COVID-NET CT-2: ENHANCED DEEP NEURAL NETWORKS FOR DETECTION OF COVID-19 FROM CHEST CT IMAGES THROUGH BIGGER, MORE DIVERSE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Computed tomography (CT) imaging has been widely used as a COVID-19 screening tool to complement RT-PCR testing. To assist radiologists with CT-based COVID-19 screening, a number of computer-aided systems using deep learning have been proposed, however many proposed systems are trained using CT data which is limited in both quantity and diversity. In this study, we attempt to address these limitations through the introduction of COVID-Net CT-2, enhanced deep neural networks for COVID-19 detection from chest CT images which are trained using a large, diverse, multinational patient cohort. We accomplish this by introducing two new benchmark CT image datasets, the largest of which comprises a multinational cohort of 4,501 patients from at least 15 countries. To the best of our knowledge, this represents the largest, most diverse multinational cohort for COVID-19 CT images in open-access form. The COVID-Net CT-2 neural networks achieved accuracy, COVID-19 sensitivity, and positive predictive value, of 98.1%/96.2%/96.7% and 97.9%/95.7%/96.4%, respectively. Moreover, explainability-driven performance validation shows that COVID-Net CT-2's decision-making behaviour is consistent with radiologist interpretation, as assessed by two board-certified radiologists with over 10 and 30 years of experience, respectively. The results are promising and suggest the strong potential of deep neural networks as an effective tool for computer-aided COVID-19 assessment.

1 INTRODUCTION

Real-time reverse transcription polymerase chain reaction (RT-PCR) testing remains the primary screening tool for COVID-19. However, despite being highly specific, the sensitivity of RT-PCR can be relatively low (Fang et al., 2020; Li et al., 2020) and can vary greatly depending on the time since symptom onset as well as sampling method (Yang et al., 2020; Li et al., 2020; Ai et al., 2020). To address this, computed tomography (CT) imaging has been explored as a complementary screening tool alongside RT-PCR (Fang et al., 2020; Ai et al., 2020; Xie et al., 2020). Early studies have shown that a number of potential indicators for COVID-19 infections may be present in chest CT images (Guan et al., 2020; Wang et al., 2020; Chung et al., 2020; Pan et al., 2020; Fang et al., 2020; Ai et al., 2020; Xie et al., 2020), but may also be present in non-COVID-19 infections. This can lead to challenges for radiologists in distinguishing COVID-19 infections from non-COVID-19 infections using chest CT (Bai et al., 2020; Mei et al., 2020).

To assist radiologists with CT-based COVID-19 screening, a number of computer-aided systems using deep learning have been proposed, but many of these systems are trained using CT data which is limited in both quantity and diversity. As such, in this study we introduce COVID-Net CT-2, enhanced deep convolutional neural networks for COVID-19 detection from chest CT images that are trained on a large, diverse, multinational patient cohort. We accomplish this through the introduction of two new benchmark CT image datasets (COVIDx CT-2A and COVIDx CT-2B), the largest of which comprises a multinational cohort of 4,501 patients from at least 15 countries. Finally, we leverage explainability to investigate the decision-making behaviour of COVID-Net CT-2 to ensure decisions are based on relevant visual indicators in CT images, with the results for select patient cases being reviewed and reported on by two board-certified radiologists with 10 and 30 years of experience, respectively.

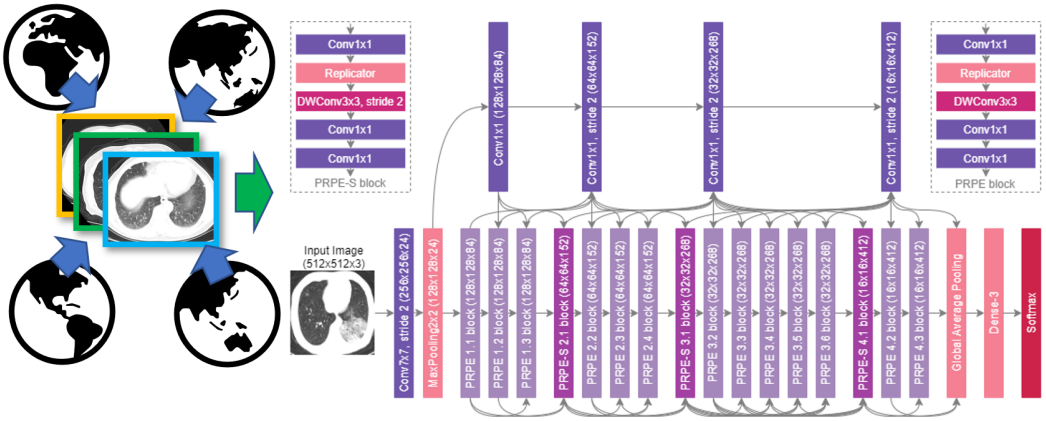

Figure 1: COVID-Net CT-2 network design and COVIDx CT-2 benchmark (COVID-Net CT-2 L network shown in figure).

2 METHODS

2.1 COVIDX CT-2 BENCHMARK DATASET

In this study, we created two diverse, large-scale benchmark datasets by unifying CT imaging data from several cohorts from around the world:

- **COVIDx CT-2A**: This benchmark dataset comprises 194,922 CT images from a multinational cohort of 3,745 patients between 0 and 93 years old (median age of 51) with strongly clinically-verified findings. The patient cohort consists of cases collected by seven different organizations and initiatives from around the world (Zhang et al., 2020; An et al., 2020; Rahimzadeh et al., 2020; Ning et al., 2020; Jun et al., 2020; Armato III et al., 2015; rad).
- **COVIDx CT-2B**: This benchmark dataset comprises 201,103 CT images from a multinational cohort of 4,501 patients between 0 and 93 years old (median age of 51) with a mix of strongly verified findings and weakly verified findings. COVIDx CT-2B consists of the multinational patient cohort we leveraged to construct COVIDx CT-2A combined with additional patient cases with weakly verified findings collected by Morozov et al. (2020). Notably, these additional cases are only included in the training dataset, and as such the validation and test datasets are identical to those of COVIDx CT-2A.

In both COVIDx CT-2 benchmark datasets, the findings for the chest CT volumes correspond to three different infection types: (1) novel coronavirus pneumonia due to SARS-CoV-2 viral infection (NCP), (2) common pneumonia (CP), and (3) normal controls. For CT volumes labelled as NCP or CP, slices containing abnormalities were identified and assigned the same labels as the CT volumes. The rationale for creating two different COVIDx CT-2 benchmark datasets stems from the availability of weakly verified findings (i.e., findings not based on RT-PCR test results or final radiology reports), which can be useful for further increasing the quantity and diversity of patient cases.

2.2 COVID-NET CT-2 CONSTRUCTION AND LEARNING

Two COVID-Net CT-2 networks (COVID-Net CT-2 L and COVID-Net CT-2 S) are created using the COVIDx CT-2 benchmark datasets introduced in the previous section, with both sharing the same macroarchitecture design but different numbers of parameters. The COVID-Net CT-2 L architecture is shown in Figure 1, and the networks are made publicly available[1]. We leverage the COVID-Net CT network architecture design proposed in Gunraj et al. (2020) as the core of the architecture designs of the COVID-Net CT-2 networks. These architecture designs were discovered automatically via a machine-driven design exploration process using generative synthesis (Wong et al., 2018). The result is highly customized architecture designs that strike a strong balance between complexity and representational power beyond what a human designer can achieve alone.

[1] https://anonymous_for_review

2.3 EXPLAINABILITY-DRIVEN PERFORMANCE VALIDATION

As with COVID-Net CT (Gunraj et al., 2020), we utilize GSInquire (Lin et al., 2019) to conduct explainability-driven performance validation. GSInquire provides an explanation of how a model makes a decision based on an input CT image by identifying a set of critical factors within the image that significantly impact the decision-making process of the deep neural network. This form of performance validation is particularly important in a clinical context, as predictions of patients' conditions can affect the health of patients via treatment and care decisions. The results obtained via GSInquire are reviewed and reported on by two board-certified radiologists. The first radiologist has over 10 years of experience, while the second radiologist has over 30 years of radiology experience.

3 RESULTS

3.1 QUANTITATIVE ANALYSIS

The test accuracy of the COVID-Net CT-2 networks and COVID-Net CT-1 (Gunraj et al., 2020) are shown in Table 1. It can be observed that COVID-Net CT-2 L and COVID-Net CT-2 S achieved strong test accuracies of 98.1% and 97.9%, respectively, on the COVIDx CT-2 test dataset, while at the same time possessing low architectural complexity ($\sim$1.4M parameters and $\sim$0.45M parameters, respectively) and low computational complexity ($\sim$4.18 GFLOPs and $\sim$1.94 GFLOPs). Compared to COVID-Net CT-1, it can be observed that COVID-Net CT-2 L and COVID-Net CT-2 S achieved 3.6% and 3.4% higher accuracy, respectively.

The sensitivity and positive predictive value (PPV) for each infection type on the COVIDx CT-2 test dataset is shown in Table 1. It can be observed that COVID-Net CT-2 L and COVID-Net CT-2 S were able to achieve both high COVID-19 sensitivity (96.2% and 95.7%, respectively) and high COVID-19 PPV (96.7% and 96.4%, respectively). Compared to COVID-Net CT-1, it can be observed that COVID-Net CT-2 L and COVID-Net CT-2 S achieved 16% and 15.5% higher COVID-19 sensitivity, respectively. At the cost of significantly lower COVID-19 sensitivity, COVID-Net CT-1 is able to achieve 0.9% and 1.2% higher COVID-19 PPV than COVID-Net CT-2 L and COVID-Net CT-2 S, respectively. From a clinical perspective, high sensitivity ensures few false negatives which would lead to missed patients with COVID-19 infections, whereas high PPV ensures few false positives which add an unnecessary burden on the healthcare system, which is already stressed due to the ongoing pandemic.

3.2 QUALITATIVE ANALYSIS AND RADIOLOGIST FINDINGS

The critical factors identified by GSInquire for example chest CT images from COVID-19-positive cases are shown in Figure 2. Overall, the GSInquire-generated visual explanations indicate that COVID-Net CT-2 L is mainly utilizing visible lung abnormalities to distinguish between COVID-19-positive and COVID-19-negative cases. In all four cases, COVID-Net CT-2 L detected them to be novel coronavirus pneumonia due to SARS-CoV-2 viral infection, which was clinically confirmed. The expert radiologist findings and observations with regards to the example images and critical factors are as follows:

Case 1 (leftmost of Figure 2). The first radiologist noted the presence of bilateral peripheral mixed ground-glass and patchy opacities with subpleural sparing, which is consistent with the identified critical factors leveraged by COVID-Net CT-2 L. The absence of large lymph nodes and effusion further helped the radiologist point to COVID-19 pneumonia. The degree of severity is observed to be moderate to high. It was confirmed by the second radiologist that the identified critical factors

Table 1: Accuracy, sensitivity and positive predictive value (PPV) for each infection type at the image level on the COVIDx CT-2 benchmark test dataset. Best results highlighted in **bold**.

		Sensitivity (%)			**PPV (%)**		
Network	**Accuracy (%)**	Normal	CP	NCP	Normal	CP	NCP
COVID-Net CT-2 L	**98.1**	**99.0**	98.2	**96.2**	**99.4**	**97.2**	96.7
COVID-Net CT-2 S	97.9	98.9	98.1	95.7	99.3	97.0	96.4
COVID-Net CT-1 (Gunraj et al., 2020)	94.5	98.8	**99.0**	80.2	96.1	90.2	**97.6**

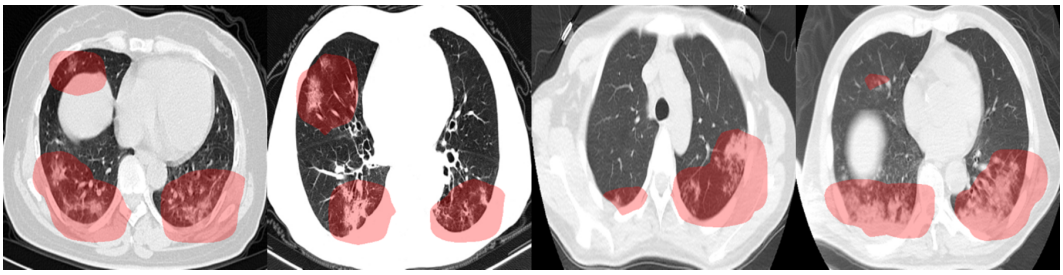

Figure 2: Example chest CT images from four COVID-19 cases and the associated critical factors (highlighted in red) as identified by GSInquire (Lin et al., 2019) for COVID-Net CT-2 L.

are correct areas of concern and represent areas of consolidation with a geographic distribution that is in favour of COVID-19 pneumonia.

Case 2 (second from left of Figure 2). The first radiologist noted the presence of bilateral peripherally-located ground-glass opacities with subpleural sparing, which is consistent with the identified critical factors leveraged by COVID-Net CT-2 L. As in Case 1, the absence of large lymph nodes and large effusion further helped the radiologist point to COVID-19 pneumonia. The degree of severity is observed to be moderate to high. It was confirmed by the second radiologist that the identified critical factors leveraged by COVID-Net CT-2 L are correct areas of concern and represent areas of consolidation with a geographic distribution that is in favour of COVID-19 pneumonia.

Case 3 (second from right of Figure 2). The first radiologist noted the presence of peripheral bilateral patchy opacities, which is consistent with the identified critical factors leveraged by COVID-Net CT-2 L. Unlike the first two cases, there is small right effusion. However, as in Cases 1 and 2, the absence of large effusion further helped the radiologist point to COVID-19 pneumonia. Considering that the opacities are at the base, a differential of atelectasis change was also provided. The degree of severity is observed to be moderate. It was confirmed by the second radiologist that the identified critical factors leveraged by COVID-Net CT-2 L are correct areas of concern and represent areas of consolidation.

Case 4 (rightmost of Figure 2). The first radiologist noted the presence of peripherally located asymmetrical bilateral patchy opacities, which is consistent with the identified critical factors leveraged by COVID-Net CT-2 L. As in Cases 1 and 2, the absence of lymph nodes and large effusion further helped the radiologist point to COVID-19 pneumonia, but a differential of bacterial pneumonia was also provided considering the bronchovascular distribution of patchy opacities. In addition, there is no subpleural sparing. This highlights the potential difficulties in differentiating between novel coronavirus pneumonia and common pneumonia. It was confirmed by the second radiologist that the identified critical factors leveraged by COVID-Net CT-2 L are correct areas of concern and represent areas of consolidation with a geographic distribution that is in favour of COVID-19 pneumonia.

4 Conclusions

In this work, we introduced COVID-Net CT-2, enhanced deep convolutional neural networks for COVID-19 detection from chest CT images trained on the largest quantity and diversity of multinational patient cases in research literature. Two new benchmark CT image datasets were introduced and used to facilitate the learning of COVID-Net CT-2, and these datasets represent the largest, most diverse, multinational cohorts of their kind available in open access form, spanning cases from at least 15 countries. Experimental results show that the COVID-Net CT-2 networks are capable of achieving strong test accuracy, sensitivity, and positive predictive value. These experimental results are particularly promising in terms of model generalization and applicability for use in different clinical environments given the much more diverse nature of the COVIDx CT-2 test dataset. Moreover, the explainability-driven validation process shows consistency between the decision-making process of COVID-Net CT-2 and radiologist interpretation. These results are promising and suggest the strong potential of deep neural networks as an effective tool for computer-aided COVID-19 assessment.

REFERENCES

Covid-19. *Radiopaedia*. URL https://radiopaedia.org/articles/covid-19-4.

Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: A report of 1014 cases. *Radiology*, 296(2):E32–E40, 2020. doi: 10.1148/radiol.2020200642. PMID: 32101510.

P. An, S. Xu, S. A. Harmon, E. B. Turkbey, T. H. Sanford, A. Amalou, M. Kassin, N. Varble, M. Blain, V. Anderson, F. Patella, G. Carrafiello, B. T. Turkbey, and B. J. WoodJ. Ct images in covid-19 [data set]. In *The Cancer Imaging Archive*, 2020. URL https://doi.org/10.7937/tcia.2020.gqry-nc81.

SG Armato III, G McLennan, L Bidaut, MF McNitt-Gray, CR Meyer, AP Reeves, B Zhao, DR Aberle, CI Henschke, Eric A Hoffman, EA Kazerooni, H MacMahon, EJR van Beek, D Yankelevitz, AM Biancardi, PH Bland, MS Brown, RM Engelmann, GE Laderach, D Max, RC Pais, DPY Qing, RY Roberts, AR Smith, A Starkey, P Batra, P Caligiuri, Ali Farooqi, GW Gladish, CM Jude, RF Munden, I Petkovska, LE Quint, LH Schwartz, B Sundaram, LE Dodd, C Fenimore, D Gur, N Petrick, J Freymann, J Kirby, B Hughes, AV Casteele, S Gupte, M Sallam, MD Heath, MH Kuhn, E Dharaiya, R Burns, DS Fryd, M Salganicoff, V Anand, U Shreter, S Vastagh, BY Croft, and LP. Clarke. Data from lidc-idri. In *The Cancer Imaging Archive*, 2015. URL http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX.

Harrison X. Bai, Ben Hsieh, Zeng Xiong, Kasey Halsey, Ji Whae Choi, Thi My Linh Tran, Ian Pan, Lin-Bo Shi, Dong-Cui Wang, Ji Mei, Xiao-Long Jiang, Qiu-Hua Zeng, Thomas K. Egglin, Ping-Feng Hu, Saurabh Agarwal, Fang-Fang Xie, Sha Li, Terrance Healey, Michael K. Atalay, and Wei-Hua Liao. Performance of radiologists in differentiating covid-19 from non-covid-19 viral pneumonia at chest ct. *Radiology*, 296(2):E46–E54, 2020. doi: 10.1148/radiol.2020200823. PMID: 32155105.

Michael Chung, Adam Bernheim, Xueyan Mei, Ning Zhang, Mingqian Huang, Xianjun Zeng, Jiufa Cui, Wenjian Xu, Yang Yang, Zahi A. Fayad, Adam Jacobi, Kunwei Li, Shaolin Li, and Hong Shan. Ct imaging features of 2019 novel coronavirus (2019-ncov). *Radiology*, 295(1):202–207, 2020. doi: 10.1148/radiol.2020200230. PMID: 32017661.

Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji. Sensitivity of chest ct for covid-19: Comparison to rt-pcr. *Radiology*, 296(2):E115–E117, 2020. doi: 10.1148/radiol.2020200432. PMID: 32073353.

Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu, Hong Shan, Chun-liang Lei, David S.C. Hui, Bin Du, Lan-juan Li, Guang Zeng, Kwok-Yung Yuen, Ru-chong Chen, Chun-li Tang, Tao Wang, Ping-yan Chen, Jie Xiang, Shi-yue Li, Jin-lin Wang, Zi-jing Liang, Yi-xiang Peng, Li Wei, Yong Liu, Ya-hua Hu, Peng Peng, Jian-ming Wang, Ji-yang Liu, Zhong Chen, Gang Li, Zhi-jian Zheng, Shao-qin Qiu, Jie Luo, Chang-jiang Ye, Shao-yong Zhu, and Nan-shan Zhong. Clinical characteristics of coronavirus disease 2019 in china. *New England Journal of Medicine*, 382(18):1708–1720, 2020. doi: 10.1056/NEJMoa2002032.

Hayden Gunraj, Linda Wang, and Alexander Wong. COVIDNet-CT: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images. *Frontiers in Medicine*, 2020. doi: 10.3389/fmed.2020.608525. URL https://doi.org/10.3389/fmed.2020.608525.

Ma Jun, Wang Yixin, An Xingle, Ge Cheng, Yu Ziqi, Chen Jianan, Zhu Qiongjie, Dong Guoqiang, He Jian, He Zhiqiang, Ni Ziwei, and Yang Xiaoping. Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation. *arXiv preprint arXiv:2004.12537*, 2020.

Yafang Li, Lin Yao, Jiawei Li, Lei Chen, Yiyan Song, Zhifang Cai, and Chunhua Yang. Stability issues of rt-pcr testing of sars-cov-2 for hospitalized patients clinically diagnosed with covid-19. *Journal of Medical Virology*, 92(7):903–908, 2020. doi: 10.1002/jmv.25786.

Zhong Qiu Lin, Mohammad Javad Shafiee, Stanislav Bochkarev, Michael St. Jules, Xiao Yu Wang, and Alexander Wong. Do explanations reflect decisions? a machine-centric strategy to quantify the performance of explainability algorithms, 2019.

Xueyan Mei, Hao-Chih Lee, Kai-yue Diao, Mingqian Huang, Bin Lin, Chenyu Liu, Zongyu Xie, Yixuan Ma, Philip Robson, Michael Chung, Adam Bernheim, Venkatesh Mani, Claudia Calcagno, Kunwei Li, Shaolin Li, Hong Shan, Jian Lv, Tongtong Zhao, Junli Xia, and Yang Yang. Artificial intelligence–enabled rapid diagnosis of patients with covid-19. *Nature Medicine*, pp. 1–5, 05 2020. doi: 10.1038/s41591-020-0931-3.

S.P. Morozov, A.E. Andreychenko, N.A. Pavlov, A.V. Vladzymyrskyy, N.V. Ledikhova, V.A. Gombolevskiy, I.A. Blokhin, P.B. Gelezhe, A.V. Gonchar, and V.Yu. Chernina. Mosmeddata: Chest ct scans with covid-19 related findings dataset. *medRxiv*, 2020. doi: 10.1101/2020.05.20.20100362. URL https://www.medrxiv.org/content/early/2020/05/22/2020.05.20.20100362.

W. Ning, S. Lei, J. Yang, et al. Open resource of clinical data from patients with pneumonia for the prediction of covid-19 outcomes via deep learning. *Nature Biomedical Engineering*, 4:1197–1207, 2020.

Feng Pan, Tianhe Ye, Peng Sun, Shan Gui, Bo Liang, Lingli Li, Dandan Zheng, Jiazheng Wang, Richard L. Hesketh, Lian Yang, and Chuansheng Zheng. Time course of lung changes at chest ct during recovery from coronavirus disease 2019 (covid-19). *Radiology*, 295(3):715–721, 2020. doi: 10.1148/radiol.2020200370. PMID: 32053470.

Mohammad Rahimzadeh, Abolfazl Attar, and Seyed Mohammad Sakhaei. A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset. *medRxiv*, 2020. doi: 10.1101/2020.06.08.20121541. URL https://www.medrxiv.org/content/early/2020/06/12/2020.06.08.20121541.

Dawei Wang, Bo Hu, Chang Hu, Fangfang Zhu, Xing Liu, Jing Zhang, Binbin Wang, Hui Xiang, Zhenshun Cheng, Yong Xiong, Yan Zhao, Yirong Li, Xinghuan Wang, and Zhiyong Peng. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. *JAMA*, 323(11):1061–1069, 03 2020. ISSN 0098-7484. doi: 10.1001/jama.2020.1585.

Alexander Wong, Mohammad Javad Shafiee, Brendan Chwyl, and Francis Li. Ferminets: Learning generative machines to generate efficient neural networks via generative synthesis, 2018.

Xingzhi Xie, Zheng Zhong, Wei Zhao, Chao Zheng, Fei Wang, and Jun Liu. Chest ct for typical coronavirus disease 2019 (covid-19) pneumonia: Relationship to negative rt-pcr testing. *Radiology*, 296(2):E41–E45, 2020. doi: 10.1148/radiol.2020200343. PMID: 32049601.

Yang Yang, Minghui Yang, Chenguang Shen, Fuxiang Wang, Jing Yuan, Jinxiu Li, Mingxia Zhang, Zhaoqin Wang, Li Xing, Jinli Wei, Ling Peng, Gary Wong, Haixia Zheng, Mingfeng Liao, Kai Feng, Jianming Li, Qianting Yang, Juanjuan Zhao, Zheng Zhang, Lei Liu, and Yingxia Liu. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-ncov infections. *medRxiv*, 2020. doi: 10.1101/2020.02.11.20021493.

Kang Zhang, Xiaohong Liu, Jun Shen, Zhihuan Li, Ye Sang, Xingwang Wu, Yunfei Zha, Wenhua Liang, Chengdi Wang, Ke Wang, Linsen Ye, Ming Gao, Zhongguo Zhou, Liang Li, Jin Wang, Zehong Yang, Huimin Cai, Jie Xu, Lei Yang, Wenjia Cai, Wenqin Xu, Shaoxu Wu, Wei Zhang, Shanping Jiang, Lianghong Zheng, Xuan Zhang, Li Wang, Liu Lu, Jiaming Li, Haiping Yin, Winston Wang, Oulan Li, Charlotte Zhang, Liang Liang, Tao Wu, Ruiyun Deng, Kang Wei, Yong Zhou, Ting Chen, Johnson Yiu-Nam Lau, Manson Fok, Jianxing He, Tianxin Lin, Weimin Li, and Guangyu Wang. Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. *Cell*, 18 (6):1423–1433, 2020.