End-to-end COVID-19 screening with 3D deep learning on chest computed tomography

Kun Yang (yangkun209@mails.ucas.ac.cn)
Xingyi Normal University for Nationalities https://orcid.org/0000-0002-3802-2091
Xinfeng Liu
Guizhou Provincial People's Hospital
Yingli Yang
University of Chinese Academy of Sciences
Xiangjun Liao
Xingyi Normal University for Nationalities
Rongpin Wang (wangrongpin@126.com)
Guizhou Provincial People's Hospital
Xianchun Zeng
Guizhou Provincial People's Hospital
Yuxiang Wang
Ezhou Central Hospital
Mudan Zhang
Guizhou Provincial People's Hospital
Tijiang Zhang
Affiliated Hospital of Zunyi Medical University

Research Article

Keywords: COVID-19, 3D CNN, Chest CT, Rapid diagnosis

Posted Date: June 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36433/v1

Abstract

The outbreak of an acute respiratory syndrome (called novel coronavirus pneumonia, NCP) caused by SARS-CoV-2 virus has now progressed to a pandemic, and became the most common threat to public death worldwide[i],[ii]. COVID-19 screening using computed tomography (CT) can perform a quick diagnosis and identify high-risk NCP patients[iii]. Automated screening using CT volumes is a challenging task owing to inter-grader variability and high false-positive and false-negative rates. We propose a three dimensional (3D) deep learning convolutional neural networks (CNN) that use a patient's CT volume to predict the risk of COVID-19, trained end-to-end from CT volumes directly, using only images and disease labels as inputs. Our model achieves a state-of-the-art performance (95.78% overall accuracy, 99.4% area under the curve) on a dataset of 1,684 COVID-19 patients, nearly twice larger than previous datasets[3], and performs similarly on an independent clinical validation set of 121 cases. We tested its performance against six radiologists on clinical confirmed patient' CT volumes, our model outperformed all six radiologists with absolute reductions of 7% in false positives and 35.9% in false negatives, demonstrating artificial intelligence (AI) capable to optimize the COVID-19 screening process via computer assistance and automation with a level of competence comparable to radiologists. While the vast majority of patients remain unscreened, we show the potential for AI to increase the accuracy and consistency of COVID-19 screening with CT.

Introduction

The World Health Organization (WHO) officially declared the outbreak of a novel coronavirus, SARS-CoV–2, as a global pandemic. The virus is termed COVID–19, can cause fever, cough, and other flu-like symptoms, and the patients died over 60% once they progress rapidly into severe acute respiratory failure stage[4]. COVID–19 diagnosis is confirmed by a real-time molecular polymerase chain reaction (PCR) testing, but conservative estimates of the detection rate are low and several negative tests might be required in a single case to be confident about excluding this disease[5]. Chest CT radiography is an important tool for diagnosis of lung diseases, capable for COVID–19 screening. CT scanning procedure has a faster turnaround time than a viral PCR test in the screening of suspected cases. The majority of COVID–19 cases have similar morphology features and a peripheral lung distribution on CT images including ground-glass opacities (GGO) in the early stage and pulmonary consolidation in the late stage[5]. The CT images of various viral pneumonia are similar and they overlap with other infectious and inflammatory lung diseases. Therefore, it is difficult for radiologists to distinguish NCP from other common pneumonia (OCP) such as other viral pneumonia, bacterial pneumonia, and mycoplasma pneumonia.

Deep learning algorithms offered an exciting potential to automate analyze complex CT images, have recently been shown the potential to assist radiologists to improve diagnostic efficiency and accuracy in the COVID–19 screening[6]. Previous works in computer-aided COVID–19 screening had lacked the

generalization capability of medical practitioners owing to insufficient data and a focus on hand-engineered features such as GGO, consolidation, and fibrosis[3]".

Methods

AI system framework. In this paper, we aimed to build an end-to-end approach to extract features of full fine information contained within 3D structure of CT images, because the deep CNN has been shown superior to hand-engineered features in many competitions[7]. We used a 3D-CNN to perform COVID–19 risk categorization tasks with the CT images input alone, and comparison to radiologist based strictly on image classification. During inference, the model outputs a probability distribution over the three classes (NCP, OCP, and normal controls (NCs)). Fig. 1 shows the AI system. We utilize an inflated Inception v1 CNN architecture[8] that was pretrained on approximately 1.28 million images (1,000 object categories) from the ImageNet dataset, and train it on our dataset using transfer learning.

*Training algorithm.*We used 3D inflated Inception V1 architecture pretrained on ImageNet dataset. We removed the final classification layer from the network and retrained it with our dataset, fine-tuning the parameters across all layers to predict the probability of NCP, OCP, and NCs. We then used the last layer before the final probability, which contains 1,024 units. We took these 1,024 numbers as the outputs for this model, and used them as features later on. Our CNN was trained using backpropagation. All layers of the network were finetuned using the same global learning rate of 0.0001 and a decay factor of 0.1 every 3000 epochs. During training, we resized each image to 224 × 224 pixels in order to make it compatible with the original dimensions of i3d network architecture.

*Human Subjects.*The ethics committee of the Guizhou Provincial People's Hospital approved this retrospective study (Approval number: [2020]01), and waived the requirement for informed consent. All procedures on this study were performed in accordance with the relevant guidelines and regulations, and compliance with the principles of the Helsinki Declaration. All of the CT images used in this study were collected from public datasets and private datasets. Public datasets consists of LUNA and CC-CCII, which are publicly available: LUNA: https://luna16.grand-challenge.org/data/. CC-CCII: http://ncov-ai.big.ac.cn/download?lang = en. Private dataset collected CT volumes from 1,144 patients, in which includes 175 cases of typical viral pneumonia and the other 915 cases from four hospitals (Guizhou Provincial People's Hospital, Affiliated Hospital of Zunyi Medical University, Guizhou Jiangjunshan Hospital and Ezhou Central Hospital) with positive RT-PCT of SARS-COV–2. The private medical datasets were used with permission for this study, and not publicly available. Some of the data may be available for reasonable request and with permission of the Guizhou Provincial People's Hospital.

Experimental Data. We excluded scans without serial information, or containing less than 50 slices, or without qualified image resolution. Fig. 2 shows the detail exclusion criterion. After excluding, our dataset contained a total of 942,056 CT slices of 4,342 CT scans from 3,253 patients, and composed of 3,132 patients for training/internal validation, including 572,220 slices from 1,684 NCP patients, 218,987 slices from 1,055 other common pneumonia patients, and 240,200 slices from 914 normal controls (see Table

1). Scans were randomly assigned to a training set and an internal validation set with 10-fold cross validation method. NCP patient was given when a patient had pneumonia with a positive viral PCR test. The other common pneumonia patient was given when a patient had viral pneumonia (including influenza, parainfluenza pneumonia, adenoviral, and epstein-barr virus), bacterial pneumonia, and mycoplasma pneumonia. The normal controls were given by a public dataset LUNA, which patients were diagnosed without pneumonia.

Table 1 | CT dataset characteristics in identifying NCP from other common pneumonia and normal controls.

	NCP	Other common pneumonia	Normal controls
Patients	1,684	1,055	914
Scans	2,479	1,701	914
Slices	572,220	218,987	240,200
Independent dataset			
Patients	52	49	20
Scans	52	49	20
Slices	15,972	13,735	9,662

k-fold cross validation. We employed the 10-fold stratified cross validation method to estimate the our AI model uncertainties. The total dataset (4,221 scans) was randomly shuffled and partitioned into ten equal sized sub-datasets, with each sub-dataset containing equal number of data. Of the ten sub-datasets, a single sub-dataset (422 scans) was retained as the validation data for testing the model, and the remaining nine sub-dataset (3,799 scans) were used as training data. The cross-validation process was then repeated ten times, with each of the ten sub-datasets used exactly once as the validation data. The models were trained and tested ten times to obtain the average prediction accuracy and standard deviation.

Results

We validated the performances of the AI model in three ways. First, we validated the effectiveness using 10-fold cross-validation with a three-class disease partition, which represent NCP, OCP and NCs. In this task, the model achieved 95.78 ± 0.87% (mean ± s.d.) overall accuracy (the macro-average of individual inference class accuracies), and the macro-average area under the ROC curve (AUC) of 0.994 on an internal validation dataset (Fig. 3).

Second, We validated the performances of our model on an independent dataset (Fig. 3). The independent test dataset which consisted of a total of 39,369 CT slice images from 121 patients including 52 NCP patients, 49 other common pneumonia patients, and 20 normal controls. Our AI system achieved overall accuracy of 89.6% in the independent test dataset for three-way classification, and overall accuracy of 93.3% in the independent test dataset for NCP versus other two groups. The AI model exhibits generalization capability of COVID–19 screening when tested on independent dataset.

Third, We compared the performances of our model with six practicing radiologists on the same independent test dataset (Fig. 4). We employed six radiologists, all with more than 5 years of clinical experience. For each case, previously unseen, RT-PCT proven images were displayed, and radiologists were asked which case they thought was NCP, other common pneumonia, or normal control. A radiologist outputs a single prediction per cases. The green points in Fig. 4(a) is the average of the radiologists (average sensitivity and specificity of all solid points in Fig. 4(b)), with error bars denoting one standard deviation. Fig. 4(c) and Fig. 4(d) are the confusion matrixes. The 3D-CNN outperforms any radiologist whose sensitivity and specificity point falls below the blue curve of the model. Our AI system achieved overall accuracy = 93.30%, sensitivity = 98.08%, specificity = 91.30%, AUC = 0.994 for diagnosis of NCP versus other classes, outperformed all six radiologists with deductions of 7% in false positives and 35.9% in false negatives (example model false positives in Fig. 5), demonstrating end-to-end 3D-CNN capable to optimize the COVID–19 screening process.

We also examined the features in last hidden layer learned by our AI model with the t-SNE (t-distributed Stochastic Neighbour Embedding) method (see Fig. 6). We projected the 1,024-dimensional features of a chest CT volume into two dimensions, and represented as a single solid point. The red points represented NCP class, and were clustering in the lower left-hand side. In contrast, the blue points represented OCP class, and were clustering on the lower right-hand side. Similarly, the green points represented normal controls, and were clustering at the top.

Discussion

We demonstrated the effectiveness in diagnosing pneumonia with deep learning, a technique that we apply to chest whole–3D volume. Using an trained 3D-CNN model, we compared the performances with six dermatologists tested across three critical diagnostic tasks: NCP, OCP, and NCs classification. Our studies demonstrated the potential of 3D-CNN into diagnostic NCP which was found have improved clinical diagnostic efficiency and accuracy significantly in COVID–19 screening. Scalable application holds the potential for substantial clinical impact, including assist radiologists and physicians in performing a quick diagnosis and clinical decision-making for radiologists especially in this pandemic.

Further research is necessary to validate the performances of the AI model in real clinical situation, and develop a free access diagnose system. While we acknowledge that radiologists make their diagnoses based on factors such as clinical context rather than visual inspection of lesions in isolation, the ability of

our model to classify NCP classes accurately has the potential to assist radiologists and clinicians coping with the COVID–19 epidemic.

References

Abstract References

[i]. Munster, V. J., Koopmans, M., Van Doremalen, N., Van Riel, D., & de Wit, E. A novel coronavirus emerging in China - key questions for impact assessment. *N. Engl. J. Med.* **382**, 692-694, https://doi.org/10.1056/NEJMp2000929 (2020).

[ii]. Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470-473, https://doi.org/10.1016/S0140-6736(20)30185-9 (2020).

[iii]. Zhang, K. *et al*. Clinically applicable AI system for accurate diagnosis, quantitative measurements and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **181**, 1-11, https://doi.org/10.1016/j.cell.2020.04.045 (2020).

Main Text References

1. Munster, V. J., Koopmans, M., Van Doremalen, N., Van Riel, D., & de Wit, E. A novel coronavirus emerging in China - key questions for impact assessment. *N. Engl. J. Med.* **382**, 692-694, https://doi.org/10.1056/NEJMp2000929 (2020).
2. Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470-473, https://doi.org/10.1016/S0140-6736(20)30185-9 (2020).
3. Zhang, K. *et al*. Clinically applicable AI system for accurate diagnosis, quantitative measurements and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **181**, 1-11, https://doi.org/10.1016/j.cell.2020.04.045 (2020).
4. Guan, W. J. *et al*. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708-1720, https://doi.org/10.1056/NEJMoa2002032 (2020).
5. Huang, C. *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497-506, https://doi.org/10.1016/S0140-6736(20)30183-5 (2020).
6. Corman, V. M. *et al*. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro. Surveill.* **25**, https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045 (2020).
7. Chu, D. K. W. *et al*. Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. *Clin. Chem.* **66**,549-555, https://doi.org/10.1093/clinchem/hvaa029 (2020).
8. Zhang, N. *et al*. Recent advances in the detection of respiratory virus infection in humans. *J. Med. Virol.* **92**, 408-417, https://doi.org/10.1002/jmv.25674 (2020).
9. Jeffrey, P. K., Brent P. L., Jonathan H. C., Brett M. E. & Loren H. K. Essentials for radiologists on COVID-19: an update - radiology scientific expert panel. *Radiology*, https://doi.org/10.1148/radiol.2020200527 (2020).

10. Fang, Y. *et al*. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*, https://doi.org/10.1148/radiol.2020200432 (2020).
11. Shi, H. *et al*. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect. Dis.* **20**, 425-434, https://doi.org/10.1016/S1473-3099(20)30086-4 (2020).
12. Chung, M. *et al*. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology*, https://doi.org/10.1148/radiol.2020200230 (2020)
13. Gulshan, V. *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *J. Am. Med. Assoc.* **316**, 2402–2410, https://doi.org/10.1001/jama.2016.17216 (2016).
14. Esteva, A., Kuprel, B., Novoa, R. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542,** 115–118, https://doi.org/10.1038/nature21056 (2017).
15. Ardila, D. *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **25,** 954–961, https://doi.org/10.1038/s41591-019-0447-x (2019).
16. Huang, P. K. *et al*. Use of chest CT in combination with negative RT-PCR assay for the 2019 novel coronavirus but high clinical suspicion. *Radiology* **295**, 22-23, https://doi.org/10.1148/radiol.2020200330 (2020).
17. Shen C. *et al*. Quantitative computed tomography analysis for stratifying the severity of coronavirus disease 2019. *J. Pharm. Anal.* **10**, 123-129, https://doi.org/10.1016/j.jpha.2020.03.004 (2020).
18. Butt, C. Gill, J. Chun, D. & Babu, B. A. Deep learning system to screen coronavirus disease 2019 pneumonia. *Appl. Intell.* **22**, 1-7, https://doi.org/10.1007/s10489-020-01714-3 (2020).
19. Pinsky, P. F. *et al*. Performance of lung-RADS in the national lung screening trial: a retrospective assessment. *Ann. Intern. Med.* **162**, 485-491, https://doi.org/10.7326/m14-2086 (2015).
20. Liao, F., Liang, M., Li, Z., Hu, X. & Song, S. Evaluate the malignancy of pulmonary nodules using the 3D deep leaky noisy-or network. *IEEE T. Neur. Net. Lear.* **30**, 3484-3495, https://doi.org/10.1109/tnnls.2019.2892409 (2017).
21. Carreira, J. & Zisserman, A. Quo vadis, action recognition? A new model and the kinetics dataset. *IEEE CVPR 2017*, 4724-4733, https://doi.org/10.1109/cvpr.2017.502 (2017).
22. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception architecture for computer vision. *IEEE CVPR 2016*, 2818-2826, https://doi.org/10.1109/cvpr.2016.308 (2016).
23. Russakovsky, O. *et al*. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252, https://doi.org/10.1007/s11263-015-0816-y (2015).
24. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data. Eng.* **22**, 1345–1359, https://doi.org/10.1109/TKDE.2009.191 (2010).
25. Van der Maaten, L., & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605, http://www.jmlr.org/papers/v9/vandermaaten08a.html (2008).

Declarations

Acknowledgements

This study was supported by Guizhou Provincial Key Laboratory of Artificial Intelligence and Medical Diagnosis (KY [2018]006), Guizhou Science and Technology Project (QKHZC[2020]4Y002), and the Guiyang Science and Technology Project (ZKXM[2020]4). We thank many volunteers and physicians for curating grading CT scans and medical records.

Author Contributions

K.Y., Y.Y. and X.L. conceptualized and trained the algorithms. X.L. and R.W. collected data and oversaw the medical tasks, and recruited radiologists. K.Y. and R.W. supervised the project and designed the study. All authors contributed to data interpretation, and reviewed and approved the final version.

Competing interests

The authors have no competing interests to disclose.

Code availability

All experiments and implementation details are described in sufficient detail in the Methods section to allow independent replication with non-proprietary libraries. Trained models parameters are available in open source repositories: https://github.com/YangKun1988/COVID19; Several major components of our work are available in open source repositories: Tensorflow: https://www.tensorflow.org; Inflated Inception: https://github.com/deepmind/kinetics-i3d—the full-volume model were trained using this feature extractor.

Figures

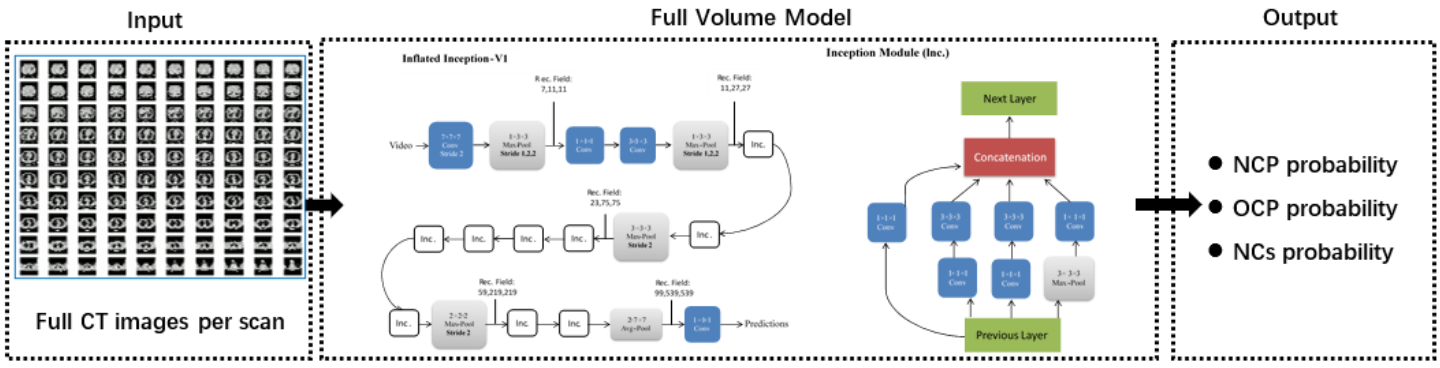

Figure 1

Overall modeling framework. Data flow is from left to right. For each patient, the AI model uses a full CT images and class label as inputs, then partition the whole CT images into a probability distribution over three classes (NCP, OCP, and NCs) using i3d architecture pretrained on the ImageNet dataset and fine-tuned on our own datasets, and outputs an overall probabilities for the case. The predicted label of an inference class is calculated by the most probable class.

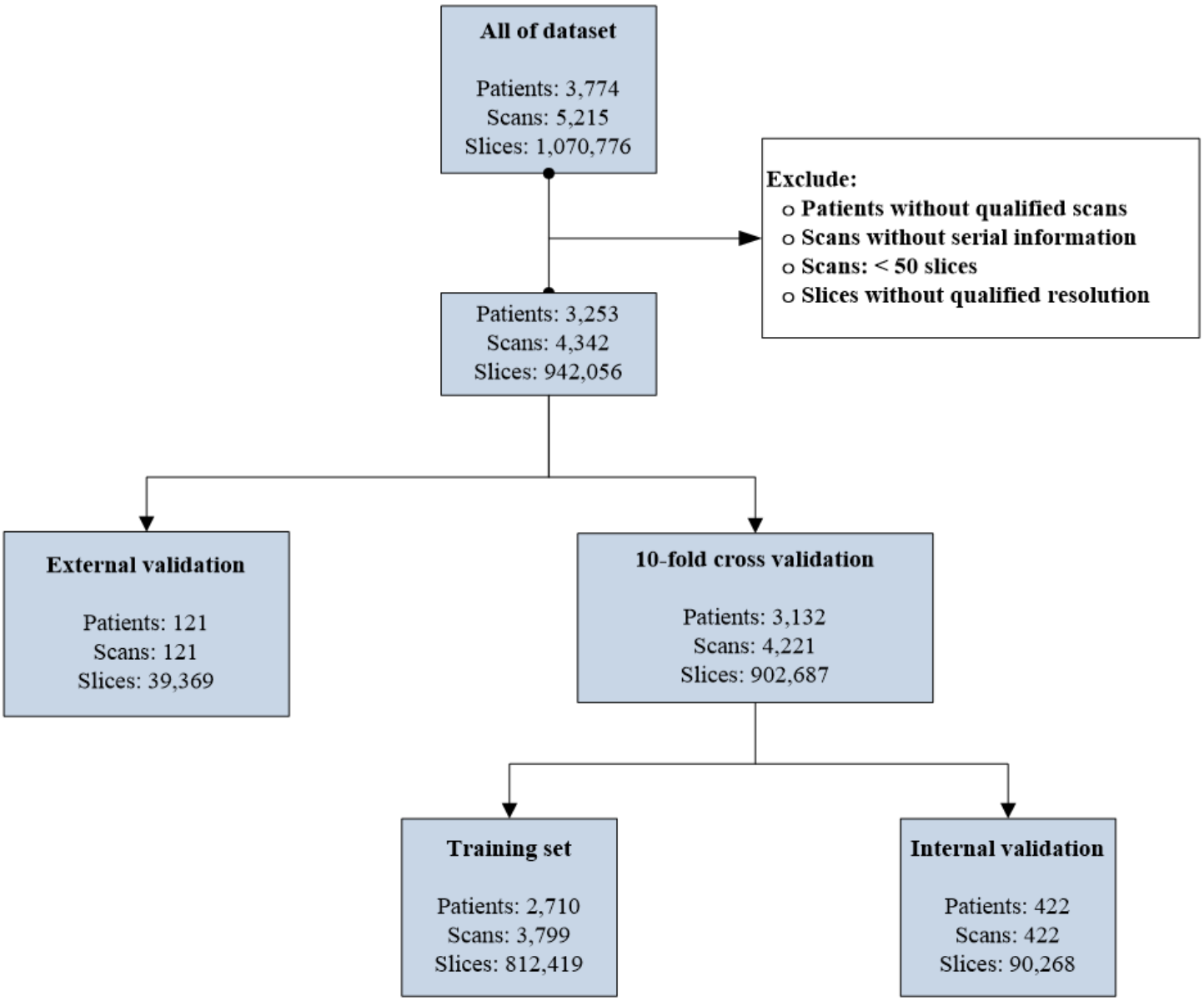

Figure 2

Diagram describing exclusions made in our analysis.

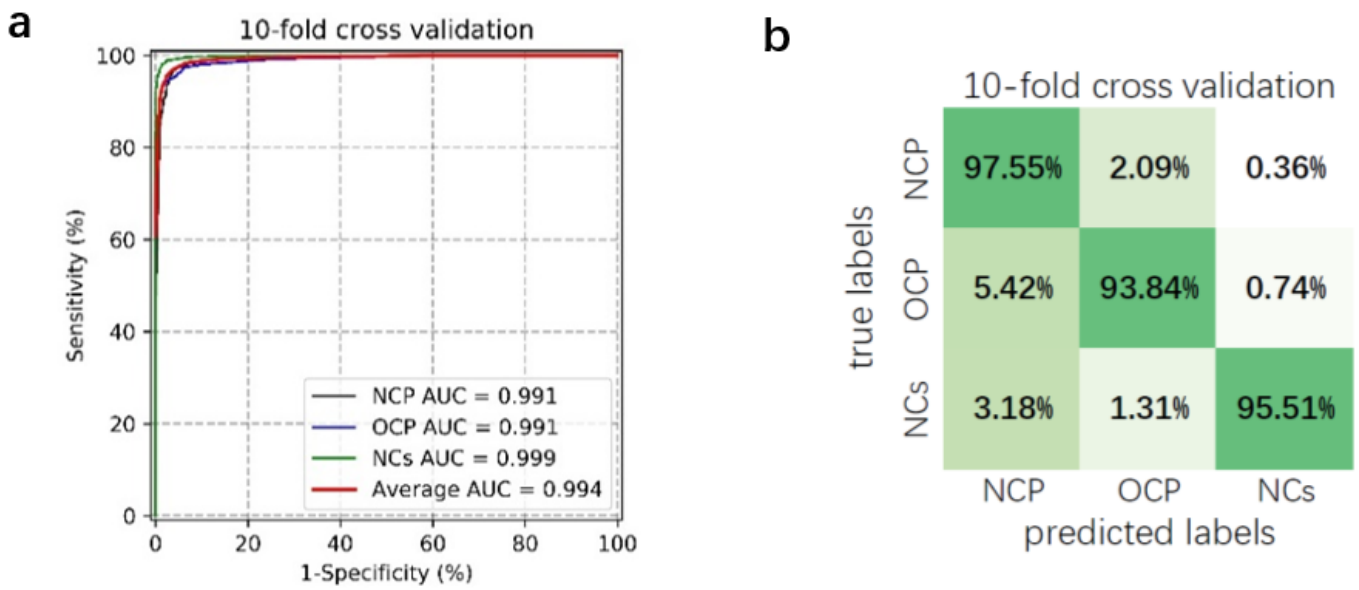

Figure 3

COVID-19 screening performance of our AI model using chest CT volume. a, The area under the ROC curve (AUC) of the deep learning 3D-CNN model performance. The red curve denoted macro-average AUC of NCP versus other two classes, including other common pneumonia (OCP) and normal controls (NCs). The value of AUC with the full validation dataset for NCP, OCP, and NCs classed was 0.991, 0.991, and 0.999 respectively. We observed negligible AUC changes (<0.004) in 10-fold cross validation, validating the reliability of our results on a larger dataset. b, Normalized confusion matrix of the AI model for identifying NCP cases from other common pneumonia (OCP) and normal controls (NCs). Our AI system achieved overall accuracy of 95.78 ± 0.87% in 10-fold cross validation for three-way classification, and overall accuracy of 96.32% ± 1.28% in 10-fold cross validation for NCP versus all other groups. The AI model exhibits reliable COVID-19 screening when tested on our dataset.

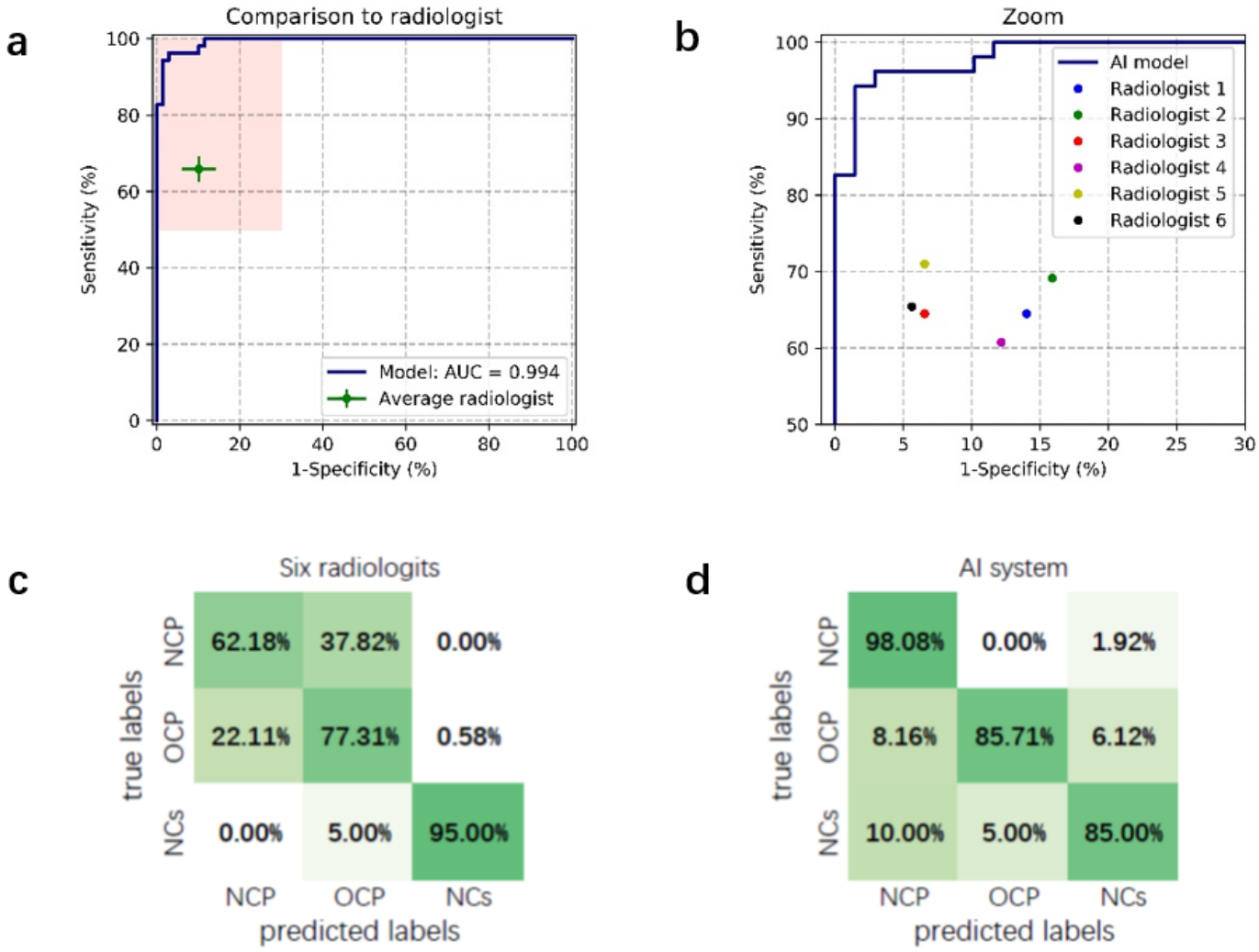

Figure 4

Comparing the performance of our AI model with six radiologists on an independent test dataset. a, Performance of model (blue line) versus average radiologist for three classes using a single CT volume. The length of the crosses represents the one standard deviation. b, The previous highlighted mistyrose area is magnified to show the performance of each of the six radiologists at various classes. Each solid point on the plots represents the sensitivity and specificity of a single radiologist. c, Confusion matrix of the mean diagnostic performance of six radiologists. d, Confusion matrix of the AI system performance comparable to that of senior practicing radiologists.

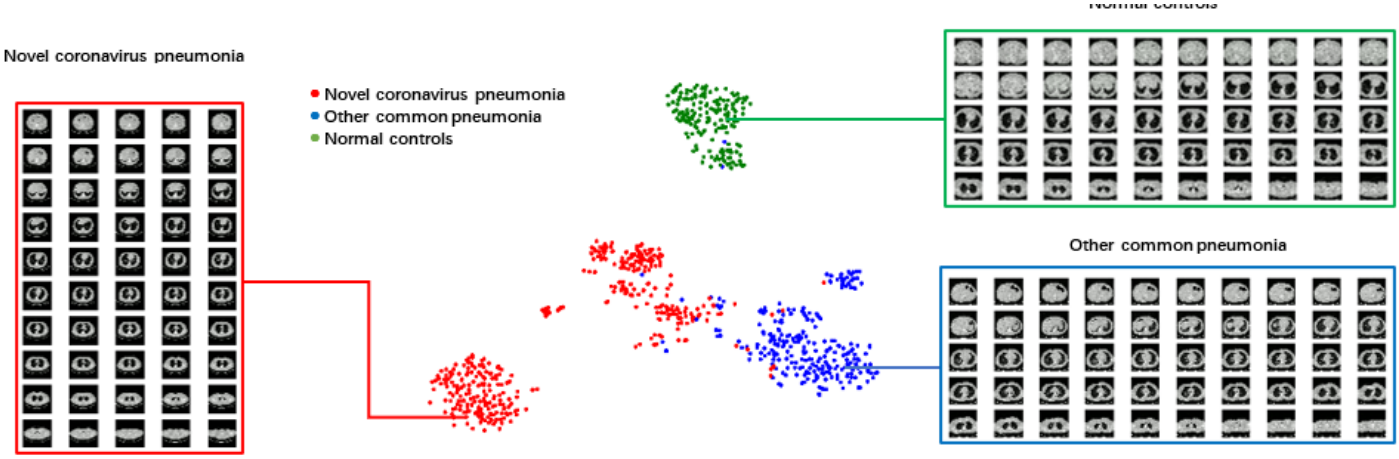

Figure 5

Examples of cases where comparing the model prediction with RT-PCR proven volumes. a, Example of COVID-19 positive case. b, Example of COVID-19 negative case. c, Example of COVID-19 false positive case. d, Example of COVID-19 false negative case.

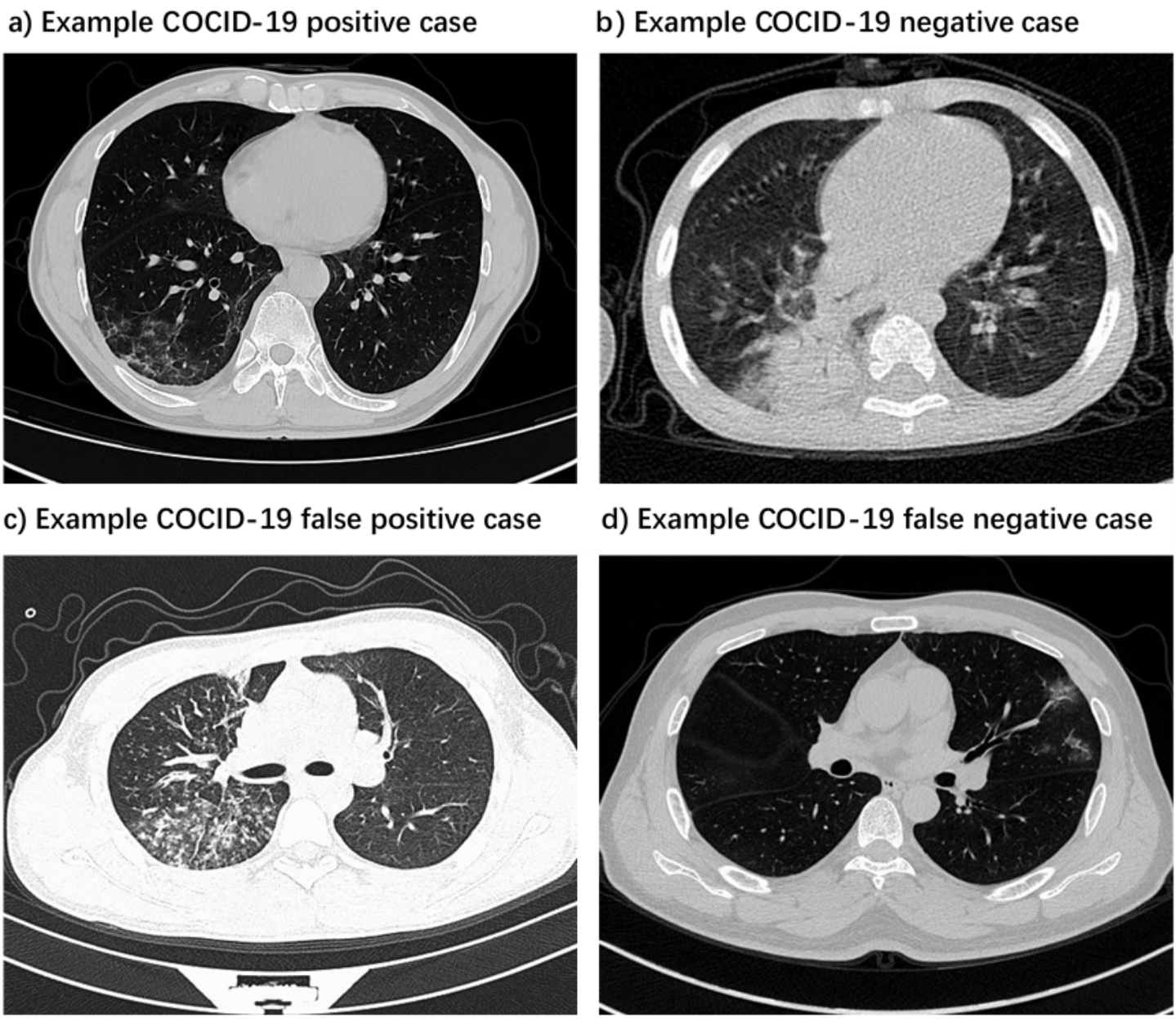

Figure 6

t-SNE visualization of the last hidden layer representations in the CNN for four disease classes.