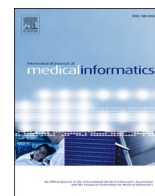




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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



Review article

# The accuracy of machine learning approaches using non-image data for the prediction of COVID-19: A meta-analysis

Kuang-Ming Kuo<sup>a</sup>, Paul C. Talley<sup>b</sup>, Chao-Sheng Chang<sup>c,\*</sup><sup>a</sup> Department of Business Management, National United University, Miaoli, Taiwan, ROC<sup>b</sup> Department of Applied English, I-Shou University, Kaohsiung City, Taiwan, ROC<sup>c</sup> Department of Emergency Medicine, E-Da Hospital, Kaohsiung City, Taiwan, ROC

## ARTICLE INFO

## Keywords:

COVID-19

Diagnostic test accuracy

Machine learning approach

Meta-analysis

## ABSTRACT

**Objective:** COVID-19 is a novel, severely contagious disease with enormous negative impact on humanity as well as the world economy. An expeditious, feasible tool for detecting COVID-19 remains yet elusive. Recently, there has been a surge of interest in applying machine learning techniques to predict COVID-19 using non-image data. We have therefore undertaken a meta-analysis to quantify the diagnostic performance of machine learning models facilitating the prediction of COVID-19.

**Materials and methods:** A comprehensive electronic database search for the period between January 1st, 2021 and December 3rd, 2021 was undertaken in order to identify eligible studies relevant to this meta-analysis. Summary sensitivity, specificity, and the area under receiver operating characteristic curves were used to assess potential diagnostic accuracy. Risk of bias was assessed by means of a revised Quality Assessment of Diagnostic Studies. **Results:** A total of 30 studies, including 34 models, met all of the inclusion criteria. Summary sensitivity, specificity, and area under receiver operating characteristic curves were 0.86, 0.86, and 0.91, respectively. The purpose of machine learning models, class imbalance, and feature selection are significant covariates useful in explaining the between-study heterogeneity, in terms of both sensitivity and specificity.

**Conclusions:** Our study findings show that non-image data can be used to predict COVID-19 with an acceptable performance. Further, class imbalance and feature selection are suggested to be incorporated whenever building models for the prediction of COVID-19, thus improving further diagnostic performance.

## 1. Introduction

Coronavirus Disease 19 (COVID-19), caused by severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) [1], has posed tremendous challenge due to the pandemic declared after March 2020 [2]. As of December 23, 2021, more than 27 million cases have been confirmed, including over 5 million deaths [3]. Currently, widely accepted management strategies for minimizing the spread of COVID-19 include forced lockdowns, travel restrictions, quarantines, social distancing, isolation, and infection-control measures [4]. As for those individuals who were infected, supportive care is the primary treatment available since specific effective and curative therapeutics remain elusive [4]. Commonly, COVID-19 adverse prognosis includes hospitalization, transfer to intensive care units, or even mortality [5,6]. Further, advanced COVID-19 is combined with heterogeneous clinical features [5]. A large number of those infected remain asymptomatic due to the

nature of COVID-19 symptomatology [7]. Efficient diagnosis of COVID-19 is difficult to achieve. The lack of optimal sensitivity and specificity in clinical detection methods has been shown to be a significant reason behind the rapid spread of COVID-19 [8]. The use of real-time, reverse transcription polymerase chain reaction (rRT-PCR) is presently the diagnostic gold standard used to confirm COVID-19 infection [1]. Materials required for this assay however are reportedly in short supply, leading to possible delays in diagnostic results throughout the pandemic period [9]. In view of such complex circumstances, a rapid and early diagnostic tool, or a ready system efficacious to identify the infected individuals, thus plays a vital role in managing the spreading COVID-19 pandemic.

Compared with traditional model-building approaches, machine learning techniques can model data without strict statistical assumptions [10], and there have been significant advances in modeling clinical data for predicting diagnosis or prognosis of differing diseases [11–13],

\* Corresponding author.

E-mail addresses: [kuangmingkuo@gmail.com](mailto:kuangmingkuo@gmail.com) (K.-M. Kuo), [atlanta.ga@msa.hinet.net](mailto:atlanta.ga@msa.hinet.net) (P.C. Talley), [zincfinger522@yahoo.com.tw](mailto:zincfinger522@yahoo.com.tw) (C.-S. Chang).<https://doi.org/10.1016/j.ijmedinf.2022.104791>

Received 23 December 2021; Received in revised form 8 April 2022; Accepted 9 May 2022

Available online 13 May 2022

1386-5056/© 2022 Elsevier B.V. All rights reserved.

COVID-19 being no exception. Two types of clinical data are usually employed to establish COVID-19 predictive models by means of machine learning techniques: structured data and/or image data. Structured data may include demographic information [14–21], medical histories [22–24], symptoms [14,25–28], vital signs [20,25,29–32], or laboratory tests [6,15–19,32]; and, image data may include lung computed tomography (CT) scan [33,34] and/or chest X-rays (CXR) [35,36]. Imagery data, especially CT scans, are found to be the most accurate diagnostic modality of SARS-CoV-2 according to several case reports, reviews and meta-analyses [8,37–42]. Compared with image data, structured clinical data can be easily obtained after patients' first encounters with frontline healthcare professionals, and thus can be incorporated into machine learning models to predict COVID-19 more efficiently.

Based on the notion that machine learning techniques have been emerging as a potential tool for healthcare professionals to accelerate their process of decision-making and to improve diagnostic accuracy. A meta-analysis to investigate the potential of using machine learning to identify COVID-19 is thus not only essential but also quite timely. Until recently, a systematic understanding of how machine learning techniques using non-imagery-based data can be used to predict COVID-19 is still lacking, not to mention a thorough meta-analysis of extant diagnostic accuracy. To fill this research gap, the objectives of this study are, as follows: 1) to meta-analyze the accuracy of the diagnosis of and prognosis of COVID-19 based on non-image data via machine learning techniques; and, 2) to compare and to contrast the diagnostic accuracy of plausible covariates that can account for the heterogeneity found among selected studies. Hence, two research questions (RQ) were proposed: 1) RQ1: What is the diagnostic accuracy of machine learning models, based on non-image data, for the diagnosis and prognosis of COVID-19 patients?; and, 2) RQ2: What covariates may contribute to the heterogeneity between these selected studies?

The remainder of the article will be organized accordingly. The "Material and methods" section describes the research method used in this study. The "Results" section presents the analytical results, the "Discussion" section discusses the significance of the findings, and the "Conclusions" section summarizes the findings of the current study.

## 2. Material and methods

This section describes the search strategy and selection process for the literature used to make up this study, as well as the method of extraction for required information taken from that literature. In this section, we also describe the tools for quality assessment and statistical techniques employed.

### 2.1. Search strategy and selection process

A comprehensive search of electronic databases was made. It included PubMed, ScienceDirect, and SpringerLink, and it was carried out between 1st January 2020 and 3rd December 2021 using the keyword combinations of COVID-19, machine learning, deep learning, and artificial intelligence. We did not search other databases such as Scopus or Web of Science for two principal considerations: 1) PubMed primarily focuses on medicine and biomedical sciences, which is more specific to this study, while Scopus and Web of Science cover multi-disciplinary fields [43]. 2) PubMed is free and easier to use than Scopus and Web of Science [43]. In addition, we used Google Scholar as a supplementary source to search articles. Despite Google Scholar not being suggested as stand-alone for purposes of a literature search [44], Google Scholar possesses sufficient stability in terms of article coverage to be used compared with either Scopus or Web of Science [45]. Detailed search queries for each database are shown in Table 1. Studies to be considered relevant were expected to meet the following criteria: 1) studies must have investigated the predictive accuracy of COVID-19; 2) studies should have leveraged structured data as features; 3) studies

**Table 1**

Search strategy for each database.

Database	Search strategy
PubMed	COVID-19[Title/Abstract] AND ((machine learning[Title/Abstract]) OR (deep learning[Title/Abstract]) OR (artificial intelligence[Title/Abstract]))
ScienceDirect	Title, abstract or author-specified keywords: COVID-19 AND ((machine learning) OR (deep learning) OR (artificial intelligence))
SpringerLink	"COVID-19" AND ("machine learning" OR "deep learning" OR "artificial intelligence")
Google Scholar	COVID-19 machine learning deep learning artificial intelligence

should have used artificial intelligence to predict the spread of the COVID-19 disease; 4) studies should have provided sufficient outcomes of predictive models; and, 5) studies taken from the literature must have been written in English and peer-reviewed. Studies meeting the following criteria were excluded: 1) studies using images (CT or CXR), image-associated reports, or unstructured data used as predictive features; 2) studies irrelevant to our research goal; and, 3) studies where full texts were unavailable for purposes of examination. Based on the stated inclusion and exclusion criteria, the data were first assessed by the first author (K.M.) and cross-checked by the third author (C.S.). Any discrepancies were resolved between both authors through a consensus discussion to ensure database accuracy/consistency. Finally, we located 30 studies including 34 models that predicted COVID-19 (see Fig. 1). Among the 30 studies, 17 studies including 20 models were extracted from PubMed, ScienceDirect, and SpringerLink, and 13 studies [6,14,17,19,21,26–29,46–49] including 14 models were identified by Google Scholar. Furthermore, three studies [21,30,32] included more than one model.

### 2.2. Data extraction

From each of the included studies, the following information was extracted: purposes of the predictive models, types of prognostic predictive models, types of features for establishing predictive models, geographic areas of the samples used to build predictive models, machine learning techniques adopted to build the predictive models, whether class imbalance issues were handled substantively, and whether extra feature selection strategies were adopted. We extracted or calculated the original *true/false* positives and *true/false* negatives from each study to derive summary outcome measures.

## 3. Methodological analysis

Diagnostic accuracy studies are often at the risk of being biased since they originate from differences in methodology, sample recruitment, or data collection [50]. We therefore assessed the quality of studies according to the revised Quality of Diagnostic Studies (QUADAS-2) guidelines, including four domains: sample selection, index test, reference standard, flow, and timing [51].

### 3.1. Statistical analysis

We meta-analyzed the diagnostic accuracy by using lme4 [52], mada [53], and meta [54] packages for R statistics. Sensitivity and specificity were pooled in accordance with a bivariate model [55]. Area under receiver operating characteristic curve (AUROC), diagnostic odds ratio (DOR), positive likelihood ratio (LR+), and negative likelihood ratio (LR-) were also estimated for purposes of this study. Forest plots were created to show heterogeneity among the models up for further consideration. Moreover, a summary receiver operating characteristic curve with 95% confidence interval (CI) and 95% prediction interval (PI) were employed to assess the existence of a threshold effect among the models.

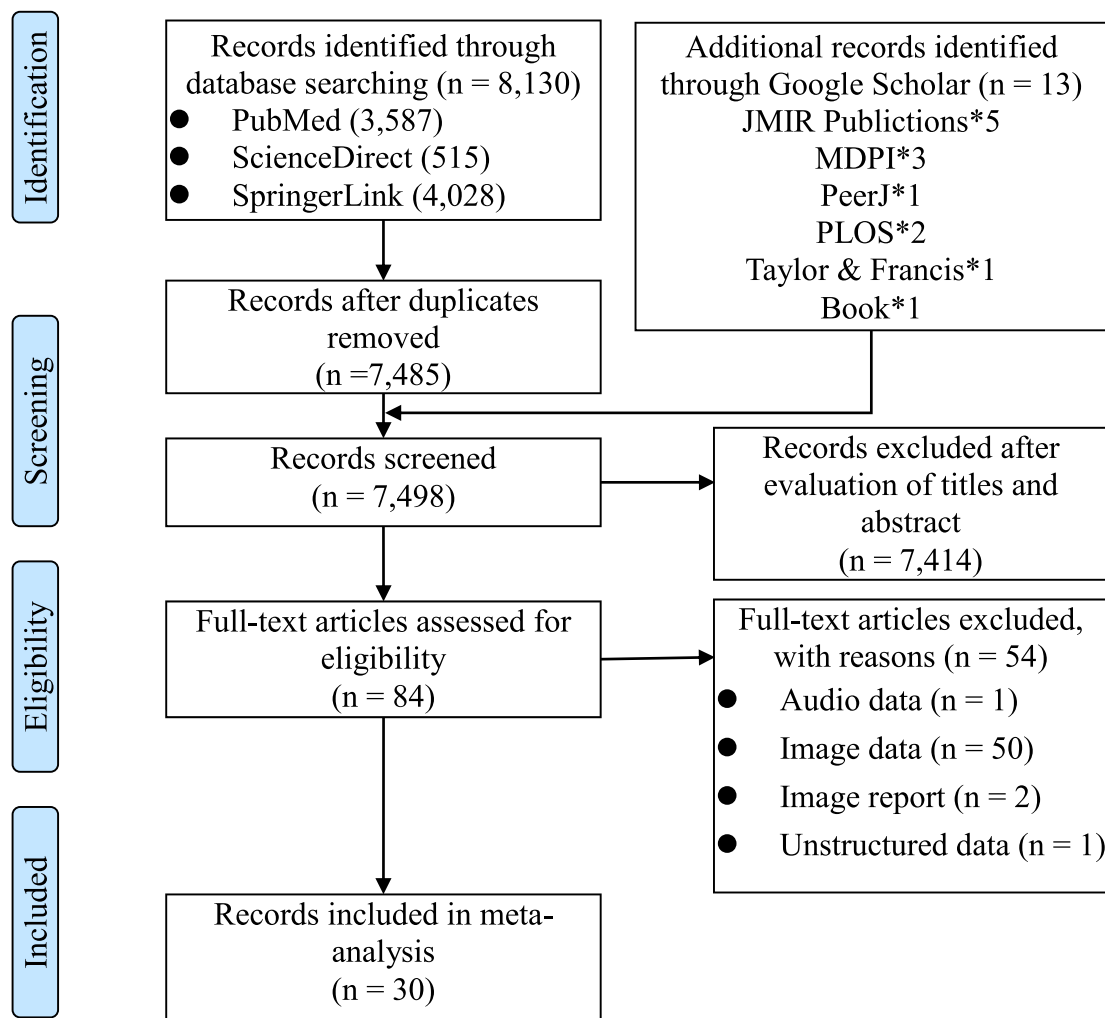


Fig. 1. PRISMA flow diagram.

According to prior suggestions about possible sources of heterogeneity between the selected studies [50], meta-regression was undertaken with three plausible types of covariates: 1) model purpose related covariates: including purpose of predictive models, and types of prognosis predicted; 2) sample related covariates: including feature type (e. g., demographic data, vital signs, laboratory data, medical history), and geographic areas of the patients; and, 3) machine learning related covariates: including types of artificial intelligence adopted, strategies for class imbalance, and feature selection strategies. Based on the due diligence performed, the Institutional Review Board of E-Da Hospital approved the study protocol (EMRP-109-158).

4. Results

In this section, we report the characteristics of included studies, as well as the results of the quality assessment made. Subsequently, we report the summary diagnostic accuracy of included studies and potential covariates used for explaining between-study heterogeneity.

4.1. General study characteristics

Among the 34 predictive models examined in this study, 14 models (41.18%) aim to diagnose the prevalent COVID-19 disease while 20 models (58.82%) aim to predict the prognosis of the COVID-19 patients (see Table 2). In the study parameters, numbers 7, 12, and 1 models serve to predict whether patient status ends-up in critical care,

Table 2 Characteristics of included models (n = 34).

Characteristics	Values	Frequency	%
Purpose	Diagnosis	14	41.18
	Prognosis	20	58.82
Prognosis (n = 20)	Critical care	7	35.00
	Mortality	12	60.00
	Hospitalization	1	5.00
Feature type	Laboratory data included	24	70.59
	None laboratory data	10	29.41
Geographic area (n = 33)	Eastern	8	24.24
	Western	25	75.76
AI techniques	Machine learning	32	94.12
	Deep learning	2	5.88
Class imbalance processed	Yes	12	35.29
	No	22	64.71
Feature selection	Yes	9	26.47
	No	25	73.53

Note: One study may be designed to predict more than one COVID-19 disease.

mortality, or hospitalization, respectively. Twenty-four models used only laboratory data or combined laboratory data with other clinical data (such as demographic information, symptoms, vital signs, or history) to predict the COVID-19 disease, while 10 models used only clinical data to predict COVID-19 disease without inclusion of any laboratory data. Most samples belong to a Western (75.76%) context, such as American or European. Thirty-two models (94.12%) were based

on machine learning techniques (e.g., random forest, support vector machine, or XGBoost), only 2 models used deep learning techniques [14,23]. About 35% of the models leveraged extra approaches (e.g., increasing the weight of the minority class or random under/over sampling) to deal with class imbalance issues. Finally, 9 models (26.47%) used additional strategies, such as recursive feature elimination or least absolute shrinkage, to select important features before training the predictive model [6,16,23,27,28,48,49,56–58].

4.2. Quality assessment

We assessed the quality of the selected studies based on QUADAS-2 [51] (see Fig. 2). For identifying bias, the 10 and 24 models were classified as having some concerns (29.41%) and low-risk of bias (70.59%) related to patient selecting domain, respectively. All 34 models were considered as low-risk of bias regarding index test and reference standard domains. Further, 13 and 21 models were regarded as generating some concerns (38.24%) and low-risk of bias (61.76%) about their flow and timing domains, respectively. As for the applicability judgment, 11 and 23 models were considered to be of some concern (32.35%) and low concern of applicability (67.65%), respectively. Finally, all 34 models were considered to be of low concern of applicability regarding the index test and reference standard domains.

4.3. RQ1: Diagnostic accuracy of non-image predictive models based on machine learning

The effect size pooled by traditional univariate meta-analysis can sometimes be misleading [59]. We therefore pooled the effect sizes based on the bivariate model [55]. As shown in Table 3, the overall pooled area under receiver operating characteristic curve for machine learning to predict the COVID-19 disease is about 0.91. Moreover, pooled sensitivity, specificity, diagnostic odds ratio, positive likelihood ratio, and negative likelihood ratio were 0.86, 0.86, 37.93, 6.20, and 0.16 respectively (see Table 3). Fig. 3 and Fig. 4 show the forest plot of sensitivity/specificity and the summary receiver operating characteristic curves with 95% confidence interval and prediction interval for the 34 predictive models, respectively. Two  $\chi^2$  tests were conducted to test for equality of sensitivity and of specificity, and these showed significant results,  $\chi^2(33) = 1090.94, p < 0.001$  and  $\chi^2(33) = 113615.20, p < 0.001$ , indicating significant between-study heterogeneity existed in terms of

Table 3

Performance of predicting COVID-19 disease by artificial intelligence.

Metrics	Performance (95% CI)
AUROC	0.91
Sensitivity	0.86 (0.81, 0.90)
Specificity	0.86 (0.79, 0.91)
DOR	37.93 (21.96, 53.90)
LR+	6.20 (3.78, 8.63)
LR-	0.16 (0.12, 0.21)

Note: AUROC: Area under receiver operating characteristic curve, DOR: Diagnostic odds ratio, LR+: Positive likelihood ratio, LR-: Negative likelihood ratio, CI: confidence interval.

both sensitivity and specificity.

4.4. RQ2: Plausible covariates explaining between-study heterogeneity

Due to the significant between-study heterogeneity for both sensitivity and specificity, we also conducted sub-group analysis by means of meta-regression to further identify potential covariates that might influence the performance of the COVID-19 disease predictive models. As shown in Table 4 and Fig. 5(a), the sensitivity was significantly ( $p = 0.002$ ) higher for the 14 models designed to diagnose COVID-19 (0.92; 95% CI, 0.88–0.95) than for the other 20 models for predicting the prognosis of COVID-19 (0.79; 95% CI, 0.71–0.86). The corresponding specificity of the 14 models for COVID-19 diagnosis (0.80; 95% CI, 0.67–0.89) was albeit lower than those of the 20 models for COVID-19 prognosis (0.89; 95% CI, 0.82–0.94), but didn't reach statistical significance ( $p = 0.144$ ). If we go deeper into the models for COVID-19 prognosis, the sensitivity of the 7 models for predicting critical care (i. e., patients getting transferred to intensive care units or using ventilation apparatus) (0.73; 95% CI, 0.49–0.88) due to COVID-19 was lower than for the 12 models for predicting mortality (0.81; 95% CI, 0.73–0.87) due to COVID-19, but didn't reach statistical significance ( $p = 0.255$ ), as shown in Table 4 and Fig. 5 (b). The corresponding specificity was proximate between the 7 models for critical care and the 12 models for mortality (0.88 vs. 0.90,  $p = 0.689$ ). There is only one model for predicting hospitalization due to COVID-19, as such we did not include that model into the sub-group analysis.

The 24 models used laboratory data with/without other data (e.g.,

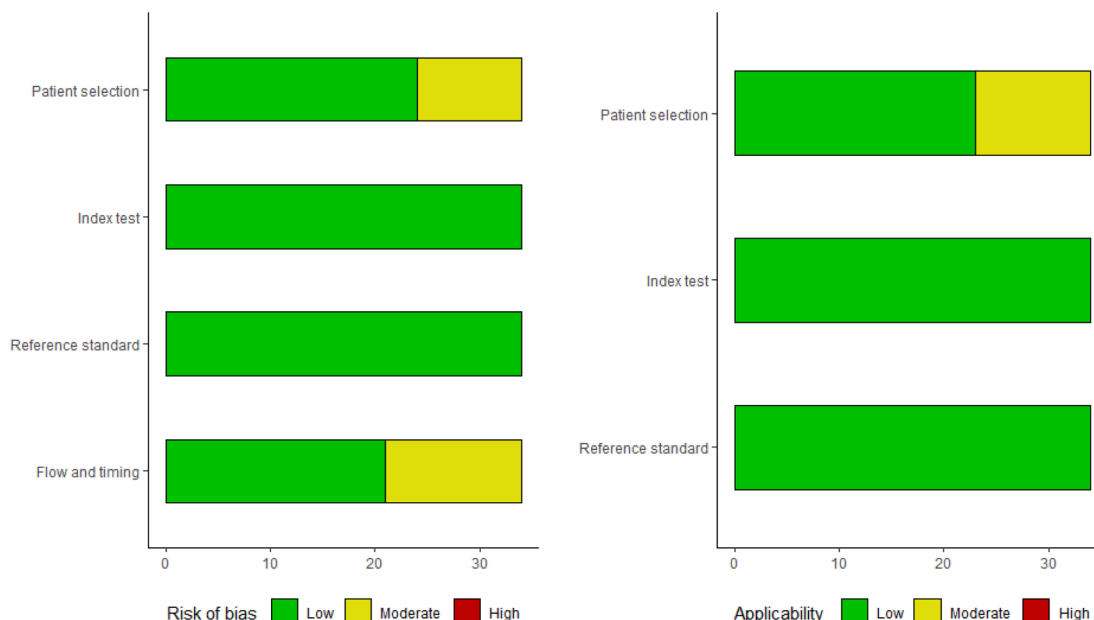


Fig. 2. Methodological assessment by QUADAS-2.



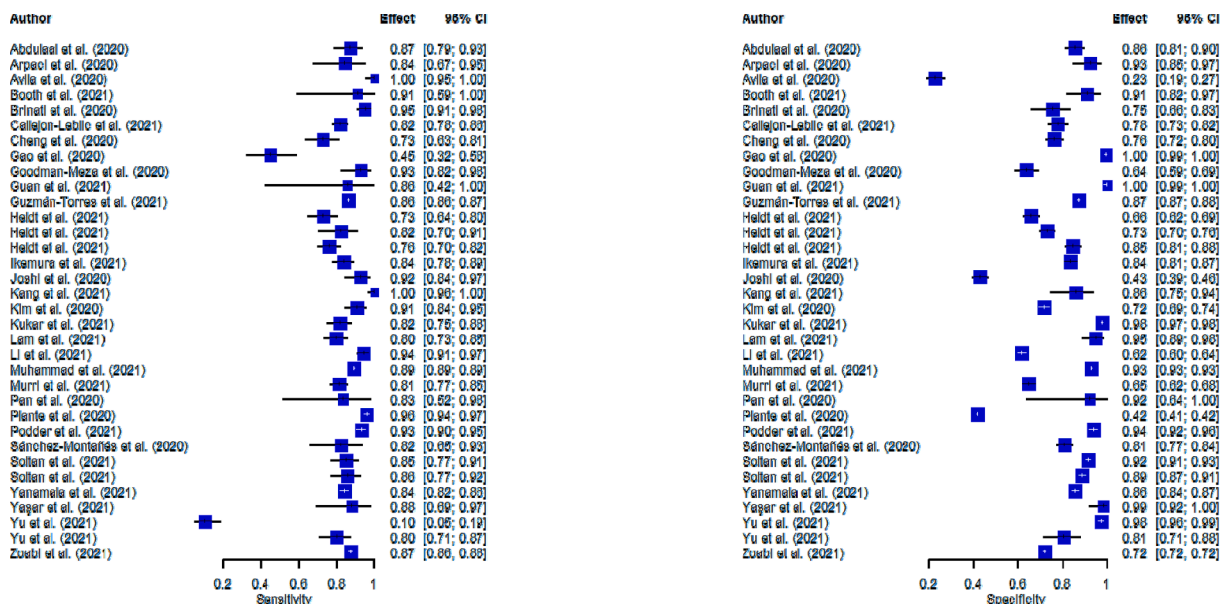


Fig. 3. Forest plot of sensitivity and specificity in this study.

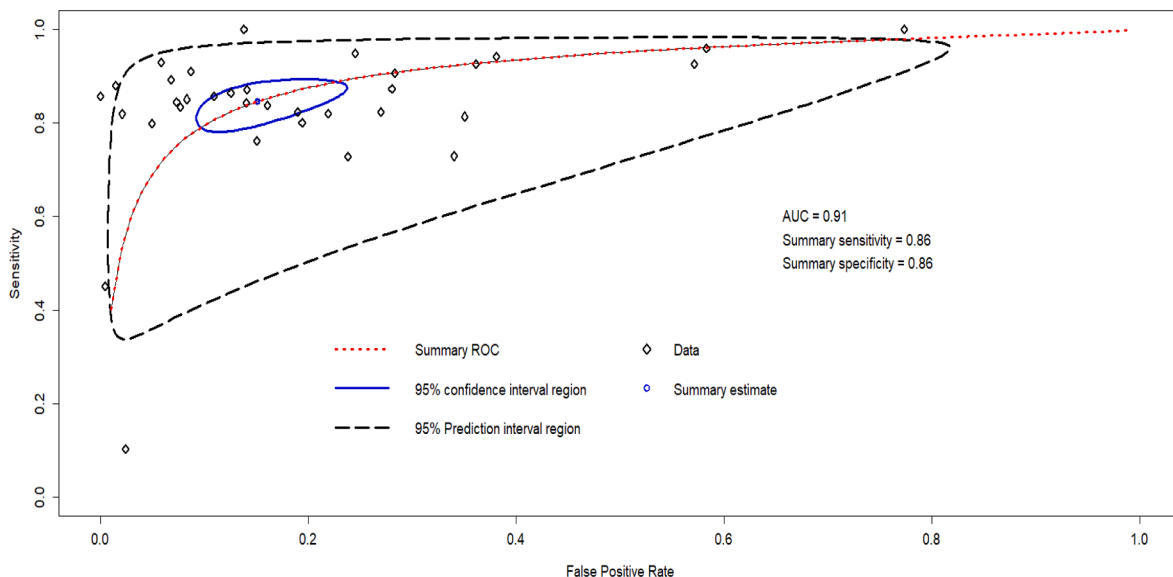


Fig. 4. Summary receiver operating characteristic curves for collected studies.

demographic information, symptoms, vital signs, or history) as features achieved a higher sensitivity (0.88; 95% CI, 0.83–0.92 vs. 0.80; 95% CI, 0.66–0.90), but didn't reach statistical significance ( $p = 0.154$ ), and a near-tie specificity (0.86; 95% CI, 0.76–0.93 vs. 0.87; 95% CI, 0.80–0.92) than models incorporating data other than laboratory test results (e.g., demographic information, symptoms, vital signs, or history) as depicted in Table 4 and Fig. 5 (c).

The models that used patients from western contexts, as demonstrated in Table 3 and Fig. 5 (d), achieved a lower sensitivity (0.86; 95% CI, 0.79–0.90 vs. 0.88; 95% CI, 0.74–0.95) and specificity (0.83; 95% CI, 0.75–0.88 vs. 0.93; 95% CI, 0.76–0.98) than models using patients from the eastern contexts, but neither reached statistical significance ( $p = 0.650$  and  $p = 0.107$ ). The 2 models that adopted deep learning techniques had a higher sensitivity (0.99; 95% CI, 0.32–1.00 vs. 0.85; 95% CI, 0.79–0.89) and a tied specificity (0.86; 95% CI, 0.81–0.90 vs. 0.86; 95% CI, 0.79–0.91) than the remaining 32 models that adopted machine learning techniques, but both did not reach statistical significance ( $p =$

0.090 and  $p = 0.780$ ), as illustrated in Table 4 and Fig. 5 (e).

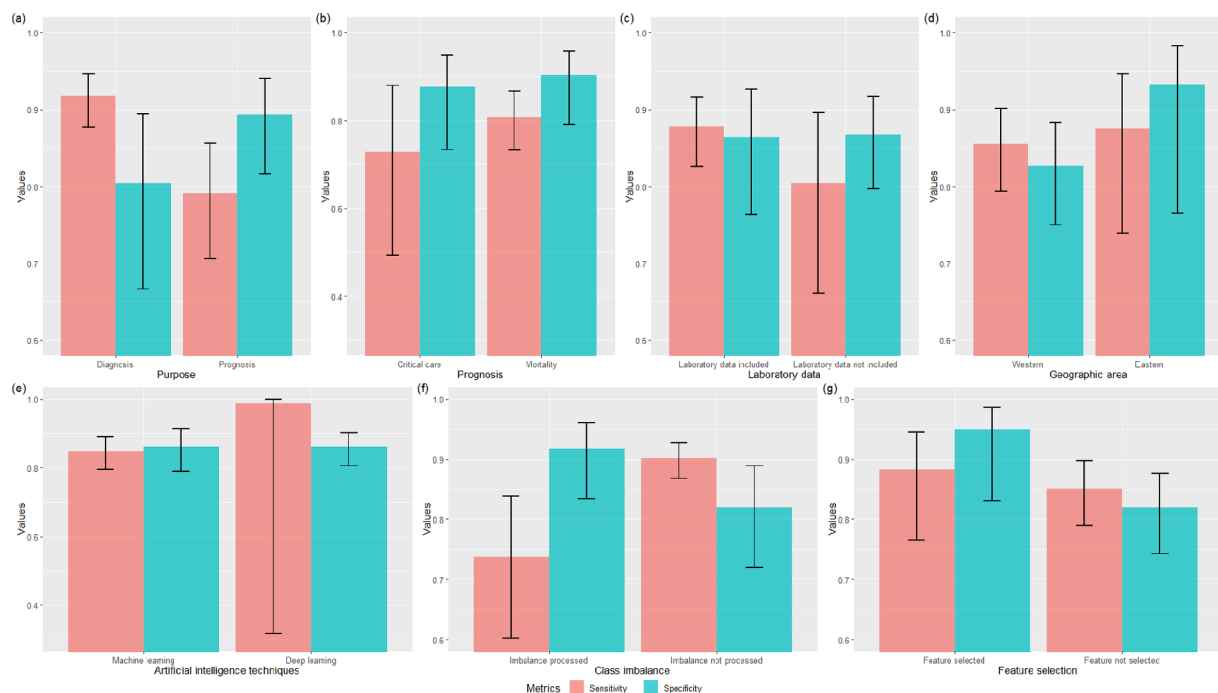
The 12 models that adopted extra strategies, as depicted in Table 4 and Fig. 5 (f), to deal with class imbalance had a lower sensitivity ( $p = 0.001$ ) than models without extra strategies for handling class imbalance (0.74; 95% CI, 0.60–0.84 vs. 0.90; 95% CI, 0.87–0.93). The specificity of the models that adopted extra strategies to deal with class imbalance was however higher than the remaining models without extra strategies for handling class imbalance (0.92; 95% CI, 0.83–0.96 vs. 0.82; 95% CI, 0.72–0.89), but no statistical difference was established ( $p = 0.076$ ).

Finally, the 9 models that leveraged feature selection strategies before building predictive models had a higher sensitivity than models without employing feature selection strategies (0.88; 95% CI, 0.77–0.95 vs. 0.85; 95% CI, 0.79–0.90), but no statistical significance was confirmed ( $p = 0.520$ ). However, the models that leveraged feature selection strategies showed a significant higher specificity ( $p = 0.022$ ) compared to models without feature selection strategies (0.95; 95% CI, 0.83–0.99 vs. 0.82; 95% CI, 0.74–0.88), as shown in Table 4 and Fig. 5

**Table 4**  
Summary estimates for sensitivity and specificity with covariates, Note: CI denotes confidence interval.

Type of covariates	Covariates	Values	Metrics	Summary estimates	95% CI	P value
Model purpose related	Overall (n = 34)		Sensitivity	0.86	[0.81, 0.90]	< 0.001
			Specificity	0.86	[0.79, 0.91]	< 0.001
	Purpose	Diagnosis (n = 14)	Sensitivity	0.92	[0.88, 0.95]	0.002
			Specificity	0.80	[0.67, 0.89]	0.144
	Prognosis	Prognosis (n = 20)	Sensitivity	0.79	[0.71, 0.86]	[Reference]
			Specificity	0.89	[0.82, 0.94]	[Reference]
Sample related	Prognosis	Critical care (n = 7)	Sensitivity	0.73	[0.49, 0.88]	0.255
			Specificity	0.88	[0.73, 0.95]	0.689
	Mortality (n = 12)	Sensitivity	0.81	[0.73, 0.87]	[Reference]	
		Specificity	0.90	[0.79, 0.96]	[Reference]	
	Data type	Lab data included (n = 24)	Sensitivity	0.88	[0.83, 0.92]	0.154
			Specificity	0.86	[0.76, 0.93]	0.754
Lab data not included (n = 10)	Sensitivity	0.80	[0.66, 0.90]	[Reference]		
	Specificity	0.87	[0.80, 0.92]	[Reference]		
Geographic area	Western (n = 25)	Sensitivity	0.86	[0.79, 0.90]	0.650	
		Specificity	0.83	[0.75, 0.88]	0.107	
	Eastern (n = 8)	Sensitivity	0.88	[0.74, 0.95]	[Reference]	
		Specificity	0.93	[0.76, 0.98]	[Reference]	
Machine learning related	AI techniques	Machine learning (n = 32)	Sensitivity	0.85	[0.79, 0.89]	0.090
			Specificity	0.86	[0.79, 0.91]	0.780
		Deep learning (n = 2)	Sensitivity	0.99	[0.32, 1.00]	[Reference]
			Specificity	0.86	[0.81, 0.90]	[Reference]
	Class imbalance processed	Yes (n = 12)	Sensitivity	0.74	[0.60, 0.84]	0.001
			Specificity	0.92	[0.83, 0.96]	0.076
		No (n = 22)	Sensitivity	0.90	[0.87, 0.93]	[Reference]
			Specificity	0.82	[0.72, 0.89]	[Reference]
	Feature selection	Yes (n = 9)	Sensitivity	0.88	[0.77, 0.95]	0.520
			Specificity	0.95	[0.83, 0.99]	0.022
		No (n = 25)	Sensitivity	0.85	[0.79, 0.90]	[Reference]
			Specificity	0.82	[0.74, 0.88]	[Reference]

Note: CI denotes confidence interval.



**Fig. 5.** Pooled sensitivity and specificity with 95% confidence interval for different covariates.

(g).

## 5. Discussion

The global COVID-19 pandemic is a growing public health concern requiring unprecedented efforts in nearly every field of endeavor. Effective coping strategies for this disease are however still under

development or of nascent consideration. Machine learning has the potential to play a key role in this fight against the COVID-19 pandemic. However, there has been a lack of meta-analysis studies that are focused on the diagnostic accuracy of COVID-19 casework based on non-image data. Based on such an understanding, our study investigated the performance of machine learning approaches based on non-image data for predicting COVID-19 incidence, undertaking a bivariate meta-analysis.

The results demonstrate strong diagnostic performance with a pooled sensitivity of 0.86, a pooled specificity of 0.86, and an AUC of 0.91, respectively. Prior meta-analysis [42] shown the pooled sensitivity and specificity of artificial intelligence for CT scan was 0.90 and 0.91, respectively which is higher than those of artificial intelligence based on non-image data. Nonetheless, non-image data are often far more obtainable than image data among those hospitals with limited material resources.

The purpose of predictive models, type of prognosis, feature type, geographic area, type of AI techniques, whether class imbalance issues were dealt with, and where extra feature selection strategies were implemented, were further included in bivariate meta-regression. This was done to account for potential instances of heterogeneity among the primary studies made. The findings demonstrated that sensitivity was significantly dependent on the purpose of predictive models and upon whether class imbalance issues were handled. It may be noted that specificity was significantly dependent on whether the extra-strategies technique was used to select features before training predictive models.

In terms of predictive models purposefulness, the sensitivity of diagnostic models is significantly higher than that of models for prognosis (0.92 vs. 0.79) while specificity of models for diagnosis is lower than models for prognosis (0.80 vs. 0.90). There was however no significant difference between the specificity of models for diagnosis and prognosis. It may be reasoned that models for diagnosis had a higher sensitivity may be because infection by COVID-19 can be confirmed by a comparison of the rRT-PCR testing results [60]. The prognosis of the COVID-19 disease however is more complicated since it relates to various factors such as age, gender, obesity, comorbidities, or time of anti-viral treatment [61,62]. For example, there is prior evidence [62] that COVID-19 patients experiencing diabetes and hypertension or some other comorbidities such as cardiovascular disease, chronic obstructive pulmonary disease, and cancer are more likely to have adverse outcomes. Further, the time span between onset of severe outcomes and anti-viral treatment application for COVID-19 are major factors influencing the prognosis [62].

We further analyzed the 19 models for predicting two types of prognoses: critical care and mortality. The pooled sensitivity and specificity is 0.73 and 0.88 when the models are used to predict critical care regimens for COVID-19 patients. These two figures are lower than those of models used for the prediction of patient risk-of-mortality, but no statistically significant difference was confirmed. The plausible reason that mortality predictive models reached a higher sensitivity and specificity than did critical care models may be due in large part to the fact that the situations of patients close to mortality are simpler to explain than those of critical care patients. This situation includes various patient situations, such as being transferred to intensive care units or using ventilation apparatus, that are readily apparent. Especially during the apex of the epidemic, the number of patients exceeded the service capacity of most primary healthcare facilities; as such, the criteria for defining critically ill patients would thus be different from the more stable pandemic contagion periods. Furthermore, the variant of the SARS-CoV-2 virus continues to mutate [63] making it difficult to estimate its exact impact on current patient-loads. Available evidence [64] suggests that the incidence of SARS-CoV-2 should be closely monitored since patients from different locations have already shown different mutated COVID-19 sequences.

In addition to rRT-PCR test, laboratory data (e.g., transaminases, lymphocytes, eosinophils, calcium, and separate aminotransferase) with/without other demographic and clinical data (e.g., symptoms, vital signs, or medical histories) predicted COVID-19 in 24 models, while the remaining models used only demographic/clinical data as features for similarly predicting COVID-19. Our meta-analysis showed that models that included laboratory data performed better than with models without laboratory data. This included heightened sensitivity (0.88 vs. 0.80). Models without any laboratory data included slightly outperformed models with laboratory data included in terms of specificity

(0.87 vs. 0.86). Both sensitivity and specificity however did not reach any real statistical significance. Previous studies [65–67] found that laboratory data can provide useful information for COVID-19 diagnostics. For example, prior evidence [67] has found that the platelet count can dynamically reflect patho-physiological changes prevalent in COVID-19 patients. Other evidence [68] however found that some laboratory results involved with COVID-19 patients are different between pregnant women, children, and other members of a general population. Including and testing a wider variety of laboratory data may be required to achieve a more stable predictive platform when dealing with COVID-19 as a whole. Further, before the widely acknowledged rRT-PCR test [69] becomes universal, other data such as laboratory tests that have shown potential should be identified in order to effectively widen the prediction of COVID-19 incidence. It would be more helpful if machine learning models can incorporate routinely available laboratory tests to correctly predict COVID-19, which would streamline the diagnosis and treatment of COVID-19 patients, saving considerable time and decision-making.

Since the first case of COVID-19 was reported in Wuhan, China and then the rest of the world, knowledge about COVID-19 has altered and expanded to a certain extent. Hence, the question remains of whether it is possible that the performance of predictive models based on eastern countries may be different from western-based predictive models through a variety of circumstances (i.e., transparency of research procedures, availability of data, reliability of findings, geo-political considerations). We therefore conducted a sub-group analysis based on samples from different geographic areas to make such a determination. The sub-group analysis showed that models using samples from the western contexts had a lower sensitivity (0.86 vs. 0.88) and specificity (0.83 vs. 0.93) than models using samples taken from eastern contexts. The plausible reason for this result is complex; so, we suspect it may be due to the algorithms adopted by these models. In the eastern category, eight models adopted only two major types of algorithms, including ensemble learning and deep learning. With appropriate configurations, these two types of machine learning models are generally considered to have a better predictive performance when compared with other algorithms [70,71]. On the other hand, the western group, consisting of 25 models, applied seven different types of algorithms in the mix. Such a variety of different algorithms may thus contribute to a higher variation status in the models' predictive performance, which may explain why the pooled sensitivity and collective specificity of the western group appeared lower than that of the eastern group.

The pooled sensitivity and specificity are (0.85, 0.86) and (0.99, 0.86) respectively when machine learning and deep learning techniques were in use. Deep learning outperformed machine-learning in terms of sensitivity, but it tied with machine-learning in specificity. There was however no significant difference between these two techniques shown. Despite the sensitivity of deep learning being quite high (0.99) in our study, its 95% confidence interval however is also very wide (0.32–1), indicating the sample sizes were too small, which is just the case in our study ( $n = 2$  for deep-learning). Still more deep learning studies are required to verify if its true performance in predicting the COVID-19 is based on non-image data.

In regards to classification tasks, the receiver operator characteristic (ROC) plot and the AUROC delineate how an adaptable threshold causes changes in two types of errors: false-positives and false-negatives [72]. However, the ROC curve and AUROC are only partially informative whenever used with imbalanced data [72]. The explainability, traceability, and interpretability of performance measures will have greater future importance in dealing with imbalanced data. Hence, problems relevant to class imbalance are often dealt with by use of various strategies such as with a synthetic minority over-sampling technique [73]. Our study demonstrates that the pooled sensitivity for models without extra-strategies for class imbalance is significantly higher than that of models with extra-strategies for class imbalance (0.90 vs. 0.74). This may be compared to the pooled specificity for models with extra-



strategies being higher than that of models without extra-strategies for class imbalance (0.92 vs. 0.82), but it did not show a statistically significant difference. Ramezankhani, Pournik, Shahrabi, Azizi, Hadaegh and Khalili [74] adopted an over-sampling strategy to deal with class imbalance problems for predicting type 2 diabetes. It was found that the original training dataset had a higher sensitivity, and a lower specificity, than a balanced-training dataset, therefore indicating such a strategy does not guarantee for better performance. Our study however showed a lower sensitivity and a higher specificity which may be due to the re-sampling strategies propounded by Ramezankhani, Pournik, Shahrabi, Azizi, Hadaegh and Khalili [74] and applied only in the training dataset. The performance data that our study collected was mainly adapted from the test dataset, and that may indicate class-imbalance handling strategies cannot necessarily guarantee the overall performance of test dataset.

In order to enhance the performance of predictive models, pre-processing such as a feature-selection aspect can be adopted before training machine learning models [10]. Our study showed that 9 models with feature-selection had a higher sensitivity (0.88 vs. 0.85) and specificity (0.95 vs. 0.82) than did 25 models that were without feature-selection, but only specificity reached any statistical significance. Based on the findings of our meta-analysis, the importance of feature-selection should not be overlooked during preliminary model-building processes.

The findings based on our study identified some current gaps in the state-of-the-art and in future research challenges. First, despite more studies based on machine learning leveraging non-image-specific data for the prediction of COVID-19, the number of studies remain less than those studies utilizing image data. More studies are thus required to better investigate the potential of non-image data for predicting COVID-19. Second, the paucity of studies using deep learning techniques for non-image data is another plausible issue since deep learning is considered better-suited for use with image data than other machine learning techniques. Future studies can be used to leverage various deep learning techniques for predicting COVID-19 based on non-image data. Third, the diagnostic accuracy achieved by machine learning based on non-image data is still lower than that of diagnostic accuracy based on image data. Future studies may combine image and non-image data to establish a sufficient model that can achieve better diagnostic performance.

Our findings may have important implications for the medical practice as well. First, hospitals that are short of material and staff resources can adopt these machine learning models based on non-image data routinely available to assist those who are identifying possible COVID-19 patients. By doing so, the contact risk of COVID-19 infection due to a lack of rRT-PCR or CT testing measures may be diminished. Second, developers of machine learning models can consider adopting strategies for feature-selection and class-imbalance features during model-building planning and formulation. By doing so, a predictive model with better performance, to support informed decision-making by healthcare professionals, may be established. Further, models based on machine learning techniques may be applied to predict other epidemics and/or diseases in future times. To achieve this purpose, specific features can be carefully selected based on the specific pandemic/disease, and then different types of machine learning algorithms can be compared. In this way, the best performed algorithm can be determined based on their demonstrated learning capabilities.

## 6. Conclusions

Our study aims to meta-analyze the accuracy of diagnostic tests of artificial intelligence techniques to confront the COVID-19 pandemic. By searching multiple electronic databases, 30 studies including 34 predictive models were included in this meta-analysis. A bivariate meta-analysis of diagnostic test accuracy was conducted to estimate sensitivity, specificity, and summary receiver operating characteristic curve. Strong diagnostic performances were obtained with the models used in

this study. These findings may indicate that machine learning models that use non-image data can be implemented in hospital settings, especially in diminished-resource locations, in order to effectively predict the incidence or prevalence of COVID-19. These models show the potential of becoming more accurate and further representative as data sets increase in terms of their size. Furthermore, covariates including diagnosis purpose, whether class-imbalance issues are processed, and whether extra-feature selection strategies being adopted were found to partially explain some of the heterogeneity found among the primary studies evaluated.

## 7. Summary points

What was already known on the topic?

- COVID-19 has had a serious impact on human lives and upon economic livelihoods; however, a quick and feasible tool for detecting COVID-19 incidence remains elusive.
- Real-time reverse transcription polymerase chain reaction is currently the “gold standard” for diagnosing COVID-19, but it requires a longer turn-around time in terms of efficacy.
- Pulmonary computed-tomography scan and chest radiography can be used to complement the practical diagnosis COVID-19.

What this study added to our knowledge?

- Strong diagnostic test accuracy of COVID-19 can be achieved by using non-image data.
- Non-image data, taken as predictive features, can assist hospitals with limited financial and human resources to identify cases of COVID-19.
- Class-imbalance and feature-selection strategies may be considered before building predictive models useful for diagnosing COVID-19.

## References

- [1] A. Tahamtan, A. Ardebili, Real-time RT-PCR in COVID-19 detection: issues affecting the results, *Expert Rev. Mol. Diagnost.* 20 (2020) 453–454.
- [2] World Health Organization Regional Office for Europe, WHO announces COVID-19 outbreak a pandemic, World Health Organization Regional Office for Europe, UN City, Denmark, 2020.
- [3] World Health Organization, WHO Coronavirus (COVID-19) Dashboard, World Health Organization, Geneva, Switzerland, 2021.
- [4] L.R. Baden, E.J. Rubin, Covid-19 — the search for effective therapy, *N. Engl. J. Med.* 382 (2020) 1851–1852.
- [5] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *The Lancet*, 395 (2020) 497–506.
- [6] T.B. Plante, A.M. Blau, A.N. Berg, A.S. Weinberg, I.C. Jun, V.F. Tapsos, T. S. Kanigan, A.B. Adib, Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: a large, Multicenter, Real-World Study, *J. f Med. Int. Res.* 22 (2020), e24048.
- [7] M. Day, Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village, *BMJ* 368 (2020), m1165.
- [8] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases, *Radiology*, 296 (2020) E32–E40.
- [9] M.L. Ranney, V. Griffeth, A.K. Jha, Critical supply shortages — the need for ventilators and personal protective equipment during the Covid-19 pandemic, *N. Engl. J. Med.* 382 (2020), e41.
- [10] M. Kuhn, K. Johnson, *Applied Predictive Modeling*, Springer, New York, 2013.
- [11] B. Wernly, B. Mamandipoor, P. Baldia, C. Jung, V. Osmani, Machine learning predicts mortality in septic patients using only routinely available ABG variables: a multi-centre evaluation, *Int. J. Med. Inf.* 145 (2021), 104312.
- [12] A. Udreia, G.D. Mitra, D. Costea, E.C. Noels, M. Wakkee, D.M. Siegel, T.M. de Carvalho, T.E.C. Nijsten, Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms, *J. Eur. Acad. Dermatol. Venereol.* 34 (2020) 648–655.
- [13] D. Pei, Y. Gong, H. Kang, C. Zhang, Q. Guo, Accurate and rapid screening model for potential diabetes mellitus, *BMC Med. Inf. Decis. Making* 19 (2019) 41.

- [14] A. Abdulaal, A. Patel, E. Charani, S. Denny, N. Mughal, L. Moore, Prognostic modeling of COVID-19 using artificial intelligence in the United Kingdom: model development and validation, *J. Med. Internet Res* 22 (2020), e20259.
- [15] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, F. Cabitza, Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study, *J. Med. Syst.* 44 (2020) 135.
- [16] Y. Gao, G.Y. Cai, W. Fang, H.Y. Li, S.Y. Wang, L. Chen, Y. Yu, D. Liu, S. Xu, P.F. Cui, S.Q. Zeng, X.X. Feng, R.D. Yu, Y. Wang, Y. Yuan, X.F. Jiao, J.H. Chi, J.H. Liu, R. Y. Li, X. Zheng, C.Y. Song, N. Jin, W.J. Gong, X.Y. Liu, L. Huang, X. Tian, L. Li, H. Xing, D. Ma, C.R. Li, F. Ye, Q.L. Gao, Machine learning based early warning system enables accurate mortality risk prediction for COVID-19, *Nat. Commun.* 11 (2020) 5033.
- [17] D. Goodman-Meza, A. Rudas, J.N. Chiang, P.C. Adamson, J. Ebinger, N. Sun, P. Botting, J.A. Fulcher, F.G. Saab, R. Brook, A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity, *PLoS ONE* 15 (2020), e0239474.
- [18] R.P. Joshi, V. Pejavar, N.E. Hammarlund, H. Sung, S.K. Lee, A.o. Furmanchuk, H.-Y. Lee, G. Scott, S. Gombar, N. Shah, S. Shen, A. Nassiri, D. Schneider, F.S. Ahmad, D. Liebovitz, A. Kho, S. Mooney, B.A. Pinsky, N. Banaei, A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results, *Journal of Clinical Virology*, 129 (2020) 104502.
- [19] M. Sánchez-Montañés, P. Rodríguez-Belenguier, A.J. Serrano-López, E. Soria-Olivas, Y. Alakhdar-Mohmara, Machine learning for mortality analysis in patients with COVID-19, *Int. J. Environ. Res. Public Health* 17 (2020) 8386.
- [20] N. Yanamala, N.H. Krishna, Q.A. Hathaway, A. Radhakrishnan, S. Sunkara, H. Patel, P. Farjo, B. Patel, P.P. Sengupta, A vital sign-based prediction algorithm for differentiating COVID-19 versus seasonal influenza in hospitalized patients, *NPJ Digital Med.* 4 (2021) 95.
- [21] L. Yu, A. Halalau, B. Dalal, A.E. Abbas, F. Ivascu, M. Amin, G.B. Nair, Machine learning methods to predict mechanical ventilation and mortality in patients with COVID-19, *PLoS ONE* 16 (2021), e0249285.
- [22] S. Li, Y. Lin, T. Zhu, M. Fan, S. Xu, W. Qiu, C. Chen, L. Li, Y. Wang, J. Yan, J. Wong, L. Naing, S. Xu, Development and external evaluation of predictions models for mortality of COVID-19 patients using machine learning method, *Neural Comput. Appl.* (2021).
- [23] J. Kang, T. Chen, H. Luo, Y. Luo, G. Du, M. Jiming-Yang, Machine learning predictive model for severe COVID-19, *Infect. Genet. Evol.* 90 (2021), 104737.
- [24] L.J. Muhammad, E.A. Algehyne, S.S. Usman, A. Ahmad, C. Chakraborty, I. A. Mohammed, Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset, *SN Comput. Sci.* 2 (2020) 11.
- [25] Y. Zoabi, S. Deri-Rozov, N. Shomron, Machine learning-based prediction of COVID-19 diagnosis based on symptoms, *NPJ Digital Med.* 4 (2021) 3.
- [26] M.A. Callejon-Leblic, R. Moreno-Luna, A. Del Cuvillo, I.M. Reyes-Tejero, M. A. Garcia-Villaran, M. Santos-Peña, J.M. Maza-Solano, D.I. Martin-Jimenez, J. M. Palacios-Garcia, C. Fernandez-Velez, J. Gonzalez-Garcia, J.M. Sanchez-Calvo, J. Solanellas-Soler, S. Sanchez-Gomez, Loss of smell and taste can accurately predict COVID-19 infection: a machine-learning approach, *J. Clin. Med.* 10 (2021) 570.
- [27] K. Ikemura, E. Bellin, Y. Yagi, H. Billett, M. Saada, K. Simone, L. Stahl, J. Szymanski, D.Y. Goldstein, M. Reyes Gil, Using automated machine learning to predict the mortality of patients with COVID-19: prediction model development study, *J. Med. Internet Res.* 23 (2021), e23458.
- [28] X. Guan, B. Zhang, M. Fu, M. Li, X. Yuan, Y. Zhu, J. Peng, H. Guo, Y. Lu, Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study, *Ann. Med.* 53 (2021) 257–266.
- [29] F.Y. Cheng, H. Joshi, P. Tandon, R. Freeman, D.L. Reich, M. Mazumdar, R. Kohli-Seth, M.A. Levin, P. Timsina, A. Kia, Using machine learning to predict ICU transfer in hospitalized COVID-19 patients, *J. Clin. Med.* 9 (2020) 1668.
- [30] F.S. Heldt, M.P. Vizcaychipi, S. Peacock, M. Cinelli, L. McLachlan, F. Andreotti, S. Jovanović, R. Dürichen, N. Lipunova, R.A. Fletcher, A. Hancock, A. McCarthy, R. A. Pointon, A. Brown, J. Eaton, R. Liddi, D. Mackillop, L. Tarassenko, R.T. Khan, Early risk assessment for COVID-19 patients from emergency department data using machine learning, *Sci. Rep.* 11 (2021) 4200.
- [31] C. Lam, J. Calvert, A. Siefkas, G. Barnes, E. Pellegrini, A. Green-Saxena, J. Hoffman, Q. Mao, R. Das, Personalized stratification of hospitalization risk amidst COVID-19: a machine learning approach, *Health Pol. Technol.* 10 (2021), 100554.
- [32] A.A.S. Soltan, S. Kouchaki, T. Zhu, D. Kiyasseh, T. Taylor, Z.B. Hussain, T. Peto, A. J. Brent, D.W. Eyre, D.A. Clifton, Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test, *The Lancet Digital Health* 3 (2021) e78–e87.
- [33] S. Sharma, Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients, *Environ. Sci. Pollut. Res.* 27 (2020) 37155–37163.
- [34] N. Lessmann, C.I. Sánchez, L. Beenen, L.H. Boulogne, M. Brink, E. Calli, J.-P. Charbonnier, T. Dofferhoff, W.M. van Everdingen, P.K. Gerke, B. Geurts, H. A. Gietema, M. Groeneveld, L. van Harten, N. Hendrix, W. Hendrix, H.J. Huisman, I. Išgum, C. Jacobs, R. Kluge, M. Kok, J. Krdzalic, B. Lassen-Schmidt, K. van Leeuwen, J. Meakin, M. Overkamp, T. van Rees Vellinga, E.M. van Rikxoort, R. Samperna, C. Schaefer-Prokop, S. Schalekamp, E.T. Scholten, C. Sital, J. L. Stöger, J. Teuwen, K.V. Venkadesh, C. de Vente, M. Vermaat, W. Xie, B. de Wilde, M. Prokop, B. van Ginneken, Automated assessment of COVID-19 reporting and data system and chest CT severity scores in patients suspected of having COVID-19 using artificial intelligence, *Radiology* 298 (2020) E18–E28.
- [35] K. Murphy, H. Smits, A.J.G. Knoop, M.B.J.M. Korst, T. Samson, E.T. Scholten, S. Schalekamp, C.M. Schaefer-Prokop, R.H.H.M. Philipsen, A. Meijers, J. Melendez, B. van Ginneken, M. Rutten, COVID-19 on chest radiographs: a multireader evaluation of an artificial intelligence system, *Radiology* 296 (2020) E166–E172.
- [36] S. Vaid, R. Kalantar, M. Bhandari, Deep learning COVID-19 detection bias: accuracy through artificial intelligence, *Int. Orthop.* 44 (2020) 1539–1542.
- [37] N. Benameur, R. Mahmoudi, S. Zaid, Y. Arous, B. Hmdia, M.H. Bedoui, SARS-CoV-2 diagnosis using medical imaging techniques and artificial intelligence: a review, *Clin. Imaging* 76 (2021) 6–14.
- [38] C. Jalaber, T. Lapotre, T. Morcet-Delattre, F. Ribet, S. Jouneau, M. Lederlin, Chest CT in COVID-19 pneumonia: a review of current knowledge, *Diagn. Intervent. Imag* 101 (2020) 431–437.
- [39] H. Mohammad-Rahimi, M. Nadimi, A. Ghalyanchi-Langeroudi, M. Taheri, S. Ghafouri-Fard, Application of machine learning in diagnosis of COVID-19 through X-Ray and CT images: a scoping review, *Front. Cardio. Med.* 8 (2021) 185.
- [40] I. Ozsahin, B. Sekeroglu, M.S. Musa, M.T. Mustapha, D. Uzun Ozsahin, Review on diagnosis of COVID-19 from chest CT images using artificial intelligence, *Comput. Math. Methods Med.* (2020) 9756518.
- [41] B. Xu, Y. Xing, J. Peng, Z. Zheng, W. Tang, Y. Sun, C. Xu, F. Peng, Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy, *Eur. Radiol.* 30 (2020) 5720–5727.
- [42] M. Moezzi, K. Shirbandi, H.K. Shahvandi, B. Arjmand, F. Rahim, The diagnostic accuracy of artificial intelligence-assisted CT imaging in COVID-19 disease: a systematic review and meta-analysis, *Inf. Med. Unlocked* 24 (2021), 100591.
- [43] M.E. Falagas, E.I. Pitsouni, G.A. Malietzis, G. Pappas, Comparison of PubMed, Scopus, web of science, and google scholar: strengths and weaknesses, *FASEB J.* 22 (2008) 338–342.
- [44] D. Giustini, M.N.K. Boulos, Google Scholar is not enough to be used alone for systematic reviews, *Online J Public Health Inform* 5 (2013) 214.
- [45] A.-W. Harzing, S. Alakangas, Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison, *Scientometrics* 106 (2016) 787–804.
- [46] E. Avila, A. Kahmann, C. Alho, M. Dorn, Hemogram data as a tool for decision-making in COVID-19 management: applications to resource scarcity scenarios, *PeerJ* 8 (2020), e9482.
- [47] H.J. Kim, D. Han, J.H. Kim, D. Kim, B. Ha, W. Seog, Y.K. Lee, D. Lim, S.O. Hong, M. J. Park, J. Heo, An easy-to-use machine learning model to predict the prognosis of patients with COVID-19: retrospective cohort study, *J Med Internet Res* 22 (2020), e24225.
- [48] P. Pan, Y. Li, Y. Xiao, B. Han, L. Su, M. Su, Y. Li, S. Zhang, D. Jiang, X. Chen, F. Zhou, L. Ma, P. Bao, L. Xie, Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: model development and validation, *J Med Internet Res* 22 (2020), e23128.
- [49] P. Podder, S. Bharati, M.R.H. Mondal, U. Kose, in: Application of machine learning for the diagnosis of COVID-19, *Data Science for COVID-19*, Academic Press, 2021, pp. 175–194.
- [50] J.F. Cohen, D.A. Korevaar, D.G. Altman, D.E. Bruns, C.A. Gatsonis, L. Hooft, L. Irwig, D. Levine, J.B. Reitsma, H.C.W. de Vet, P.M.M. Bossuyt, STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration, *BMJ Open* 6 (2016), e012799.
- [51] P.F. Whiting, A.W.S. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M. G.G. Leeflang, J.A.C. Sterne, P.M.M. Bossuyt, Q.-G. the, QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies, *Annals of Internal Medicine*, 155 (2011) 529–536.
- [52] D. Bates, M. Mächler, B. Bolker, S. Walker, fitting linear mixed-effects models using lme4, *J Stat Softw* 67 (2015) 1–48.
- [53] P. Doebler, P.-C. Bürkner, G. Rücker, Statistical Packages for Diagnostic Meta-Analysis and Their Application, in: G. Biondi-Zoccai (Ed.), *Diagnostic Meta-Analysis*, Springer, Cham, Switzerland, 2018, pp. 161–182.
- [54] G. Schwarzer, Package meta, the comprehensive R Archive, Network (2019).
- [55] J.B. Reitsma, A.S. Glas, A.W.S. Rutjes, R.J.P.M. Scholten, P.M. Bossuyt, A. H. Zwinderman, Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, *J. Clin. Epidemiol.* 58 (2005) 982–990.
- [56] J.A. Guzmán-Torres, E.M. Alonso-Guzmán, F.J. Domínguez-Mota, G. Tinoco-Guerrero, Estimation of the main conditions in (SARS-CoV-2) Covid-19 patients that increase the risk of death using Machine learning, the case of Mexico, *Results Phys.* 27 (2021), 104483.
- [57] V. Schöning, E. Liakoni, C. Baumgartner, A.K. Exadaktylos, W.E. Hautz, A. Atkinson, F. Hammann, Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital, *J. Translat. Med.* 19 (2021) 56.
- [58] Ş. Yaşar, C. Çolak, S. Yoloğlu, Artificial intelligence-based prediction of Covid-19 severity on the results of protein profiling, *Comput. Methods Programs Biomed.* 202 (2021), 105996.
- [59] C. Gatsonis, P. Paliwal, Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer, *Am. J. Roentgenol.* 187 (2006) 271–281.
- [60] S. Bolourani, M. Brenner, P. Wang, T. McGinn, J.S. Hirsch, D. Barnaby, T.P. Zanos, A Machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: model development and validation, *J. Med. Int. Res.* 23 (2021), e24246.
- [61] X. Fang, S. Li, H. Yu, P. Wang, Y. Zhang, Z. Chen, Y. Li, L. Cheng, W. Li, H. Jia, X. Ma, Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis, *Aging (Albany NY)* 12 (2020) 12493–12503.
- [62] M. Li, Y. Dong, H. Wang, W. Guo, H. Zhou, Z. Zhang, C. Tian, K. Du, R. Zhu, L. Wang, L. Zhao, H. Fan, S. Luo, D. Hu, Cardiovascular disease potentially

- contributes to the progression and poor prognosis of COVID-19, *Nutrit. Metabol. Cardiovascular Diseases* 30 (2020) 1061–1067.
- [63] C. del Rio, P.N. Malani, S.B. Omer, Confronting the Delta Variant of SARS-CoV-2, Summer 2021, *JAMA* 326 (2021) 1001–1002.
- [64] S. Akhter, S. Akhtar, Emerging coronavirus diseases and future perspectives, *VirusDisease* 31 (2020) 113–120.
- [65] Y. Gao, T. Li, M. Han, X. Li, D. Wu, Y. Xu, Y. Zhu, Y. Liu, X. Wang, L. Wang, Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19, *J. Med. Virol.* 92 (2020) 791–796.
- [66] H. Yun, Z. Sun, J. Wu, A. Tang, M. Hu, Z. Xiang, Laboratory data analysis of novel coronavirus (COVID-19) screening in 2510 patients, *Clin. Chim. Acta* 507 (2020) 94–97.
- [67] X. Zhao, K. Wang, P. Zuo, Y. Liu, M. Zhang, S. Xie, H. Zhang, X. Chen, C. Liu, Early decrease in blood platelet count is associated with poor prognosis in COVID-19 patients—indications for predictive, preventive, and personalized medical approach, *EPMA J.* 11 (2020) 139–145.
- [68] S. Vakili, A. Savardashtaki, S. Jamalnia, R. Tabrizi, M.H. Nematollahi, M. Jafarinia, H. Akbari, Laboratory findings of COVID-19 infection are conflicting in different age groups and pregnant women: a literature review, *Arch. Med. Res.* 51 (2020) 603–607.
- [69] B. Böger, M.M. Fachi, R.O. Vilhena, A.F. Cobre, F.S. Tonin, R. Pontarolo, Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19, *Am. J. Infect. Control* 49 (2021) 21–29.
- [70] J. Brownlee, *Deep learning with Python: Develop Deep Learning Models On Theano And TensorFlow Using Keras*, 2016.
- [71] J. Brownlee, *Ensemble Learning Algorithms with Python*, 2020.
- [72] A.M. Carrington, P.W. Fieguth, H. Qazi, A. Holzinger, H.H. Chen, F. Mayr, D. G. Manuel, A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms, *BMC Med. Inf. Decis. Making* 20 (2020) 4.
- [73] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.
- [74] A. Ramezankhani, O. Pournik, J. Shahrabadi, F. Azizi, F. Hadaegh, D. Khalili, The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of Type 2 diabetes, *Med. Decis. Making* 36 (2014) 137–144.