and Youden Index. TP and TN represent the number of true positive and true negative samples. FP and FN mean the number of false positive false negative samples. Sensitivity measures the ratio of positives that are correctly discriminated. Specificity measures the ratio of negatives that are correctly discriminated. AUC is an index to measure the performance of the classifier. NPV was used to evaluate the algorithm for screening and PPV was the probability of getting a disease when the diagnostic index is positive. Youden Index was the determining exponent of the optimal bound. F1 score was a measure of the accuracy of a binary model. Additionally, performance was evaluated with F-measure (F1) to compare the similarity and diversity of performance. $F_1 = 2g\frac{PregSen}{Pre+Sen}$

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Results

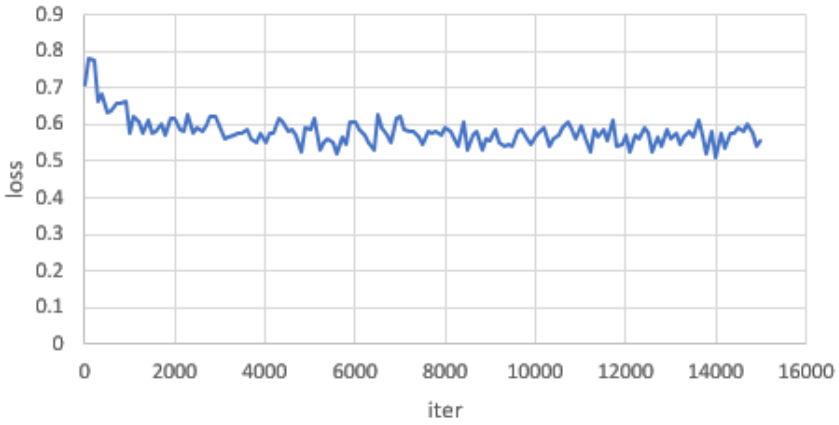
Algorithm development

In order to develop a deep learning algorithm for the identification of viral pneumonia images, we retrospectively enrolled 99 patients, in which the cohort includes 55 cases of typical viral pneumonia that were diagnosed previously before the COVID-19 outbreak. These patients are termed COVID-19 negative in the cohort. The other 44 cases are from the three hospitals with confirmed nucleic acid testing of SARS-COV-2, termed COVID-19 positive. Two radiologists were asked to review the images and sketched a total of 453 representative images (258 for COVID-19 negative and 195 for COVID-19 positive) for analysis. These images were randomly divided into a training set and a validation set. First, we randomly selected two or three images from all of the patients' images for training and the remaining images were used for internal validation. The model training has been iterated for 15,000 times with a step size of 0.01. The training loss curve is shown in **Figure 2A.**

To test the stability and generalization of the model, we also randomly selected 237 images (118 images from COVID-19 negative and 119 images from COVID-19 positive) to construct the model. The remaining images were then used for external validation. The model training has been iterated for 15,000 times with a step size of 0.01. The training loss curve is shown in **Figure 2B.**

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

A.



B.

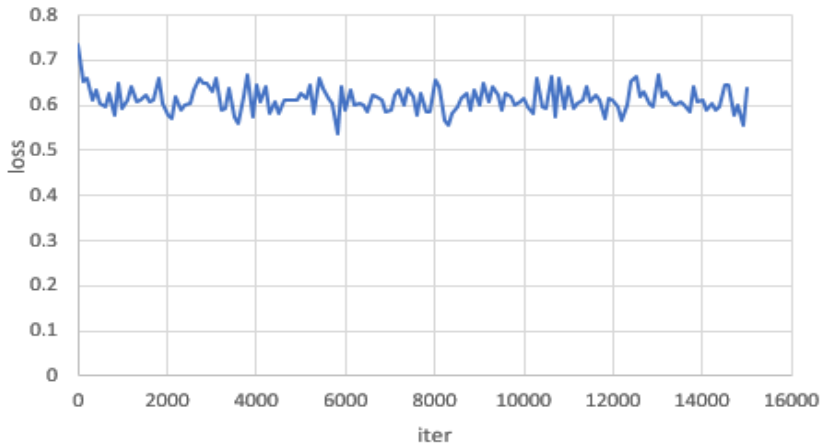


Figure 2 The training loss curve of the models on internal and external. The loss curve tends to be stable after descending, indicating that the training process converges.

Deep learning performance

The DL algorithm yielded an AUC of 0.90 (95% CI, 0.86 to 0.94) on the internal validation and 0.78 (95% CI, 0.71 to 0.84) on the external validation. The AUC was shown in **Figure 3.** Using the maximized Youdenindex threshold probability, the sensitivity was 80.5% and 67.1%, specificity 84.2% and 76.4%, theaccuracy was 82.9% and 73.1%, the negative prediction value was 0.88 and 0.81, the Youden index was 0.69 and 0.44 and F1 score was 0.77 and

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

0.64 on the internal and external datasets, respectively (**Table 2**). The algorithm was executed at a rate of xxxx seconds per case on the graphics processing unit.

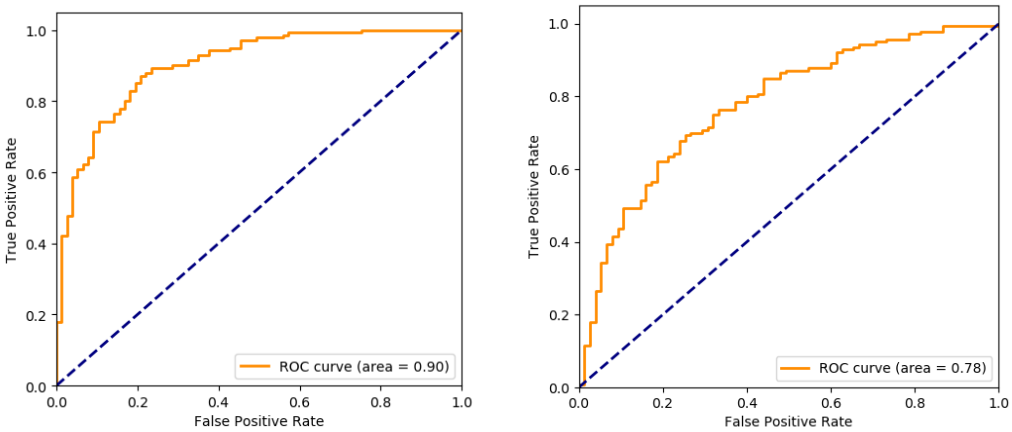


Figure 3 Receiver operating characteristic plots for pneumonia identification for deep learning (Inception) algorithm.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 2. Deep learning Algorithm Performance

Performance Metric	Internal	External
AUC（95%CI）	0.90(0.86 to 0.94)	0.78(0.71 to 0.84)
Accuracy, %	82.9	73.1
Sensitivity	0.81	0.67
Specificity	0.84	0.76
PPV	0.73	0.61
NPV	0.88	0.81
F1 score	0.77	0.63
Yoden index	0.69	0.44

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Discussion

Timely diagnosis and triaging of PUIs are crucial for the control of emerging infectious diseases such as the current COVID-19. Due to the limitation of nucleic acid -based laboratory testing, there is an urgent need to look for fast alternative methods that can be used by front-line health care personal for quickly and accurately diagnosing the disease. In the present study, we have developed an AI program by analyzing representative CT images using a deep learning method. This is a retrospective, multicohort, diagnostic study. We constructed an Inception migration neuro network that achieved 82.9% accuracy. Moreover, the high performance of the deep learning model we developed in this study was tested using external samples with 73% accuracy. These findings have demonstrated the proof of principle that deep learning can extract CT image features of COVID-19 for diagnostic purposes. Further developing this system can significantly shorten the diagnosis time for disease control. In addition, it can reduce the diagnostic workload of physicians in the field. Our study represents the first study to apply artificial intelligence technologies to CT images for effectively screening for COVID-19.

The gold standard for COVID-19 diagnosis has been nucleic acid based detection for the existence of specific sequences of the SARS-COV-2 gene. While we still value the importance of nucleic acid detection in the diagnosis of the viral infection, we must also note that the significantly high number of

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

false negatives due to several factors such as methodological disadvantages, disease stages, and methods for specimen collection might delay diagnosis and disease control. Recent data have suggested that the accuracy of nucleic acid testing is only about 30-50%. Using CT imaging feature extraction, we are able to achieve above 83% accuracy, significantly outplaying nucleic acid testing. In addition, this method is non-invasive with minimal cost. Although we are satisfied with the initial results, we believe that with more CT images included in the training, we will achieve higher accuracy. Therefore, further optimizing and testing this system is warranted. To achieve this, we have generated a webpage that licensed healthcare personnel can access to upload CT images for testing. The webpage information is as following: https://ai.nscc-tj.cn/thai/deploy/public/pneumonia_ct.

There are some limitations to our study. CT images present a difficult classification task due to the relatively large number of variable objects, specifically the imaged areas outside the lungs that are irrelevant to the diagnosis of pneumonia [11]. In our study, only one radiologist was involved in outlining the ROI area. In addition, the training data set is relatively small. The performance of this system is expected to increase when the training volume is increased. It should also be noted that, the features of the CT images we analyzed are from patients with severe lung lesions at later stages of disease development. A study to associate this with the progress and all pathologic stages of COVID-19 is necessary to optimize the diagnostic system.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

In future, we intend to link hierarchical features of CT images to features of other factors such as genetic, epidemiological and clinical information for multi-omics and multi-modeling analysis for enhanced disease diagnosis. The artificial intelligence system developed in our study could significantly contribute to COVID-19 disease control by reducing the number of PUIs for timely quarantine and treatment.

Reference

[1] Yang Y,Lu Q,,Liu M, Wang Y,Zhang A, Jalali N, et al. 2020, Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv doi: 10.1101/2020.02.10.20021675

[2] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama 2020.

[3] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020.

[4] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. The New England journal of medicine 2020.

[5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020.

[6] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 2020;25.

[7] Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, et al. Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clinical chemistry 2020.

[8] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 2020:200230.

[9] Gomez P, Semmler M, Schutzenberger A, Bohr C, Dollinger M. Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. Med Biol Eng Comput 2019;57:1451-63.

[10] Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. Radiology 2019;292:365-73.

[11] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 2018;172:1122-31 e9.

[12] Negassi M, Suarez-Ibarrola R, Hein S, Miernik A, Reiterer A. Application of

medRxiv preprint doi: https://doi.org/10.1101/2020.02.14.20023028; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects. World J Urol 2020.

[13] Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng 2018;2:741-8.

[14] Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH. Radiographic and CT Features of Viral Pneumonia. Radiographics 2018;38:719-39.