Application of Artificial Intelligence for Screening COVID-19 Patients Using Digital Images: A Meta-Analysis

Tahmina Nasrin Poly, Md.Mohaimenul Islam, Belal Alsinglawi, Min-Huei Hsu, Wen Shan Jian, Hsuan-Chia Yang, Yu-Chuan (Jack) Li

Submitted to: JMIR Medical Informatics
on: June 16, 2020

Table of Contents

Application of Artificial Intelligence for Screening COVID-19 Patients Using Digital Images: A Meta-Analysis

Tahmina Nasrin Poly[1, 2, 3] MSc; Md.Mohaimenul Islam[1, 2, 3] MSc; Belal Alsinglawi[4] MSc; Min-Huei Hsu[5] PhD; Wen Shan Jian[6] PhD; Hsuan-Chia Yang[1, 2, 3] PhD; Yu-Chuan (Jack) Li[1, 2, 3, 7, 8] MD, PhD

[1]Graduate Institute of Biomedical Informatics College of Medical Science and Technology Taipei Medical University Taipei TW
[2]International Center for Health Information Technology (ICHIT) Taipei Medical University Taipei TW
[3]Research Center of Big Data and Meta-Analysis Wan Fang Hospital Taipei Medical University Taipei TW
[4]School of Computer, Data, and Mathematical Science Western Sydney University Sydney AU
[5]Graduate Institute of Data Science Taipei Medical University Taipei TW
[6]School of Health Care Administration Taipei Medical University Taipei TW
[7]Department of Dermatology Wan Fang Hospital Taipei Medical University Taipei TW
[8]TMU Research Center of Cancer Translational Medicine Taipei Medical University Taipei TW

Corresponding Author:
Yu-Chuan (Jack) Li MD, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
15 F, No. 172-1, Sec. 2, Kellung Rd, Da'an dist
Taipei
TW

Abstract

Background: The Coronavirus Disease 2019 (COVID-2019) outbreak has spread rapidly and hospitals are overwhelmed with COVID-19 patients. While using swabs from patients is the main way for detecting coronavirus, analyzing chest images could offer an alternative to hospitals where healthcare personnel and testing kits are scarce. Deep learning, in particular, has shown impressive performances for analyzing medical images including COVID-19 pneumonia.

Objective: To perform a systematic review with a meta-analysis of relevant studies to quantify the performance of the DL algorithms for automatic stratification of COVID-19 using chest images.

Methods: A search strategy for use of PubMed, Scopus, Google Scholar, and Web of Science was developed (between January 1, 2020, and April 25) using the key terms COVID-19, coronavirus, SARS-CoV-2, novel corona, 2019-ncov and deep learning. Two authors independently extracted data on study characteristics, methods, risk of bias, and outcomes. Any disagreement between them was resolved by consensus.

Results: Sixteen studies were included in the meta-analysis, including 5,896 chest images of COVID-19. The pooled sensitivity and specificity of DL for detecting COVID-19 was 0.95 (95%CI: 0.94-0.95), and 0.96 (95%CI: 0.96-0.97), respectively, with an SROC of 0.98. The positive likelihood, negative likelihood, and diagnostic odds ratio were 19.02 (12.83-28.19), 0.06(95%CI:0.04-0.10), and 368.07 (95%CI: 162.30-834.75), respectively. The pooled sensitivity and specificity for detecting Pneumonia was 0.93 (95%CI:0.92-0.94), and 0.95(95%CI: 0.94-0.95). The performance of radiologists for detecting COVID-19 was lower than DL; however, the performance of junior radiologists was improved when they used DL-based prediction tools.

Conclusions: Our study findings show that deep learning models have immense potential accurately stratified COVID-19, and correctly differentiate from other pneumonia and normal patients. Implementation of deep learning-based tools can assist radiologists to correctly and quickly detect COVID-19 and to combat the COVID-19 pandemic. Clinical Trial: N/a

(JMIR Preprints 16/06/2020:21394)
DOI: https://doi.org/10.2196/preprints.21394

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Application of Artificial Intelligence for Screening COVID-19 Patients Using Digital Images: A Meta-Analysis

Tahmina Nasrin **Poly**[1, 2,5]; Md.Mohaimenul **Islam**[1,2,5]; Belal **Alsinglawi**[6]; Min-Huei **Hsu**[7]; Wen-Shan **Jian**[8]; Hsuan-Chia **Yang**[1,2,5,ϕ,*];Yu-Chuan **(Jack) Li**[1, 2, 3, 4,5,ϕ*]

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan.
[2]International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei, Taiwan.
[3]Department of Dermatology, Wan Fang Hospital, Taipei, Taiwan.
[4]TMU Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei, Taiwan.
[5]Research Center of Big Data and Meta-Analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan.
[6]School of Computer, Data, and Mathematical Science, Western Sydney University, Sydney, Australia.
[7]Graduate Institute of Data Science, Taipei Medical University, Taipei, Taiwan.
[8]School of Health Care Administration, Taipei Medical University, Taipei, Taiwan

Φ= Equal Contribution
***Corresponding Author:**

Yu-Chuan (Jack) Li, MD, PhD

Professor, Graduate Institute of Biomedical Informatics, College of Medicine Science and Technology, Taipei Medical University, Taipei, Taiwan;

International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei, Taiwan.

President-Elect, American Medical Informatics Association

Address: 250 Wu-Hsing St., Taipei 110 (Taiwan)

Tel.: +886-2-27361661 ext.7600

Fax: +886-2-6638-7537;

Email: jack@tmu.edu.tw, jaak88@gmail.com

Abbreviations:
AUROC= Area Under the receiver operating characteristic curve
CI= Confidence Interval
COVID-19= Coronavirus disease 2019
SARS-CoV-2= Severe acute respiratory syndrome coronavirus 2
2019-ncov= Coronavirus disease 2019

MERS-CoV= Middle East Respiratory Syndrome Coronavirus
RT-PCR= Reverse transcription-polymerase chain reaction
DOR= Diagnostic odds ratio

Abstract:
Background: The Coronavirus Disease 2019 (COVID-2019) outbreak has spread rapidly and hospitals are overwhelmed with COVID-19 patients. While using swabs from patients is the main way for detecting coronavirus, analyzing chest images could offer an alternative to hospitals where healthcare personnel and testing kits are scarce. Deep learning, in particular, has shown impressive performances for analyzing medical images including COVID-19 pneumonia.
Purpose: To perform a systematic review with a meta-analysis of relevant studies to quantify the performance of the DL algorithms for automatic stratification of COVID-19 using chest images.
Methods: A search strategy for use of PubMed, Scopus, Google Scholar, and Web of Science was developed (between January 1, 2020, and April 25) using the key terms COVID-19, coronavirus, SARS-CoV-2, novel corona, 2019-ncov and deep learning. Two authors independently extracted data on study characteristics, methods, risk of bias, and outcomes. Any disagreement between them was resolved by consensus.
Results: Sixteen studies were included in the meta-analysis, including 5,896 chest images of COVID-19. The pooled sensitivity and specificity of DL for detecting COVID-19 was 0.95 (95%CI: 0.94-0.95), and 0.96 (95%CI: 0.96-0.97), respectively, with an AUROC of 0.98. The positive likelihood, negative likelihood, and diagnostic odds ratio were 19.02 (12.83-28.19), 0.06(95%CI:0.04-0.10), and 368.07 (95%CI: 162.30-834.75), respectively. The pooled sensitivity and specificity for detecting Pneumonia was 0.93 (95%CI:0.92-0.94), and 0.95(95%CI: 0.94-0.95). The performance of radiologists for detecting COVID-19 was lower than DL; however, the performance of junior radiologists was improved when they used DL-based prediction tools.
Conclusion: Our study findings show that deep learning models have immense potential accurately stratified COVID-19, and correctly differentiate from other pneumonia and normal patients. Implementation of deep learning-based tools can assist radiologists to correctly and quickly detect COVID-19 and to combat the COVID-19 pandemic.
Keywords: COVID-19, SARS-CoV-2, Pneumonia, artificial intelligence, deep learning.

Introduction:
The coronavirus disease 2019 (COVID-19) is a serious global infectious disease, spreading at an unprecedented level worldwide[1, 2]. The WHO has declared this infectious disease as a pandemic and public health emergency. SARS-CoV-2 infection is even more contagious than SARS-CoV or MERS-CoV and is sometimes being undetected due to having asymptomatic or mild symptoms [3, 4]. Earlier detection, paired with aggressive public health steps such as social distancing, isolation of suspected or sick patients can help to tackle the crisis [5]. Presently reverse-transcription polymerase chain reaction (RT-PCR), gene sequencing, or blood specimens are considered as gold standard methods for detecting COVID-19; however, the performance of these methods (sensitivity:~73% for nasal swabs and ~61% for throat swabs) is not satisfactory [6, 7]. Since hospitals are overwhelmed by COVID-19 patients; patients with severe acute respiratory illness (SARI) are given priority over others with mild symptoms. Therefore, a large number of undiagnosed patients may lead to a

serious risk of cross-infection.
Chest radiography imaging (e.g. X-ray, CT-scan) is often used as an effective tool for quick diagnosis of pneumonia [8, 9]. The CT-scan image of COVID-19 patients shows multibolar involvement and peripheral airspace opacities (mostly are ground-glass)[10, 11]. Moreover, asymmetric patchy or diffuse airspace opacities have also been reported in patients with SARS-CoV-2 infection [12]. These changes in CT-scan images can be easily interpreted by a trained or experienced radiologist. Automatic classification of COVID-19 patients, however, has huge benefits such as increasing efficiency, wide-coverage, reducing the barrier to access, and improving patient outcomes. Several studies showed the application of deep learning techniques to identify and detect novel COVID-19 using radiography images[13, 14].
Herein, we report the results of a comprehensive systematic review of DL algorithms studies that investigated the performance of DL algorithms for COVID-19 classification from chest radiography imaging. Our main objective was to quantify the performance of DL methods for COVID-19 classification that might encourage healthcare policymakers to implement DL-based automated tools in the real-world clinical setting. The DL-based automated tool can help to reduce radiologist's work-load, as DL can help maintain diagnostic radiology support in real-time and with increased sensitivity.

Methods:
Experimental section:
The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), which is based on the Cochrane's Handbook for Systematic Reviews, was used to conduct the current study [15].
Literature Search:
We searched electronic databases such as PubMed, Scopus, Google Scholar, and Web of Science between January 1, 2020, and April 25, 2020. We developed search strategy using combinations of the following Medical Subject Headings: "COVID-19", or "coronavirus", or "SARS-CoV-2", or "novel corona", or "2019-ncov" and "deep learning", or "artificial intelligence", "automatic detection". Reference lists of the retrieved articles and relevant reviews were also checked for additional eligible articles.
Eligibility Criteria:
During the first screening, two authors (MMI, TNP) assessed the titles and abstracts of each article and excluded the irrelevant articles. To include the eligible articles, those two authors examined the full-text articles and evaluated whether the article fulfills the inclusion criteria to be included in this study. Disagreement during this selection process was resolved by consensus or if necessary, the main investigator (YCL) was consulted. We included articles if they met the following criteria: a) Published in English, b) Peer-reviewed journal, c) Assessed performance of deep learning model to detect COVID-19, and d) Provided a clear description of the methodology and the total number of images. We excluded studies if they published in Rxiv, and in the form of review, letter to the editor.
Data Extraction and Synthesis:
Two authors (MMI, TNP) independently screened all titles and abstracts of retrieved articles. Most relevant studies were selected based on the predefined

selection criteria. Any disagreement during the screening process was resolved by discussion with other authors; an unsettled issue was settled by discussion with the study supervisor (YC, L). The same two authors' cross-checked studies for duplication by comparing author name, publication date, and journal name. They excluded all duplicate studies. Afterward, they collected data from the selected studies such as author name, publication year, location, model description, total number of images, total number of COVID-19, modality of images, total number of patients, sensitivity, specificity, accuracy, AUROC, and database.

Assessment of Bias Risk:

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to assess the quality of selected studies [16]. The QUADAS-2 scale comprises four domains: patient selection, index test, reference standard, and flow and timing. The first three domains are used to evaluate the risk of bias in terms of concerns regarding applicability. The overall risk of bias was categorized into three groups (low, high, and unclear risk bias).

Statistical Analysis:

Meta-Disc (Version: 1.4) software was used to calculate the evaluation metrics of the DL model. It also utilized to a) perform statistical pooling from each study, b) assess the homogeneity with a variety of statistics including chi-square ($\chi 2$) and I^2. The sensitivity and specificity with 95% Confidence Intervals (CIs) of COVID-19, Pneumonia, and normal were calculated. The pooled receiver operating curve (ROC) was plotted and the area under the curve (AUC) was calculated with 95% CIs based on the DerSimonian-Laird (REM) method. Diagnostic odds ratio (DOR) Moses' constant of the linear model. Diagnostic tests where the DOR is constant regardless of the diagnostic threshold have symmetrical curves around the "Sensitivity =Specificity" line. In these situations, it is possible to combine DOR's by the DerSimonian Laird methods to estimate the overall DOR and hence to determine the best-fitting ROC curve [17]. The mathematical equation is given below:

$$Sensitivity = \frac{1}{1+\frac{1}{DOR_T \times \left(\frac{1-Specificity}{Specificity}\right)}} \quad \text{.... (1)}$$

When the DOR changes with diagnostic threshold, the ROC curve is asymmetrical. To fit DOR variation based on different threshold, Moses-Shapiro-Littenberg method was used. It consists of observing the relationship by fitting the straight line-

$$D = a + bS \quad \text{... (2)}$$

Where D is the log of DOR and S a measure of threshold given by

$$S = \ln\left(\frac{Sensitivity}{1-Sensitivity} \times \frac{1-Specificity}{Specificity}\right) \quad \text{...... (3)}$$

Estimates of parameters a, and b and their standard errors and covariance are obtained by ordinary or weighted least squares method using the NAG C library.

The ROC curve is the area under the curve (AUC) that summarized the diagnostic performance as a single number: an AUC close to 1 is considered as a perfect curve, and AUC close to 0.5 is considered as poor [18]. The AUC is computed by numeric integration of the curve equation by the trapezoidal method [19]. The Q* index is defined by the point where sensitivity and specificity are equal, which is the point closest to the ideal top-left corner of the ROC space. It is calculated by-

$$Q^{¿}=\frac{\sqrt{DOR_T}}{1+\sqrt{DOR_T}} \quad \text{......... (4)}$$

Moreover, standard error of the AUROC is calculated by following equation:

$$SE\left(AUC_{sym}\right)=\frac{DOR_T}{\left(DOR_T-1\right)3}\left[\left(DOR_T+1\right)\ln DOR_T-2\left(DOR_T-1\right)\right]SE\left(\ln DOR_T\right) \quad \text{.... (5)}$$

The standard error of Q* is calculated by following equation:

$$SE\left(Q^{¿}\right)=\frac{\sqrt{DOR_T}}{2\left(1+\sqrt{DOR_T}\right)2}SE\left(\ln DOR_T\right) \quad \text{..... (6)}$$

Results:

Selection Criteria:

Figure 1 shows the process of identifying relevant deep learning studies. A total of 562 studies were retrieved by searching electronic databases and by reviewing their references list. We excluded 435 duplicate studies and an additional 104 studies that did not fulfill the selection criteria. We reviewed 23 full-text studies and further excluded 7 studies because of several reasons shown in the figure. We finally included 16 studies in the meta-analyses [13, 14, 20-33].

Characteristics of Included Studies:

Among sixteen deep learning-based COVID-19 detection studies, we identified 5,896 digital images for COVID-19, and 645,825 images for non-COVID-19 including viral pneumonia, and normal patients. Included studies used DL algorithms like CNN, MobileNet v2, and COVNet for stratifying COVID-19 patients with higher accuracy. The range of accuracy for detecting COVID-19 correctly was 76 to 99.51. Eight studies used CT images, eight studies used X-ray images. The characteristics of the included studies in the meta-analysis show in **Table 1**.

Table 1: Characteristics of included studies in the meta-analysis.

Author	Modality	Methods	Number of images	Number of COVID-19 images	[a]SN	[b]SP	[c]AC
D. Apostolopoulos1	X-ray	MobileNet v2	1,428	224	98.66	96.46	99.18
Butt	CT	CNN	618	219	98.20	92.2	N/A
D. Apostolopoulos	X-ray	MobileNet v2	3,905	463	97.36	99.42	96.78
Li	CT	COVNet	4,356	127	90.00	95.00	N/A
Ucar	X-rays	[e]CNN	4,608[d]	15,36[d]	-	99.13	98.30
Ozturk	X-rays	CNN DarkNet	1,186	108	95.13	95.30	98.08
Bai	CT	EfficientNet	1,186	521	95.00	96.00	96.0

Zhang	CT	DeepLabv3	617,775		94.93	91.13	92.49
Asnaoui	X-ray	Inception_Resnet_V2	6,087	231	92.11	96.06	N/A
Ardakani	CT	ResNet-101	1,020	510	100	99.02	99.51
Pathak	CT	CNN	852	413	91.45	94.77	93.01
Wu	CT	ResNet50	495	368	81.10	61.50	76.00
Togaçar	X-rays	SqueezeNet	458	295	100	100	100
Waheed	X-ray	[f]ACGAN	1,124	403	90.00	97.00	95.00
Asif	X-rays	Xception	1,251	284	99.30	98.60	99.00
Wang	CT	DenseNet	5,372	102	80.39	76.66	78.32
Wang*	CT	DenseNet	5,372	92	79.35	81.16	80.12

[a]*SN= Sensitivity;* [b]*SP= Specificity;* [c]*AC= Accuracy; d=augmented data;* [e]CNN= Convolutional neural network; [f]*ACGAN= Auxiliary Classifier Generative Adversarial Network*

Model Performance:

Based on the sixteen studies, the performance of deep learning algorithms for detecting COVID-19 is summarized in **Table 2**. The pooled sensitivity and specificity of DL for detecting COVID-19 was 0.95 (95%CI: 0.94-0.95), and 0.96 (95%CI: 0.96-0.97), respectively, with a SROC of 0.98 (**Figure 2**). The pooled sensitivity and specificity were shown in **Figure 3**.

DL can correctly distinguish Pneumonia from COVID-19 with SROC of 0.98 [sensitivity: 0.93 (95%CI: 0.92-0.94), specificity: 0.95 (95%CI:0.94-0.95)]. The positive likelihood, negative likelihood, and diagnostic odds ratio were 22.45 (95%CI: 12.86-39.19), 0.06 (95%CI: 0.03-0.13), and 461.81 (95%CI: 134.96-1580.24). Moreover, DL model showed good performance for correctly stratifying normal patients with SROC of 0.99 [sensitivity: 0.95 (95%CI: 0.94-0.96), specificity: 0.98 (95% CI: 0.97-0.98)]. The positive likelihood, negative likelihood, and diagnostic odds ratio were 47.47 (95%CI: 20.70-108.86), 0.04 (95%CI: 0.02-0.08), and 1524.81 (95%CI: 625.29-3718.34).

Table 2: Performance comparison between deep learning models and radiologists.

Class	**Method**	[a]**N**	[b]**SN**	[c]**SP**	[d]**LR+**	[e]**LR-**	[f]**AUROC / AC***
COVID-19	**DL**	17	0.95 (0.94-0.95)	0.96 (0.96-0.97)	19.02 (12.83-28.19)	0.06 (0.04-0.10)	0.98
	Radiologists◊	**6**	0.79 (0.64-0.89)	0.88 (0.78-0.94)	-	-	0.85*
	Junior	3	0.80 (0.72-0.87)	0.88 (0.83-0.92)	-	-	-
	Senior	3	0.78 (0.70-0.85)	0.87 (0.82-0.91)	-	-	-
	Junior +AI	-	0.88 (0.81-0.93)	0.93 (0.89-0.96)	-	-	-
	Senior +AI	-	0.88 (0.81-0.93)	0.89 (0.84-0.93)			
	Radiologists^Δ	**15**	0.75(0.65-0.84)	0.90 (0.86-0.94)	-	-	-
	Junior	4	0.65 (0.48-0.79)	0.89 (0.81-0.94)	-	-	0.82*
	Senior	4	0.85 (0.70-0.94)	0.91 (0.85-0.96)	-	-	0.90*
	Junior + AI	-	0.80 (0.64-0.90)	0.94 (0.88-0.97)	-	-	0.90*
	Radiologist^α	1	0.89 (0.81-0.94)	0.83 (0.74-0.89)			0.86*

Pneumonia	**[g]DL**	7	0.93 (0.92-0.94)	0.95 (0.94-0.95)	22.45 (12.86-39.19)	0.06 (0.03-0.13)	0.98
Normal	**DL**	6	0.95 (0.94-0.96)	0.98 (0.97-0.98)	47.47 (20.70-108.86)	0.04 (0.02-0.08)	0.99

[a]N= Number of articles/radiologists;
[b]SN= Sensitivity;
[c]SP= Specificity;
[d]LR⁺: Positive likelihood ratio;
[e] LR⁺: Negative Likelihood ratio;
[f]AC= Accuracy;
[g]DL=Deep learning; (= Accuracy).(◊= Bai et al. ; Δ= Zhang et al.; α= Ardakani et al.) [Junior radiologists: experienced 5-15 years; senior radiologists: experienced 15-25 years.*

Performance of Radiologists:

Three studies compared the performance of DL with radiologists [34-36]. Zhang *et al.* [35] included eight radiologists with an experience of 5-25 years, and categorized into two groups (junior radiologist: 5-15 years' experience, and senior radiologist: 15-20 years' experience). Bai *et al.* [34] compared performance with 6 radiologists with an experience of 10, 10, 20, 20, 20 and 10 years (3 junior and 3 senior radiologists based on experience). Finally, Ardakani *et al*. [36] compared the performance of deep learning with one radiologist (senior) with an experience of 15 years. The performance of fifteen radiologists was evaluated to detect COVID-19; the pooled range of sensitivity and specificity for detecting COVID-19 was 0.75~0.89, and 0.83~ 0.90. With the assist of DL-based artificial intelligent (AI) tools, the performance of junior radiologists improved (sensitivity improved: 0.08-0.15, specificity improved: 0.05).

Sensitivity Analysis: Eight studies evaluated the performance of DL algorithms for detecting COVID-19 using X-ray photographs. The pooled sensitivity and specificity of DL for detecting COVID-19 was 0.96 (95%CI: 0.95-0.97), and 0.97 (95%CI: 0.97-0.98), respectively, with a SROC of 0.99. Moreover, eight studies assessed the performance of DL for classifying COVID-19 using CT images. The pooled sensitivity and specificity was 0.94 (95%CI: 0.94-0.95), and 0.95 (95%CI: 0.95-0.96), respectively, with a SROC of 0.96 (Supplementary Figure S1-S12).

Risk of Bias and Applicability: In this meta-analysis, we also assessed heterogeneous findings that originated from included studies based on the QUADAS-2 tool (**Table 3**). The risk of bias for patient's selection was unclear for sixteen studies. All studies had an unclear risk of bias for flow and timing and index test. Moreover, all studies had a high risk of bias for the reference standard. In case of applicability, all studies had a low risk of bias for the patient selection. Although, it was uncertain risk of index test and applicability concern for the reference standard.

Table 3: Quality Assessment of Diagnostic Accuracy Studies-2 for Included Studies

Study	**Risk of Bias**				**Applicability Concerns**		
	Patients selection	**Index Test**	**Reference Standard**	**Flow and Timing**	**Patients Selection**	**Index Test**	**Reference Standard**

D. Apostolopoulos1	☹	?	☹	?	☺	?	?
Butt	☹	?	☹	?	☺	?	?
D. Apostolopoulos	☹	?	☹	?	☺	?	?
Li	☹	?	☹	?	☺	?	?
Ucar	☹	?	☹	?	☺	?	?
Ozturk	☹	?	☹	?	☺	?	?
Bai	☹	?	☹	?	☺	?	?
Zhang	☹	?	☹	?	☺	?	?
Asnaoui	☹	?	☹	?	☺	?	?
Ardakani	☹	?	☹	?	☺	?	?
Pathak	☹	?	☹	?	☺	?	?
Wu	☹	?	☹	?	☺	?	?
Togaçar	☹	?	☹	?	☺	?	?
Waheed	☹	?	☹	?	☺	?	?
Asif	☹	?	☹	?	☺	?	?
Wang	☹	?	☹	?	☺	?	?

☹= High risk bias; ☺ = Low risk bias; ?= uncertain risk bias.

Discussion:

Main Findings:

In this study, we evaluated the performance of the DL model to detect COVID-19 automatically using chest images to assist proper diagnosis and prognosis. The findings of our study showed that the DL model achieved high sensitivity and specificity (95% and 96%) of detecting COVID-19. The pooled SROC value of both COVID-19 and Pneumonia was 98%. The performance of the DL model was comparable to that of experienced radiologists (clinical experience at least 10 years) and could improve the performance of junior radiologists.

Clinical Implications:

The rate of COVID-19 cases has been mounting day by day; therefore, it is important to fast and accurate diagnose of patients for combating this pandemic. However, screening an increased number of chest images is challenging for the radiologists and the number of trained radiologists is not sufficient, especially in under-developed and developing countries [37]. The recent success of the deep learning applications in imaging analysis of CT-scan, X-Ray in automatic segmentation and classification in the radiology domain has interested healthcare providers and researchers in exploiting the advancement of deep neural networks in other applications [38]. DL models are trained to assist radiologists to achieve higher interrater reliability during their years of experience in clinical practice.

Since the start of COVID-19 Pandemic, lots of efforts have been made by artificial intelligence researchers and AI modelers, to help radiologists for rapid diagnosis and combatting the COVID-19 pandemic [33, 39]. Developing an accurate automated AI COVID-19 detection tool is deemed as essential to reduce unnecessary waiting time, shortening screening and examination time, and improve performance. Moreover, it would help to shorten radiologist's workload and allow them to respond rapidly and cost-effectively in emergency conditions [25]. RT-PCR is considered as a gold standard; however, findings of our study showed that chest CT could be used as a reliable and rapid approach for screening of COVID-19. Our findings also showed that the DL model can discriminate COVID-19 from other pneumonia with higher sensitivity and specificity which is a challenging task for radiologists [32].

Strengths and Limitations: Our study has several strengths. First, this is the first meta-analysis that evaluate the performance of DL model to classify COVID-19 patients. Second, we considered only peer-reviewed articles to be included in our study because non peer-reviewed articles might have some bias and still has room to improve. Third, we compared the performance of DL model with radiologists (senior vs junior) which would be help for policy maker to consider automated classification system in the real-world clinical settings to speed up routine examination.

However, our study has some limitations that also need to be addressed. First, only 16 studies were used to evaluate the performance but inclusion of more studies would be given more specific findings. Second, some studies added similar dataset that may create some bias but they have optimized algorithms to improve performance. Third, two different kinds of digital photographs (CT scan and X-ray) were used to develop and evaluate the performance of DL to classify COVID-19; however, the performance of DL was almost similar in both cases. Finally, all studies had lack of external validation; therefore, performance could vary if those model implement in other clinical settings.

Future Perspective: The primary objective of prediction models is to quick screening COVID-19 patients, and to help physicians to take appropriate decision. Missing diagnosis could have destructive effect on society as it spreads infected people to healthy people. Therefore, it is important to select a target population in which automated tool serve a clinical need, and a representative dataset on which model can be trained, developed and validated internally and externally. All the studies included in this meta-analysis had high risk of bias for patients' selection and reference standards. Moreover, generalizability was lacking in those newly developed classification models. Model without proper evidences and lack of external validation is not appropriate for clinical practice because it might cause more harm than good. Since, the number of cases are mounting each day and spreading it to all continent; therefore, it is important to develop a model to assist in quick and efficient screening of patients during covid-19 pandemic might encourage clinicians and policymakers to prematurely implement prediction models without sufficient documentation and validation. Although, all studies showed promising discrimination in their training and testing/validation cohorts but future studies should focus on external validation and comparing their findings to other dataset. Interpretability of DL systems is more important to a healthcare professional than to an AI expert. Proper interpretation/explanation of algorithms will more likely to be acceptable to physicians. More clinical research are needed to determine the tangible benefits for patients in terms of high performance of model. High sensitivity and specificity do not necessarily represent clinical efficacy, and the higher value of AUROC is not always best metric to exhibit clinical applicability. All the papers should follow standard guidelines and also present positive and negative predictive values to make a fair comparison. Although all the included studies used a significant amount of data to show model performance and compared with radiologists, they used only retrospective data to train model which might give worse performance in the real-world clinical setting (data complexity is different). Therefore, prospective evaluation is needed in future before considering implementation in clinical setting. AI model always consists of potential flaws including, the inapplicability of new data, reliability and bias. Generalization of the model is important to present real performance because the rate of sensitivity and specificity varies across the studies (0.79-1.0 and 0.62-1.0). Higher false negative will help to make the situation worse, and higher false-negative will waste healthcare resources.

Conclusion:

Our study shows that deep learning has immense potential to stratify COVID-19, with high sensitivity and specificity, from other types of Pneumonia and normal patients. Deep learning-based tools could assist radiologists in fast screening COVID-19 and classifying potential high-risk patients, which have clinical significance for early management and optimize medical resources. As higher false-negative can cause a devastating effect on society; therefore, it is crucial to test the performance of models to other unknown datasets. Retrospective evaluation and reliable interpretation are warranted to consider AI models in real-world clinical settings.

Conflict of Interest: None.

Funding's: This research is sponsored in part by the Ministry of Education (MOE) under grant MOE 108-6604-001-400 and DP2-109-21121-01-A-01 and Ministry of Science and Technology (MOST) under grant MOST 108-2823-8-038-002- and 109-2222-E-038-002-MY2.

References:

1. Sun J, He W-T, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J: COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. *Trends Mol Med* 2020.
2. Sun J, He W-T, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou JJTimm: COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. 2020, 26(5):483-495.
3. Surveillances V: The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly* 2020, 2(8):113-122.
4. Rothan HA, Byrareddy SN: The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* 2020:102433.
5. Lipsitch M, Swerdlow DL, Finelli L: Defining the epidemiology of Covid-19—studies needed. *N Engl J Med* 2020, 382(13):1194-1196.
6. Kojima N, Turner F, Slepnev V, Bacelar A, Deming L, Kodeboyina S, Klausner JD: Self-Collected Oral Fluid and Nasal Swabs Demonstrate Comparable Sensitivity to Clinician Collected Nasopharyngeal Swabs for Covid-19 Detection. *medRxiv* 2020.
7. Rhoads DD, Cherian SS, Roman K, Stempak LM, Schmotzer CL, Sadri N: Comparison of Abbott ID Now, Diasorin Simplexa, and CDC FDA EUA methods for the detection of SARS-CoV-2 from nasopharyngeal and nasal swabs from individuals diagnosed with COVID-19. *J Clin Microbiol* 2020.
8. Jones BP, Tay ET, Elikashvili I, Sanders JE, Paul AZ, Nelson BP, Spina LA, Tsung JW: Feasibility and safety of substituting lung ultrasonography for chest radiography when diagnosing pneumonia in children: a randomized controlled trial. *Chest* 2016, 150(1):131-138.

9. Ye X, Xiao H, Chen B, Zhang S: Accuracy of lung ultrasonography versus chest radiography for the diagnosis of adult community-acquired pneumonia: review of the literature and meta-analysis. *PLoS One* 2015, 10(6).
10. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, Pan I, Shi L-B, Wang D-C, Mei J: Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* 2020:200823.
11. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L: Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 2020:200642.
12. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C: Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet Infectious Diseases* 2020.
13. Butt C, Gill J, Chun D, Babu BA: Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence* 2020:1.
14. Apostolopoulos ID, Mpesiana TA: Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* 2020:1.
15. Moher D, Liberati A, Tetzlaff J, Altman DG, med PGJP: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. 2009, 6(7):e1000097.
16. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011, 155(8):529-536.
17. Moses LE, Shapiro D, Littenberg BJSim: Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. 1993, 12(14):1293-1316.
18. Islam MM, Yang H-C, Poly TN, Jian W-S, Li Y-CJJCM, Biomedicine Pi: Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. 2020, 191:105320.
19. Walter SJSim: Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. 2002, 21(9):1237-1256.
20. Elasnaoui K, Chawki Y: Using x-ray images and deep learning for automated detection of coronavirus disease. *J Biomol Struct Dyn* 2020(just-accepted):1-22.
21. Apostolopoulos ID, Aznaouridis SI, Tzani MA: Extracting possibly representative COVID-19 Biomarkers from X-Ray images with Deep Learning approach and image data related to Pulmonary Diseases. *Journal of Medical and Biological Engineering* 2020:1.
22. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A: 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* 2020:103795.
23. Khan AI, Shah JL, Bhat MM: Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed* 2020:105581.
24. Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, Tran TML, Choi JW, Wang D-C, Shi L-B: AI augmentation of radiologist performance in

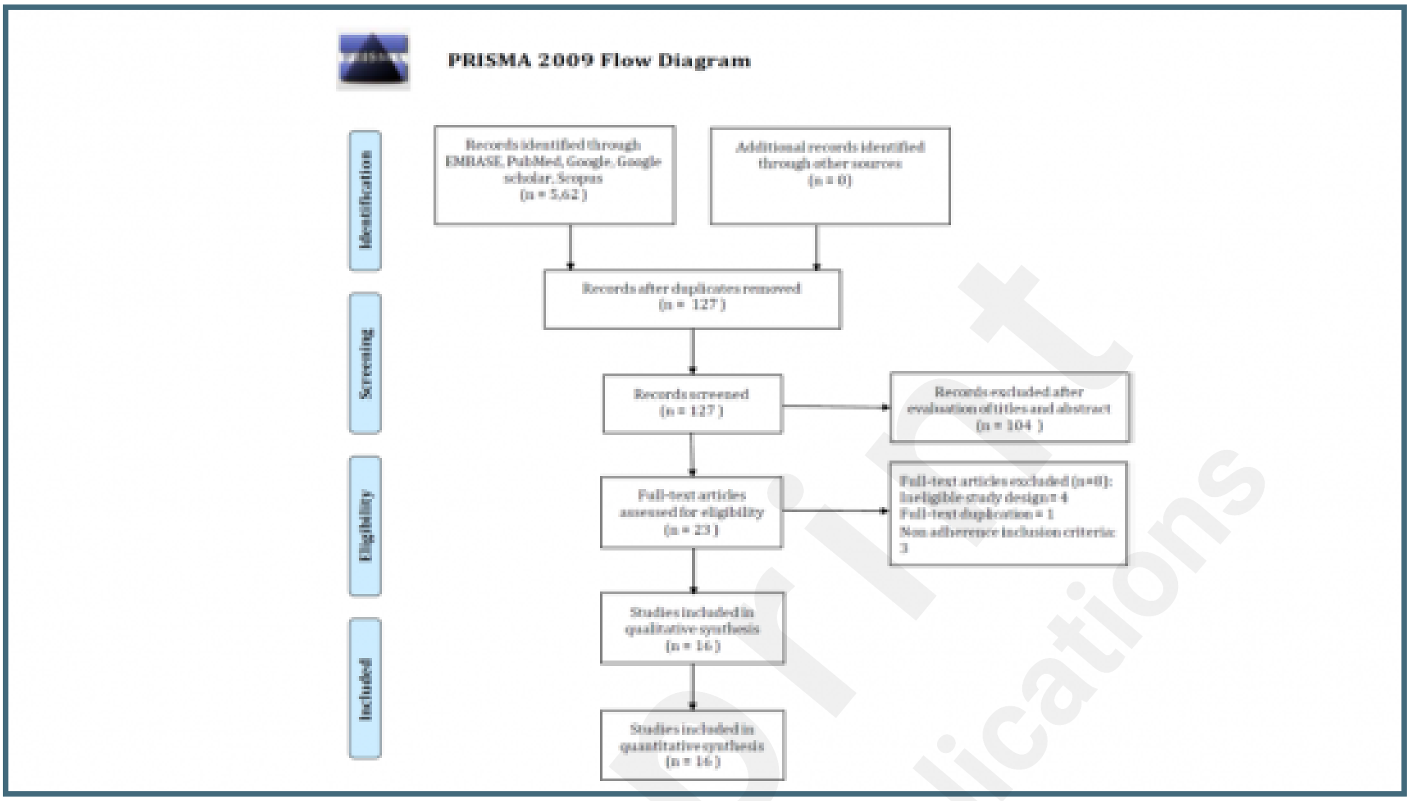
distinguishing COVID-19 from pneumonia of other etiology on chest CT. *Radiology* 2020:201491.
25. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q: Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology* 2020:200905.
26. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR: Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med* 2020:103792.
27. Pathak Y, Shukla PK, Tiwari A, Stalin S, Singh S, Shukla PK: Deep Transfer Learning based Classification Model for COVID-19 Disease. *IRBM* 2020.
28. Toğaçar M, Ergen B, Cömert Z: COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Comput Biol Med* 2020:103805.
29. Ucar F, Korkmaz D: COVIDiagnosis-Net: Deep Bayes-SqueezeNet based Diagnostic of the Coronavirus Disease 2019 (COVID-19) from X-Ray Images. *Med Hypotheses* 2020:109761.
30. Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR: Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access* 2020, 8:91916-91923.
31. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, Wang M, Qiu X, Li H, Yu H: A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020.
32. Wu X, Hui H, Niu M, Li L, Wang L, He B, Yang X, Li L, Li H, Tian J: Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. *Eur J Radiol* 2020:109041.
33. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, Zha Y, Liang W, Wang C, Wang K: Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 2020.
34. Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, Tran TML, Choi JW, Wang D-C, Shi L-BJR: Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. 2020, 296(3):E156-E165.
35. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, Zha Y, Liang W, Wang C, Wang KJC: Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. 2020, 181(6):1423-1433. e1411.
36. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi AJCiB, Medicine: Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. 2020, 121:103795.
37. Liedenbaum MH, Bipat S, Bossuyt PM, Dwarkasing RS, de Haan MC, Jansen RJ, Kauffman D, van der Leij C, de Lijster MS, Lute CC: Evaluation of a standardized CT colonography training program for novice readers. *Radiology* 2011, 258(2):477-487.
38. Ker J, Wang L, Rao J, Lim T: Deep learning applications in medical image analysis. *Ieee Access* 2017, 6:9375-9389.
39. Hemdan EE-D, Shouman MA, Karar ME: Covidx-net: A framework of deep

learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:200311055* 2020.

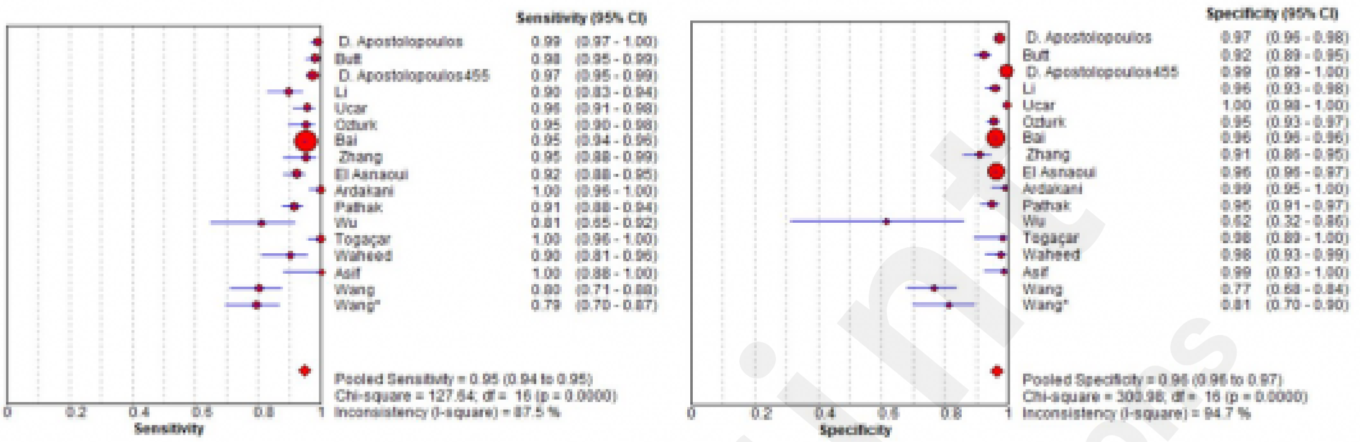
Supplementary Files

Figures

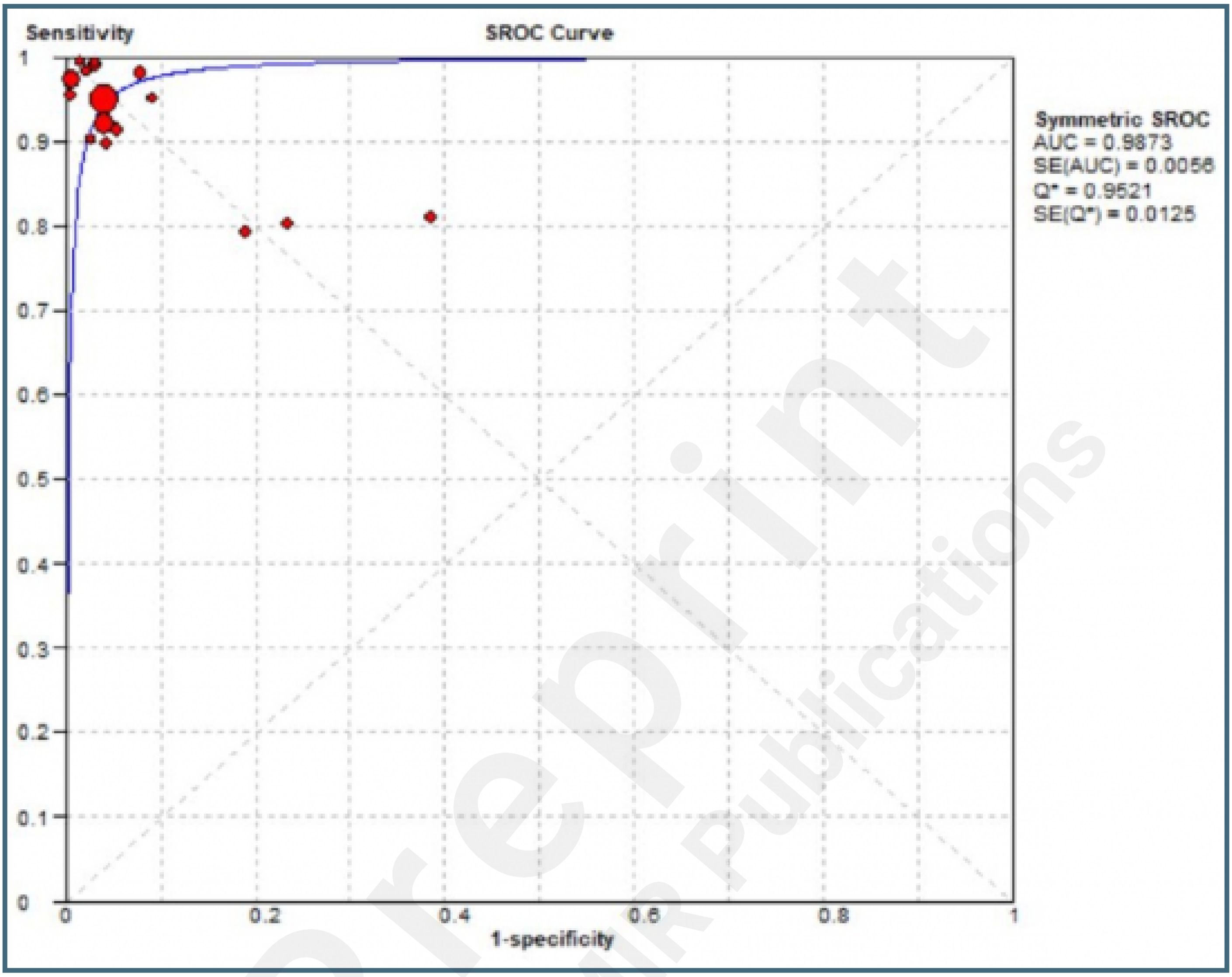
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for study selection.



Performance of the DL model for detecting COVID-19: Sensitivity (Left), and Specificity (right).



ROC curve of DL.



Multimedia Appendixes

Supplementary figures.
URL: http://asset.jmir.pub/assets/0dd2bedd49bba6d53b038140e3e18d93.docx