Applications of artificial intelligence and machine learning in diagnosis and prognosis of COVID-19 infection: A systematic review

Mahdieh Montazeri¹ , Ali Afraz¹ , Mitra Montazeri¹ , Sadegh Nejatizadeh² , Fatemeh Rahimi³ , Mohsen Taherian⁴ , Mohadeseh Montazeri⁵ , Leila Ahmadian^{1*}

¹Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

²Yasuj University of Medical Science, Yasuj, Iran

³PhD Candidate in Medical Informatics, Department of Health Information Technology and Management, Shahid Beheshti University of Medical Science, Tehran, Iran

⁴Young Researchers and Elite Club, Yasuj Branch, Islamic Azad University, Yasuj, Iran

⁵Technical Instructor of Computer Engineering, Department of Computer, Faculty of Fatimah, Kerman Branch Technical and Vocational University, Kerman, Iran

Article Info

Article type:
Review

Article History:

Received: 2021-07-27

Accepted: 2021-09-13

Published: 2021-10-03

* Corresponding author:

Leila Ahmadian

Medical Informatics Research
Center, Institute for Futures Studies
in Health, Kerman University
of Medical Sciences, Kerman, Iran

Email: l.Ahmadian@kmu.ac.ir

Keywords:

Artificial Intelligence

Machine Learning

Diagnosis

Prognosis

COVID-19

Coronavirus Disease 2019

ABSTRACT

Introduction: Our aim in this study was to summarize information on the use of intelligent models for predicting and diagnosing the Coronavirus disease 2019 (COVID-19) to help early and timely diagnosis of the disease.

Material and Methods: A systematic literature search included articles published until 20 April 2020 in PubMed, Web of Science, IEEE, ProQuest, Scopus, bioRxiv, and medRxiv databases. The search strategy consisted of two groups of keywords: A) Novel coronavirus, B) Machine learning. Two reviewers independently assessed original papers to determine eligibility for inclusion in this review. Studies were critically reviewed for risk of bias using prediction model risk of bias assessment tool.

Results: We gathered 1650 articles through database searches. After the full-text assessment 31 articles were included. Neural networks and deep neural network variants were the most popular machine learning type. Of the five models that authors claimed were externally validated, we considered external validation only for four of them. Area under the curve (AUC) in internal validation of prognostic models varied from .94 to .97. AUC in diagnostic models varied from 0.84 to 0.99, and AUC in external validation of diagnostic models varied from 0.73 to 0.94. Our analysis finds all but two studies have a high risk of bias due to various reasons like a low number of participants and lack of external validation.

Conclusion: Diagnostic and prognostic models for COVID-19 show good to excellent discriminative performance. However, these models are at high risk of bias because of various reasons like a low number of participants and lack of external validation. Future studies should address these concerns. Sharing data and experiences for the development, validation, and updating of COVID-19 related prediction models is needed.

Cite this paper as:

Montazeri M, Afraz A, Montazeri M, Nejatizadeh S, Rahimi F, Taherian M, Montazeri M, Ahmadian L. Applications of artificial intelligence and machine learning in diagnosis and prognosis of COVID-19 infection: A systematic review. Front Health Inform. 2021; 10: 93. DOI: [10.30699/fhi.v10i1.321](https://doi.org/10.30699/fhi.v10i1.321)

INTRODUCTION

Currently, the coronavirus disease 2019 (COVID-19) is a serious threat to global health. Since the outbreak of the disease in December 2019 in Wuhan, China, the number of people with the disease worldwide has exceeded 194 million (up to July 26, 2021) [1]. The world health organization (WHO) has identified the disease as a major concern that requires

international cooperation [2, 3].

The prevalence of this disease has led to an increase in demand for hospital beds and a shortage of medical equipment. Despite many efforts to control this disease, many countries are facing a crisis of intensive care [4-6]. On the other hand, the contamination of medical staff has doubled the health system's problems. In order to reduce the burden on the health

care system and, at the same time, provide the best possible care for patients, timely diagnosis, and effective prognosis of this disease is important and necessary. The disease's prediction at an early stage helps providers prevent delays for the best possible treatment. Disease prediction models by combining several variables or characteristics can diagnose or estimate the risk of people developing the disease, which can be of great help in the field of health in resource allocation [6].

Computed tomography (CT), as a non-invasive imaging technique, can detect some of the symptoms associated with the COVID-19 disease in a patient's lungs [7, 8]. Therefore, CT can be used as an effective method for early screening and the diagnosis of COVID-19. Despite the advantages of this method, in some cases, this method cannot differentiate some imaging features between COVID-19 and other types of pneumonia, which makes it difficult to diagnose the disease. On the other hand, diagnosing the specific features of the disease in CT images may be challenging for novice medical staff.

Artificial intelligence has achieved great success in medical imaging and image analysis using deep learning technology due to its high feature extraction capability [9, 10]. This technique can help medical staff in early diagnosis and prepare reports that are time-consuming tasks, by providing automatic reports [11, 12]. Several studies have suggested the use of this technique in the diagnosis of diseases. For example, some studies have used deep learning techniques to diagnose and differentiate bacterial and viral pneumonia in pediatric chest radiography [13, 14]. There has also been a great deal of effort in diagnosing various chest CT imaging features in different diseases [15, 16]. Artificial intelligence has also been used successfully in epidemic expansion prediction methods [17].

Various models from rule-based systems to advanced machine learning models (deep learning) have been published in the field of the diagnosis and prognosis of the COVID-19 disease, which had greatly contributed to the diagnosis and treatment of this disease in the field of health and preservation of human life [18]. Our aim in this study was to conduct a systematic review and report on intelligent models (all types of methods and algorithms that were considered an artificial intelligence or machine learning method) to predict and diagnose the COVID-19 disease to help early and timely diagnosis of the disease.

MATERIAL AND METHODS

Information source and search

We performed a systematic search of the PubMed, Web of Science, IEEE, ProQuest, Scopus, bioRxiv, and

medRxiv databases for studies done up to 20 April 2020. To report this study, PRISMA guidelines were followed [19]. The search strategy consisted of two groups of keywords: A) novel coronavirus, B) machine learning (Fig 1).

```
(COVID-19 OR "Novel coronavirus" OR
"2019 nCoV" OR "Wuhan coronavirus" OR
"2019-nCoV" OR "2019-nCoV disease" OR
"coronavirus disease-19" OR COVID19 OR
SARS-CoV-2 OR "severe acute respiratory
syndrome coronavirus 2")
AND
("Machine Learning" OR "Artificial
Intelligence" OR "Natural Language
Processing" OR "Neural Networks
Computer" OR "Support Vector Machine" OR
Machine learning OR "Artificial Intelligence"
OR "Naive Bayes" OR "bayesian learning" OR
```

Fig 1: Search strategy

Inclusion and exclusion criteria

All studies applying intelligent models for predicting and diagnosing the COVID-19 disease were considered. We included original studies regardless of their languages. Editorials, all review types, commentaries, letters, books, presentations, and conference papers were excluded.

Study selection

After removing duplicate studies, two authors independently reviewed titles and abstracts of all identified studies. After this initial screening, the same authors independently reviewed the full-text of the remaining studies. The disagreements regarding the selection of the studies were resolved through discussion. During the screening of the studies, the reviewers documented the reasons for the exclusion of each study. We used a free web and mobile application platform for paper screening (Rayyan QCRI systematic review software) [20].

Data extraction and synthesis

Reviewers used a standardized data extraction form based on the critical appraisal and data extraction for systematic reviews of prediction modeling studies (CHARMS) checklist [21]. Five reviewers used a data extraction form to extract specific details about each study. This form consisted of setting, data source and outcome, sample size, machine learning model, the model's performance, and evaluation type.

Risk of bias assessment

The risk of bias was assessed using the prediction model risk of bias assessment tool (PROBAST) [22, 23].

RESULTS

Study selection

We retrieved 1650 articles through database searches. After title and abstract screening, 67 articles were identified for full-text assessment. The full-text assessment excluded 36 studies due to various reasons. The detailed process regarding the identification and selection of studies is presented in Fig 2. Twenty-nine studies of 31 included studies developed or analyzed machine learning-based models for the diagnose [24-51] and three of them for the prognoses of the COVID-19 disease [52-54]. These studies were selected for data extraction and critical appraisal (Table 1). Among the three prognosis studies, two studies tried to determine the risk factors of COVID-19 disease progression [54], another predicting mortality risk in patients with COVID-19.

Fig 2: Study identification and selection process.

Data types

Twenty-three studies used CT and x-ray images to train models for the diagnosis of COVID-19 [26-35, 38-49, 51], and one study used CT and x-ray images for the prognosis of COVID-19 [31]. Five studies used demographic, epidemiological data (e.g., travel history, and contact information), and other clinical information (e.g. symptoms, signs, comorbidities, laboratory data like hemoglobin, platelets, red blood cells, red cell distribution width, leukocytes, lymphocytes) to train models for the diagnosis of the COVID-19 patients [24, 25, 36, 37, 50] and three studies used demographic, epidemiological data (e.g. travel history, and contact information), and other clinical information (e.g. symptoms, symptom onset

date) for the prognosis of the COVID-19 patients [52-54]. Because two-thirds of algorithms were some variants of Neural Network algorithms and some other studies lack reporting about selected features' effects, we did not report feature importance in the results.

Datasets

Thirteen studies used datasets from one or more institutions from China [28, 29, 32-39, 50, 51, 54]. Because the authors did not always report institutions' names, we could not reliably identify the overlaps between datasets. Thirteen studies used the public available COVID-19 datasets from GitHub or Kaggle [26-28, 40-47, 49, 52, 53]. Most of these publicly available international datasets do overlap, and many studies used the same repositories. But because of different access times and continuous updating of these data, they may have varied among different studies. Two studies used data obtained from a few institutions in Italy [48, 54], and two other studies used data from Brazil, one from the Brazilian COVID-19 national surveillance system and other data were obtained from Hospital Israelite Albert Einstein in São Paulo [24, 25]. One study secured data from Montgomery County hospital, USA [27].

Algorithms

In the included studies, 71 different machine learning models were used for the diagnosis and the prognosis of the COVID-19 patients. Neural networks and deep neural network variants used in 49 models (69%) were the most popular machine learning types, of which 45 of them were convolutional networks based on algorithms used for image processing. In neural networks and deep neural network variants, ResNet with 14 variants (20%), and VGGNet with 4 variants (5%) were the most frequent deep learning sub-type models. Other employed machine learning models were support vector machine (5 variants (7%)), logistic regression (6 variants (8%)), K-nearest neighbor, decision tree, random forest, AdaBoost, and a few others. Area under the curve (AUC) in internal validation of prognostic models varied from 0.94 to 0.97, and there was no external validation for prognostic models. AUC in internal validation of diagnostic models ranged from 0.84 to 0.99, and AUC in external validation of diagnostic models went from 0.73 to 0.94.

Validations

From all 31 included studies, there were only five models that authors claimed were validated externally [24, 31, 32, 36, 37]. But two studies used data from the same source in different time periods as training data for external validation [31, 36]. One study used data from 8 centers in China for training and validation but randomly divided for internal and

external validation [37]. Another study divided data from 3 hospitals randomly into three sets of training, internal validation, and external validation [32]. We

considered external validation of all four above mentioned studies biased.

Table1: Characteristics of included studies

Ref.	Setting, data source and outcome	Sample size	Predictors	Machine Learning Model(s)	Performance of the (best) model	Evaluation Type	ROB
[24]	Predict SARS-CoV-2 infection in suspected patients reported to the Brazilian surveillance system	5739	Demographic and clinical information	logistic regression	Internal: Sensitivity .87, Specificity .92, AUC .95 External: Sensitivity .46, Specificity .79, AUC .73	Internal and external	Low
[25]	Predict the risk of positive COVID-19 in emergency care admission suspected patients in Sao Paulo, Brazil	256	Demographic and clinical symptoms and laboratory results	SVM, Random Forests, Neural Networks, Logistic Regression, Grad. Boost. Trees	Sensitivity .67, Specificity .85, AUC .84	Internal	High
[26]	Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub	126	Image features	SVM	Sensitivity .91, Specificity .98, AUC .99	Internal	High
[27]	Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub and Montgomery County hospital	40	Image features	SVM	Sensitivity .95, Specificity .99	Internal	High
[52]	Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub	>117,000	Image features	SVM	AUC .95	Internal	High
[28]	Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub and NIH Chest X-ray Dataset	278	Image features	ResNet-50-convolutional neural network	Sensitivity .98, F1 score .98	Internal	High
[53]	Predict mortality of COVID-19 patients using Kaggle Novel Corona Virus 2019 Dataset	1085	Demographic and clinical information, Date of hospital visits, relation to Wuhan of China, death or recovery	Random forest	AUC .97	Internal	High
[29]	Predict COVID-19 and other Common Infectious Diseases of the Lung using Chest CT Scan Images from seven hospitals in Wuhan, China	89,628	Image features	Deep CNN	Sensitivity .98, Specificity .98, AUC .99	Internal	High
[30]	Differentiating novel coronavirus pneumonia and influenza pneumonia using Chest CT Scan Image from eight tertiary referral centers in China	1641	Image features	VGGNet	AUC .99	Internal	High
[31]	Predict COVID-19 and using Chest X-rays images from two hospitals in Lombardy, Italy	1110	Image features	Ensemble of ten convolutional neural networks	Internal: Sensitivity .78, Specificity .82, AUC .90 External:	Internal and External	High

Ref.	Setting, data source and outcome	Sample size	Predictors	Machine Learning Model(s)	Performance of the (best) model	Evaluation Type	ROB
					Sensitivity .80, Specificity .81, AUC .80		
[32]	Predict COVID-19 and using Chest CT images from three hospitals China	1065	Image features	Ensemble of Deep learning Neural Networks	Internal: Sensitivity .88, Specificity .87, AUC .95 External: Sensitivity .83, Specificity .67, AUC .81	Internal and External	Low
[33]	Predict COVID-19 and using Chest CT images from three hospitals, China	275	Image features	Deep Learning algorithms: VGG16, DenseNet, ResNet, DRE-Net	Sensitivity .93, AUC .99	Internal	High
[34]	Predict COVID-19 and using Chest CT images from one hospital, china	630	Image features	Deep CNN	Sensitivity .90, Specificity .91, AUC .95	Internal	High
[35]	Predict COVID-19 and using patients information from one hospital, Wuhan China	106	imaging features, symptoms, demographics	Deep Learning(UNet++)	Sensitivity 1, Specificity .93	Internal	High
[36]	Predict COVID-19 and using patients information from some institutions in China	620	Demographics and lab results	multivariate logistic regression	Internal: AUC .89 External: AUC .87	Internal and External	High
[37]	Predict COVID-19 and using patients information from one hospital, Beijing China	164	Demographic information, symptoms, lab results	logistic regression with LASSO, logistic regression with Ridge regularization, decision tree, Adaboost algorithms	External: Sensitivity 1, Specificity .72, AUC .84 External: Sensitivity 1, Specificity .78, AUC .94	Internal and External	High
[38]	Predict COVID-19 and using Chest CT images from three hospitals, Wuhan China	1418	Imaging features	Deep residual network (ResNet), Inception networks, Dual path network (DPN-92), Residual attention network (Attention ResNet)	AUC .99	Internal	High
[54]	Predict COVID-19 and using patients information from one hospital, Wuhan China	-	Clinical and laboratory data, chest CT images	PCA+LDA PCA+SVM MLP MLP+LSTM	Sensitivity .93, Specificity .86, AUC .95	Internal	High
[39]	Predict COVID-19 and using Chest CT images from six medical centers in China	4356	Imaging features	Modified RestNet50	Sensitivity .90, Specificity .96, AUC .96	Internal	High
[40]	Predict COVID-19 and using Chest CT images from publicly available Chest X-ray and CT	170 X-ray images, 361 CT images (only 42 COVID-19 image from 22 patients)		Deep learning (CNN), modified AlexNet	Sensitivity 1, Specificity .96	Internal	High
[41]	Predict COVID-19 and using Chest X-ray images	5941 (limited)	Imaging features	ResNet 50	Sensitivity 1, F1 Score 1	Internal	High

Ref.	Setting, data source and outcome	Sample size	Predictors	Machine Learning Model(s)	Performance of the (best) model	Evaluation Type	ROB
	from chest radiography images(GitHub)	Covid-19 patients)					
[42]	Predict COVID-19 and using Chest X-ray and CT images from chest radiography images(GitHub+kaggle)	16,756	Imaging features	Deep CNN	Sensitivity .80	Internal	High
[43]	Predict COVID-19 and using Chest X-ray images from chest radiography images(GitHub+kaggle)	1531	Imaging features	Deep learning[SGD algorithm+Inference algorithm with Threshold(0.15-0.50)]	Sensitivity .97, Specificity .95	Internal	High
[44]	Predict COVID-19 and using Chest X-ray images from publicly available data(GitHub+kaggle+SIRM)	2876 (only 190 Covid-19)	Imaging features	AlexNet ResNet18 DenseNet201 SqueezeNet	Sensitivity .97, Specificity .99, AUC .99	Internal	High
[45]	Predict COVID-19 and using Chest X-ray images from few publicly available datasets on Kaggle and GitHub	2855	Imaging features	Deep Learning Algorithms: VGG19, Inception, Xception, MobileNet, Inception_Resnet_V2	Sensitivity .93, Specificity .99	Internal	High
[46]	Predict COVID-19 and using Chest X-ray images from two publicly available datasets on Kaggle and GitHub	50	Imaging features	Deep Learning Algorithms: VGG19, InceptionV3,Xception, DensNet201, MobileNet_V2, Inception_Resnet_V2, Resnet_V2	Sensitivity 1, AUC .90	Internal	High
[47]	Predict COVID-19 and using Chest X-ray images from publicly available datasets on Kaggle and GitHub	100	Imaging features	Deep Learning (Inception_V3) Deep Learning (ResNet50) Deep Learning (Inception_Resnet_V2)	Sensitivity .96, Specificity 1	Internal	High
[48]	Predict COVID-19 and using Chest CT images from one medical center in Italy	150 CT belong to 53 cases	Imaging features	SVM	Sensitivity .97, Specificity .99	Internal	High
[49]	Predict COVID-19 and using Chest X-ray images from publicly available datasets on GitHub	196	Imaging features	Modified ResNet18	Sensitivity .98, Specificity .92, AUC .94	Internal	High
[50]	Predict COVID-19 and using Chest CT images from three medical centers in China	2685 (1658 covid19)	Imaging features	Modified VB-Net	Sensitivity .91, Specificity .83, AUC .94	Internal	High
[51]	Predict COVID-19 and using Chest CT images from designated covid19 Zhejiang Province, China	219 CT covid-19, 224 CT Influenza-A, pneumonia, and 175 Ct healthy	Imaging features	ResNet, ResNet with location-attention mechanism	Sensitivity .76, F1 Score .76	Internal	High

Risk of bias

We critically reviewed studies for risk of bias using PROBAST. It includes 20 signaling questions across 4 domains (participants, predictors, outcome, and analysis). This explanation and elaboration document describes the rationale for including each domain and signaling question and guides researchers, reviewers, readers, and guideline

developers in how to use them to assess risk of bias and applicability concerns. Our analysis revealed that except for two, all studies had some sort of bias due to various reasons like a low number of participants, lack of external validation, and failure to meet study's goal. Moreover, insufficient reporting of performance parameters in result, not enough information about where the participant's data came

from (e.g. from what institutions), or too limited COVID-19 patients were other reasons for bias within the conducted studies.

DISCUSSION

In this review, we systematically reviewed intelligent models in the diagnosis and prognosis of the COVID-19 disease. The results show that many studies used artificial intelligence techniques for predicting diagnosis and prognosis of COVID-19. We identified and evaluated 31 studies that presented diagnostic and predictive models, of which 28 were related to the diagnosis of the disease, and 3 were related to the prognosis of the disease (Fig 2). We found that all models in this review reported good to excellent predictive performance.

After appraising the quality of studies based on PROBAST. All studies except for two were appraised to have a high risk of bias owing to a combination of poor reporting and poor methodological conduct for statistical methods used, participant selection, predictor description statistical methods used, low number of participants and lack of external validation.

The high risk of bias indicates that these models' performance may not be as good as the performance reported in these studies. We recommend that future studies follow the TRIPOD guidelines [55] to improve the descriptions of the study population and their modeling results.

In this study, 71 variations of machining learning models were applied to the COVID-19 datasets to find the best model for diagnosing and predicting this disease. In most studies, more than one Machine learning method was used, so that in 5 studies, 5 Machine Learning methods or more were used separately on the COVID-19 datasets [25, 37, 38, 45, 46]. Our results showed that neural network and deep neural network were the most commonly used algorithms. According to the performance of these algorithms, it seems they act better than other algorithms.

The inconsistency in the relative frequencies of the outcomes that were predicted presents an important barrier to the prediction modeler. In a setting with various relative frequency of the outcome, a prediction model might produce mis-calibrated predictions [56], and before it can be applied in new settings, it might need to be updated [57, 58].

As applying individual participants' data from multiple settings and countries might allow a better understanding of the prediction models across different settings and populations, it is recommended to apply the prediction models in data obtaining from various data sources. This could increase the strength of prediction models in actual care [59, 60].

Various studies have shown that the sample size affects the accuracy of Machine learning algorithms. The small sample size reduced the accuracy of the algorithm [61-63]. Most of the datasets in this review had an acceptable number of samples with appropriate performance. In a number of studies, the sample size was small, and that might lead to an increase in the risk of overfitting the model [64]. Almost all of the studies have been done on small datasets or local datasets. The low sample size in these studies is due to the fact that the data related to COVID-19 are being completed and updated due to the novelty of this disease. The data were often limited to data from China, Italy, and in few cases from international records. This may be due to limited time and time constraints due to the urgency of the issue and also the difficulty of obtaining the COVID-19 data as much as needed to perform such a study because of privacy issues.

We reviewed 24 studies that used advanced machine learning methodology on chest CT and X-ray scans to diagnose the COVID-19 disease, COVID-19 related pneumonia, or to assist in the segmentation of lung images. The predictive performance measures showed the almost perfect ability to identify COVID-19.

The main goal of predictive models is to support medical decision making. Therefore, identifying the target population is very important. This target population must also be carefully described so that the performance of the developed or valid model can be evaluated in the text, and users can know who is being used when preparing predictions. Because of lack of external validation in many of the studies included in this review (Table 1), the generalizability of the researches results are in doubt.

Common metrics reported were accuracy sensitivity, specificity, and AUC. However, accuracy is minor than AUC, particularly when imbalanced datasets are used [65]. The imbalanced data do not influence the AUC measure, but precision-recall curves may reflect model performance more accurately [66]. In this study, none of the included studies utilized and integrated optimization techniques, such as genetic algorithms and particle swarm optimization, to their systems. Moreover, no study in the literature exploited clustering algorithms for the detection and diagnosis of the COVID-19.

With the fast publication of new articles on diagnostic and predictive models of COVID-19, this systematic review cannot be considered as an up-to-date list of all COVID-19 related diagnostic and predictive models. However, when creating and using a model for diagnosis and prediction, it is necessary to review previous studies and use the opinion of experts. A strength of this study was reviewing a database including unpublished studies, besides reviewing published studies.

Challenges and opportunities

It is important to identify the target population and describe it accurately, to evaluate and validate the predictive model, and for users to know what population these predictive models are applied to. However, the studies included in this systematic review often lacked a sufficient and adequate description of the study population, which could cast doubt on the models' use. To improve the reporting of this type of studies it is recommended that future studies follow the TRIPOD guidelines, which can help them to better describe the study population and select the appropriate model [55].

The absence of a large dataset in the literature for COVID-19 is considered another challenging task because it prevents the understanding of viral features and patterns. Although there is an abundance of data about COVID-19 outbreak statistics like the number of infected people, recovered people, and mortality, there is a shortage of more detailed data like radiology reports and images, infected people symptoms on large scale.

According to various algorithms in this systematic review, it is suggested that factors including age, body temperature, and signs and symptoms be identified to diagnose COVID-19 disease. For prognostic models, useful variables area: age, sex, C reactive protein, lactic dehydrogenase, lymphocyte count, and potentially features derived from CT scoring. For diagnostic models, these include age, body temperature, and signs and symptoms. Predictors like albumin, direct bilirubin, and red blood cell distribution width in both a diagnostic and prognosis

model should be considered.

CONCLUSION

Diagnostic and prognostic models for COVID-19 show good to excellent discriminative performance. However, these models are at high risk of bias because of various reasons like a low number of participants, unclear population definition, and lack of external validation.

Future studies should address these concerns. Sharing data and experiences for the development, validation, and updating of COVID-19 related prediction models is needed.

AUTHOR'S CONTRIBUTION

Mahdieh M and AA were the major contributors in writing the manuscript and screening the papers. Mitra M, SN, FR, MT, Mohadeseh M helped in writing the manuscript. LA was the research supervisor and final editor.

All authors contributed to the literature review, design, data collection and analysis, drafting the manuscript, read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

- World Health Organization. Coronavirus disease (COVID-2019) situation reports 2020 [Internet]. 2020 [cited: 27 Jul 2021]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg*. 2020; 76: 71-6. PMID: 32112977 DOI: 10.1016/j.ijsu.2020.02.034 [PubMed]
- World Health Organization. Coronavirus disease (COVID-19) pandemic 2020 [Internet]. 2020 [cited: 27 Jul 2021]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
- Arabi YM, Murthy S, Webb S. COVID-19: A novel coronavirus and a novel challenge for critical care. *Intensive Care Med*. 2020; 46(5): 833-6. PMID: 32125458 DOI: 10.1007/s00134-020-05955-1 [PubMed]
- Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. *JAMA*. 2020; 323(16): 1545-6. PMID: 32167538 DOI: 10.1001/jama.2020.4031 [PubMed]
- Xie J, Tong Z, Guan X, Du B, Qiu H, Slutsky AS. Critical care crisis and some recommendations during the COVID-19 epidemic in China. *Intensive Care Med*. 2020; 46(5): 837-40. PMID: 32123994 DOI: 10.1007/s00134-020-05979-7 [PubMed]
- Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology*. 2020; 295(1): 202-7. PMID: 32017661 DOI: 10.1148/radiol.202002030 [PubMed]
- Yang X, Yu Y, Xu J, Shu H, Liu H, Wu Y, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *Lancet Respir Med*. 2020; 8(5): 475-81. PMID: 32105632 DOI: 10.1016/S2213-2600(20)30079-5 [PubMed]
- Xia C, Li X, Wang X, Kong B, Chen Y, Yin Y, et al. A multi-

- modality network for cardiomyopathy death risk prediction with CMR images and clinical information. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2019.
10. Kong B, Wang X, Bai J, Lu Y, Gao F, Cao K, et al. Learning tree-structured representation for 3D coronary artery segmentation. *Comput Med Imaging Graph*. 2020; 80: 101688. PMID: 31926366 DOI: 10.1016/j.compmedimag.2019.101688 [PubMed]
 11. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz C. Deep learning in neuroradiology. *AJNR Am J Neuroradiol*. 2018; 39(10): 1776-84. PMID: 29419402 DOI: 10.3174/ajnr.A5543 [PubMed]
 12. Yuan J, Liao H, Luo R, Luo J. Automatic radiology report generation based on multi-view image fusion and medical concept enrichment. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2019.
 13. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018; 172(5): 1122-31. PMID: 29474911 DOI: 10.1016/j.cell.2018.02.010 [PubMed]
 14. Rajaraman S, Candemir S, Kim I, Thoma G, Antani S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Appl Sci (Basel)*. 2018; 8(10): 1715. PMID: 32457819 DOI: 10.3390/app8101715 [PubMed]
 15. Depeursinge A, Chin AS, Leung AN, Terrone D, Bristow M, Rosen G, et al. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution CT. *Invest Radiol*. 2015; 50(4): 261-7. PMID: 25551822 DOI: 10.1097/RLI.0000000000000127 [PubMed]
 16. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging*. 2016; 35(5): 1207-16. PMID: 26955021 DOI: 10.1109/TMI.2016.2535865 [PubMed]
 17. Pal R, Sekh AA, Kar S, Prasad DK. Neural network based country wise risk prediction of COVID-19. *Applied Sciences*. 2020; 10(18): 6448.
 18. Carr D. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak [Internet]. 2020 [cited: 17 Jul 2021]. Available from: <https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak>
 19. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD statement. *JAMA*. 2015; 313(16): 1657-65. PMID: 25919529 DOI: 10.1001/jama.2015.3656 [PubMed]
 20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan: A web and mobile app for systematic reviews. *Syst Rev*. 2016; 5(1): 210. PMID: 27919275 DOI: 10.1186/s13643-016-0384-4 [PubMed]
 21. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med*. 2014; 11(10): e1001744. PMID: 25314315 DOI: 10.1371/journal.pmed.1001744 [PubMed]
 22. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Annals of Internal Medicine*. 2019; 170: W1-33.
 23. Wynants L, van Calster B, Bonten MM, Collins GS, Debray TP, de Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal. *BMJ*. 2020; 369: m1328. PMID: 32265220 DOI: 10.1136/bmj.m1328 [PubMed]
 24. Diaz-Quijano FA, da Silva JMN, Ganem F, Oliveira S, Vesga-Varela AL, Croda J. A model to predict SARS-CoV-2 infection based on the first three-month surveillance data in Brazil. *Trop Med Int Health*. 2020; 25(11): 1385-94. PMID: 32790891 DOI: 10.1111/tmi.13476 [PubMed]
 25. de Moraes Batista AF, Miraglia JL, Donato THR, Chiavegatto Filho ADP. COVID-19 diagnosis prediction in emergency care patients: A machine learning approach [Internet]. 2020 [cited: 14 Apr 2020]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.04.04.20052092>
 26. Ozturk S, Ozkaya U, Barstugan M. Classification of coronavirus images using Shrunken features. *Int J Imaging Syst Technol*. 2020; Online ahead of print. PMID: 32904960 DOI: 10.1002/ima.22469 [PubMed]
 27. Mahdy LN, Hassanien AE, Ezzat KA, Elmousalami HH, Aboul Ella H. Automatic X-ray COVID-19 lung image classification system based on multi-level thresholding and support vector machine [Internet]. 2020 [cited: 14 Apr 2020]. Available from: <https://www.medrxiv.org/content/10.1101/2020.03.30.20047787v1>
 28. Bukhari SUK, Bukhari SSK, Syed A, Shah SSH. The diagnostic evaluation of convolutional neural network (CNN) for the assessment of chest X-ray of patients infected with COVID-19 [Internet]. 2020 [cited: 14 Apr 2020]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.03.26.20044610>
 29. Fu M, Yi S-L, Zeng Y, Ye F, Li Y, Dong X, et al. Deep learning-based recognizing COVID-19 and other common infectious diseases of the lung by chest CT scan images [Internet]. 2020 [cited: 14 Apr 2020]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.03.28.20046045>
 30. Zhou M, Chen Y, Wang D, Xu Y, Yao W, Huang J, et al. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. *Ann Transl Med*. 2021; 9(2): 111. PMID: 33569413 DOI: 10.21037/atm-20-5328 [PubMed]

31. Castiglioni I, Ippolito D, Interlenghi M, Monti CB, Salvatore C, Schiaffino S, et al. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: A first experience from Lombardy, Italy. *Eur Radiol Exp*. 2021; 5(1): 7. PMID: 33527198 DOI: 10.1186/s41747-020-00203-z [PubMed]
32. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *Eur Radiol*. 2021; 31(8): 6096-104. PMID: 33629156 DOI: 10.1007/s00330-021-07715-1 [PubMed]
33. Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z, et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Trans Comput Biol Bioinform*. 2021; Online ahead of print. PMID: 33705321 DOI: 10.1109/TCBB.2021.3065361 [PubMed]
34. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Transactions on Medical Imaging*. 2020; 39(8): 2615-25.
35. Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study. *Sci Rep*. 2020; 10(1): 19196. PMID: 33154542 DOI: 10.1038/s41598-020-76282-0 [PubMed]
36. Meng Z, Wang M, Song H, Guo S, Zhou Y, Li W, et al. Development and utilization of an intelligent application for aiding COVID-19 diagnosis [Internet]. 2020 [cited: 14 Apr 2020]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.03.18.20035816>
37. Feng C, Huang Z, Wang L, Chen X, Zhai Y, Zhu F, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics. *Ann Transl Med*. 2021; 9(3): 201. PMID: 33708828 DOI: 10.21037/atm-20-3073 [PubMed]
38. Jin S, Wang B, Xu H, Luo C, Wei L, Zhao W, et al. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. *Appl Soft Comput*. 2021; 98: 106897. PMID: 33199977 DOI: 10.1016/j.asoc.2020.106897 [PubMed]
39. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*. 2020; 200905. PMID: 32191588 DOI: 10.1148/radiol.2020200905 [PubMed]
40. Maghddid HS, Asaad AT, Ghafoor KZ, Sadiq AS, Khan MK. Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms [Internet]. 2020 [cited: 15 Apr 2020]. Available from: <http://arxiv.org/abs/2004.00038>
41. Farooq M, Hafeez A. COVID-ResNet: A deep learning framework for screening of COVID19 from radiographs [Internet]. 2020 [cited: 15 Apr 2020]. Available from: <http://arxiv.org/abs/2003.14395>
42. Wang L, Wong A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. *Scientific Reports*. 2020; 10: 19549.
43. Zhang J, Xie Y, Li Y, Shen C, Xia Y. COVID-19 screening on chest X-ray images using deep learning based anomaly detection [Internet]. 2020 [cited: 15 Apr 2020]. Available from: <http://arxiv.org/abs/2003.12338>
44. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access*. 2020; 8: 132665-76.
45. Apostolopoulos ID, Bessiana T. Covid-19: Automatic detection from X-Ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med*. 2020; 43(2): 635-40. PMID: 32524445 DOI: 10.1007/s13246-020-00865-4 [PubMed]
46. Hemdan EE-D, Shouman MA, Karar ME. COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images [Internet]. 2020 [cited: 15 Apr 2020]. Available from: <http://arxiv.org/abs/2003.11055>
47. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Anal Appl*. 2021; Online ahead of print. PMID: 33994847 DOI: 10.1007/s10044-021-00984-y [PubMed]
48. Barstugan M, Ozkaya U, Ozturk S. Coronavirus (COVID-19) classification using CT images by machine learning methods [Internet]. 2020 [cited: 15 Apr 2020]. Available from: <http://arxiv.org/abs/2003.09424>
49. Abbas A, Abdelsamea M, Gaber M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Appl Intell*. 2021; 51: 854-64.
50. Shi F, Xia L, Shan F, Wu D, Wei Y, Yuan H, et al. Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. *Phys Med Biol*. 2021; Online ahead of print. PMID: 33607630 DOI: 10.1088/1361-6560/abe838 [PubMed]
51. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. A deep learning system to screen coronavirus disease 2019 pneumonia. *Engineering (Beijing)*. 2020; 6(10): 1122-9. PMID: 32837749 DOI: 10.1016/j.eng.2020.04.010 [PubMed]
52. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. *Smart Health (Amst)*. 2021; 20: 100178. PMID: 33521226 DOI: 10.1016/j.smhl.2020.100178 [PubMed]
53. Sarkar J, Chakrabarti P. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with COVID-19 [Internet]. 2020 [cited: 14 Apr 2020]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.03.25.20043331>
54. Fang C, Bai X, Zhou Y, Bai S, Liu Z, Chen Q, et al. Deep

- learning for predicting COVID-19 malignant progression. *Medical Image Analysis*. 2021; 72: 102096.
55. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*. 2015; 350: g7594. PMID: 25569120 DOI: 10.1136/bmj.g7594 [[PubMed](#)]
 56. van Calster B, McLernon D, van Smeden M, Wynants L, Steyerberg E. Calibration: The achilles heel of predictive analytics. *BMC Med*. 2019; 17(1): 230. PMID: 31842878 DOI: 10.1186/s12916-019-1466-7 [[PubMed](#)]
 57. Steyerberg EW. Clinical prediction models: Practical approach to development, validation, and updating. SpringerLink; 2019.
 58. Steyerberg EW. Evaluation of clinical usefulness. In: Steyerberg EW. Clinical prediction models: Practical approach to development, validation, and updating. SpringerLink; 2019.
 59. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. *BMJ*. 2016; 353: i3140. PMID: 27334381 DOI: 10.1136/bmj.i3140 [[PubMed](#)]
 60. Wynants L, Kent D, Timmerman D, Lundquist C, van Calster B. Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting. *Diagn Progn Res*. 2019; 3: 6. PMID: 31093576 DOI: 10.1186/s41512-019-0046-9 [[PubMed](#)]
 61. Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*. 2018; 178: 622-37. PMID: 29870817 DOI: 10.1016/j.neuroimage.2018.06.001 [[PubMed](#)]
 62. Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C, Alzheimer's Disease Neuroimaging Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*. 2012; 60(1): 59-70. PMID: 22166797 DOI: 10.1016/j.neuroimage.2011.11.066 [[PubMed](#)]
 63. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*. 2017; 145: 137-65. PMID: 27012503 DOI: 10.1016/j.neuroimage.2016.02.079 [[PubMed](#)]
 64. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020; 368: m441. PMID: 32188600 DOI: 10.1136/bmj.m441 [[PubMed](#)]
 65. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*. 2005;17(3): 299-310.
 66. Jeni LA, Cohn JF, de la Torre F. Facing imbalanced data: Recommendations for the use of performance metrics. *Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE; 2013.