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

Mahdieh Montazeri¹, Ali Afraz¹, Mitra Montazeri³, Sadegh Nejatzadeh², Fatemeh Rahimi³, Mohsen Taherian¹, Mohadesheh Montazeri⁵, Leila Ahmadian¹*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

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 .84 to .99, and AUC in external validation of diagnostic models varied from .73 to .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:

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).


("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"

**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 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 prognostics studies, two studies tried to determine the risk factors of COVID-19 disease progression [54], another predicting mortality risk in patients with COVID-19.

![Study identification and selection process.](image)

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
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.

Table 1: Characteristics of included studies

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Setting, data source and outcome</th>
<th>Sample size</th>
<th>Predictors</th>
<th>Machine Learning Model(s)</th>
<th>Performance of the (best) model</th>
<th>Evaluation Type</th>
<th>ROB</th>
</tr>
</thead>
<tbody>
<tr>
<td>[24]</td>
<td>Predict SARS-CoV-2 infection in suspected patients reported to the Brazilian surveillance system</td>
<td>5739</td>
<td>Demographic and clinical information</td>
<td>logistic regression</td>
<td>Internal: Sensitivity .87, Specificity .92, AUC .95 External: Sensitivity .46, Specificity .79, AUC .73</td>
<td>Internal and external</td>
<td>Low</td>
</tr>
<tr>
<td>[25]</td>
<td>Predict the risk of positive COVID-19 in emergency care admission suspected patients in Sao Paulo, Brazil</td>
<td>256</td>
<td>Demographic and clinical symptoms and laboratory results</td>
<td>SVM, Random Forests, Neural Networks, Logistic Regression, Grad. Boost. Trees</td>
<td>Sensitivity .67, Specificity .85, AUC .84</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[26]</td>
<td>Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub</td>
<td>126</td>
<td>Image features</td>
<td>SVM</td>
<td>Sensitivity .91, Specificity .98, AUC .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[27]</td>
<td>Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub and Montgomery County hospital</td>
<td>40</td>
<td>Image features</td>
<td>SVM</td>
<td>Sensitivity .95, Specificity .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[28]</td>
<td>Predict COVID-19 using CT and X-ray using COVID-19 image data collection in GitHub</td>
<td>&gt;117,000</td>
<td>Image features</td>
<td>SVM</td>
<td>AUC .95</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[30]</td>
<td>Predict mortality of COVID-19 patients using Kaggle Novel Corona Virus 2019 Dataset</td>
<td>1085</td>
<td>Demographic and clinical information, Date of hospital visits, relation to Wuhan of China, death or recovery</td>
<td>Random forest</td>
<td>AUC .97</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[31]</td>
<td>Predict COVID-19 and other Common Infectious Diseases of the Lung using Chest CT Scan Images from seven hospitals in Wuhan, China</td>
<td>89,628</td>
<td>Image features</td>
<td>Deep CNN</td>
<td>Sensitivity .98, Specificity .98, AUC .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[32]</td>
<td>Differentiating novel coronavirus pneumonia and influenza pneumonia using Chest CT Scan Image from eight tertiary referral centers in China</td>
<td>1641</td>
<td>Image features</td>
<td>VGGNet</td>
<td>AUC .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[33]</td>
<td>Predict COVID-19 and using Chest X-rays images from two hospitals in Lombardy, Italy</td>
<td>1110</td>
<td>Image features</td>
<td>Ensemble of ten convolutional neural networks</td>
<td>Internal: Sensitivity .78, Specificity .82, AUC .90</td>
<td>Internal and External</td>
<td>High</td>
</tr>
<tr>
<td>Ref.</td>
<td>Setting, data source and outcome</td>
<td>Sample size</td>
<td>Predictors</td>
<td>Machine Learning Model(s)</td>
<td>Performance of the (best) model</td>
<td>Evaluation Type</td>
<td>ROB</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>---------------------------</td>
<td>-------------------------------</td>
<td>-----------------</td>
<td>-----</td>
</tr>
<tr>
<td>[32]</td>
<td>Predict COVID-19 and using Chest CT images from three hospitals, China</td>
<td>1065</td>
<td>Image features</td>
<td>Ensemble of Deep learning Neural Networks</td>
<td>Sensitivity:.80, Specificity:.81, AUC:.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[33]</td>
<td>Predict COVID-19 and using Chest CT images from three hospitals, China</td>
<td>275</td>
<td>Image features</td>
<td>Deep Learning algorithms: VGG16, DenseNet, ResNet, DRE-Net</td>
<td>Sensitivity:.93, AUC:.99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[34]</td>
<td>Predict COVID-19 and using Chest CT images from one hospital, China</td>
<td>630</td>
<td>Image features</td>
<td>Deep.CNN</td>
<td>Sensitivity:.90, Specificity:.91, AUC:.95</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[35]</td>
<td>Predict COVID-19 and using patients information from one hospital, Wuhan China</td>
<td>106</td>
<td>imaging features, symptoms, demographics</td>
<td>Deep Learning(UNet++)</td>
<td>Sensitivity:.91, Specificity:.93</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[36]</td>
<td>Predict COVID-19 and using patients information from some institutions in China</td>
<td>620</td>
<td>Demographics and lab results</td>
<td>multivariate logistic regression</td>
<td>Internal: AUC:.89, External: AUC:.87</td>
<td>Internal and External</td>
<td>High</td>
</tr>
<tr>
<td>[37]</td>
<td>Predict COVID-19 and using patients information from one hospital, Beijing China</td>
<td>164</td>
<td>Demographic information, symptoms, lab results</td>
<td>logistic regression with LASSO, logistic regression with Ridge regularization, decision tree, Adaboost algorithms</td>
<td>External: Sensitivity:.72, AUC:.84, Internal: Sensitivity:.78, AUC:.94</td>
<td>Internal and External</td>
<td>High</td>
</tr>
<tr>
<td>[38]</td>
<td>Predict COVID-19 and using Chest CT images from three hospitals, Wuhan China</td>
<td>1418</td>
<td>Imaging features</td>
<td>Deep residual network (ResNet), Inception networks, Dual path network (DPN-92), Residual attention network (Attention ResNet)</td>
<td>AUC:.99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[39]</td>
<td>Predict COVID-19 and using patients information from one hospital, Wuhan China</td>
<td>-</td>
<td>Clinical and laboratory data, chest CT images</td>
<td>PCA+LDA, PCA+SVM, MLP, MLP+LSTM</td>
<td>Sensitivity:.93, Specificity:.86, AUC:.95</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[40]</td>
<td>Predict COVID-19 and using Chest CT images from six medical centers in China</td>
<td>4356</td>
<td>Imaging features</td>
<td>Modified RestNet50</td>
<td>Sensitivity:.90, Specificity:.96, AUC:.96</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>[41]</td>
<td>Predict COVID-19 and using Chest X-ray images</td>
<td>5941 (limited)</td>
<td>Imaging features</td>
<td>ResNet 50</td>
<td>Sensitivity 1, F1 Score 1</td>
<td>Internal</td>
<td>High</td>
</tr>
</tbody>
</table>
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 chest radiography images (GitHub) Covid-19 patients).

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Predicting COVID-19 and using Chest X-ray and CT images from chest radiography images (GitHub+kaggle)</th>
<th>16,756</th>
<th>Imaging features</th>
<th>Deep CNN</th>
<th>Sensitivity .80</th>
<th>Internal</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>Predicting COVID-19 and using chest X-ray images from chest radiography images (GitHub+kaggle)</td>
<td>1531</td>
<td>Imaging features</td>
<td>Deep learning + SGD algorithm + Inference algorithm with Threshold (0.15-0.50)</td>
<td>Sensitivity .97, Specificity .95</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>44</td>
<td>Predicting COVID-19 and using publicly available data (GitHub+kaggle+SIIM)</td>
<td>2876 (only 190 Covid-19)</td>
<td>Imaging features</td>
<td>AlexNet ResNet18 DenseNet201 SqueezeNet</td>
<td>Sensitivity .97, Specificity .99, AUC .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>45</td>
<td>Predicting COVID-19 and using Chest X-ray images from few publicly available datasets on Kagg le and GitHub</td>
<td>2855</td>
<td>Imaging features</td>
<td>Deep Learning Algorithms: VGG19, Inception, Xception, MobileNet, Inception, ResNet_V2</td>
<td>Sensitivity .93, Specificity .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>46</td>
<td>Predicting COVID-19 and using Chest X-ray images from two publicly available datasets on Kagg le and GitHub</td>
<td>50</td>
<td>Imaging features</td>
<td>Deep Learning Algorithms: VGG19, InceptionV3, Xception, DenseNet201, MobileNet_V2, Inception, ResNet_V2, ResNet_V2</td>
<td>Sensitivity 1, AUC .90</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>47</td>
<td>Predicting COVID-19 and using Chest X-ray images from publicly available datasets on Kaggle and GitHub</td>
<td>100</td>
<td>Imaging features</td>
<td>Deep Learning (InceptionV3) Deep Learning (ResNet50) Deep Learning (Inception, Resnet_V2)</td>
<td>Sensitivity .96, Specificity 1</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>48</td>
<td>Predicting COVID-19 and using Chest CT images from one medical center in Italy</td>
<td>150 CT belong to 53 cases</td>
<td>Imaging features</td>
<td>SVM</td>
<td>Sensitivity .97, Specificity .99</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>49</td>
<td>Predicting COVID-19 and using Chest X-ray images from publicly available datasets on GitHub</td>
<td>196</td>
<td>Imaging features</td>
<td>Modified ResNet18</td>
<td>Sensitivity .98, Specificity .92, AUC .94</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>50</td>
<td>Predicting COVID-19 and using Chest CT images from three medical centers in China</td>
<td>2605 (1658 covid19)</td>
<td>Imaging features</td>
<td>Modified VB-Net</td>
<td>Sensitivity .91, Specificity .83, AUC .94</td>
<td>Internal</td>
<td>High</td>
</tr>
<tr>
<td>51</td>
<td>Predicting COVID-19 and using Chest CT images from designated covid19 Zhejiang Province, China</td>
<td>219 CT covid19, 224 CT Influenza-A, pneumonia, and 175 CT healthy</td>
<td>Imaging features</td>
<td>ResNet, ResNet with location-attention mechanism</td>
<td>Sensitivity .76, F1 Score .76</td>
<td>Internal</td>
<td>High</td>
</tr>
</tbody>
</table>
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


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]
