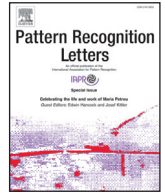




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A practical artificial intelligence system to diagnose COVID-19 using computed tomography: A multinational external validation study

Ali Abbasian Ardakani^a, Robert M. Kwee^b, Mohammad Mirza-Aghazadeh-Attari^c,
Horacio Matías Castro^d, Taha Yusuf Kuzan^e, Kübra Murzoğlu Altintoprak^e, Giulia Besutti^{f,g},
Filippo Monelli^{f,g}, Fariborz Faeghi^a, U Rajendra Acharya^{h,i,j}, Afshin Mohammadi^{k,*}

^a Radiology Technology Department, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

^b Department of Radiology, Zuyderland Medical Center, Heerlen/Sittard-Geleen, the Netherlands

^c Medical Radiation Sciences Research Group, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

^d Pulmonology Department, Hospital Italiano de Buenos Aires, Argentina

^e Department of Radiology, Sancaktepe Sehit Prof. Dr. Ilhan Varank Training and Research Hospital, Istanbul, Turkey

^f Radiology Department, Azienda USL - IRCCS di Reggio Emilia, Reggio Emilia, Italy

^g Clinical and Experimental Medicine PhD program, University of Modena and Reggio Emilia, Modena, Italy

^h Ngee Ann Polytechnic, Department of Electronics and Computer Engineering, 599489, Singapore

ⁱ Department of Biomedical Engineering, School of Science and Technology, SUSS University, Singapore

^j Department of Biomedical Informatics and Medical Engineering, Asia University, Taichung, Taiwan

^k Department of Radiology, Faculty of Medicine, Urmia University of Medical Science, Urmia, Iran

ARTICLE INFO

Article history:

Received 8 April 2021

Revised 14 August 2021

Accepted 16 September 2021

Available online 23 September 2021

Edited by Maria De Marsico

Keywords:

Artificial intelligence

Coronavirus infections

Machine learning

Pneumonia

Tomography, X-ray computed

ABSTRACT

Computed tomography has gained an important role in the early diagnosis of COVID-19 pneumonia. However, the ever-increasing number of patients has overwhelmed radiology departments and has caused a reduction in quality of services. Artificial intelligence (AI) systems are the remedy to the current situation. However, the lack of application in real-world conditions has limited their consideration in clinical settings. This study validated a clinical AI system, COVIDiag, to aid radiologists in accurate and rapid evaluation of COVID-19 cases. 50 COVID-19 and 50 non-COVID-19 pneumonia cases were included from each of five centers: Argentina, Turkey, Iran, Netherlands, and Italy. The Dutch database included only 50 COVID-19 cases. The performance parameters namely sensitivity, specificity, accuracy, and area under the ROC curve (AUC) were computed for each database using COVIDiag model. The most common pattern of involvement among COVID-19 cases in all databases were bilateral involvement of upper and lower lobes with ground-glass opacities. The best sensitivity of 92.0% was recorded for the Italian database. The system achieved an AUC of 0.983, 0.914, 0.910, and 0.882 for Argentina, Turkey, Iran, and Italy, respectively. The model obtained a sensitivity of 86.0% for the Dutch database. COVIDiag model could diagnose COVID-19 pneumonia in all of cohorts with AUC of 0.921 (sensitivity, specificity, and accuracy of 88.8%, 87.0%, and 88.0%, respectively). Our study confirmed the accuracy of our proposed AI model (COVIDiag) in the diagnosis of COVID-19 cases. Furthermore, the system demonstrated consistent optimal diagnostic performance on multinational databases, which is critical to determine the generalizability and objectivity of the proposed COVIDiag model. Our results are significant as they provide real-world evidence regarding the applicability of AI systems in clinical medicine.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a "once in a generation" pandemic. Conventionally the disease and its corresponding condition termed coro-

navirus disease 2019 (COVID-19) was diagnosed early by first-line clinicians or health workers and was confirmed with reverse transcription-polymerase chain reaction (RT-PCR) [1]. The high specificity of RT-PCR is guaranteed by the available diagnostic kits that target specific regions on the ribonucleic acid (RNA) of the virus, which could not be found on other subfamilies of the virus [2]. However, this comes at the cost of reduced sensitivity, as the detection of these sequences depends on high levels of technical and procedural efficacy. Early studies suggested

* Corresponding author at: Department of Radiology, Urmia University of Medical Sciences, Modarress Blvd, Ershad Ave, Urmia 12468, Iran.

E-mail address: afshin.mohadi@gmail.com (A. Mohammadi).

that RT-PCR's sensitivity was below 70 percent, as a subgroup of patients tested negative for the virus when asymptomatic, and showed positive tests after becoming symptomatic; furthermore, some patients who were symptomatic with COVID-19 never had positive RT-PCR test result [3–6].

To overcome these limitations and address the accessibility and availability of commercial tests, a handful of other diagnostic methods were considered, especially when RT-PCR is not available or the results are delayed [7]. One of these is chest CT, which was shown to be sensitive in diagnosing COVID-19 pneumonia and may be used as an adjunct to RT-PCR [8–10]. Early chest CT of the disease showed a dominant pattern, which consisted of bilateral ground-glass opacities (GGOs) located at the periphery of the lungs. In later stages, the patients developed airspace consolidation and fibrous band formation in some cases. Although no specific signs were seen in CT images, the distribution of opacities especially during the disease peak, when prevalence was high, made diagnosis possible by radiologists [11].

Chest CT was considered for initial triage of patients and was later incorporated into treatment guidelines as a proxy indicator of disease severity. Some studies also suggested a prognostic role for initial CT imaging obtained from the patients [12]. This overreliance on imaging, highlighted the inherent limitations of radiology practice. First of all, limited access to the proper imaging machines and expert radiologists may limit the wide-scale utilization of chest imaging in the frontlines [13]. Second, correct diagnosis of condition may be challenging as a wide range of conditions can possibly have similar radiologic presentations, such as other viral and atypical pneumonias [14]. Finally, an overload of patients and imaging outputs may heavily burden the radiology units, and cause attrition of infrastructure and human resources, reducing the quality of image interpretation [15,16]. To mitigate these limitations, a novel computer-aided diagnosis (CAD) systems is proposed.

CT images are used as an input to the CAD systems, as they are widely used in clinical settings and have achieved acceptable accuracy in detecting COVID-19. These systems have been fed by either images or the interpretation of the images [17,18]. Such a system can be an optimal tool in aiding the diagnosis of COVID-19, as it relies on limited information from human source, is fast, accurate, and would omit subjective assessment of relations between different inputs, as radiologists often do. In fact, objectifying this step of diagnosis may prove to be detrimental in any form of evidence-based clinical decision-making.

In order to achieve a desirable level of generalizability for any CAD system, it needs to be validated using multinational external databases in order to determine the equal performance in different clinical settings, imaging protocols, populations, and countries. Furthermore, these systems should be compared to radiologists, to see if their relative efficacy would help in real clinical situations [19]. Importantly, most of the AI systems currently being considered for COVID-19 detection, have not been compared with radiologists [20]. In the present study, we validate the clinical CAD system (COVIDiag) that yielded better diagnostic performance than radiologists. Also, we have tested the performance of the model, using databases from different countries to determine its efficiency in daily routine practices [17]. In this work, the COVIDiag model was validated using five databases from five different countries. The proposed COVIDiag software is made freely available and attached to the article.

2. Patients and Methods

The present study reports the results of the real-world application of an AI system based on CT findings of COVID-19 and non-COVID-19 patients. The COVIDiag model was built and validated (internal validation) in our previous publication [17]. For external

validation purposes, the databases used in this study were gathered independently from different countries and were interpreted by independent readers.

2.1. Patients

This study was approved by the Institutional Review Board of respective centers. In this study, laboratory-confirmed COVID-19 and non-COVID-19 patients from five centers from different countries and continents were included: 1- Hospital Italiano de Buenos Aires, city of Buenos Aires, Argentina; 2- University of Health Sciences, Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital, Istanbul, Turkey; 3- Imam-Asadabadi Hospital, Tabriz University of Medical Sciences, Tabriz, Iran; 4- Zuyderland Medical Center, Heerlen/Sittard-Geleen, The Netherlands; and 5- AUSL-IRCCS di Reggio Emilia, Reggio Emilia, Italy. The COVID-19 confirmation was done by RT-PCR test based on nasopharyngeal and/or pharyngeal swabs. Fig. 1. demonstrates the geographical distribution of the databases.

Exclusion criteria consisted of those individuals with suggestive clinical symptoms but negative RT-PCR test results, or those who had pre-existing chronic lung conditions, such as silicosis, idiopathic fibrosis, cystic fibrosis, etc. Patients with a time gap between PCR and CT examination greater than 3 days were excluded. The non-COVID-19 group consisted of those individuals with other causes of atypical and viral pneumonia, acquired from before the COVID-19 pandemic to eliminate all uncertainties and confounding factors.

2.2. CT imaging and features extraction

All cases had undergone high-resolution CT (HRCT) imaging. The HRCT protocols of each imaging center are presented in Table 1. Fifteen CT features, which were useful to diagnose COVID-19 cases [17], were extracted by an expert pulmonologist with 6 years (Argentinian), and expert radiologist with 5 years (Dutch), 6 years (Turkish), 5 years (Iranian), 5 years (Italian) of experience in thoracic imaging:

- a) Location of the lesion, 1 (side of involvement: unilateral vs. bilateral)
- b) Location of the lesion, 2 (position of involved lobes: lower, upper, or both lobes)
- c) Distribution pattern (central, peripheral or both)
- d) Number of lesions (single, if one patch of a lesion existed; multiple, if 2–4 patches of lesions existed; diffuse, if the entire lobe affected bilaterally)
- e) GGO (hazy, ill-defined opacities which do not obscure the underlying lung parenchyma)
- f) Consolidation (an opacification of the lung field which obscures the underlying lung parenchyma)
- g) Reticular opacity (linear lesions with a thickness equal or less than 3mm)
- h) Nodules (round or oval lesions with well-defined margins, without regard to the diameter of the lesion)
- i) Bronchial wall thickening (abnormal thickening of bronchial walls usually due to an inflammatory response)
- j) Air bronchogram (the visibility of air-filled bronchi by the opacification of alveoli)
- k) Cavity
- l) Crazy-paving (the simultaneous appearance of GGO and interlobar and intralobular septal thickening)
- m) Pleural effusion (abnormal accumulations of fluid within the pleural space. This term does not consider the characteristics of the fluid accumulated)
- n) Pleural thickening (thickening of either the parietal or visceral pleura, which is less than 5mm in most benign causes)

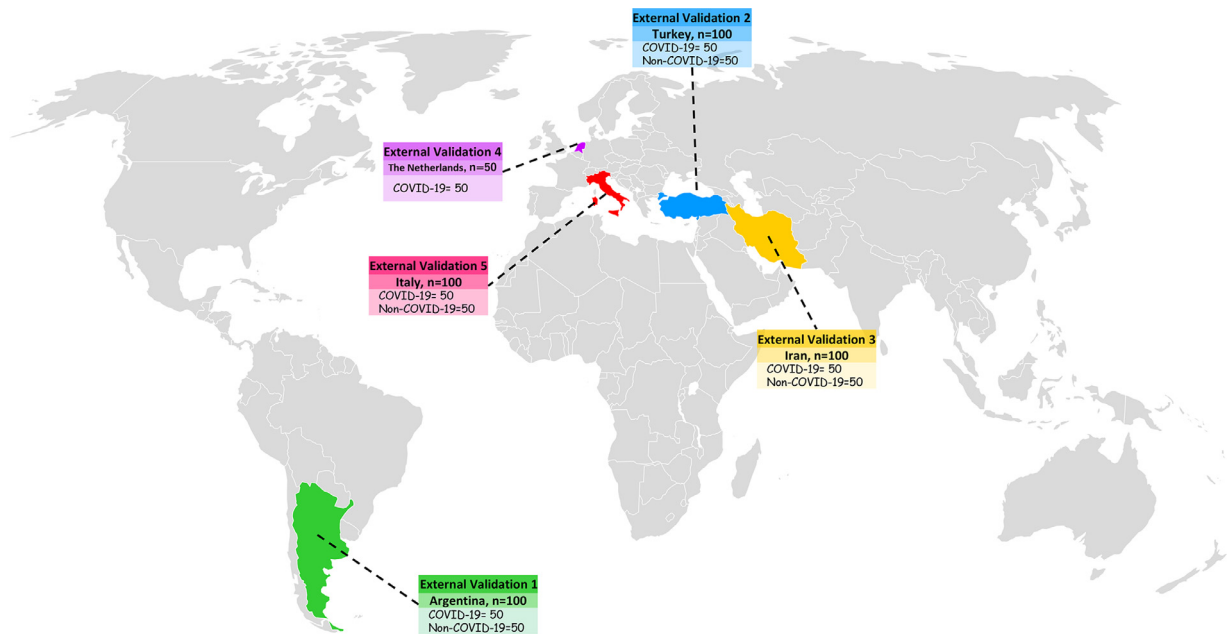


Fig. 1. Geographical distribution of databases used in this study.

Table 1

Computed tomography parameters of five centers used in this study.

Center	Scanner	Tube voltage, kVp	Tube current, mAs	Pitch	Matrix	Reconstruction slice thickness, mm	Reconstruction algorithm
Argentina	Canon Aquilion 64	120	50–100	1.5	512 × 512	1	Adaptive iterative Dose Reduction 3D (AIDR3D)
	Canon Activion 16	120	50–100	0.9	512 × 512	2	Adaptive iterative Dose Reduction 3D (AIDR3D)
Turkey	GE Optima 520	120	100–200	0.8–2.0	512 × 512	1.25	Adaptive Statistical Iterative Reconstruction (ASIR)
Iran	Siemens SOMATOM scope	120	50–100	0.8–1.5	512 × 512	1.5	Model-based iterative reconstruction (MBIR)
The Netherlands	Philips Incisive	120	73	1.0	512 × 512	1	Iterative
	Siemens SOMATOM Definition Flash	120	85	1.2	512 × 512	1	Iterative
Italy	Siemens SOMATOM Definition Edge	120	50–150	1.2	512 × 512	1	Advanced Modeled Iterative Reconstruction (ADMIRE)

o) Lymphadenopathy (the existence of a lymph node with a size greater or equal to 10 mm in the short axis)

2.3. Artificial intelligence technique

In this study, the COVIDiag model was used to evaluate its validity to differentiate COVID-19 from non-COVID-19 cases. The COVIDiag model is based on ensemble learning, which succeeded in diagnosing COVID-19 and non-COVID-19 cases with an area under the ROC curve (AUC) of 0.988 and 0.965 in the training and testing dataset of Ardakani et al. study [17]. All the fifteen CT findings were fed into the COVIDiag model in the proposed order from each database. The COVIDiag software is made freely available and attached to the article.

2.4. Statistical analysis and performance analysis

To evaluate the model, the following parameters were determined for each of the databases:

$$\text{Sensitivity (Sen)} : \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$\text{Specificity (Spc)} : \frac{N_{TN}}{N_{TN} + N_{FP}}$$

$$\text{Accuracy (Acc)} : \frac{N_{TN} + N_{TP}}{N_{TN} + N_{FN} + N_{TP} + N_{FP}}$$

where N_{TP} and N_{TN} are the number of cases diagnosed correctly as COVID-19 and non-COVID-19 pneumonia, respectively. In addition, non-COVID-19 cases wrongly diagnosed as COVID-19 and incorrectly diagnosed COVID-19 cases are assigned as N_{FP} and N_{FN} , respectively. The ROC curve analysis was used to determine the AUC with 95% confidence interval (CI) and evaluate the algorithm's performance on each of the database. A p-value less than 0.05 was considered significant. All statistical analysis was performed by the SPSS software (version 24, IBM Corporation).

3. Results

The performance of COVIDiag model was evaluated by databases from five different countries. Each of database from Argentina, Turkey, Iran, and Italy had 50 COVID-19 and 50 non-COVID-19 pneumonia cases. The database from the Netherlands included only 50 COVID-19 pneumonia subjects.

Table 2
CT chest findings of COVID-19 and non-COVID-19 groups based on each center.

CT Findings	Database									
	Argentina		Turkey		Iran		Italy		The Netherlands	
	C	NC	C	NC	C	NC	C	NC	C	
Location 1										
Unilateral	1 (2.0)	14 (28.0)	8 (16.0)	17 (34.0)	7 (14.0)	19 (38.0)	0 (0.0)	9 (18.0)	5 (10.0)	
Bilateral	49 (98.0)	36 (72.0)	42 (84.0)	33 (66.0)	43 (86.0)	31 (62.0)	50 (100)	41 (82.0)	45 (90.0)	
Location 2										
Lower Lobe	15 (30.0)	18 (36.0)	17 (34.0)	10 (20.0)	5 (10.0)	12 (24.0)	0 (0.0)	10 (20.0)	5 (10.0)	
Upper Lobe	0 (0.0)	5 (10.0)	2 (4.0)	6 (12.0)	4 (8.0)	25 (50.0)	1 (2.0)	2 (4.0)	0 (0.0)	
Both Lobes	35 (70.0)	27 (54.0)	31 (62.0)	34 (68.0)	41 (82.0)	13 (26.0)	49 (98.0)	38 (76.0)	45 (90.0)	
Distribution										
Peripheral	31 (62.0)	6 (12.0)	34 (68.0)	3 (6.0)	37 (74.0)	8 (16.0)	5 (10.0)	13 (26.0)	26 (52.0)	
Central	0 (0.0)	17 (34.0)	0 (0.0)	1 (2.0)	4 (8.0)	22 (44.0)	0 (0.0)	3 (6.0)	0 (0.0)	
Both Central and Peripheral	19 (38.0)	27 (54.0)	16 (32.0)	46 (92.0)	9 (18.0)	20 (40.0)	45 (90.0)	34 (68.0)	24 (48.0)	
Lesion										
Single	1 (2.0)	12 (24.0)	6 (12.0)	5 (10.0)	7 (14.0)	15 (30.0)	0 (0.0)	4 (8.0)	4 (8.0)	
Multiple	34 (68.0)	33 (66.0)	40 (80.0)	26 (52.0)	35 (70.0)	29 (58.0)	2 (4.0)	23 (46.0)	26 (52.0)	
Diffuse	15 (30.0)	5 (10.0)	4 (8.0)	19 (38.0)	8 (16.0)	6 (12.0)	48 (96.0)	23 (46.0)	20 (40.0)	
GGO										
No	4 (8.0)	38 (76.0)	0 (0.0)	4 (8.0)	1 (2.0)	34 (68.0)	2 (4.0)	11 (22.0)	1 (2.0)	
Yes	46 (92.0)	12 (24.0)	50 (100)	46 (92.0)	49 (98.0)	16 (32.0)	48 (96.0)	39 (78.0)	49 (98.0)	
Consolidation										
No	29 (58.0)	11 (22.0)	33 (66.0)	8 (16.0)	29 (58.0)	26 (52.0)	31 (62.0)	15 (30.0)	28 (56.0)	
Yes	21 (42.0)	39 (78.0)	17 (34.0)	42 (84.0)	21 (42.0)	24 (48.0)	19 (38.0)	35 (70.0)	22 (44.0)	
Reticular										
No	27 (54.0)	44 (88.0)	44 (88.0)	33 (66.0)	47 (94.0)	21 (42.0)	8 (16.0)	9 (18.0)	46 (92.0)	
Yes	23 (46.0)	6 (12.0)	6 (12.0)	17 (34.0)	3 (6.0)	29 (58.0)	42 (84.0)	41 (82.0)	4 (8.0)	
Nodule										
No	48 (96.0)	21 (42.0)	48 (96.0)	21 (42.0)	42 (84.0)	32 (64.0)	48 (96.0)	26 (52.0)	49 (98.0)	
Yes	2 (4.0)	29 (58.0)	2 (4.0)	29 (58.0)	8 (16.0)	18 (36.0)	2 (4.0)	24 (48.0)	1 (2.0)	
Bronchial Wall Thickening										
No	46 (92.0)	37 (74.0)	47 (94.0)	11 (22.0)	30 (60.0)	31 (62.0)	47 (94.0)	22 (44.0)	47 (94.0)	
Yes	4 (8.0)	13 (26.0)	3 (6.0)	39 (78.0)	20 (40.0)	19 (38.0)	3 (6.0)	28 (56.0)	3 (6.0)	
Air Bronchogram										
No	43 (86.0)	38 (76.0)	47 (94.0)	22 (44.0)	34 (68.0)	48 (96.0)	36 (72.0)	22 (44.0)	39 (78.0)	
Yes	7 (14.0)	12 (24.0)	3 (6.0)	28 (56.0)	16 (32.0)	2 (4.0)	14 (28.0)	28 (56.0)	11 (22.0)	
Cavity										
No	50 (100)	43 (86.0)	50 (100)	50 (100)	50 (100)	47 (94.0)	50 (100)	46 (92.0)	48 (96.0)	
Yes	0 (0.0)	7 (14.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (6.0)	0 (0.0)	4 (8.0)	2 (4.0)	
Crazy Paving										
No	35 (70.0)	50 (100)	47 (94.0)	48 (96.0)	36 (72.0)	42 (84.0)	32 (64.0)	42 (84.0)	37 (74.0)	
Yes	15 (30.0)	0 (0.0)	3 (6.0)	2 (4.0)	14 (28.0)	8 (16.0)	18 (36.0)	8 (16.0)	13 (26.0)	
Pleural Effusion										
No	47 (94.0)	36 (72.0)	50 (100)	35 (70.0)	43 (86.0)	37 (74.0)	50 (100)	31 (62.0)	44 (88.0)	
Yes	3 (6.0)	14 (28.0)	0 (0.0)	15 (30.0)	7 (14.0)	13 (26.0)	0 (0.0)	19 (38.0)	6 (12.0)	
Pleural Thickening										
No	50 (100)	47 (94.0)	43 (86.0)	32 (64.0)	50 (100)	45 (90.0)	47 (94.0)	40 (80.0)	48 (96.0)	
Yes	0 (0.0)	3 (6.0)	7 (14.0)	18 (36.0)	0 (0.0)	5 (10.0)	3 (6.0)	10 (20.0)	2 (4.0)	
Lymphadenopathy										
No	45 (90.0)	48 (96.0)	46 (92.0)	38 (76.0)	47 (94.0)	33 (66.0)	48 (96.0)	35 (70.0)	45 (90.0)	
Yes	5 (10.0)	2 (4.0)	4 (8.0)	12 (24.0)	3 (6.0)	17 (34.0)	2 (4.0)	15 (30.0)	5 (10.0)	

The common patterns in the COVID-19 patients were bilateral involvement (229/250, 91.6%), with multiple patchy opacities (137/250, 54.8%) affecting both upper and lower segments of the lungs (201/250, 80.4%) with the lesions predominantly affecting the lungs' peripheral regions (133/250, 53.2%). In addition, ground-glass opacity and crazy-paving were seen in 242 (96.8%) and 67 (26.8%) out of 250 COVID-19 cases. Table 2 demonstrates other CT findings of patients, obtained from each of the five centers.

In the non-COVID-19 group, the common pattern consisted of bilateral involvement (141/200, 70.5%), with multiple patchy opacities (111/200, 55.5%) affecting both upper and lower segments of the lungs (112/200, 56%), with the lesions mostly occurring in both central and peripheral regions of the lobes (127/200, 63.5%). The most common imaging sign in the non-COVID19 group was consolidation (140/200, 70%), followed by ground-glass opacities (113/200, 56.5%).

The COVIDiag model could distinguish COVID-19 from non-COVID-19 pneumonia cases with AUC of 0.983 (95% CI: 0.963–

1.000; 90.0% sensitivity; 96.0% specificity), 0.914 (95% CI: 0.856–0.972; 86.0% sensitivity; 88.0% specificity), 0.910 (95% CI: 0.849–0.970; 90% sensitivity; 84% specificity), and 0.882 (95% CI: 0.810–0.953; 92% sensitivity; 80% specificity) for Argentinian, Turkish, Iranian, and Italian databases, respectively. As the database from the Netherlands consisted of COVID-19 patients, only sensitivity was determined. Hence, our model could diagnose 43 out of 50 (86.0% sensitivity) COVID-19 cases correctly (Table 3). COVIDiag model could diagnose COVID-19 pneumonia in all cohorts with AUC of 0.921 (95%CI: 0.894–0.948, 88.8% sensitivity, 87.0% specificity, 88.0% accuracy). Fig. 2 represents ROC curves and radar plots of COVIDiag for different databases.

4. Discussion

In the present study, we have validated the performance of our proposed novel AI system (COVIDiag) in differentiating COVID-19 from non-COVID pneumonia cases using CT images obtained from

Table 3

Summary of developed AI systems for COVID-19 diagnosis.

Ref	Network	Origin	Patients used for training (internal validation)	Pneumonia Groups	Number of Centers, Countries for System Development	Training/Internal validation				Patients used for external validation	External validation				
						Sen (%)	Spc (%)	Acc (%)	AUC		Number of centers, countries for external validation	Sen (%)	Spc (%)	Acc (%)	AUC
[29]	DensNet121	United States	984 (296)	COVID-19, non-COVID-19 (Viral, Bacterial, Fungal)	Six centers, four countries (China, Italy, Japan, USA)	NA/NA	NA/NA	NA/91.70	NA/NA	1337	Six centers from four countries (China, Italy, Japan, USA)	84.00	93.00	90.80	0.949
[30]	Inception	Netherlands	476 (105)	Negative, positive COVID-19	Two centers, Netherlands	NA/85.7	NA/89.8	NA/NA	NA/0.950	262	One center, Netherlands	82.00	80.50	NA	0.880
[18]	ResNet18	China	2246 (260)	COVID-19, non-COVID-19 (Viral, Bacterial, Mycoplasma), and Normal	Seven centers from China	NA/94.93	NA/91.13	NA/92.49	NA/0.979	208	Yichang, China	92.51	85.92	90.70	0.971
										242	Hefei, China	94.74	89.19	90.32	0.970
										409	Wuhan, China	94.03	88.46	91.20	0.961
										140	Guangzhou, China	90.00	84.15	84.78	0.951
										107	Ecuador	86.67	82.26	84.11	0.905
[31]	DenseNet	China	709 (NA)	COVID-19, non-COVID-19 (Viral, Bacterial, Mycoplasma, Fungal)	Two centers from China	78.93/NA	89.93/NA	81.24/NA	0.900/NA	161	Heilongjiang, China	79.35	81.16	80.12	0.880
										226	Anhui, China	80.39	76.61	78.32	0.870
[32]	U-Net based algorithm	China	2447(639)	COVID-19, non-COVID-19	Two centers from China	NA/97.30	NA/85.00	NA/NA	NA/0.985	369 (820 scans)	Xianning, China	83.90	66.00	NA	0.837
										411 (1097 Scans)	Tianyou, China	90.70	38.60	NA	0.725
										130 (203 scans)	Xiangy, China	83.30	51.70	NA	0.679
[33]	ResNet152	China	2688 (2688)	COVID-19, non-COVID-19 pneumonia (Viral and Bacterial)	Three centers from China	NA/87.30	NA/96.60	NA/NA	NA/0.974	2539	Seven centers, China	78.00	93.50	NA	0.921
This study [17]	Ensemble learning	Iran	488 (124)	COVID-19, and non-COVID-19 (Viral, Atypical)	Single center	94.67/93.54	93.03/90.32	93.85/91.94	0.988/0.965	100	Argentina	90.0	96.0	93.0	0.983
										100	Turkey	86.0	88.0	87.0	0.914
										100	Iran	90.0	84.0	87.0	0.910
										100	Italy	92.0	80.0	86.0	0.882
										50	The Netherlands	86.0	NA	NA	NA

Sen, Sensitivity; Spc, Specificity, Acc, Accuracy, AUC, Area under the ROC curve

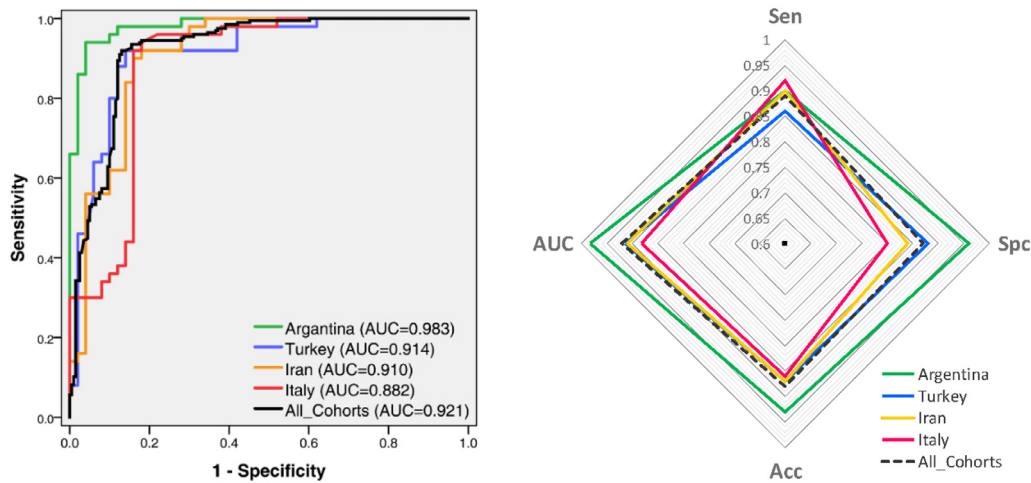


Fig. 2. (a) ROC curves and (b) radar plots of COVIDIag model on different centers. Sen, sensitivity; Spc, specificity; Acc, accuracy; AUC, area under the ROC curve.

five countries. As during the pandemic, health systems worldwide are overwhelmed by a staggering number of patients, early diagnosis is essential for authorities to manage patients in time [21,22].

Chest imaging has been used both as an initial screening method and a secondary measure to determine the extent of COVID-19 progression [23]. The inadequate infrastructure in this field has urged radiologists and medical imaging experts to actively look for new options to increase the quality and quantity of image interpretation. Hence, AI systems have been developed using CT and chest radiographic images which are the most accessible imaging modalities in almost any clinical setting [8]. Also, World Health Organization (WHO) encouraged the researchers to "study the role of artificial intelligence in chest imaging in different settings." [7]. The Italian Society of Medical and Interventional Radiology, has investigated the use of AI in diagnosing COVID-19 and has endorsed the multicenter studies by validating different AI schemes [24]. However, this particular society and other peers have not endorsed the wide-scale use of AI-based diagnosing tools to be used as the first step for screening and diagnosing [25]. Potential factors limiting wide-scale use, are lack of multicenter and multinational studies, privacy issues, and technical issues pertaining to technical specifications of CT imaging [26,27]. National and international initiatives such as the "AI-ROBOTICS vs. COVID-19 initiative" set by the European commission and academic, private, and private non-profit ones have aimed to provide such evidence [28]. The results of such research initiatives are presented in the following paragraphs.

A multicenter study performed by Harmon et al. utilized multinational databases gathered from China, Italy, Japan, and the United States to validate an AI model using CT images to detect COVID-19 and non-COVID-19 (viral, bacterial, fungal) pneumonias. Their control group consisted of a wide spectrum of subjects with either bacterial, viral, or fungal pneumonia. They reached an AUC of 0.949 and 0.947 in 3D and hybrid 3D model, respectively [29].

A similar multicenter study was Dutch authorities, using an AI tool that determined COVID-19 report and database scoring (CORADS) for each series of imaging sequences presented to the machine. The deep learning-based algorithm could diagnose positive COVID-19 cases with an AUC of 0.95 and 0.88 in the internal test and external dataset, respectively [30]. Zhang et al. validated an AI model on 1366 cases of COVID-19 and non-COVID-19 (viral, bacterial, mycoplasma) from all over China. Their AI system could diagnose COVID-19 cases with an AUC of 0.979 based on internal validation, and AUC of 0.971, 0.96, 0.970, and 0.951 for other independent Chinese Cohorts. However, the system's performance based

on an international foreign database (Ecuador) decreased to AUC of 0.905 [18]. Wang et al. developed a deep learning algorithm that automatically segmented CT images and diagnosed and differentiated COVID-19 from other pneumonias (viral, bacterial, mycoplasma, fungal). Their algorithm was based on DenseNet and could diagnose COVID-19 pneumonia with an AUC of 0.900 based on the training dataset. The external validation was done on two databases from different regions of China (Anhui, 226 patients, and Heilongjiang 161 patients). The system achieved an AUC of 0.870 and 0.880 for the first and second external validation databases, respectively [31].

A similar study was performed in China, where an AI model was developed using 2447 patients. This AI model was based on U-Net algorithm, and consecutive slices of CT imaging were fed to the system as the primary input. Internal and external validations were done by applying 639 and 910 cases from three different centers to the algorithm. The model achieved an AUC of 0.985 for the internal validation dataset. However, the performance of the model dropped to 0.725, 0.837, and 0.679 for three external validation databases, respectively [32]. Another multicenter study developed an AI system using 2688 patients with COVID-19 and other pneumonias (viral and bacterial) obtained from three Chinese centers. Their system achieved an AUC of 0.974 and 0.921 using the internal and external validation dataset, respectively [33]. Our results indicate that the performances of COVIDIag model are comparable to these studies. Also, our COVIDIag model is simple, easy to run, and does not require any high-performance computer as is based on machine learning, but the models proposed by other studies are based on deep learning. One point which needs to be noted is that these studies included all types of non-COVID-19 pneumonia, which can improve their model significantly as the differentiation between COVID-19 and non-viral pneumonia is not critical. Since the image patterns of the patients with COVID-19 and other viral or atypical pneumonias look similar, radiologists' main and vital task is to differentiate these two groups [34–36]. The COVIDIag model was developed to differentiate COVID-19 and other viral or atypical pneumonias to help the radiologists in their daily practices. Table 3 shows the summary of comparison between our proposed COVIDIag model with other similar studies.

According to the WHO advisory guide, chest imaging should be used as a primary method to diagnose COVID-19 pneumonia under three conditions: 1- RT-PCR kits are not available, 2- results of RT-PCR are not available within 24 hours, and 3- RT-PCR results are negative for patients suspected of COVID-19 pneumonia. Therefore, the COVIDIag model can be used based on CT findings for those

conditions effectively and eliminate the relevant WHO's concerns, especially in section 5.1 of the advice guide [7]. Furthermore, the model can be used in regions where there is shortage of diagnostic kits or whenever the results of the PCR test are not instantly available [37].

Although the results of COVIDiag model is encouraging, its limitations are as follows. A majority of studies reported earlier and those not mentioned in this study, only depended on images and not their interpretations. Since the COVIDiag model is based on CT findings, which are determined by radiologists, the model is prone to human errors. However, relying on basic radiologic signs, which can even be distinguished by non-radiologist clinicians such as front-line emergency room physicians, pulmonologists, and other specialists, maybe more implementable in resource-poor settings. Imaging outputs are also prone to alterations in individuals with COVID-19 undergoing treatment, those with pre-existing conditions or superimposed medical problems such as bacterial or fungal pneumonia or cardiac failure, and those being under treatment with certain medications. These conditions may cause a deviation from the typical pattern reported for COVID-19 patients and accordingly affect the performance of AI model [18,38]. Human-based interpretations of imaging signs may adjust this error to some extent, but cannot eliminate it. The main advantage of the COVIDiag model is that it cannot be affected by imaging parameters. Hence, the COVIDiag model can be readily used more than other AI models. However, other AI models based on deep learning could be affected by image characteristics such as resolution [29]. Another limitation is that our study did not involve any pediatric patients. However, according to the literature, the CT findings of pediatrics and adults are similar [39]. Therefore, we presume that our model can also be used for children, but this application needs further validation.

5. Conclusion

The significant increase in the number of COVID-19 patients undergoing CT imaging for diagnostic and prognostic reasons has led to an unbalance between radiologists' workforce and workload. One proposed method to reinstate this balance has led to incorporate AI systems in routine medical practice.

In the present study, the performance of our proposed AI system called COVIDiag, was validated using five multinational databases obtained from five different countries in three continents. We found that the proposed COVIDiag model has yielded an acceptable diagnostic performance for all databases and can perform consistently among different groups of patients from various countries. Our results are significant as they provided evidence regarding the applicability of AI systems in real-world clinical settings and integration of AI in medicine.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Institutional Review Board (IRB) approval was obtained for each center:

1. Hospital Italiano de Buenos Aires, city of Buenos Aires, Argentina: IRB #5596
2. University of Health Sciences, Hamidiye Scientific Research Ethics Committee, Istanbul, Turkey: IRB #20/135.2020
3. Imam-Asadabadi Hospital, Tabriz University of Medical Sciences, Tabriz, Iran: IRB #IR.UMSU.REC.1399.029

4. AUSL-IRCCS di Reggio Emilia, Reggio Emilia, Italy (obtained by Area Vasta Emilia Nord Ethics Committee): IRB #2020/0045199

5. Zuyderland Medical Center, Heerlen/Sittard-Geleen, The Netherlands: IRB #Z2020061

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patrec.2021.09.012.

References

- [1] A La Marca, M Capuzzo, T Paglia, L Roli, T Trenti, SM Nelson, Testing for SARS-CoV-2 (COVID-19): a systematic review and clinical guide to molecular and serological in-vitro diagnostic assays, *Reprod. Biomed. Online* 41 (2020) 483–499, doi:10.1016/j.rbmo.2020.06.001.
- [2] L Wijsman, R Molenkamp, C Reusken, A Meijer, Comparison of seven commercial RT-PCR diagnostic kits for COVID-19, *J. Clin. Virol.* 128 (2020) 104412, doi:10.1016/j.jcv.2020.104412.
- [3] S Woloshin, N Patel, AS Kesselheim, False Negative Tests for SARS-CoV-2 Infection—Challenges and Implications, *N. Engl. J. Med.* 383 (2020) e38, doi:10.1056/NEJMp2015897.
- [4] C Long, H Xu, Q Shen, et al., Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *Eur. J. Radiol.* 126 (2020) 108961, doi:10.1016/j.ejrad.2020.108961.
- [5] TY Kuzan, KM Altıntoprak, HÖ Çiftçi, et al., A comparison of clinical, laboratory and chest CT findings of laboratory-confirmed and clinically diagnosed COVID-19 patients at first admission, *Diagn. Interv. Radiol.* (2020), doi:10.5152/dir.2020.20270.
- [6] L-D Chen, H Li, Y-M Ye, et al., A COVID-19 patient with multiple negative results for PCR assays outside Wuhan, China: a case report, *BMC Infect. Dis.* 20 (2020) 517–517, doi:10.1186/s12879-020-05245-7.
- [7] World Health Organization Use of chest imaging in COVID-19: a rapid advice guide, 11 June (2020). World Health Organization <https://apps.who.int/iris/handle/10665/332326>
- [8] Y Fang, H Zhang, J Xie, et al., Sensitivity of chest CT for COVID-19: comparison to RT-PCR, *Radiology* 296 (2020) E115–E117, doi:10.1148/radiol.2020200432.
- [9] JCL Rodrigues, SS Hare, A Edey, et al., An update on COVID-19 for the radiologist - A British society of Thoracic Imaging statement, *Clin. Radiol.* 75 (2020) 323–325, doi:10.1016/j.crad.2020.03.003.
- [10] G Besutti, P Giorgi Rossi, V Iotti, et al., Accuracy of CT in a cohort of symptomatic patients with suspected COVID-19 pneumonia during the outbreak peak in Italy, *Eur. Radiol.* 30 (2020) 6818–6827, doi:10.1007/s00330-020-07050-x.
- [11] Ng Ming-Yen, EY Lee, J Yang, et al., Imaging profile of the COVID-19 infection: radiologic findings and literature review, *Radiology: Cardiothorac. Imaging* 2 (2020) e200034, doi:10.1148/ryct.2020200034.
- [12] SMH Tabatabaei, H Talari, F Moghaddas, H Rajebi, Computed tomographic features and short-term prognosis of coronavirus disease 2019 (COVID-19) pneumonia: a single-center study from Kashan, Iran, *Radiology: Cardiothorac. Imaging* 2 (2020) e200130, doi:10.1148/ryct.2020200130.
- [13] B Joob, V Wiwanitkit, Radiology Management and COVID-19 in Resource Limited Setting, *Acad. Radiol.* 27 (2020) 750 <https://dx.doi.org/10.1016%2Fj.acra.2020.03.021>.
- [14] HX Bai, B Hsieh, Z Xiong, et al., Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT, *Radiology* 296 (2020) E46–E54, doi:10.1148/radiol.2020200823.
- [15] S Taylor-Phillips, C Stinton, Fatigue in radiology: a fertile area for future research, *Brit. J. Radiol.* 92 (2019) 20190043 <https://dx.doi.org/10.1259%2Fbjr.20190043>.
- [16] T Nihashi, T Ishigaki, H Satake, et al., Monitoring of fatigue in radiologists during prolonged image interpretation using fNIRS, *Jpn. J. Radiol.* 37 (2019) 437–448, doi:10.1007/s11604-019-00826-2.
- [17] AA Ardakani, UR Acharya, S Habibollahi, A Mohammadi, COVIDiag: a clinical CAD system to diagnose COVID-19 pneumonia based on CT findings, *Eur. Radiol.* 31 (2021) 121–130, doi:10.1007/s00330-020-07087-y.
- [18] K Zhang, X Liu, J Shen, et al., Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography, *Cell* 181 (2020) 1423–1433 e1411, doi:10.1016/j.cell.2020.04.045.
- [19] S Wong, H Al-Hasani, Z Alam, A Alam, Artificial intelligence in radiology: how will we be affected? *Eur. Radiol.* 29 (2019) 141–143 <https://doi.org/10.1007/s00330-018-5644-3>.
- [20] Shoeibi A, Khodatars M, Alizadehsani R et al. Automated detection and forecasting of covid-19 using deep learning techniques: A review. (2020) arXiv preprint <https://arxiv.org/abs/2007.10785>
- [21] R Alizadehsani, Z Alizadeh Sani, M Behjati, et al., Risk factors prediction, clinical outcomes, and mortality in COVID-19 patients, *J. Med. Virol.* 93 (2021) 2307–2320, doi:10.1002/jmv.26699.
- [22] M Nicola, N O'Neill, C Sohrabi, M Khan, M Agha, R Agha, Evidence based management guideline for the COVID-19 pandemic - Review article, *Int. Surg. J.* 77 (2020) 206–216, doi:10.1016/j.ijssu.2020.04.001.

- [23] D Colombi, FC Bodini, M Petrini, et al., Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia, *Radiology* 296 (2020) 86–96, doi:[10.1148/radiol.2020201433](https://doi.org/10.1148/radiol.2020201433).
- [24] E Neri, V Miele, F Coppola, R Grassi, Use of CT and artificial intelligence in suspected or COVID-19 positive patients: statement of the Italian Society of Medical and Interventional Radiology, *Radiol. Med.* 125 (2020) 505–508, doi:[10.1007/s11547-020-01197-9](https://doi.org/10.1007/s11547-020-01197-9).
- [25] Sookman B, AI and contact tracing: How to protect privacy while fighting the COVID-19 pandemic. Macdonald-Laurier Institute Publication. <https://www.macdonaldlaurier.ca/contact-tracing-protect-privacy/> (accessed 26 July 2020)
- [26] W Naudé, Artificial intelligence vs COVID-19: limitations, constraints and pitfalls, *AI Soc.* 35 (2020) 761–765, doi:[10.1007/s00146-020-00978-0](https://doi.org/10.1007/s00146-020-00978-0).
- [27] R Zhang, X Tie, Z Qi, et al., Diagnosis of COVID-19 Pneumonia Using Chest Radiography: Value of Artificial Intelligence, *Radiology* 298 (2021) E88–E97, doi:[10.1148/radiol.2020202944](https://doi.org/10.1148/radiol.2020202944).
- [28] GD Rubin, CJ Ryerson, LB Haramati, et al., The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society, *Chest*. Chest. 158 (2020) 106–116, doi:[10.1016/j.chest.2020.04.003](https://doi.org/10.1016/j.chest.2020.04.003).
- [29] SA Harmon, TH Sanford, S Xu, et al., Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets, *Nat. Commun.* 11 (2020) 1–7, doi:[10.1038/s41467-020-17971-2](https://doi.org/10.1038/s41467-020-17971-2).
- [30] N Lessmann, CI Sánchez, L Beenen, et al., Automated Assessment of CO-RADS and Chest CT Severity Scores in Patients with Suspected COVID-19 Using Artificial Intelligence, *Radiology* 298 (2020) 18–28, doi:[10.1148/radiol.2020202439](https://doi.org/10.1148/radiol.2020202439).
- [31] S Wang, Y Zha, W Li, et al., A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis, *Eur. Respir. J.* 56 (2020) 2000775, doi:[10.1183/13993003.00775-2020](https://doi.org/10.1183/13993003.00775-2020).
- [32] M Wang, C Xia, L Huang, et al., Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation, *Lancet. Digit. Health* 2 (2020) e506–e515, doi:[10.1016/S2589-7500\(20\)30199-0](https://doi.org/10.1016/S2589-7500(20)30199-0).
- [33] C Jin, W Chen, Y Cao, et al., Development and evaluation of an artificial intelligence system for COVID-19 diagnosis, *Nat. Commun.* 11 (2020) 1–14, doi:[10.1038/s41467-020-18685-1](https://doi.org/10.1038/s41467-020-18685-1).
- [34] BA Cunha, The atypical pneumonias: clinical diagnosis and importance, *Clin. Microbiol. Infect.* 12 (2006) 12–24, doi:[10.1111/j.1469-0691.2006.01393.x](https://doi.org/10.1111/j.1469-0691.2006.01393.x).
- [35] ZY Zu, MD Jiang, PP Xu, et al., Coronavirus Disease 2019 (COVID-19): A Perspective from China, *Radiol.* 296 (2020) e15–e25, doi:[10.1148/radiol.2020200490](https://doi.org/10.1148/radiol.2020200490).
- [36] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, *Comput. Biol. Med.* 121 (2020) 103795, doi:[10.1016/j.compbiomed](https://doi.org/10.1016/j.compbiomed).
- [37] JP Kanne, BP Little, JH Chung, BM Elicker, Ketai LH Essentials for Radiologists on COVID-19: An Update-Radiology Scientific Expert Panel, *Radiology* 296 (2020) e113–e114, doi:[10.1148/radiol.2020200527](https://doi.org/10.1148/radiol.2020200527).
- [38] S Katal, L Aghaghazvini, A Gholamrezanezhad, Chest-CT findings of COVID-19 in patients with pre-existing malignancies; a pictorial review, *Clin. Imaging* 67 (2020) 121–129, doi:[10.1016/j.clinimag.2020.06.004](https://doi.org/10.1016/j.clinimag.2020.06.004).
- [39] S Katal, SK Johnston, JH Johnston, A Gholamrezanezhad, Imaging Findings of SARS-CoV-2 Infection in Pediatrics: A Systematic Review of Coronavirus Disease 2019 (COVID-19) in 850 Patients, *Academic Radiol.* 27 (2020) 1608–1621, doi:[10.1016/j.acra.2020.07.031](https://doi.org/10.1016/j.acra.2020.07.031).