



Review

Artificial intelligence model on chest imaging to diagnose COVID-19 and other pneumonias: A systematic review and meta-analysis

Lu-Lu Jia^a, Jian-Xin Zhao^a, Ni-Ni Pan^a, Liu-Yan Shi^a, Lian-Ping Zhao^c, Jin-Hui Tian^{b,1}, Gang Huang^{c,*}

^a First Clinical School of Medicine, Gansu University of Chinese Medicine, Lanzhou 73000, China

^b Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China

^c Department of Radiology, Gansu Provincial Hospital, Lanzhou 730000, China

ARTICLE INFO

Keywords:

COVID-19

Pneumonia

Artificial Intelligence

Diagnostic Imaging

Machine learning

ABSTRACT

Objectives: When diagnosing Coronavirus disease 2019(COVID-19), radiologists cannot make an accurate judgments because the image characteristics of COVID-19 and other pneumonia are similar. As machine learning advances, artificial intelligence(AI) models show promise in diagnosing COVID-19 and other pneumonias. We performed a systematic review and meta-analysis to assess the diagnostic accuracy and methodological quality of the models.

Methods: We searched PubMed, Cochrane Library, Web of Science, and Embase, preprints from medRxiv and bioRxiv to locate studies published before December 2021, with no language restrictions. And a quality assessment (QUADAS-2), Radiomics Quality Score (RQS) tools and CLAIM checklist were used to assess the quality of each study. We used random-effects models to calculate pooled sensitivity and specificity, I^2 values to assess heterogeneity, and Deeks' test to assess publication bias.

Results: We screened 32 studies from the 2001 retrieved articles for inclusion in the meta-analysis. We included 6737 participants in the test or validation group. The meta-analysis revealed that AI models based on chest imaging distinguishes COVID-19 from other pneumonias: pooled area under the curve (AUC) 0.96 (95 % CI, 0.94–0.98), sensitivity 0.92 (95 % CI, 0.88–0.94), pooled specificity 0.91 (95 % CI, 0.87–0.93). The average RQS score of 13 studies using radiomics was 7.8, accounting for 22 % of the total score. The 19 studies using deep learning methods had an average CLAIM score of 20, slightly less than half (48.24 %) the ideal score of 42.00.

Conclusions: The AI model for chest imaging could well diagnose COVID-19 and other pneumonias. However, it has not been implemented as a clinical decision-making tool. Future researchers should pay more attention to the quality of research methodology and further improve the generalizability of the developed predictive models.

1. Introduction

Beginning in 2020, the coronavirus disease 2019 (COVID-19) has spread widely around the world. As of July 15, 2022, there have been more than 557,917,904 confirmed cases of COVID-19 and 6,358,899

deaths worldwide [1]. Based on the estimated viral reproduction number (R_0), the average number of infected individuals who transmit the virus to others in a completely non-immune population is about 3.77 [2], indicating that the disease is highly contagious. Therefore, It is crucial to identify infected individuals as early as possible for quarantine

Abbreviations: AI, artificial intelligence; AUC, area under the curve; COVID-19, Coronavirus disease 2019; CT, Computed tomography; CXR, Chest X-Ray; CNN, Convolutional neural network; CRP, C-reactive protein; SARS, severe acute respiratory syndrome; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; GGO, ground-glass opacities; KNN, K-nearest neighbor; LASSO, least absolute shrinkage and selection operator; ML, machine learning; MEERS-COV, Middle East respiratory syndrome coronavirus; PLR, positive likelihood ratio; PLR, negative likelihood ratio; RT-PCR, Reverse transcriptase polymerase chain reaction; ROI, regions of interest; SROC, summary receiver operating characteristic; SVM, Support vector machine; 2D, two-dimensional; 3D, three-dimensional.

* Corresponding author.

E-mail addresses: wenjju20212021@163.com (L.-L. Jia), zjx121415@163.com (J.-X. Zhao), pnn18394130729@163.com (N.-N. Pan), sly461461@126.com (L.-Y. Shi), 370013496@qq.com (L.-P. Zhao), tianjh@lzu.edu.cn (J.-H. Tian), keen0999@163.com (G. Huang).

¹ Gang Huang, Jin-Hui Tian are contributed equally.

<https://doi.org/10.1016/j.ejro.2022.100438>

Received 20 July 2022; Accepted 16 August 2022

Available online 18 August 2022

2352-0477/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and treatment procedures.

The diagnosis of COVID-19 relies on the following criteria: clinical symptoms, epidemiological history, chest imaging, and laboratory tests [3,4]. The most common clinical symptoms were: fever, cough, dyspnea, malaise, fatigue, phlegm/discharge, among others [5]. However, these symptoms are nonspecific, and non-COVID-19 pneumonia will have similar symptoms [6]. Reverse transcriptase polymerase chain reaction (RT-PCR) is the gold standard for diagnosing COVID-19, however, it has been reported that RT-PCR may not be sensitive enough for early detection of suspected patients, and in many cases the test must be repeated multiple times to confirm the results [7–9].

Another major diagnostic tool for COVID-19 is chest imaging. Chest CT of COVID-19 is characterized by ground-glass opacities (GGO) (including crazy-paving) and consolidation [10–12]. While typical CT images may be useful for early screening of suspected cases, images of various viral pneumonias are highly similar and overlap with image features of other lung infections [13]. For example, GGOs are common in other atypical pneumonia and viral pneumonia diseases such as influenza, severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS) [14], making it difficult for radiologists to diagnose COVID-19. The results of a meta-analysis showed that chest CT can be used to rule out COVID-19 pneumonia, but cannot distinguish COVID-19 from other lung infections; The fact that both types of pneumonia can appear on chest CT as exudative lesions, GGOs, implies that CT cannot differentiate SARS-CoV-2 infection from other respiratory diseases [15].

Radiomics is an emerging field that can extract high-throughput imaging features from biomedical images and convert them into mineable data for quantitative analysis. The underlying assumption is that changes and heterogeneity of lesions at the microscopic scale (such as at the cellular or molecular level) can be reflected in the images [16], offering hope for distinguishing COVID-19 from other pneumonias. In the past 3 years, there have been many studies on the diagnosis of COVID-19 based on radiomics methods. However, there has not been any research systematically summarizing the current research on artificial intelligence(AI) models for distinguishing COVID-19 from other pneumonias on images, and the overall efficacy of this predictive model is still unknown. Additionally, because radiomics research is a multi-step, complicated process, it is crucial to evaluate the method's quality before applying it to clinical applications to assure dependable and repeatable models.

Our systematic review aimed to (1) provide an overview of radiomics studies identifying COVID-19 from other pneumonias and evaluate the efficacy of prediction models; (2) Assess methodological quality and risk of bias in radiomics workflows; (3) Determine which algorithms are most commonly used to distinguish COVID-19 from other pneumonias.

2. Materials and methods

We followed the STARD (Standards for the Reporting of Diagnostic Accuracy Studies) [17] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [18]. The registration number CRD 42021272433.

2.1. Search strategy

We searched from the databases of Pubmed, Web of Science, Embase, and Cochrane Library, for studies conducted before November 30, 2021. We searched for preprints from medRxiv and bioRxiv, using the method of combining subject words and free words. The main subject words were "COVID-19", "Artificial Intelligence", and "Diagnostic Imaging". We aimed to identify all relevant studies, regardless of language or publication status, with no language restrictions. We also filtered through the search to identify relevant systematic reviews for inclusion. Meetings, letters, short communication, opinion article were excluded. Details of the search are provided in the Table S1.

2.2. Eligibility criteria

Two doctors independently screened articles that were retrieved electronically. Articles that met all the following criteria were included: (1) the index test was studied with chest CT, chest X-ray, or lung ultrasound (2) only tests of metrics interpreted by algorithms, not human interpretations, were included. We included studies involving human interpretations if they provided data related to the diagnostic accuracy of algorithmic interpretations, (3) they had information that distinguished between COVID-19 and other pneumonia, including (community-acquired pneumonia, bacterial pneumonia, viral pneumonia, influenza, interstitial pneumonia, etc.). Exclusion criteria were as follows: (1) the included cases had normal, lung cancer, lung nodules, or other non-pneumonic cases, (2) only the training group model was included and there was no validation group or data regarding diagnostic accuracy, (3) the validation group accepted the index test and reference standard studies with less than 10 participants, (4) no exact number of cases of COVID-19 was provided or other pneumonias, and the data related to the diagnostic accuracy were calculated by the number of CT image layers.

2.3. Data extraction

We extracted the following items: date of the study, number of participants and demographic information about participants, type of common pneumonia, type of images used in the model, interest in the selection basis of the area, the diagnostic performance of the training group model, the diagnostic accuracy data of the verification model, whether there was external verification, detailed information regarding the AI algorithm, technical parameters of the index test, reference standard results, and detailed information.

Two reviewers independently assessed and extracted relevant information from each included study. For each study, we extracted 2×2 data (true positive (TP), true negative (TN), false positive (FP), false negative (FN)) for the validation group. If a study reported accuracy data for more than one model, we took the 2×2 contingency table for the model with the largest Youden index. If a study reported accuracy data for one or more radiologists and AI accuracy data, we extracted only the 2×2 contingency table corresponding to AI accuracy. If a study reported a combined model of clinical information and radiomics signature data and accuracy data for a separate radiomics data model, we only extracted the 2×2 contingency table corresponding to the radiomic model data. If both internal and external validation were reported in a study, we only extracted the 2×2 contingency table corresponding to the external validation accuracy data; if the training group, the validation group, and the test group were reported in a study, and we only extracted the test group accuracy data corresponding to the 2×2 contingency table. If a study reported accuracy data for more than one external validation, we extracted the 2×2 contingency table for the accuracy data for the validation group with the largest number of participants.

2.4. Quality assessment

The Radiomics Quality Score (RQS) [19], Checklist for Artificial Intelligence in Medical Imaging (CLAIM) checklist [20] and Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) [21] were used to assess the methodological quality and study-level risk of bias of the included studies, respectively. Studies based on machine learning(ML) methods were evaluated using the Radiomics Quality Score (RQS) (Table S2), while studies relying on deep learning(DL) methods were evaluated using the CLAIM checklist (Table S3). The Diagnostic Accuracy Study Quality Assessment 2 (QUADAS-2) standard consists of four parts: patient selection, index test, reference standard, and flow and timing (Table S4). Two graduate students independently assessed the quality and discussed disagreements with the evidence-based medicine

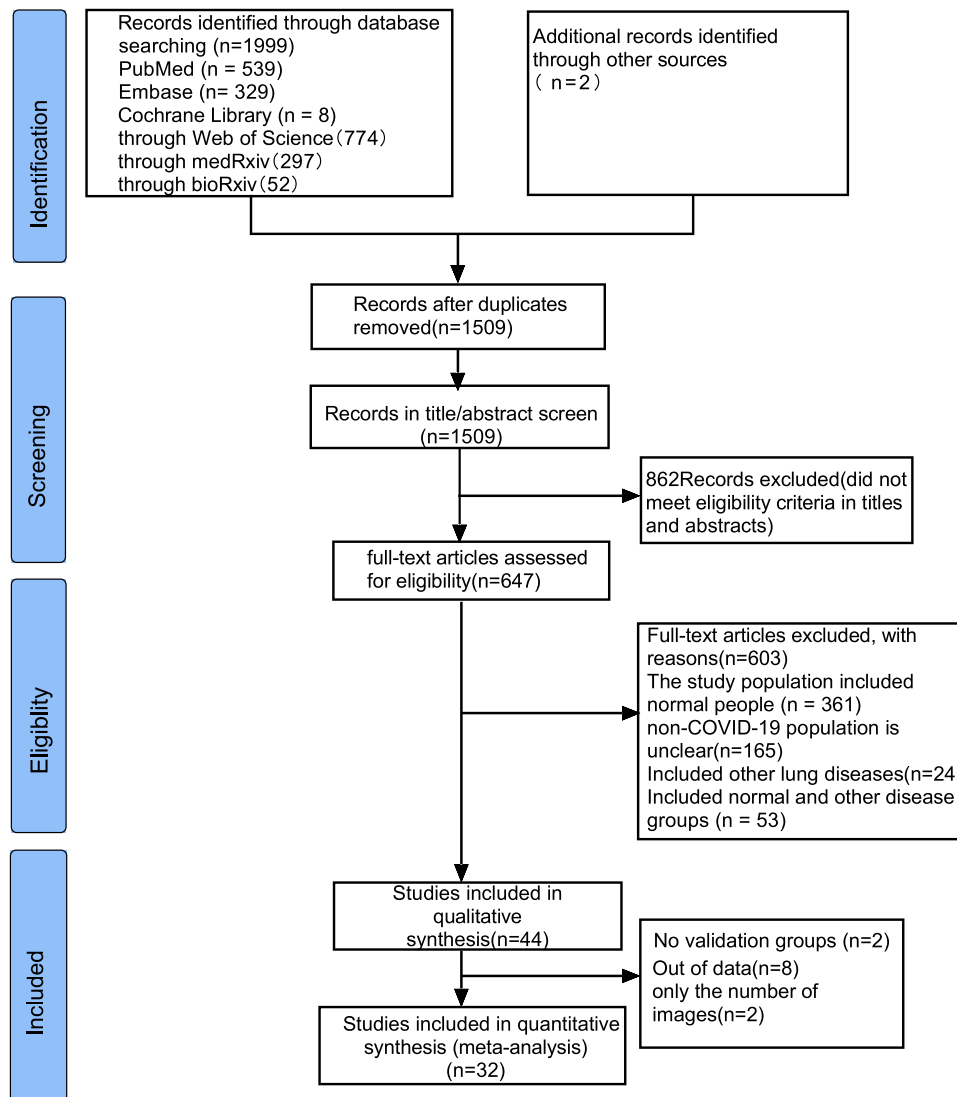


Fig. 1. Flow diagram of the study selection process for this meta-analysis.

teacher to reach a consensus.

2.5. Statistical analysis

We created a 2×2 table for each study based on data extracted directly from the article and calculated the accuracy of diagnostic tests [22] (sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (PLR), and negative likelihood ratio (PLR) with 95 % CI of each study. We analyzed the data at the participant level, rather than the image level and lesion level, which is related to the treatment of pneumonia, and which is how most studies report data.

We performed meta-analyses using a bivariate random-effects model, taking into account any correlations that may exist between sensitivity and specificity [23]. We did not perform meta-analyses if only two or three studies (less than four) were assessed for a given study. analysis, because the number of studies was too small for a reliable assessment. We assessed sensitivity and specificity with 95 % confidence intervals(CI) by plotting forest plots, and we performed meta-analyses using midas in Stata14.

We explored heterogeneity between studies by visually examining the sensitivity and specificity of forest plots and summary receiver operating characteristic (SROC) plots. If sufficient data and information

were available, we plan to perform subgroup analyses to explore study heterogeneity. We considered some sources of heterogeneity, including: comparisons between different imaging methods (CT vs. CRX), modeling methods (radiomics models vs. DL models), comparisons between different sample sizes, Comparisons between different regions of interest (Infection regions vs. Others) and different segmentation methods (2D vs. 3D). Also allow us to assess the impact of various factors on the model's diagnostic performance.

We assessed publication bias because we included more than ten studies in this systematic review. We initially assessed reporting bias using funnel plot visual asymmetry, plotting measures of effect size with measures of study precision. We then conducted a formal evaluation using Deeks' test and diagnostic odds ratio (DOR) as a measure of test accuracy [24].

3. Results

3.1. Literature search

As of November 30, 2021, we have retrieved a total of 2001 articles, and after removing duplicates, there are 1509 articles left. Two reviewers independently browsed the titles and abstracts and removed 862 articles that did not match the research topic. After evaluating 647

Table 1
Summary of general study characteristics.

Training validation/ Testing													
Study ID	Country of corresponding author	Study type	Index test	Date source	Eligibility criteria	Reference standard	Common type of pneumonia	Number of COVID-19 vs. other pneumonias	AUC	Type of validation	Number of COVID-19 vs. other pneumonias	SEN	SPC
Ardakani 2020	Iran	R	CT	Single hospital	Yes	RT-PCR	Atypical, viral pneumonia	86 vs. 69	0.999	Random split	22 vs.17	1.00	0.99
Ardakani 2021	Iran	R	CT	Single hospital	Yes	RT-PCR	Atypical and viral pneumonia	244 vs.244	0.988	Random split	62 vs.62	0.935	0.903
Ali 2021	turkey	R	CXR	Single database	No	NA	Viral pneumonia	146 vs.901	NR	3fold CV	73 vs.444	0.973	NR
Han2021	Korea	R	CT	2datasets	No	NA	Viral pneumonia, bacterial pneumonia, fungal pneumonia	164 vs.320	NR	External validation	21 vs.40	0.997	0.959
Di2020	China	R	CT	5hospitals	No	RT-PCR	CAP	1933 vs. 1064	NR	10 fold CV	215 vs.118	0.932	0.840
Bai 2020	China	R	CT	10hospitals	Yes	RT-PCR	Pneumonia of other origin	377 vs.453	NR	Random split	42 vs.77	0.950	0.960
Panwar 2020	Mexico	R	CXR	3datasets	No	NA	Pneumonia	133 vs.231	NR	Random split	29 vs.85	0.966	0.953
Kang 2020	China	R	CT	3hospitals	No	RT-PCR	CAP	1046 vs. 719	NR	Random split	449 vs.308	0.966	0.932
Liu 2021	China	R	CT	2hospitals	Yes	RT-PCR	Viral infections, mycoplasma infections, chlamydia infections, fungus infections, co-infections	66 vs.313	1.000	External validation	20 vs.20	0.850	0.900
Chen 2021	China	R	CT	Single hospital	Yes	RT-PCR	Other types of pneumonia	54 vs.60	0.984	Random split	9 vs.11	0.816	0.923
Song 2020	China	R	CT	2hospitals	Yes	RT-PCR	CAP	66 vs.66	0.979	External validation	15 vs.20	0.800	0.750
Sun 2020	China	R	CT	6hospitals	No	RT-PCR	CAP	1196 vs. 822	NR	5fold CV	299 vs.205	0.931	0.899
Wang 2021	China	R	CT	3hospitals	Yes	RT-PCR	Other types of viral pneumonia	74 vs.73	0.970	External validation	17 vs.17	0.722	0.751
Zhou 2021	China	R	CT	12hospitals	Yes	RT-PCR	Influenza pneumonia	118 vs.157	NR	External validation	57 vs.50	0.860	0.772
Azouji2021	Switzerland	R	CXR	7datasets	No	NA	MERS, SARS	338 vs.222	NR	5fold CV	85 vs.56	0.989	NR
Cardobi 2021	Italy	R	CT	Single hospital	No	swab test	Interstitial pneumonias	54 vs.30	0.830	Random split	14 vs.17	0.570	0.930
Yang 2021	China	R	CT	Single hospital	No	RT-PCR	Other pneumonias	70 vs.70	NR	10fold CV	20 vs.20	0.942	0.854
Chikontwe 2021	Korea	R	CT	Single hospital	No	RT-PCR	Bacterial pneumonia	38 vs.49	NR	Random split	30 vs.39	1.000	0.975
Zhu 2021	China	R	CT	6hospitals	No	RT-PCR	CAP	1345 vs. 924	NR	10fold CV	150 vs.103	0.913	0.910
Xie 2020	China	R	CT	5hospitals	Yes	RT-PCR	Bacterial infection, Viral infection	227 vs.153	NR	prospective RWD	243 vs.73	0.810	0.820
Qi 2021	China	R	CT	3hospitals+dataset	Yes	RT-PCR	CAP	127 vs.90	NR	10fold CV	14 vs.10	0.972	0.940
Wang 2020	China	R	CT	7hospitals	Yes	RT-PCR	Bacterial pneumonia, Mycoplasma pneumonia, Viral pneumonia, Fungal pneumonia	560 vs.149	0.900	External validation	102 vs.124	0.804	0.766
Yang 2020	China	R	CT	8hospitals	No	RT-PCR	CAP	960 vs.628	0.976	External validation	1605 vs. 452	0.869	0.901
Wu 2020	China	R	CT	3hospitals	No	RT-PCR	Other pneumonia	294 vs.101	0.767	Random split	37 vs.13	0.811	0.615
Zhang 2021	China	R	CT	3hospitals	No	RT-PCR	CAP, influenza, mycoplasma pneumonia	72 vs.127	0.987	5fold CV	31 vs.62	0.879	0.887
Xin 2021	China	R	CT	2hospitals	Yes	swab tests	CAP	34 vs.48	NR	5fold CV	9 vs.12	0.957	0.984
Guo 2020	China	R	CT	2hospitals	No	RT-PCR	Seasonal flu , CAP	8 vs.42	0.970	External validation	11 vs.44	0.889	0.935
Fang2020	China	R	CT	2hospitals	Yes	nucleic acid detection	Viral pneumonia	136 vs.103	0.959	Random split	56 vs.34	0.929	0.971
Xia 2021	China	R	CXR	2hospitals	Yes	nucleic acid	Influenza A/B pneumonia	246 vs.44	NR	Random split	266 vs.62	0.869	0.742
Huang2020	China	R	CT	15hospitals	Yes	RT-PCR	Viral pneumonia	62 vs.64	0.849	5fold CV	27 vs.28	0.778	0.786
Wu 2021	China	R	CT	Single hospital	Yes	nucleic acid	Other infectious pneumonia	76 vs.77	NR	5fold CV	19 vs.19	0.809	0.842
Chen2021	China	R	CT	2hospitals	No	RT-PCR	Viral pneumonia	81 vs.81	0.807	Random split	27 vs.19	0.733	0.822

Abbreviations: AUC: area under the curve; CAP: community acquired pneumonia;CT: Computed tomography; CV: cross-validation;CXR: Chest X-Ray; R: retrospective; RT-PCR: Reverse transcriptase polymerase chain reaction; RWD: Real-world dataset; SEN: sensitivity;SPC: specificity

Table 2

Summary of artificial intelligence-based prediction model characteristics described in included studies.

Study ID	ROI	Segmentation Style	AI Method	Labeling Procedure	Pre-Processing	Augmentations	Model Structure	Loss Function	Comparison between algorithms	AI vs. Radiologist
Ardakani 2020	Regions of infections	2D	DL	by a radiologist with more than 15 years of experience in thoracic imaging	Manual ROI extraction by cropping, Normalization, transfer-learning	NA	Ten well-known CNN	NA	Ten well-known CNN	Yes
Ardakani 2021	CT chest	2D	ML	By two radiologists	feature extraction	random scaling shearing horizontal flip	ensemble method	NA	DT, KNN, Naïve Bayes, SVM	Yes
Ali 2021	Whole image	2D	DL	NA	Normalization, transfer-learning	Horizontal, vertical flip, Zoom, Shift	ResNet50, ResNet101, Res Net 152	NA	ResNet50, ResNet101, Res Net 152	No
Han2021	CT slices	2D	DL	using the labeled COVID-19 dataset	both labeled and unlabeled data can be used	random scaling random translation, random shearing, horizontal flip	a semi-supervised deep neural network	standard cross entropy loss	Supervised learning	No
Di2020	Infected lesions	2D	ML	NA	extracted both regional and radiomics features, Segmentation	NA	UVHL	cross-entropy	SVM, MLP, iHL, tHL	No
Bai 2020	Lung regions	2D	DL	Lesions (COVID-19 or pneumonia) were manually labeled by 2 radiologists	Normalized, Segmentation	flips, scaling, rotations, random brightness and contrast manipulations, random noise, and blurring	DNN	NA	No	Yes
Panwar 2020	Whole image	2D	DL	NA	Filter, dimension reduction, deep transfer learning	Shear, Rotation Zoom, shift	A DL and Grad-CAM	binary cross-entropy loss	No	No
Kang 2020	Lesion region	3D	ML	NA	Segmentation, Feature Extraction, Normalization	NA	Structured Latent Multi-View Representation Learning	Ross-entropy loss	LR, SVM, GNB, KNN, NN	No
Liu 2021	Each pneumonia lesion	3D	ML	By three experienced radiologists	Feature Extraction, Filters	NA	LASSO regression	NA	No	Yes
Chen 2021	Consolidation and ground-glass opacity lesions	3D	ML	By fifteen radiologists	Feature Extraction, wavelet filters, Laplacian of Gaussian filters, Feature selection	NA	SVM	NA	No	No
Song 2020	CT images	2D	DL	NA	semantic feature extraction	NA	BigBiGAN	NA	SVM, KNN	Yes
Sun 2020	Infected lung regions	3D	DL	NA	Feature extraction	NA	AFS-DF	NA	LR, SVM, RF, NN	No
Wang 2021	Pneumonia lesions	3D/2D	ML	By four radiologists	manual segmentation, Feature extraction	NA	Linear, LASSO, RF, KNN	NA	Linear, LASSO, RF, KNN	Yes
Zhou 2021	Lesion regions	2D	DL	annotated by 2 radiologists	Segmentation	randomly flipped, cropped	Trinary scheme (DL)	Binary cross-entropy loss	Plain scheme(DL)	Yes
Azouji2021	X-ray images	2D	DL	NA	Resizing x-ray images, Contrast limited adaptive histogram equalization, Deep feature extraction, Deep feature fusion	Rotation, translation	LMPL classifier	hinge loss function	NaiveBayes, KNN, SVM,DT, AdaBoostM2, TotalBoost,RF, SoftMax,VGG-Net	No
Cardobi 2021	Lung area	3D	ML	NA	Segmentation, features extraction	NA	LASSO model	NA	No	No

(continued on next page)

Table 2 (continued)

Study ID	ROI	Segmentation Style	AI Method	Labeling Procedure	Pre-Processing	Augmentations	Model Structure	Loss Function	Comparison between algorithms	AI vs. Radiologist
Yang 2021	Pneumonia lesion	3D	ML	artificially delineated	Segmentation, features extraction	spatially resampled	SVM	NA	Sigmoid-SVM, Poly-SVM, Linear-SVM, RBF-SVM	No
Chikontwe 2021	CT slices	3D	DL	NA	Segmentation	random transformations, flipping	DA-CMIL	NA	DeCoVNet, MIL, DeepAttentionMIL, JointMIL	No
Zhu 2021	CT images	3D	DL	NA	Segmentation, features extraction	NA	GACDN	Binary cross entropy	SVM, KNN, NN	No
Xie 2020	CT slices	3D	DL	NA	Segmentation, extract 2D local features and 3D global features	random horizontal flip, random rotation, random scale, random translation, and random elastic transformation	DNN	NA	No	Yes
Qi 2021	Lung field	3D	DL	NA	segmentation of the lung field, Extraction of deep features, Feature representation	Image rotation, reflection, and translation	DR-MIL	NA	MResNet-50-MIL, MmedicalNet, MResNet-50-MIL-max-pooling, MResNet-50-MIL-Noisy-AND-pooling, MResNet-50-Voting, MResNet-50-Montages	Yes
Wang 2020	Lung area	3D	DL	NA	fully automatic DL model to segment, normalization, convolutional filter	NA	DL	NA	No	No
Yang 2020	Infection regions	3D	DL	NA	Class Re-Sampling Strategies, Attention Mechanism	scaling	Dual-Sampling Attention Network	binary cross entropyloss	RN34 + US, Attention RN34 + US Attention RN34 + SS Attention RN34 + DS	No
Wu 2020	CT slices	3D	DL	NA	segmentation	NA	Multi-view deep learning fusion model	NA	Single-view model	No
Zhang 2021	Major lesions	3D	DL	NA	Segmentation Feature extraction, Feature selection, Segmentation Feature extraction	scaling	DL-MLP	NA	DL-SVM, DL-LR, DL-XGBoost	Yes
Xin 2021	Lungs, lobes, and detected opacities	2D	DL	Confirmed by 3 experienced radiologists and human auditing by two radiologists	Segmentation Feature extraction	NA	LR, MLP, SVM, XGboost	NA	LR, MLP, SVM, XGboost	No
Guo 2020	NR	NA	ML	by two radiologists	Segmentation Feature extraction	NA	RF	NA	No	No
Fang2020	Primary lesion	3D/2D	ML	by two chest radiologists	Segmentation feature extraction, feature reduction and selection	NA	LASSO regression	NA	No	No
Xia 2021	Lung areas	2D	DL	NA	Segmentation feature extraction	random rotation, scale, transmit	DNN	Categorical cross-entropy	No	Yes (pulmonary physicians)
Huang2020	Pneumonia lesion	3D	ML	by two chest radiologists	Segmentation feature extraction, filter	NA	Logistic model	NA	No	No
Wu 2021	Maximal regions Involving inflammatory lesions	2D	ML	by two radiologists	feature extraction, manually delineating	NA	RF	NA	No	No
Chen2021	Lesion region	2D	ML	by two radiologists	Segmentation feature extraction, feature dimensionality reduction	NA	WSVM	NA	RF, SVM LASSO	Yes

Abbreviations: AFS-DF: adaptive feature selection guided deep forest; AI: artificial intelligence; BigBiGAN: bi-directional generative adversarial network; CT: Computed tomography; CXR: Chest X-Ray; CNN: Convolutional neural network; DA-CMIL: Dual Attention Contrastive multiple instance learning; DT: Decision tree; DNN: Deep Neural Networks; DR-MIL: deep represented multiple instance learning; DL: deep learning; RF: Random Forests; GNB: Gaussian-Naive-Bayes; Grad-CAM: Gradient Weighted Class Activation Mapping; GACDN: generative adversarial feature completion and diagnosis network; IHL: Inductive Hypergraph Learning; KNN: K-nearest neighbor; LR: Logistic-Regression; LASSO: least absolute shrinkage and selection operator; LMPL: large margin piecewise linear; ML: machine learning; MLA: Machine learning algorithms; MLP: Multilayer Perceptron; MERS: Middle East respiratory syndrome; NN: Neural-Networks; ROI: Region of interest; SVM: Support vector machine; THL: Transductive Hypergraph Learning; 2D: two-dimensional; 3D: three-dimensional; UVHL: Uncertainty Vertex-weighted Hypergraph Learning; WSVM: weighted support vector machine

AI-assisted imaging studies conducted to diagnose, classify, and detect the full text of COVID-19, we excluded 603 articles (including 165 non-new coronary pneumonia types that did not specifically account for the participants; 279 articles including COVID-19, common pneumonia, and healthy individuals; 8 articles including COVID-19, pneumonia, and other non-inflammatory lung diseases for participants; 53 articles including participants with COVID-19, pneumonia, healthy people, and other lung diseases; participants with COVID-19, 16 articles of other lung diseases; participants with COVID-19, 82 articles including healthy individuals); in the final 44 articles, we ultimately included 32 articles, because 8 articles lacked sufficient data to construct a 2×2 table; 2 articles lacked a validation group; 2 studies were conducted at the image level. The selection process is shown in Fig. 1.

We included 32 studies with a total of 6737 participants, of whom 4076 (60.5 %) were diagnosed with COVID-19, other pneumonias, including viral pneumonia (MEERS-COV, adenovirus, respiratory syncytial virus, influenza virus), bacterial pneumonia, fungal pneumonia, pneumonia caused by atypical pathogens (mycoplasma, chlamydia and legionella), pulmonary mycosis, interstitial pneumonia, and community-acquired pneumonia. The number of participants ranged from 105 to 5372. The mean age of the participants ranged from 40.92 ± 20.41 years to 61.45 ± 15.04 years. The percentage of male participants with COVID-19 ranged from 40.7 % to 62.0 %, and the percentage of male participants in other pneumonias ranged from 36.8 % to 64.0 %. The characteristics of the included studies are summarized in Table 1 and Table 2.

Most studies (20/32) included participants selected from two or more hospitals, 7 studies included participants from only one hospital, 4 studies used image data from public databases, and one study had participants from both hospitals and public database [25]. Most studies (28/32) used CT scans, and the remaining four studies used X-rays. Most of the studies (28/32) used RT-PCR as the diagnostic criteria for diagnosing SARS-CoV-2, and the diagnostic criteria of the remaining four studies were unknown [26–29].

Sixteen studies performed automatic segmentation, 12 studies performed manual segmentation, and the remaining four studies input full-slice images. Fourteen studies performed two-dimensional(2D) segmentation, 15 studies performed three-dimensional(3D) segmentation, two studies performed both 2D segmentation and 3D segmentation [30, 31], and the remaining one study did not describe the segmentation method [32]. Fifteen studies used the infected lesions as regions of interest (ROI), 10 studies used the entire image level as the ROI input models, 6 studies used the entire lung region as ROI, and the remaining one study ROI was not described [32].

Pyradiomics (6/32) was the most often used software for extracting image characteristics, followed by MatLab (4/32), PyTorch (3/32) and Python (3/32).

With 13 studies using radiomics models and 19 employing DL models, feature selection and dimensionality reduction are essential to prevent overfitting when developing radiomics models since radiomics characteristics typically exceed the sample size [33]. Least Absolute Shrinkage and Selection Operator (LASSO) regression is the most used algorithm.

Twenty studies used two or more models, and 12 studies used a single model. The three most common models include convolutional neural network(CNN), support vector machine (SVM), K-nearest neighbor (KNN). Twenty studies only calculated the diagnostic performance of the AI model, 11 studies compared the AI model with the diagnostic performance of radiologists, and one study compared the AI model with the diagnostic level given by pulmonary physicians [34]. The results of these 12 studies all showed that the diagnostic performance of the AI model in distinguishing other pneumonias of the SARS-CoV-2 was higher than that of radiologists or pulmonary physicians.

3.2. Risk of bias assessment

The mean RQS score of the included 13 studies was 7.8, accounting for 22 % of the total score. The highest RQS score was 13 (full score was 36), seen in only one study [35], and the lowest RQS score was 4 [32, 36]. Since no study considered the six items "Phantom study", "Imaging at multiple time points", "Biological correlates", "Cut-off analyses", "Prospective study" and "Comparison to 'goldstandard'", these six items received a score of zero. Other underperforming items included "Multivariable analysis with nonradiomics features", "Calibration statistics" and "Potential clinical utility", "Cost-effectiveness analysis", "Open science and data", where each item had an average score below 15 % (Fig. 2). Table S5 provides a detailed description of the RQS scores. The average CLAIM score of the 19 included studies using the DL approach was 20, slightly less than half (48.24 %) of the ideal score of 42.00, the highest score was 29 [37] and the lowest was 14 [28] (Fig. 3, Table S6).

Risk of bias and applicability issues for 32 diagnostic-related studies according to QUADAS-2 are shown in Fig. S1. Overall, the methods of the 32 selected studies were of poor quality. Most studies showed unclear risk or high of bias in each domain (Table S7). Regarding patient selection, 22 studies were considered to be at high or unclear risk of bias due to unclear how participants were selected and/or unclear detailed exclusion criteria. With regard to the index test, 30 studies were considered to be at high or unclear risk of bias, because it was unclear whether a threshold was used or the threshold was not pre-specified. Regarding reference standards, 5 studies were considered to be at high or unclear risk of bias because reference standards were not described. Regarding the flow and timing, 30 studies were considered to be at high or unclear risk of bias, due to unclear time intervals between indicator tests and reference standards and/or to clarify whether all participants received the same reference standards.

3.3. Data analysis

A total of 32 studies were included in the meta-analysis, and for the validation or test group of all studies, the pooled values and 95 % CI for sensitivity, specificity, PLR, NLR, and AUC were 0.92 (95 % CI, 0.88–0.94), 0.91, (95 % CI, 0.87–0.93), 9.7 (95 % CI, 6.8–13.9), 0.09 (95 % CI, 0.06–0.13), 0.96 (95 % CI, 0.94–0.98), respectively. When calculating pooled estimates, We observed great heterogeneity between studies in terms of sensitivity ($I^2 = 84.7$ %), specificity ($I^2 = 81.1$ %). The forest plot is shown in Fig. 4, and we can also see the obvious difference between the 95% confidence and 95 % prediction regions from the SROC curve in Fig. 5, indicating a high possibility of heterogeneity across the studies.

3.4. Subgroup analysis

We performed subgroup analyses including five different conditions and ten subgroups. Different imaging methods (CT, CRX), modeling methods (radiomics and deep learning), sample size (whether greater than 100), regions of interest (infection and others) and segmentation methods (2D and 3D) moderate to high diagnostic value was shown in each subgroup. The results are shown in Table 3.

3.5. Publication bias

We assessed publication bias for the 3 included studies, first observing that the funnel plots (Fig. 6) were symmetric and uniformly distributed along the x and y axes. Second, we formally assessed using Deeks' test and observed that the slope coefficients were not statistically significant, ($P = 0.89$) indicating that the data were symmetric, suggesting a low possibility of publication bias.

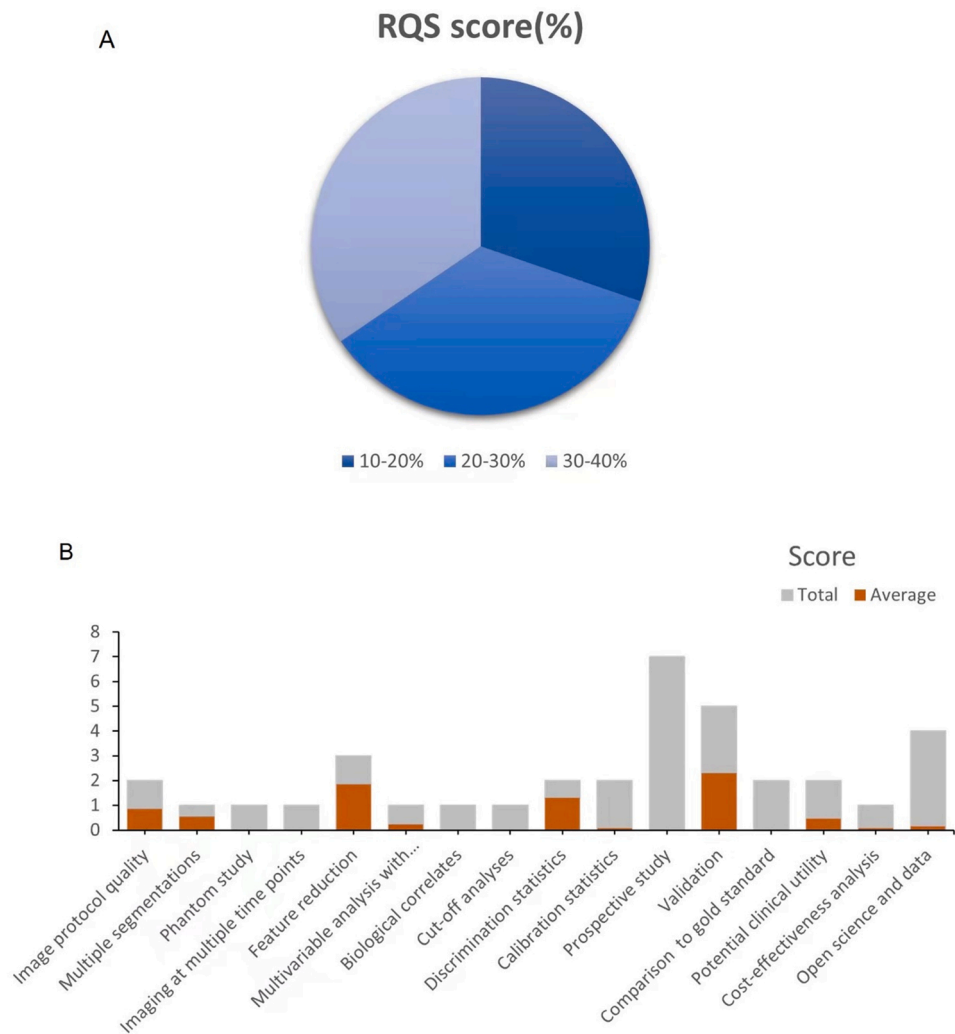


Fig. 2. Methodological quality evaluated by using the Radiomics Quality Score (RQS) tool. (A). Proportion of studies with different RQS percentage score. (B). Average scores of each RQS item (gray bars stand for the full points of each item, and red bars show actual points).

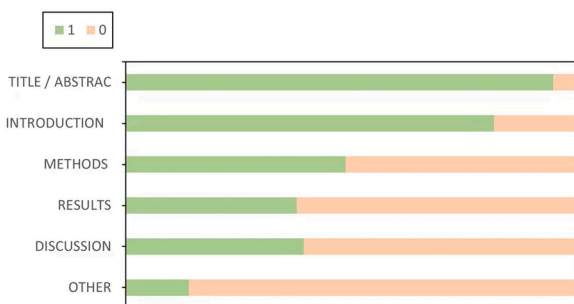


Fig. 3. CLAIM items of the 19 included studies expressed as percentage of the ideal score according to the six key domains. CLAIM, Checklist for Artificial Intelligence in Medical Imaging.

4. Discussion

In this systematic review, we aimed to determine the diagnostic accuracy of chest imaging-based AI models in distinguishing COVID-19 from other pneumonias, using the QUADAS-2, RQS tool, and the CLAIM checklist assess the quality of included studies. Furthermore, our meta-analysis is the first to quantitatively combine and interpret data from different independent surveys, potentially providing key clues for

its clinical application and further research. Despite the favorable results, pooled sensitivity, specificity, and AUC were 0.92 (95 % CI, 0.88–0.94), 0.91 (95 % CI, 0.87–0.93), and 0.96 (95 % CI, 0.94–0.98), but due to the immature stage and relatively poor methodological quality, these imaging studies did not provide clear conclusions for clinical implementation and widespread use.

In this review, the combination of the complete RQS tool, CLAIM checklist, and QUADAS-2 assessments revealed several common methodological limitations, some of which apply to both DL and ML studies.

The majority of studies (13/32) did not have images segmented by multiple radiologists, however, due to inter-observer heterogeneity, unavoidable even among experienced radiologists [38], this also limits the generalizability of the developed predictive models. Some studies applied automatic segmentation, which overcomes the differences introduced by human factors. However, models created utilizing various segmentations would undoubtedly perform differently even when trained on the same dataset and using the same AI techniques, adding another level of heterogeneity to the field.

More than half of the studies did not describe algorithms and software in sufficient detail to replicate the study. Only six percent of the studies published the codes for the models, indicating that readers have access to the full protocol, i.e., code availability. Open data and code facilitate independent researchers using the same methodology and same/different datasets to validate results, with the aim of making research findings more robust. However, only two studies published

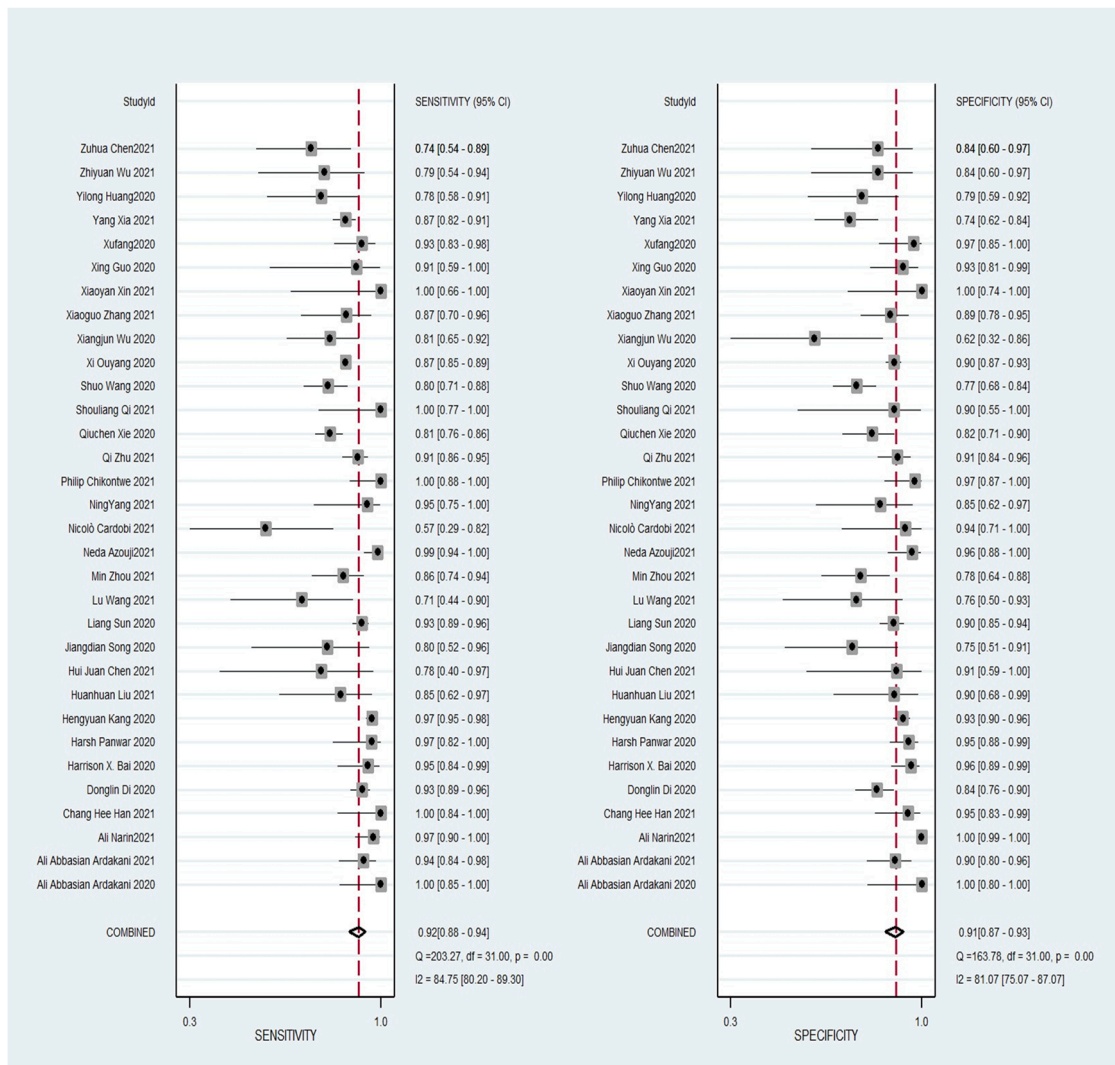


Fig. 4. Coupled forest plots of pooled sensitivity and specificity of diagnostic performance of chest imaging for distinguished COVID-19 and other pneumonias. The numbers are pooled estimates with 95 % CIs in parentheses; horizontal lines indicate 95 % CIs.

small amounts of data [37,39]. Therefore, it is hypothesized that some practical issues, such as reproducibility and generalizability of AI models, should be well resolved before translating these models into routine clinical applications.

We know that the typical imaging manifestations of SARS-CoV-2 are ground-glass opacities and consolidation foci, GGO is an indistinct increase in attenuation that occurs in various interstitial and alveolar processes while sparing bronchial [40] and vascular margins, while consolidation is an area of opacity obscuring the margins of the vessel and airway walls [41]. However, other types of pneumonia may share some similar CT imaging features with SARS-CoV-2, especially other viral pneumonias [42–44], This confuses radiologists when diagnosing SARS-CoV-2, unable to correctly diagnose whether it is SARS-CoV-2 or other pneumonia. A total of 11 studies in our systematic review also assessed the diagnostic performance of radiologists, and one study assessed the diagnostic performance of pulmonologists [34]. Then compared it with the diagnostic accuracy of AI models. all studies have shown that the diagnostic performance of AI models is higher than that of radiologists/pulmonologists. Shows that AI models have great

potential in diagnosing SARS-CoV-2 and other pneumonias.

We performed a subgroup analysis using five key factors, and in the subgroup analysis of different imaging modalities, the diagnostic performance of the chest X-rays -based AI models were better than that of the CT-based models, but only four studies focusing on chest X-rays (including 453 COVID-19 patients out of 1100 subjects) were included, and all studies used deep learning models. Therefore, the pooled results showing that chest X-rays is superior to chest CT are not entirely convincing. Another subgroup analysis showed that studies using DL models were slightly more valuable than those using ML. The main disadvantage of ML algorithm is that the method is based on hand-crafted feature extractors, which requires a lot of manpower and effort [45]. Furthermore, radiomic signatures are contrived and rely on domain-specific expertise [46]. The advantage of DL is that it does not need to manually extract features during the learning process, avoiding the defects of artificially designed features in radiomics analysis [47]. Since the classifier training, feature selection, and classification of DL model occur simultaneously, researchers only need to input images, not clinical data, or radiomics features. The most commonly used DL model

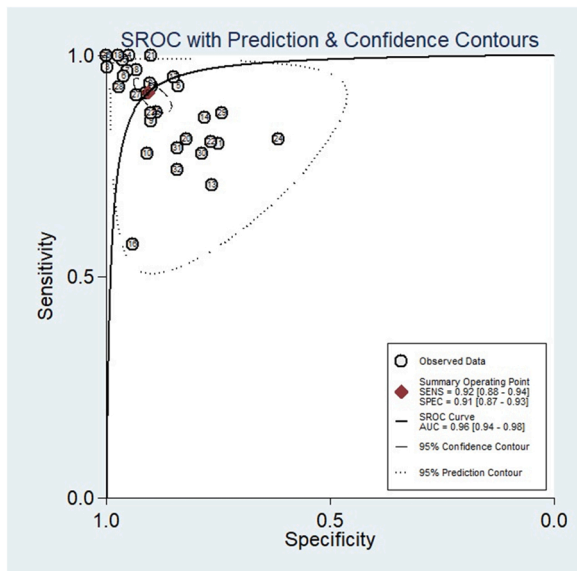


Fig. 5. Diagnostic performance of SROC curve of an artificial intelligence model for distinguishing COVID-19 from other pneumonias on chest imaging. There was an obvious difference between the 95 % confidence and 95 % prediction regions, indicating a high possibility of heterogeneity across the studies.

in research is CNN, which inspired by the biological natural visual cognition mechanism, build by convolutional layer, rectified linear units layer, pooling layer and fully-connected layer [48,49]. For example, VGG and ResNet are adjusted and combined by simple CNN [50]. In addition, the results showed that studies with large sample sizes had better diagnostic accuracy than studies with small sample sizes. Therefore, in future studies, increasing the sample size will improve the ability to diagnose SARS-CoV-2 and various other pneumonias.

Limitations of this review. First, many articles published in authoritative journals using AI models to diagnose COVID-19 were not included because the models were not validated. Unvalidated models have limited value, and validation is an integral part of a complete radiomics analysis [19]. Models must be validated internally or externally. Second, the heterogeneity of studies was evident, we performed subgroup analyses to explore sources of heterogeneity, but this was limited, and in fact, heterogeneity is a recognized feature in a review of diagnostic test accuracy [23], and it is impossible to know the source of all the heterogeneity.

To date, no systematic review or meta-analysis has been performed

Table 3
The results of subgroup analysis.

Subgroup	Number of study	Sensitivity (95 % CI)	I ² (%)	Specificity	I ² (%)	PLR	I ² (%)	NLR	I ² (%)	AUC
Imaging modality										
CRX	4	0.91(0.88,0.94)	85.6	0.96(0.95,0.98)	95.3	26.04(3.73,181.94)	93.3	0.04(0.00,0.41)	92.6	0.9914
CT	28	0.89(0.88,0.90)	78.9	0.89(0.87,0.90)	62.1	6.92(5.35,8.96)	69.5	0.14(0.11,0.19)	80.0	0.9427
Modeling methods										
Radiomic algorithm	13	0.92(0.90,0.94)	78.4	0.90(0.87,0.92)	36.8	7.16(4.96,10.33)	53.0	0.15(0.08,0.28)	85.6	0.9446
Deep learning	19	0.88(0.87,0.89)	78.0	0.91(0.90,0.92)	88.5	8.32(5.69,12.18)	82.5	0.12(0.09,0.17)	76.9	0.9702
sample size										
< 100	18	0.87(0.83,0.90)	65.4	0.89(0.86,0.92)	47.8	6.50(4.42,9.58)	49.3	0.18(0.12,0.28)	59.0	0.9371
> 100	14	0.89(0.88,0.90)	87.0	0.91(0.90,0.92)	90.8	8.81(6.02,12.89)	86.2	0.10(0.07,0.14)	88.6	0.9725
ROI										
Infection regions	15	0.89(0.88,0.90)	81.0	0.89(0.88,0.91)	48.8	6.89(5.20,9.12)	58.0	0.14(0.09,0.20)	81.3	0.9409
others	16	0.88(0.86,0.90)	80.4	0.92(0.90,0.94)	89.5	9.33(5.64,15.45)	83.3	0.11(0.07,0.19)	83.2	0.9691
segmentation										
2D	14	0.91(0.89,0.93)	71.6	0.93(0.91,0.95)	88.9	9.71(5.78,16.33)	79.3	0.10(0.06,0.17)	77.3	0.9740
3D	15	0.88(0.87,0.90)	85.1	0.89(0.87,0.90)	64.8	6.77(4.79,9.57)	76.6	0.15(0.10,0.22)	85.9	0.9386

Abbreviations: AUC: area under the curve; CT: Computed tomography; CXR: Chest X-Ray; NLR: negative likelihood ratio; PLR: positive likelihood ratio; ROI: Region of interest; 2D: two-dimensional; 3D: three-dimensional

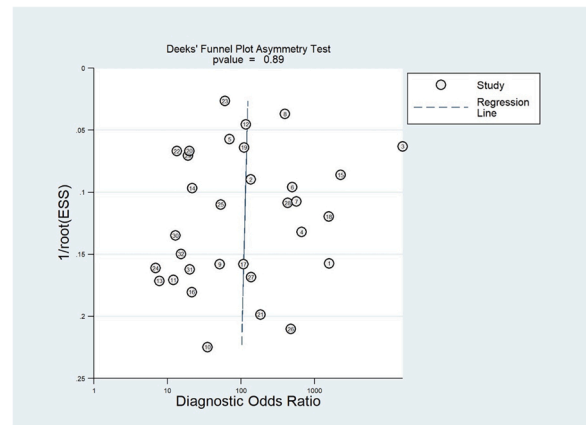


Fig. 6. Effective sample size (ESS) funnel plots and the associated regression test of asymmetry, as reported by Deeks et al. A p value < 0.10 was considered evidence of asymmetry and potential publication bias.

that includes all types of imaging techniques to diagnose COVID-19 and other pneumonias. Kao et al. [51] evaluated the CT-based radiomics signature model to successfully distinguish COVID-19 from other viral pneumonias, and came to similar conclusions as ours, with high study heterogeneity. They assessed studies up to February 26, 2021, so only 6 studies were included, and all studies were conducted in China. However, there are several systematic reviews on the diagnosis of COVID-19 based on AI models [52–55]. The participants in their studies included a series of non-pneumonic participants including lung cancer patients, lung nodules patients, and normal healthy people. These non-pneumonic chest images each have their own typical features. The imaging features are significantly different from those of COVID-19, and radiologists can easily distinguish them, so we did not include such articles in our study.

In conclusion, the artificial intelligence approach shows potential for diagnosing COVID-19 and other pneumonias. However, the immature stage and unsatisfactory quality of the research means that the proposed model cannot currently be used for clinical implementation. Before the AI models can be successfully introduced into the clinical environment of COVID-19, we need further large-sample multi-center research, open science and data, to increase the universality of the model. Furthermore, there are some technical hurdles that should be faced when considering the application of image mining tools into daily practice. Persistent efforts are required to make this tool widely available in clinical practice.

Ethical statement

This manuscript has not been published or presented elsewhere and is not under consideration for publication elsewhere. All the authors have approved the manuscript and agree with submission to your esteemed journal. There are no conflicts of interest to declare.

Funding

This study was supported by the Health Commission of Gansu Province, China [GSWSKY2020–15]. The funder has no role in the initial plan of the project, designing, implementing, data analysis, interpretation of data and in writing the manuscript.

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.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejro.2022.100438](https://doi.org/10.1016/j.ejro.2022.100438).

References

- [1] World Health Organization. <https://covid19.who.int/>
- [2] T.P. Velavan, C.G.J. Tm Meyer, i. health, The COVID-19 epidemic, 25(3) (2020) 278.
- [3] B. Hu, H. Guo, P. Zhou, Z.-L. Shi, Characteristics of SARS-CoV-2 and COVID-19, *N. R.M.* 19 (3) (2021) 141–154.
- [4] B.D. Kevadiya, J. Machhi, J. Herskovitz, M.D. Oleynikov, W.R. Blomberg, N. Bajwa, D. Soni, S. Das, M. Hasan, M.J.Nm Patel, Diagnostics for SARS-CoV-2 infections, *Nat. Mater.* 20 (5) (2021) 593–605.
- [5] T. Struyf, J.J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, M.M. Leeflang, R. Spijker, L. Hooft, D. Emperador, J. Domen, Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19, *Cochrane Database Syst. Rev.* 5 (2022).
- [6] K. McIntosh, S.J.M. Perlman, Douglas, b.s. principles, p.o.i. diseases, Coronaviruses, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), (2015) 1928.
- [7] P. Winichakoon, R. Chaiwarith, C. Liwsrisakun, P. Salee, A. Goonna, A. Limsukon, Q. Kaewpoowat, Negative nasopharyngeal and oropharyngeal swabs do not rule out COVID-19, *J. Clin. Microbiol.* 58 (5) (2020).
- [8] Z. Chen, Y. Li, B. Wu, Y. Hou, J. Bao, X. Deng, A patient with COVID-19 presenting a false-negative reverse transcriptase polymerase chain reaction result, *Korean J. Radiol.* 21 (5) (2020) 623–624.
- [9] N. Sethuraman, S.S. Jeremiah, A. Ryo, Interpreting diagnostic tests for SARS-CoV-2, *JAMA* 323 (22) (2020) 2249–2251.
- [10] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R.L. Hesketh, L.J. R. Yang, Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia, *Radiology* 295 (3) (2020).
- [11] D. Caruso, M. Zerunian, M. Polici, F. Pucciarelli, T. Polidori, C. Rucci, G. Guido, B. Bracci, C. De Dominicis, A.J.R. Laghi, Chest CT features of COVID-19 in Rome, Italy, (2020).
- [12] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L.J.R. Xia, Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, (2020).
- [13] D.S. Hui, E.I. Azhar, Z.A. Memish, A. Zumla, Human coronavirus infections—severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and SARS-CoV-2, *Encycl. Respir. Med.* 2 (2022) 146.
- [14] M. Hosseiny, S. Kooraki, A. Gholamrezaezhad, S. Reddy, L. Myers, Radiology perspective of coronavirus disease 2019 (COVID-19): lessons from severe acute respiratory syndrome and middle east respiratory syndrome, *AJR Am. J. Roentgenol.* 214 (5) (2020) 1078–1082.
- [15] N. Islam, S. Ebrahimzadeh, J.P. Salameh, S. Kazi, N. Fabiano, L. Treanor, M. Absi, Z. Hallgrimson, M.M. Leeflang, L. Hooft, C.B. van der Pol, R. Prager, S.S. Hare, C. Dennie, R. Spijker, J.J. Deeks, J. Dinnes, K. Jenniskens, D.A. Korevaar, J. F. Cohen, A. Van den Bruel, Y. Takwoingi, J. van de Wijgert, J.A. Damen, J. Wang, M.D. McInnes, Thoracic imaging tests for the diagnosis of COVID-19, *The, Cochrane Database Syst. Rev.* 3 (3) (2021) Cd013639.
- [16] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, *Radiology* 278 (2) (2016), 563–77.
- [17] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L. Irwig, J.G. Lijmer, D. Moher, D. Rennie, H.C. de Vet, H.Y. Kressel, N. Rifai, R.M. Golub, D.G. Altman, L. Hooft, D.A. Korevaar, J.F. Cohen, STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies, *BMJ (Clinical research ed.)* 351 (2015) h5527.
- [18] M.D.F. McInnes, D. Moher, B.D. Thombs, T.A. McGrath, P.M. Bossuyt, T. Clifford, J. F. Cohen, J.J. Deeks, C. Gatsonis, L. Hooft, H.A. Hunt, C.J. Hyde, D.A. Korevaar, M. M.G. Leeflang, P. Macaskill, J.B. Reitsma, R. Rodin, A.W.S. Rutjes, J.P. Salameh, A. Stevens, Y. Takwoingi, M. Tonelli, L. Weeks, P. Whiting, B.H. Willis, Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement, *JAMA* 319 (4) (2018) 388–396.
- [19] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J.E. Wildberger, S. Walsh, Radiomics: the bridge between medical imaging and personalized medicine, *Nat. Rev. Clin. Oncol.* 14 (12) (2017) 749–762.
- [20] J. Mongan, L. Moy, C.E. Kahn Jr., Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers, *Radiol. Artif. Intell.* 2 (2) (2020).
- [21] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M. M. Leeflang, J.A. Sterne, P.M. Bossuyt, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, *Ann. Intern. Med.* 155 (8) (2011), 529–36.
- [22] R. Jaeschke, G.H. Guyatt, D.L. Sackett, G. Guyatt, E. Bass, P. Brill-Edwards, G. Browman, D. Cook, M. Farkouh, H.J.J. Gerstein, Users' guides to the medical literature: III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? *JAMA* 271 (9) (1994) 703–707.
- [23] J.B. Reitsma, A.S. Glas, A.W. Rutjes, R.J. Scholten, P.M. Bossuyt, A.H.J. Joce. Zwiderman, Bivariate analysis of sensitivity and specificity produces informative summary data in diagnostic reviews, 58(10) (2005) 982–990.
- [24] W.A. van Enst, E. Ochodo, R.J. Scholten, L. Hooft, M.M.J.Bmrm. Leeflang, Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study, 14(1) (2014) 1–11.
- [25] S. Qi, C. Xu, C. Li, B. Tian, S. Xia, J. Ren, L. Yang, H. Wang, H.J.Cm Yu, p.i. biomedicine, DR-MIL: deep represented multiple instance learning distinguishes COVID-19 from community-acquired pneumonia in CT images, 211 (2021) 106406.
- [26] A. Narin, Y.J.JotFo.E. Isler, A.o.G. University, Detection of new coronavirus disease from chest x-ray images using pre-trained convolutional neural networks, 36(4) (2021) 2095–2107.
- [27] C.H. Han, M. Kim, J.T. Kwak, Semi-supervised learning for an improved diagnosis of COVID-19 in CT images, *PLoS One* 16 (4) (2021), e0249450.
- [28] H. Panwar, P. Gupta, M.K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, V.J.C. Singh, Solitons, Fractals, A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images, 140 (2020) 110190.
- [29] N. Azouji, A. Sami, M. Taheri, H.J.Ci.B. Müller, Medicine, A large margin piecewise linear classifier with fusion of deep features in the diagnosis of COVID-19, 139 (2021) 104927.
- [30] L. Wang, B. Kelly, E.H. Lee, H. Wang, J. Zheng, W. Zhang, S. Halabi, J. Liu, Y. Tian, B.J.Efor. Han, Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable radiomics features, 136 (2021) 109552.
- [31] X. Fang, X. Li, Y. Bian, X. Ji, J.J.Er Lu, Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2, 30(12) (2020) 6888–6901.
- [32] X. Guo, Y. Li, H. Li, X. Li, X. Chang, X. Bai, Z. Song, J. Li, K. Li, An improved multivariate model that distinguishes COVID-19 from seasonal flu and other respiratory diseases, *Aging* 12 (20) (2020) 19938–19944.
- [33] M.E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypiński, P. Gibbs, G. Cook, Introduction to radiomics, *J. Nucl. Med.: Off. Publ., Soc. Nucl. Med.* 61 (4) (2020) 488–495.
- [34] Y. Xia, W. Chen, H. Ren, J. Zhao, L. Wang, R. Jin, J. Zhou, Q. Wang, F. Yan, B. Zhang, J. Lou, S. Wang, X. Li, J. Zhou, L. Xia, C. Jin, J. Feng, W. Li, H. Shen, A rapid screening classifier for diagnosing COVID-19, *Int. J. Biol. Sci.* 17 (2) (2021) 539–548.
- [35] H. Liu, H. Ren, Z. Wu, H. Xu, S. Zhang, J. Li, L. Hou, R. Chi, H. Zheng, Y. Chen, S. Duan, H. Li, Z. Xie, D. Wang, CT radiomics facilitates more accurate diagnosis of COVID-19 pneumonia: compared with CO-RADS, *J. Transl. Med.* 19 (1) (2021) 29.
- [36] A. Abbasian Ardakani, U.R. Acharya, S. Habibollahi, A. Mohammadi, COVIDiag: a clinical CAD system to diagnose COVID-19 pneumonia based on CT findings, *Eur. Radiol.* 31 (1) (2021) 121–130.
- [37] H.X. Bai, R. Wang, Z. Xiong, B. Hsieh, K. Chan, Q.H. Halsey, T.M.L. Tran, J.W. Choi, D.C. Wang, L.B. Shi, J. Mei, X.L. Jiang, I. Pan, Q.H. Zeng, P.F. Hu, Y.H. Li, F.X. Fu, R.Y. Huang, R. Sebros, Q.Z. Yu, M.K. Atalay, W.H. Liao, Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT, *Radiology* 296 (3) (2020) E156–e165.
- [38] D. Bellini, N. Panvini, M. Rengo, S. Vicini, M. Lichtner, T. Tieghi, D. Ippoliti, F. Giulio, E. Orlando, M.J.Er Iozzino, Diagnostic accuracy and interobserver variability of CO-RADS in patients with suspected coronavirus disease-2019: a multireader validation study 31 (4) (2021) 1932–1940.
- [39] J. Song, H. Wang, Y. Liu, W. Wu, G. Dai, Z. Wu, P. Zhu, W. Zhang, K.W. Yeom, K. Keng, End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT, *Eur. J. Nucl. Med. Mol. Imaging* 47 (11) (2020) 2516–2524.
- [40] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z.A. Fayad, A. Jacobi, K. Li, S. Li, H. Shan, C.T. Imaging, Features of 2019 Novel Coronavirus (2019-nCoV), *Radiology* 295 (1) (2020) 202–207.

- [41] D.M. Hansell, A.A. Bankier, H. MacMahon, T.C. McLoud, N.L. Müller, J. Remy, Fleischner society: glossary of terms for thoracic imaging, *Radiology* 246 (3) (2008) 697–722.
- [42] H.J. Koo, S. Lim, J. Choe, S.-H. Choi, H. Sung, K.-H.J.R. Do, Radiographic and CT features of viral pneumonia 38 (3) (2018) 719–739.
- [43] P. Reittner, S. Ward, L. Heyneman, T. Johkoh, N.L.J. Müller, Pneumonia: high-resolution CT findings in 114 patients, 13(3) (2003) 515–521.
- [44] K.T. Shiley, V.M. Van Deerlin, W.T. Miller Jr., Chest CT features of community-acquired respiratory viral infections in adult inpatients with lower respiratory tract infections, *J. Thorac. Imaging* 25 (1) (2010) 68–75.
- [45] V. Kumar, Y. Gu, S. Basu, A. Berglund, S.A. Eschrich, M.B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D.J. Mri Fenstermacher, *Radio.: Process Chall.* 30 (9) (2012) 1234–1248.
- [46] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. Van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A.J. E. Dekker, *Radiomics: extracting more information from medical images using advanced feature analysis*, 48(4) (2012) 441–446.
- [47] M. Wainberg, D. Merico, A. Delong, B.J. Frey, *Deep learning in biomedicine*, *Nat. Biotechnol.* 36 (9) (2018) 829–838.
- [48] L.C. Yan, B. YOSHUA, H. J. Geoffrey, *Deep Learn.* 521 (7553) (2015) 436–444.
- [49] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L.J. JobD. Farhan, Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions, 8(1) (2021) 1–74.
- [50] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Noguees, J. Yao, D. Mollura, R.M.J. Itomi. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, 35(5) (2016) 1285–1298.
- [51] Y.S. Kao, K.T. Lin, A meta-analysis of computerized tomography-based radiomics for the diagnosis of COVID-19 and viral pneumonia, *Diagnostics* 11 (6) (2021).
- [52] Y. Bouchareb, P. Moradi Khaniabadi, F. Al Kindi, H. Al Duhli, I. Shiri, H. Zaidi, A. Rahmim, Artificial intelligence-driven assessment of radiological images for COVID-19, *Comput. Biol. Med.* 136 (2021), 104665.
- [53] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19, *IEEE Rev. Biomed. Eng.* 14 (2021) 4–15.
- [54] A. Gudigar, U. Raghavendra, S. Nayak, C.P. Ooi, W.Y. Chan, M.R. Gangavarapu, C. Dharmik, J. Samanth, N.A. Kadri, K. Hasikin, P.D. Barua, S. Chakraborty, E. J. Ciaccio, U.R. Acharya, Role of artificial intelligence in COVID-19 detection, *Sensors* 21 (23) (2021).
- [55] J.S. Suri, S. Agarwal, S.K. Gupta, A. Puvvula, M. Biswas, L. Saba, A. Bit, G. S. Tandel, M. Agarwal, A. Patrick, G. Faa, I.M. Singh, R. Oberleitner, M. Turk, P. S. Chadha, A.M. Johri, J. Miguel Sanches, N.N. Khanna, K. Viskovic, S. Mavrogeni, J.R. Laird, G. Pareek, M. Miner, D.W. Sobel, A. Balestrieri, P.P. Sfikakis, G. Tsoulfas, A. Protogerou, D.P. Misra, V. Agarwal, G.D. Kitas, P. Ahluwalia, J. Teji, M. Al-Maini, S.K. Dhanjil, M. Sockalingam, A. Saxena, A. Nicolaidis, A. Sharma, V. Rathore, J.N.A. Ajuluchukwu, M. Fatemi, A. Alizad, V. Viswanathan, P. K. Krishnan, S. Naidu, A narrative review on characterization of acute respiratory distress syndrome in COVID-19-infected lungs using artificial intelligence, *Comput. Biol. Med.* 130 (2021), 104210.