Annals of Clinical Epidemiology 2022;4(4):110–119

Annals of Clinical Epidemiology

ORIGINAL ARTICLE

Development and external validation of a deep learning-based computed tomography classification system for COVID-19

Yuki Kataoka[1,2,3,4], Tomohisa Baba[5], Tatsuyoshi Ikenoue[6,7], Yoshinori Matsuoka[3,8], Junichi Matsumoto[9], Junji Kumasawa[6,10], Kentaro Tochitani[11], Hiraku Funakoshi[12], Tomohiro Hosoda[13], Aiko Kugimiya[14], Michinori Shirano[15], Fumiko Hamabe[16], Sachiyo Iwata[17], Yoshiro Kitamura[18], Tsubasa Goto[18], Shingo Hamaguchi[12], Takafumi Haraguchi[19], Shungo Yamamoto[11], Hiromitsu Sumikawa[20], Koji Nishida[21], Haruka Nishida[8], Koichi Ariyoshi[8], Hiroaki Sugiura[16], Hidenori Nakagawa[15], Tomohiro Asaoka[15], Naofumi Yoshida[22], Rentaro Oda[23], Takashi Koyama[24], Yui Iwai[24], Yoshihiro Miyashita[25], Koya Okazaki[26], Kiminobu Tanizawa[27], Tomohiro Handa[27,28], Shoji Kido[29], Shingo Fukuma[6], Noriyuki Tomiyama[30], Toyohiro Hirai[27], Takashi Ogura[5]

ABSTRACT

BACKGROUND

We aimed to develop and externally validate a novel machine learning model that can classify CT image findings as positive or negative for SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR).

METHODS

We used 2,928 images from a wide variety of case-control type data sources for the development and internal validation of the machine learning model. A total of 633 COVID-19 cases and 2,295 non-COVID-19 cases were included in the study. We randomly divided cases into training and tuning sets at a ratio of 8:2. For external validation, we used 893 images from 740 consecutive patients at 11 acute care hospitals suspected of having COVID-19 at the time of diagnosis. The dataset included 343 COVID-19 patients. The reference standard was RT-PCR.

RESULTS

In external validation, the sensitivity and specificity of the model were 0.869 and 0.432, at the low-level cutoff, 0.724 and 0.721, at the high-level cutoff. Area under the receiver operating characteristic was 0.76.

CONCLUSIONS

Our machine learning model exhibited a high sensitivity in external validation datasets and may assist physicians to rule out COVID-19 diagnosis in a timely manner at emergency departments. Further studies are warranted to improve model specificity.

KEY WORDS

COVID-19, Diagnosis, Computer-Assisted, Tomography, X-Ray Computed, Deep Learning, COVID-19 Nucleic Acid Testing

[1] Department of Internal Medicine, Kyoto Min-Iren Asukai Hospital
[2] Section of Clinical Epidemiology, Department of Community Medicine, Kyoto University Graduate School of Medicine
[3] Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health
[4] Scientific Research Works Peer Support Group (SRWS-PSG)
[5] Department of Respiratory Medicine, Kanagawa Cardiovascular and Respiratory Center
[6] Human Health Sciences, Kyoto University Graduate School of Medicine
[7] Graduate School of Data Science, Shiga University
[8] Department of Emergency Medicine, Kobe City Medical Center General Hospital
[9] Department of Emergency and Critical Care Medicine, St. Marianna University School of Medicine
[10] Department of Critical Care Medicine, Sakai City Medical Center
[11] Department of Infectious Diseases, Kyoto City Hospital
[12] Department of Emergency and Critical Care Medicine Department of Interventional Radiology, Tokyo Bay Urayasu Ichikawa Medical Center
[13] Department of Infectious Disease, Kawasaki Municipal Kawasaki Hospital
[14] Department of Respiratory Medicine, Yamanashi Prefectural Central Hospital
[15] Department of Infectious Diseases, Osaka City General Hospital
[16] Department of Radiology, National Defense Medical College
[17] Division of Cardiovascular Medicine, Hyogo Prefectural Kakogawa Medical Center
[18] Imaging Technology Center, Fujifilm Corporation
[19] Department of Radiology, St. Marianna University School of Medicine
[20] Department of Diagnostic Radiology, Sakai City Medical Center
[21] Department of Respiratory Medicine, Sakai City Medical Center
[22] Division of Cardiovascular Medicine, Department of Internal Medicine, Kobe University Graduate School of Medicine
[23] Department of Infectious Diseases, Tokyo Bay Urayasu Ichikawa Medical Center
[24] Department of Infectious Diseases, Hyogo Prefectural Amagasaki General Medical Center
[25] Department of Respiratory Medicine, Yamanashi Prefectural Central Hospital
[26] Department of Respiratory Medicine, Hyogo Prefectural Amgasaki General Medical Center
[27] Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University
[28] Department of Advanced Medicine for Respiratory Failure, Graduate School of Medicine, Kyoto University
[29] Department of Artificial Intelligence Diagnostic Radiology, Graduate School of Medicine, Osaka University
[30] Department of Diagnostic and Interventional Radiology, Osaka University Graduate School of Medicine

Corresponding author: Yuki Kataoka
Department of Internal Medicine, Kyoto Min-Iren Asukai Hospital, Tanaka Asukai-cho 89, Sakyo-ku, 606-8226 Kyoto, Japan
E-mail: youkiti@gmail.com

Received: March 8, 2022
Accepted: May 31, 2022
J-STAGE Advance published date: July 8, 2022

No. 22014

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has become a global pandemic. The disease is typically confirmed by reverse-transcription polymerase chain reaction (RT-PCR) testing; however, reports have suggested that the sensitivity of RT-PCR might be insufficient for early detection, and false-negative results can occur [1]. In addition, RT-PCR has a long turnaround time in facilities that require specimen transport [2]. To overcome these shortcomings, computed tomography (CT) has attracted attention as a complementary diagnostic method to RT-PCR.

Machine learning models can assist physicians in performing a rapid diagnosis during a pandemic, especially if the reviewer is not a senior radiologist [3]. Numerous studies over the past 2 years have focused on building machine learning models that detect COVID-19 based on CT images [4–8]. However, the actual usability of almost all presented methods in clinical practice is limited by methodological flaws or underlying biases [9]. One major pitfall is that machine learning can easily overfit the given datasets. For instance, when a model is trained with a single site or a few datasets, the datasets are more easily classified based on differences in imaging protocols rather than disease findings, resulting in very low classification accuracy for other datasets. This problem is known to occur even for large datasets [10]. Hence, it is necessary to employ a wide variety of datasets to develop and conduct external validation that uses data separate to that used during development in order to ensure proper evaluation of machine learning models [11].

In this study, we aimed to develop a novel machine learning model that can classify COVID-19 CT image findings and to externally validate the model using consecutive real-world data with rigorous methodology.

METHODS

We present the following article in accordance with the TRIPOD reporting checklist [11]. The study process is illustrated in **Fig. 1**. The Institutional Review Board of Osaka University Hospital (20511) and Kanagawa Cardiovascular and Respiratory Center (KCRC-20-0005) approved the development process protocol, and the need for written informed consent was waived. The Institutional Review Board of Hyogo Prefectural Amagasaki General Medical Center approved the external validation process, and the requirement for written informed consent was waived (2-96).

DEVELOPMENT

Development dataset

We curated 2,928 non-contrast-enhanced three-dimensional (3D) volumetric CT images from two medical centers and four publicly available datasets. We used COVID-19-positive images that were confirmed as positive by RT-PCR examination within 1 week of the imaging day. A total of 633 COVID-19 (positive) images from 576 patients were acquired from three facilities: Kanagawa Cardiovascular and Respiratory Center, Medical Imaging Databank of the Valencia Region (BIMCV) [12] public dataset, and CT images in COVID-19 [13] of the Cancer Imaging Archive (TCIA) [14] public dataset. In total, 2,295 non-COVID-19 (negative) images from 2,294 patients were obtained from the Osaka University Hospital, the National Lung Screening Trial (NLST) [15] public dataset, and Thoracic Volume and Pleural Effusion Segmentations in Diseased Lungs for Benchmarking Chest CT Processing Pipelines (PleThora) [16, 17] of the TCIA public dataset.

These datasets were randomly divided at a ratio of 8:2 into a multiple-site training dataset and tuning dataset that were used for training and selecting the best model, respectively. The details of the datasets, number of CT images, patient information, and imaging protocol of each dataset are presented in **Tables S1** and **S2**.

Deep learning model

We developed a 3D deep neural network model employing a U-net-like structure [18] to classify image findings as COVID-19 or non-COVID-19 (**Fig. S1**). The details are described in the supplement.

Internal validation

In order to clarify the importance of developing models using a variety of datasets, we evaluated the performance of two different models that were trained with single- or multiple-site datasets. The single-site model used only the Osaka University Hospital dataset for COVID-19-negative cases, while the multiple-site model used all datasets for training. Both models used the same datasets for COVID-19-positive cases. The performance was evaluated by the receiver operating characteristic (ROC) and specificity using the best model that outputs the highest area under the curve (AUC) value in the tuning dataset.

EXTERNAL VALIDATION PART A

We conducted external validation to evaluate the model's predictive performance in independent participant data

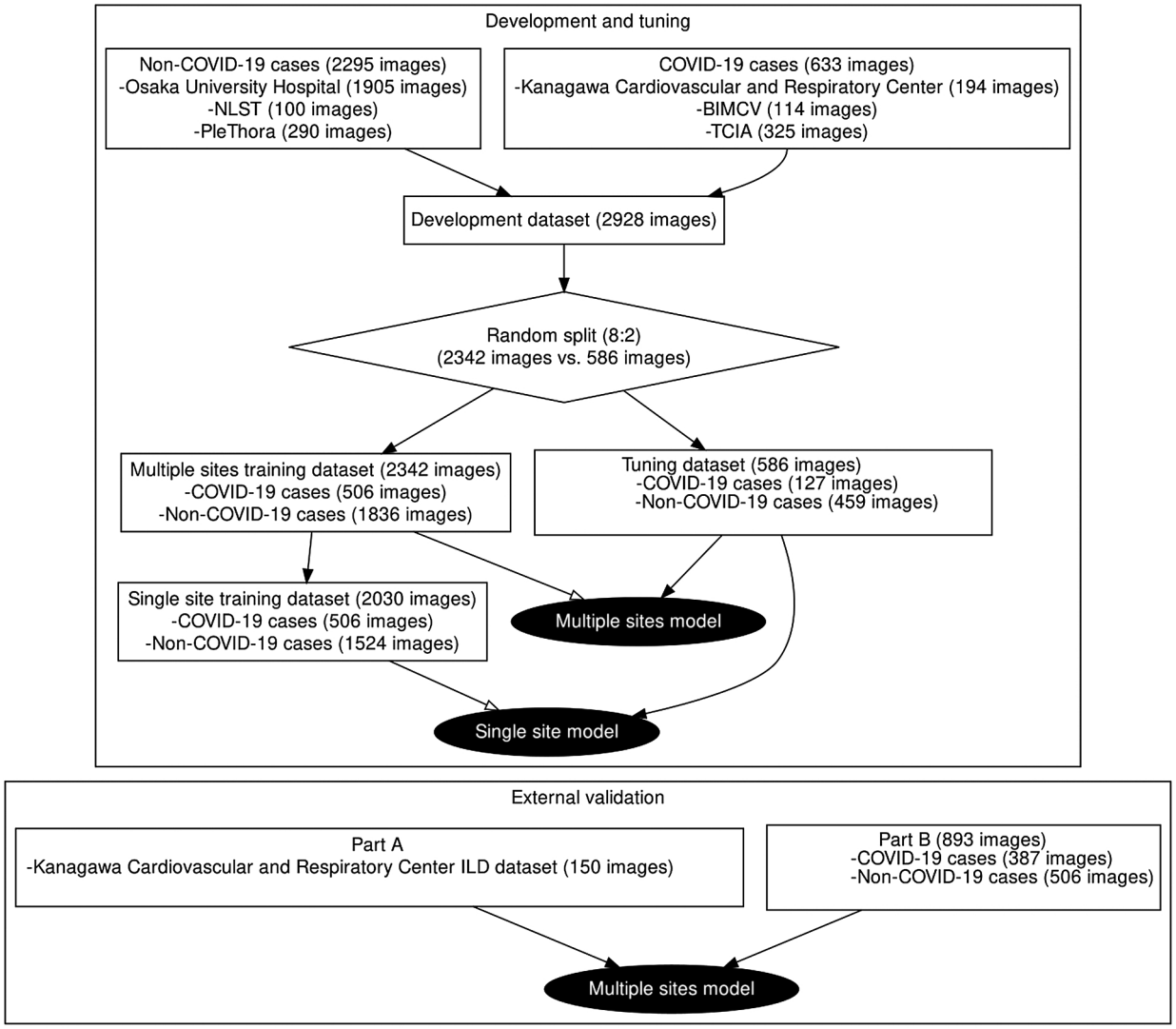

Fig. 1 **The whole study process**

Abbreviations: NLST, National Lung Screening Trial; PleThora, Thoracic Volume and Pleural Effusion Segmentations in Diseased Lungs for Benchmarking Chest CT Processing Pipelines; BIMCV, Medical Imaging Databank of the Valencia Region; TCIA, Cancer Imaging Archive; ILD, interstitial lung disease

that were not used for developing the model [11]. In addition to the training datasets described in **Tables S1** and **S2**, 150 CT images of patients with interstitial lung disease (ILD) acquired at Kanagawa Cardiovascular and Respiratory Center were used for the evaluation. These images were acquired before the pandemic and were thus considered COVID-19-negative. The breakdown of the image diagnosis was 50% for UIP, 30% for NSIP, and OP patterns. Receiver operating characteristic (ROC) curve, area under the curve (AUC), and specificity with a sensitivity of over 90% were used to assess model performance.

EXTERNAL VALIDATION PART B

We used the collected datasets to validate another COVID-19 diagnostic system (23,24). The external validation cohort consisted of 11 Japanese tertiary care facilities that played key roles in the treatment of COVID-19 cases in each area. We included patients who underwent both RT-PCR and chest CT for the diagnosis of COVID-19. Participants were selected on the basis that the physicians ordered both RT-PCR examinations and chest CT when the patients were suspected of having COVID-19 or were a close contact of COVID-19 patients. We included patients who underwent consecutive

sampling methods between January 1, 2020 and May 30, 2020. RT-PCR results were extracted from the patients' medical records at each facility. Patients were excluded if the time interval between chest CT and the first RT-PCR assay was longer than 7 days. We conducted source document verification (SDV) on all patients at three of the 11 hospitals. SDV is a verification of the conformity of the data presented in case report forms with source data [21].

Image analysis

Image analysis was performed using the aforementioned model that was trained with multiple-site datasets. The deep learning model was equipped with two cutoff values: the lower value was used to classify images into low or high (middle and high) confidence levels of COVID-19 image findings (referred to as low-level cutoff value), whereas the higher value was used to classify images into lower (low and middle) or high confidence levels (referred to as high-level cutoff value). The two values were determined as follows. First, an expert radiologist read each tuning dataset image and assigned the RSNA expert consensus statement, which categorizes images into four grades (1, typical; 2, indeterminate; 3, atypical; and 4, negative for pneumonia) following the criteria of CT findings (25). Next, we calculated a threshold value at which the deep learning model correctly classified grade 1 images with a sensitivity >90%; this was set as the high-level cutoff value. The low-level cutoff value corresponded to the threshold at which grade 1, 2, and 3 image sets were classified with a sensitivity >90%. In this step, we did not use the reference standard information.

Reference standard

RT-PCR tests were used as the main reference standard. For patients with a suspected COVID-19 infection, the diagnosis was made using RT-PCR to detect COVID-19 nucleic acid in sputum, throat swabs, and saliva samples. If more than one test was performed for a single episode, the patient was considered to have COVID-19 if any of the tests were positive. In this step, we did not use any other information.

Statistical analysis

ROC curves were plotted. AUC, sensitivity, and specificity were calculated, and a complete case analysis was conducted. We performed the main analysis by employing the thinnest slice for each analyzable case. We performed separate subgroup analyses for slices ≥5 mm and slices <5 mm, as well as for participants who underwent SDV. We performed a sensitivity analysis that considered the initial RT-PCR findings as the reference standard to evaluate the influence of the initial viral load. Statistical analyses were conducted using R statistical software, version 4.1.0 (R Foundation for Statistical Computing).

RESULTS

INTERNAL VALIDATION

Fig. S2 and **Table 1** present the classification performance of the two models trained with single- or multiple-site datasets in the tuning dataset. The multiple-site model exhibited high classification accuracy (ROC-AUC: 0.978) and high specificity for each dataset. In contrast, the performance of the single-site model was comparatively lower in the tuning dataset (ROC-AUC: 0.962). In particular, the specificities of the NLST and PleThora datasets were low. As each dataset included the same disease categories, we conjecture that the difference in performance was predominantly due to differences in imaging protocols. In this regard, most of the manufacturers of Osaka University Hospital were TOSHIBA. The average radiation dosages were lower in NLST and PleThora than in Osaka University Hospital (**Tables S1** and **S2**), which resulted in differential appearance of CT images, as illustrated in **Fig. S3**.

EXTERNAL VALIDATION PART A

We then evaluated how often the multiple-site model generated false positives for ILD images that were all COVID-19-negative. We calculated specificities for each ILD type by counting the number of images classified as COVID-19-positive when the sensitivity for COVID-19

Table 1 Summary of the results of deep learning models for tuning datasets

	ROC-AUC	Sensitivity	Specificity		
			Osaka University Hospital	NLST	PleThora
Multiple	0.978	0.906	0.944	0.931	0.986
Single	0.962	0.906	0.979	0.483	0.784

Abbreviations: ROC-AUC, Receiver Operating Characteristic – Area Under the Curve; NLST, National Lung Screening Trial; PleThora, Thoracic Volume and Pleural Effusion Segmentations in Diseased Lungs for Benchmarking Chest CT Processing Pipelines

CT images of the model was 0.906 (**Table 2**). Specificities for ILDs were high, with the exception of OP.

EXTERNAL VALIDATION PART B

Study population characteristics

The patient flow diagram is presented in **Fig. S4**. A total of 893 images from 740 patients were included in the study. **Table 3** presents the characteristics of the study population. In total, 343 patients were confirmed to have COVID-19 using RT-PCR. We were unable to obtain a diagnostic probability for 20 patients (comprising 21 images) due to abnormalities such as missing slices (n = 12, 50%), lack of DICOM tag information required for analysis (n = 5, 21%), and the inability to extract lung fields due to technical problems (n = 4, 17%).

Model performance

Multiple-site model performance for external validation datasets tested in 720 patients is presented in **Table 4** and **Fig. 2**. The sensitivity analysis is presented in **Table S5** and **Fig. S5**.

Table 2 **Classification results of the multiple-site model for interstitial lung disease images**

	Type of interstitial lung disease of processed CT image		
	UIP	NSIP	OP
Classified as COVID	1	9	18
Classified as non-COVID	69	41	12
Specificity	0.986	0.820	0.400

Abbreviations: UIP, usual interstitial pneumonia; NSIP, nonspecific interstitial pneumonia; OP, organizing pneumonia

Table 3 **Patient characteristics for external validation in consecutive sampling dataset**

Characteristic	RT-PCR positive, n = 343	RT-PCR negative, n = 377	Excluded, n = 20
Gender			
Female	135 (39%)	157 (42%)	8 (40%)
Male	208 (61%)	220 (58%)	12 (60%)
Age (years)	56 (44, 70)	68 (45, 79)	53 (45, 68)
Symptom	205 (60%)	194 (51%)	9 (45%)
Oxygen supplementation	86 (25%)	119 (32%)	7 (35%)
Positive with high-level cutoff	247	106	
Positive with low-level cutoff	298	215	

[1]n (%); median (interquartile range)
Abbreviations: RT-PCR, reverse transcription polymerase chain reaction.

Table 4 **Model performance for external validation datasets**

	Overall (n = 720)		≤5 mm (n = 524)		>5 mm (n = 343)		SDV (n = 217)	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
High-level cutoff	0.723	0.721	0.700	0.737	0.721	0.751	0.737	0.718
Low-level cutoff	0.869	0.432	0.876	0.423	0.883	0.487	0.877	0.379

Abbreviations: SDV, source data verification

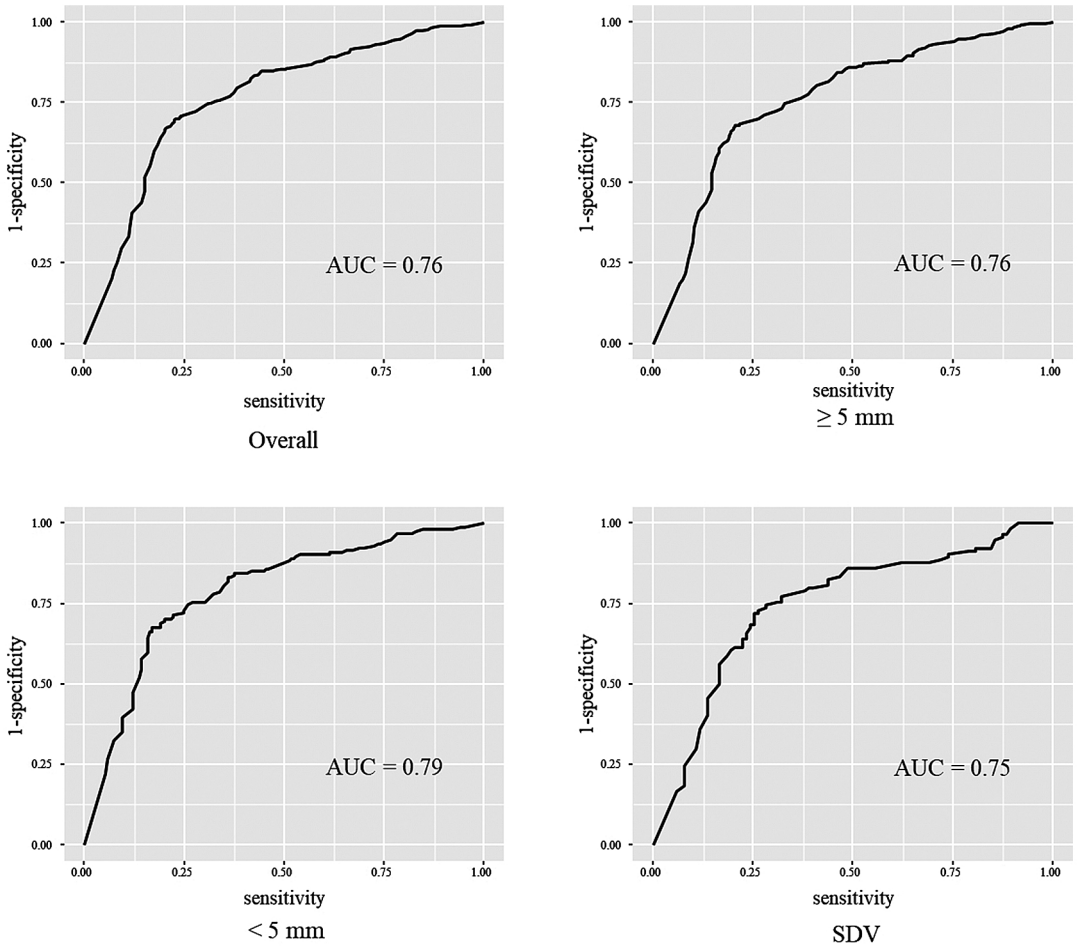

Fig. 2 **Receiver operating characteristic curves for external validation datasets**

Abbreviations: SDV, source data verification

DISCUSSION

In this study, we developed a novel machine learning model that was able to classify COVID-19 CT image findings. The internal validation revealed the superiority of the model trained using multiple datasets compared to that trained using a limited dataset. The model exhibited high specificity in the two patterns of interstitial pneumonia CT images, with the exception of OP. External validation revealed high sensitivity and moderate AUC. Collectively, these findings highlight the potential of this machine learning model to improve diagnostic speed for COVID-19. The AUC of the external validation dataset was lower than the AUC of the tuning dataset. We consider the reason is the difference of patient background between the cohorts. Since, most images of the external validation showed moderate to severe pneumonia, the difficulty of classification was very high.

Several studies using deep learning approaches to diagnose COVID-19 have been performed. Zhao et al. [4] reported sensitivity and specificity of approximately 90%, but they used two-gate datasets that would have resulted in overestimation [11]. Li et al. [6] achieved a sensitivity of 90%. However, they used randomly split internal validation, which would also have led to overestimation [11]. Silva et al. [5] and Wang et al. [7] conducted external validation, and their model achieved a sensitivity of <80%. Nevertheless, they only adopted a two-gate dataset for external validation. In this study, we conducted external validation of another deep learning model using the same dataset. The model achieved a sensitivity similar to that of our model [19]. In sum, a sensitivity of 90% may be the limit for CT diagnosis of COVID-19 to ensure specificity.

The high sensitivity of our model may be useful to rule out COVID-19 in a rapid and timely manner. Although antigen tests and RT-PCR tests are widely employed, our model can further reduce the probability of COVID-19 diagnosis, which is difficult to rule out based on a single test. In Japan, the lack of beds for patients with COVID-19 has led to a collapse of the healthcare system [23]. In this regard, our model may assist in the appropriate allocation of medical resources. Nevertheless, our model requires a certain radiation dosage when obtaining CT images to ensure good performance. In the internal validation, single-site model performance was poorer for the US open data (NLST and PleThora) than for Japanese university hospitals (Osaka University Hospital). In contrast, in the external validation study, the datasets from acute care hospitals in Japan exhibited a constant performance regardless of slice thickness. These findings support the applicability of our model to CT imaging with an appropriate radiation dose.

Our model could be applicable to patients with respiratory failure who are indicated for CT imaging at the emergency department. Due to the low prevalence and potential harm, chest imaging is not indicated as a screening test for COVID-19 in asymptomatic patients or in patients with mild respiratory symptoms [24]. On the contrary, in patients suspected with pneumonia visiting the emergency department, CT imaging would assist clinical management [25, 26]. In the "With-COVID-19 Era", the differentiation of COVID-19 pneumonia is essential among patients with infiltration shadow [27]. However, our external validation used an early 2020 dataset. Further studies accounting for the impact of vaccination and/or SARS-CoV-2 variants are required. In addition, our model is developed to classify CT images at the onset of disease and is not intended to be used in sequelae clinics to classify other ILDs.

Our study has several limitations. First, the low specificity of our model warrants improvements. In treatment strategies for COVID-19, rapid diagnosis is critical to prevent progression to severe illness [28, 29]. However, the sensitivity of CT remains low, even if COVID-19 is diagnosed by professional radiologists [30]. Several reports have indicated that supplementing CT findings with clinical information improves model accuracy [31, 32]. Further studies incorporating clinical information such as history and laboratory data are warranted to improve the specificity of our model. Second, while UIP and NSIP were classified as non-COVID-19 with high specificity, the performance for OP was very low. Bilateral multifocal consolidation accompanied by GGO is the main characteristic feature of cryptogenic OP, but such radiologic findings are also observed in late-phase complications of COVID-19 [33]. As such, it is difficult to distinguish OP from COVID-19 solely based on CT imaging findings. This is consistent with the comparison of chest CT findings for COVID-19 and OP by radiologists [34].

CONCLUSIONS

In conclusion, we have developed a CT-based machine learning model that is highly sensitive for diagnosing COVID-19. This diagnostic system may assist physicians to rule out COVID-19 in a timely manner at emergency departments. Further studies are warranted to improve the specificity of our model incorporating updated datasets. Another direction is the use of clinical information. Developing a multi-modality machine learning model would improve specificity.

CONFLICT OF INTEREST DISCLOSURES

All authors have completed the ICMJE uniform disclosure form.

Yoshiro Kitamura and Tsubasa Goto are employees of Fujifilm Corporation.

Shoji Kido is an endowed chair funded by Fujifilm Corporation.

Tomohisa Baba, Tomohiro Handa, Toyohiro Hirai, Noriyuki Tomiyama, Takashi Ogura, and Kiminobu Tanizawa received research grants from FUJIFILM Corporation.

Tomohiro Handa is an employee of the Collaborative Research Laboratory funded by Teijin Pharma Co., Ltd.

ACKNOWLEDGMENTS

The authors would like to thank the Multi-National NIH Consortium for CT AI in COVID-19 and the National Cancer Institute for access to NCI data collected by the National Lung Screening Trial (NLST). The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by the NCI. We also thank Ms. Kyoko Wasai, who assisted in data retrieval. We would like to thank Editage (www.editage.com) for English language editing. To access the system please contact Fujifilm Corporation (https://global.fujifilm.com/en/all-regions).

The development and external validation of the machine learning model presented in this article were sponsored by Fujifilm Corporation. Authors who were employed by the study sponsor were involved in the design of the study, preparation of the manuscript, and the decision to submit the manuscript for publication. Medical writing and editorial support were funded

by the sponsoring companies. The costs for source document verification (SDV) were paid for by Fujifilm Corporation. A contract research organization independent of Fujifilm Corporation conducted the SDV.

The first and second authors contributed equally to this work.

REFERENCES

1. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* 2020;296:E32–40. https://doi.org/10.1148/radiol.2020200642.

2. Kohli A, Joshi A, Shah A, Jain RD, Gorlawar A, Dhapare A, et al. Does CT help in reducing RT-PCR false negative rate for COVID-19? *Indian J Radiol Imaging* 2021; 31:S80–86. https://doi.org/10.4103/ijri.IJRI_739_20.

3. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 2020; 181:1423–33.e11. https://doi.org/10.1016/j.cell.2020.04.045.

4. Zhao W, Jiang W, Qiu X. Deep learning for COVID-19 detection based on CT images. *Sci Rep* 2021;11:14353. https://doi.org/10.1038/s41598-021-93832-2.

5. Silva P, Luz E, Silva G, Moreira G, Silva R, Lucio D, et al. COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. *Informatics Med Unlocked* 2020;20:100427. https://doi.org/10.1016/j.imu.2020.100427.

6. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 2020;296:E65–71. https://doi.org/10.1148/radiol.2020200905.

7. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020;56:2000775. https://doi.org/10.1183/13993003.00775-2020.

8. Alsharif MH, Alsharif YH, Yahya K, Alomari OA, Albreem MA, Jahid A. Deep learning applications to combat the dissemination of COVID-19 disease: A review. *Eur Rev Med Pharmacol Sci* 2020;24:11455–60.

9. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell* 2021;3:199–217. https://doi.org/10.1038/s42256-021-00307-0.

10. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. *Med Image Anal* 2021;72:102125. https://doi.org/10.1016/j.media.2021.102125.

11. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;162:55. https://doi.org/10.7326/M14-0697.

12. Vayá M de la I, Saborit JM, Montell JA, Pertusa A, Bustos A, Cazorla M, et al. BIMCV COVID-19+: a large annotated dataset of RX and CT images from COVID-19 patients. *ArXiv* 2020.

13. An P, Xu S, Harmon SA, Turkbey EB, Sanford TH, Amalou A, et al. CT Images in COVID-19 [Data set]. *Cancer Imaging Arch* 2020. https://doi.org/10.7937/TCIA.2020.GQRY-NC81.

14. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging* 2013;26:1045–57. https://doi.org/10.1007/s10278-013-9622-7.

15. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med* 2011;365:395–409. https://doi.org/10.1056/NEJMoa1102873.

16. Kiser KJ, Ahmed S, Stieb SM, Mohamed ASR, Elhalawani H, Park PYS, et al. Data from the Thoracic Volume and Pleural Effusion Segmentations in Diseased Lungs for Benchmarking Chest CT Processing Pipelines[Data set]. *Cancer Imaging Arch* 2020. https://doi.org/10.7937/tcia.2020.6c7y-gq39.

17. Aerts HJWL, Wee L, Rios Velazquez E, Leijenaar RTH, Parmar C, Grossmann P, et al. Data From NSCLC-Radiomics [Data set]. *Cancer Imaging Arch* 2020. https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI.

18. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation BT—Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors., Cham: Springer International Publishing; 2015, p. 234–41.

19. Ikenoue T, Kataoka Y, Matsuoka Y, Matsumoto J, Kumasawa J, Tochitatni K, et al. Accuracy of deep learning-based computed tomography diagnostic system for COVID-19: A consecutive sampling external validation cohort study. *PLoS One* 2021;16:e0258760. https://doi.org/10.1371/journal.pone.0258760.

20. Kataoka Y, Kimura Y, Ikenoue T, Matsuoka Y, Kumasawa J, Tochitatni K, et al. Integrated Model for COVID-19 Diagnosis Based on Computed Tomography AI, and Clinical Features: A Multicenter Cohort Study. *Res Sq* 2021. https://doi.org/10.21203/rs.3.rs-979599/v1.

21. Andersen JR, Byrjalsen I, Bihlet A, Kalakou F, Hoeck HC, Hansen G, et al. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. *Br J Clin Pharmacol* 2015;79:660–8. https://doi.org/10.1111/bcp.12531.

22. Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, et al. Radiological Society of North America Expert Consensus Document on Reporting Chest CT Findings Related to COVID-19: Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. *Radiol Cardiothorac Imaging* 2020;2:e200152. https://doi.org/10.1148/ryct.2020200152.

23. Ministry of Health L and W. The 50th Meeting of the Advisory Board for Countermeasures to Combat Novel Coronavirus Infections (September 1, 2021) n.d. https://www.mhlw.go.jp/content/10900000/000823680.pdf (accessed November 8, 2021).

24. Kwee TC, Kwee RM. Chest CT in COVID-19: What the Radiologist Needs to Know. *RadioGraphics* 2020;40:1848–65. https://doi.org/10.1148/rg.2020200159.

25. Claessens Y-E, Debray M-P, Tubach F, Brun A-L, Rammaert B, Hausfater P, et al. Early Chest Computed Tomography Scan to Assist Diagnosis and Guide Treatment Decision for Suspected Community-acquired Pneumonia. *Am J Respir Crit Care Med* 2015;192:974–82. https://doi.org/10.1164/rccm.201501-0017OC.

26. The Japanese Respiratory Society. The JRS guidelines for the management of pneumonia in Adults 2017. Tokyo: 2017.

27. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The Role of Chest Imaging in Patient Management During the COVID-19 Pandemic: A Multinational Consensus Statement From the Fleischner Society. *Chest* 2020;158:106–16. https://doi.org/10.1016/j.chest.2020.04.003.

28. Gupta A, Gonzalez-Rojas Y, Juarez E, Crespo Casal M, Moya J, Falci DR, et al. Early Treatment for Covid-19 with SARS-CoV-2 Neutralizing Antibody Sotrovimab. *N Engl J Med* 2021;385:1941–50. https://doi.org/10.1056/NEJMoa2107934.

29. Hammond J, Leister-Tebbe H, Gardner A, Abreu P, Bao W, Wisemandle W, et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N Engl J Med* 2022. https://doi.org/10.1056/NEJMoa2118542.

30. Liu J, Yang X, Zhu Y, Zhu Y, Liu J, Zeng X, et al. Diagnostic value of chest computed tomography imaging for COVID-19 based on reverse transcription-polymerase chain reaction: a meta-analysis. *Infect Dis Poverty* 2021; 10:126. https://doi.org/10.1186/s40249-021-00910-8.

31. Kataoka Y, Kimura Y, Ikenoue T, Matsuoka Y, Matsumoto J, Kumasawa J, et al. Integrated model for COVID-19 diagnosis based on

computed tomography artificial intelligence, and clinical features: a multicenter cohort study. *Ann Transl Med* 2022;10:130–. https://doi.org/10.21037/atm-21-5571.

32. Keshwani D, Kitamura Y, Li Y. Computation of Total Kidney Volume from CT Images in Autosomal Dominant Polycystic Kidney Disease Using Multi-task 3D Convolutional Neural Networks BT - Machine Learning in Medical Imaging. In: Shi Y, Suk H-I, Liu M, editors., Cham: Springer International Publishing; 2018, p. 380–8.

33. de Oliveira Filho CM, Vieceli T, de Fraga Bassotto C, da Rosa Barbato JP, Garcia TS, Scheffel RS. Organizing pneumonia: A late phase complication of COVID-19 responding dramatically to corticosteroids. *Brazilian J Infect Dis* 2021;25:101541. https://doi.org/10.1016/j.bjid.2021.101541.

34. Garrana SH, Som A, Ndakwah GS, Yeung T, Febbo J, Heeger AP, et al. Comparison of Chest CT Findings of COVID-19, Influenza, and Organizing Pneumonia: A Multireader Study. *Am J Roentgenol* 2021;217:1093–102. https://doi.org/10.2214/AJR.21.25640.