



Review

# Artificial Intelligence Applications for COVID-19 in Intensive Care and Emergency Settings: A Systematic Review

Marcel Lucas Chee <sup>1</sup>, Marcus Eng Hock Ong <sup>2,3</sup>, Fahad Javaid Siddiqui <sup>2</sup>, Zhongheng Zhang <sup>4</sup>,  
Shir Lynn Lim <sup>5</sup>, Andrew Fu Wah Ho <sup>2,3</sup> and Nan Liu <sup>2,6,7,\*</sup>

<sup>1</sup> Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria 3800, Australia; tche0014@student.monash.edu

<sup>2</sup> Duke-NUS Medical School, National University of Singapore, Singapore 169857, Singapore; marcus.ong.e.h@singhealth.com.sg (M.E.H.O.); fahad.siddiqui@duke-nus.edu.sg (F.J.S.); andrew.ho@mohh.com.sg (A.F.W.H.)

<sup>3</sup> Department of Emergency Medicine, Singapore General Hospital, Singapore 169608, Singapore

<sup>4</sup> Department of Emergency Medicine, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China; zh\_zhang1984@zju.edu.cn

<sup>5</sup> Department of Cardiology, National University Heart Centre, Singapore 119074, Singapore; shir\_lynn\_lim@nuhs.edu.sg

<sup>6</sup> Health Service Research Centre, Singapore Health Services, Singapore 169856, Singapore

<sup>7</sup> Institute of Data Science, National University of Singapore, Singapore 117602, Singapore

\* Correspondence: liu.nan@duke-nus.edu.sg



**Citation:** Chee, M.L.; Ong, M.E.H.; Siddiqui, F.J.; Zhang, Z.; Lim, S.L.; Ho, A.F.W.; Liu, N. Artificial Intelligence Applications for COVID-19 in Intensive Care and Emergency Settings: A Systematic Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4749. <https://doi.org/10.3390/ijerph18094749>

Academic Editor: Xudong Huang

Received: 25 March 2021

Accepted: 21 April 2021

Published: 29 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** *Background:* Little is known about the role of artificial intelligence (AI) as a decisive technology in the clinical management of COVID-19 patients. We aimed to systematically review and critically appraise the current evidence on AI applications for COVID-19 in intensive care and emergency settings. *Methods:* We systematically searched PubMed, Embase, Scopus, CINAHL, IEEE Xplore, and ACM Digital Library databases from inception to 1 October 2020, without language restrictions. We included peer-reviewed original studies that applied AI for COVID-19 patients, healthcare workers, or health systems in intensive care, emergency, or prehospital settings. We assessed predictive modelling studies and critically appraised the methodology and key findings of all other studies. *Results:* Of fourteen eligible studies, eleven developed prognostic or diagnostic AI predictive models, all of which were assessed to be at high risk of bias. Common pitfalls included inadequate sample sizes, poor handling of missing data, failure to account for censored participants, and weak validation of models. *Conclusions:* Current AI applications for COVID-19 are not ready for deployment in acute care settings, given their limited scope and poor quality. Our findings underscore the need for improvements to facilitate safe and effective clinical adoption of AI applications, for and beyond the COVID-19 pandemic.

**Keywords:** artificial intelligence; machine learning; COVID-19; emergency department; intensive care; critical care

## 1. Introduction

The ongoing coronavirus disease 2019 (COVID-19) pandemic has challenged healthcare systems and healthcare practitioners worldwide. Intensive care units (ICU) and emergency departments (ED) in badly afflicted areas have been overwhelmed by the surge in patients suspected or diagnosed with COVID-19 [1–3]. This exerts significant pressure on healthcare resources, necessitating novel diagnostic and care pathways to rationally deploy scarce emergency and intensive care healthcare resources. Current strategies and recommendations on clinical management and resource rationalisation draw on past pandemic experiences and expert recommendations [3–5]; however, there has been growing interest in novel applications of artificial intelligence (AI) to assist in the COVID-19 response within these settings.

AI is a branch of computer science that uses computational methods to mimic human intelligence. AI is becoming increasingly ubiquitous; from self-driving cars to Siri, drone-assisted farming to diagnosis, AI applications have become indispensable in many industries, including the healthcare sector. Exponential advances in computer processing speeds, increased access to big data, and electronic medical record usage have all been crucial factors in driving the uptake of AI applications within the medical field. In medical research, AI often refers specifically to machine learning, which is a subset of AI that focuses on automatic improvement of computer programmes through experience [6,7]. For example, simple forms of machine learning such as logistic, linear, or Cox regression have commonly been used to investigate associations between predictors and disease outcomes. More advanced machine learning models, including random forest models, neural networks, or support vector machines, are also becoming more common in the medical literature. Such models can assist with more complex tasks, for example, image analysis and diagnosis in radiology [8], dermatology [9], and ophthalmology [10], amongst others. In this review, AI is not limited to machine learning models but includes broader forms of AI such as natural language processing, decision trees, and computational decision assistance software. In pre-COVID-19 intensive care and emergency settings, AI applications have assisted with automated patient monitoring [11–14], prognostication [15], and optimisation of staffing allocations [16–19]. Given the unprecedented volume of COVID-19 patients, recent reviews have also identified resource optimisation of ICU beds as a potentially significant application of AI [20,21].

Earlier systematic reviews have identified significant issues in the quality and reporting of predictive models for COVID-19 diagnosis and prognosis [22] and AI applications for classifying COVID-19 medical images [23]. Shillian, et al. [24] in a systematic review of machine learning studies in pre-COVID-19 ICUs reported similar issues, such as limited sample size and poor validation of predictions. However, no study has evaluated the scope and quality of all available AI applications for COVID-19 in intensive care and emergency settings. We hypothesise that issues in quality and reporting will exist across all AI applications in these settings due to the urgency and limited time afforded to developing these applications during pandemic. However, we also anticipate that diverse, pre-COVID-19 applications of AI may have been repurposed for use in the COVID-19 pandemic, including diagnosis, prognostication, monitoring, and resource optimisation. An exploration of the quality and breadth of current AI applications will provide valuable insight for improving the development and deployment of AI applications in intensive care and emergency settings, for and beyond the COVID-19 pandemic. We aim to systematically review and critically appraise the current evidence on AI applications for COVID-19 in intensive care and emergency settings, focussing on methods, reporting standards, and clinical utility.

## 2. Materials and Methods

We reported this systematic review according to the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines (Supplementary File S1). A review protocol was developed but was not publicly registered.

### 2.1. Search Strategy and Selection Criteria

We searched six databases: (PubMed, Embase, Scopus, CINAHL, IEEE Xplore, and ACM Digital Library) by combining search terms related to AI, COVID-19, and intensive care or emergency settings. PubMed, Embase, Scopus, and CINAHL were chosen for their broad coverage across biomedical, nursing, allied health, and general scientific literature, while IEEE Xplore and ACM Digital Library were included for coverage of more technical AI literature. For brevity, the truncated search strategy showing only the first three terms in each concept set is as follows: (“Artificial intelligence” OR “Deep learning” OR “Machine learning” OR . . . ) AND (“COVID-19” OR “Coronavirus disease 2019” OR “2019-nCoV” OR . . . ) AND (Emergency OR “ED” OR “intensive care” OR . . . ). The complete search

strategy can be found in Supplementary File S2. We also screened the reference lists of included articles to identify additional relevant studies.

We included articles that met the following criteria: (1) applied AI; (2) investigated COVID-19 operations of ICU, ED, or emergency medical services (EMS) or analysed data from COVID-19 patients in the ED or within a prehospital setting, COVID-19 patients requiring intensive care (admission to the ICU, mechanical ventilation, or a composite including either of these outcomes), or the healthcare workers treating these patients, including ED or ICU physicians and nurses as well as paramedics; and (3) were original, peer-reviewed research articles. For this review, only conventional machine learning algorithms such as random forest models, neural networks, or support vector machines were considered AI; multivariable logistic regression predictive models (including ridge and least absolute shrinkage and selection operator (LASSO) regression) were excluded. No restrictions were placed on the language of articles; all non-English articles that appeared in our search were translated and assessed for suitability, however, we did not search the databases using equivalent non-English search terms.

## 2.2. Literature Selection and Data Extraction

We conducted an initial search on 30 August 2020 and updated the results on 1 October 2020. Articles were screened by title, abstract, and, if ambiguous, full text by two independent reviewers (MLC and NL). Subsequently, the two reviewers (MLC and NL) independently extracted data using a standardised data extraction form. Discrepancies in article selection and data extraction were resolved between reviewers through discussion.

We extracted the following data for all included articles: country of study population, outcome predicted, sample size of the training and validation datasets, AI algorithms used, discrimination (e.g., C-index, accuracy) and calibration (e.g., calibration slope, Brier loss score) of models on the strictest form of validation, features included in the final model, and transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) study type [25], if applicable.

## 2.3. Data Analysis

For studies including multivariate AI predictive models, we evaluated the risk of bias within the study methodology using prediction model risk of bias assessment tool (PROBAST) [26]. PROBAST is a structured tool comprising 20 signalling questions for assessing the risk of bias and applicability across the four domains of participants, predictors, outcome, and analysis. Applicability of included studies was not assessed as our study was not concerned with a specific application of AI predictive models. In lieu of specific reporting standards for AI studies at the time of study conception [27], we assessed the reporting quality of multivariable predictive modelling studies using an adaptation of Wang, et al. [28] modified TRIPOD statement [29] for AI models (Supplementary File S3). The TRIPOD statement is a reporting guideline for studies that develop, validate, or update a prediction model, often for diagnostic or prognostic purposes. It contains 22 checklist items for assessing the reporting quality of predictive modelling studies. For studies that could not be assessed using the above guidelines (for example, non-predictive modelling studies), we summarised the study methodology, including data sources, application of AI, and validation methods, as well as the key findings of the study.

## 3. Results

### 3.1. Study Characteristics

Our search of the six databases returned 764 studies, of which 14 were included and analysed in this review (Figure 1). Table 1 presents the main characteristics of the study. Eleven of the 14 studies investigated predictive models and were assessed according to PROBAST and TRIPOD: eight studies developed prognostic models [30–37] and three studies developed diagnostic models [38–40]. Of the remaining three studies, two evaluated the prognostic potential of existing AI-based lung segmentation software (without integration

into a multivariate predictive model) [41,42] and one investigated an AI-based system for resource optimisation in the ICU [43]. Eleven studies used patient data collected from the ICU and four studies used data from the ED. No study collected data from the prehospital setting, despite including prehospital-related search terms in the search strategy.

In terms of country of study, Italy ( $n = 3$ ) and United States ( $n = 3$ ) were represented by more than one study, while Brazil, Canada, China, France, Germany, Israel, Turkey, and the United Kingdom had one study each.

According to the TRIPOD classification of predictive models, two studies were classified as Type 2b (validation using a non-random split of data by time and/or location), three studies as Type 2a (validation using a random split of data such as a train-test split), four studies as Type 1b (validation using re-sampling techniques such as bootstrapping or k-fold cross-validation), and one study as Type 1a (no validation, only evaluation of apparent model performance on the same training dataset). One study that conducted development and validation using data from separate studies was considered Type 3.

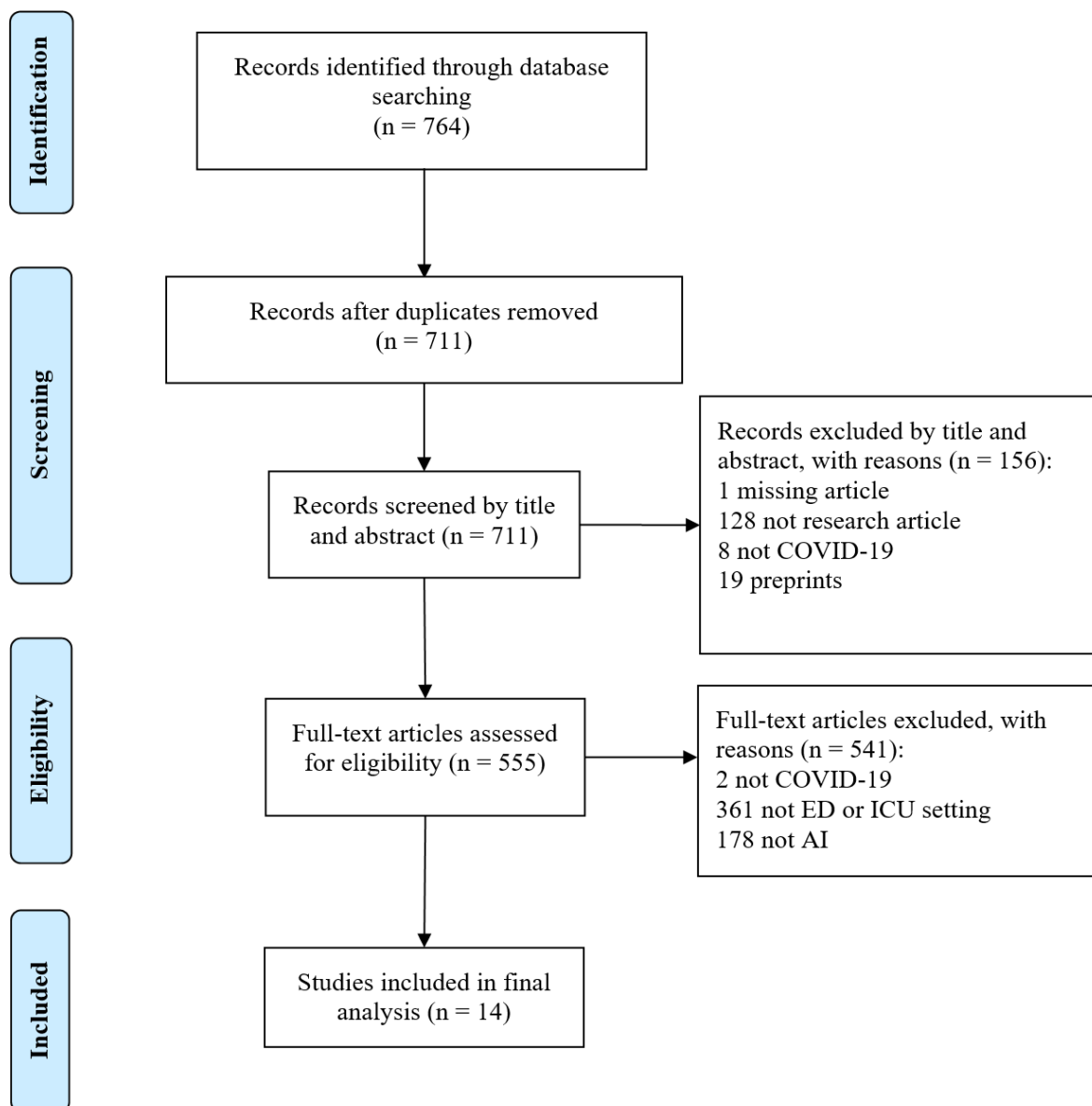


Figure 1. PRISMA flow diagram.

**Table 1.** Main study characteristics.

| Author [Reference]    | Study Type    | Country of Study Population | Relevant Setting of Collected Data (ED, ICU, or Prehospital) | Outcome Predicted   | Sample Size of Training Dataset | Sample Size of Test Dataset | Model Performance  | TRIPOD Classification |
|-----------------------|---------------|-----------------------------|--|---|---------------------------------|-----------------------------|--|-----------------------|
| Diagnostic            |               |                             |  |   |                                 |                             |  |                       |
| Brinati, et al. [38]  | Retrospective | Italy                       | ED   | Positive COVID-19 status  | 279                             | N/A (cross-validation)      | Random forest (C-index = 0.84)                           | 1b                    |
| Fraser, et al. [39]   | Prospective   | Canada                      | ICU  | Positive COVID-19 status  | 20                              | N/A (cross-validation)      | Decision tree (accuracy = 98%)                           | 1b                    |
| Vasse, et al. [40]    | Retrospective | France                      | ED   | Positive COVID-19 status  | 744                             | 2390                        | Decision tree (Sensitivity = 60.5%, Specificity = 89.7%) | 2b                    |
| Prognostic            |               |                             |  |   |                                 |                             |  |                       |
| Abdulaal, et al. [30] | Retrospective | United Kingdom              | ED   | In-patient mortality  | 318                             | 80                          | Neural network (C-index = 0.901)                         | 2a                    |
| Assaf, et al. [31]    | Retrospective | Israel                      | ED; ICU  | Critical illness (mechanical ventilation, ICU admission, multi-organ failure, and/or death) | 162                             | N/A (cross-validation)      | Random forest (C-index = 0.93)                           | 1b                    |
| Burdick, et al. [32]  | Prospective   | United States               | ICU  | Decompensation leading to mechanical ventilation within 24 h                                | 49,623                          | 197                         | Gradient boosting machine (C-index = 0.866)              | 3                     |
| Burian, et al. [33]   | Prospective   | Germany                     | ICU  | ICU admission   | 65                              | N/A (cross-validation)      | Random forest (C-index = 0.79)                           | 1b                    |
| Cheng, et al. [34]    | Retrospective | United States               | ICU  | ICU admission within 24 h   | 401                             | 521                         | Random forest (C-index = 0.799)                          | 2a                    |
| Durhan, et al. [41]   | Retrospective | Turkey                      | ICU  | ICU admission (software evaluates the extent of normal lung parenchyma)                     | 90                              | N/A                         | Deep learning software (C-index = 0.944)                 | N/A                   |
| Jackson, et al. [35]  | Retrospective | United States               | ICU  | Invasive mechanical ventilation   | 297                             | N/A                         | Fast-and-frugal decision tree (accuracy = 70%)           | 1a                    |
| Liang, et al. [36]    | Retrospective | China                       | ICU  | Critical illness (ICU admission, invasive ventilation, death)                               | 1590                            | 710                         | Deep learning survival Cox model (C-index = 0.852–0.967) | 2b                    |

Table 1. Cont.

| Author [Reference]    | Study Type    | Country of Study Population | Relevant Setting of Collected Data (ED, ICU, or Prehospital) | Outcome Predicted   | Sample Size of Training Dataset | Sample Size of Test Dataset | Model Performance  | TRIPOD Classification |
|-----------------------|---------------|-----------------------------|--|---|---------------------------------|-----------------------------|--|-----------------------|
| Prognostic            |               |                             |  |   |                                 |                             |  |                       |
| Mushtaq, et al. [42]  | Prospective   | Italy                       | ICU  | ICU admission (software evaluates the extent of lung opacity and consolidation) | 697                             | N/A                         | Deep learning software based on convolutional neural networks (C-index = 0.77) | N/A                   |
| Schwab, et al. [37]   | Retrospective | Brazil                      | ICU  | ICU admission   | 391                             | 167                         | Support vector machine (C-index = 0.98)  | 2a                    |
| Resource optimisation |               |                             |  |   |                                 |                             |  |                       |
| Belciug, et al. [43]  | Retrospective | Italy                       | ICU  | Developed a model for simulating ICU bed occupancy                              | N/A                             | N/A                         | Artificial immune system algorithm (no accuracy measure estimated)             | N/A                   |

COVID-19: coronavirus disease 2019, ED: Emergency Department, N/A: Not applicable, ICU: Intensive Care Unit; a: Performance of the best performing model is reported if multiple models were constructed. Only the performance on the strictest form of validation is reported. A range is given if the model was validated on multiple datasets. b: TRIPOD classification according to strictest validation used (higher values indicate stricter classification, i.e., type 3 is the strictest amongst included studies). 1a: Performance is evaluated directly on the same data; 1b: Performance and optimism of the model are evaluated using re-sampling techniques, such as bootstrapping or *k*-fold cross-validation; 2a: Model development and performance evaluation are done separately on a random split of the data, such as a train-test split; 2b: Model development and performance evaluation is done separately on a non-random split of the data by time, location, or both; 3: Model development and performance evaluation are conducted on separate data sets, for example, from different studies.

### 3.2. Risk of Bias

Table 2 presents the risk of bias assessment of AI predictive models according to PROBAST. All 11 predictive modelling studies had a high overall risk of bias. Two out of 11 studies had an unclear risk of bias within the participant domain. Unclear risk of bias in the participant domain was mainly due to ambiguous exclusion criteria that may lead to the study population not being representative of the intended target population [36,37].

**Table 2.** PROBAST (prediction model risk of bias assessment tool) assessment of predictive modelling studies.

| Author<br>[Reference] | Risk of Bias according to PROBAST Domain |            |          |          |         |
|-----------------------|--|------------|----------|----------|---------|
|                       | Participants                             | Predictors | Outcomes | Analysis | Overall |
| Diagnostic            |  |            |          |          |         |
| Brinati, et al. [38]  | Low                                      | Low        | Low      | High     | High    |
| Fraser, et al. [39]   | Low                                      | Unclear    | Low      | High     | High    |
| Vasse, et al. [40]    | Low                                      | Low        | Low      | High     | High    |
| Prognostic            |  |            |          |          |         |
| Abdulaal, et al. [30] | Low                                      | Low        | Low      | High     | High    |
| Assaf, et al. [31]    | Low                                      | Low        | Unclear  | High     | High    |
| Burdick, et al. [32]  | Low                                      | Low        | Unclear  | High     | High    |
| Burian, et al. [33]   | Low                                      | Low        | Low      | High     | High    |
| Cheng, et al. [34]    | Low                                      | High       | Unclear  | High     | High    |
| Jackson, et al. [35]  | Low                                      | High       | High     | High     | High    |
| Liang, et al. [36]    | Unclear                                  | High       | High     | High     | High    |
| Schwab, et al. [37]   | Unclear                                  | Unclear    | Unclear  | High     | High    |
| Vasse, et al. [40]    | Low                                      | Low        | Low      | High     | High    |

All three studies at a high risk of bias in the predictor domain were prognostic. Two studies [35,36] used retrospective, multicentre data and were at risk of bias from varying methods of predictor assessment at different centres. The remaining study [34] obtained predictor data from the most recent assessments available, instead of assessing predictors at the intended time of use. Two studies did not report adequately on the assessment of computed tomography (CT) [39] or other features [37], resulting in an unclear risk of bias.

Two and four out of 11 studies were at high and unclear risk of bias within the outcome domain, respectively. In many prognostic studies [31,32,34,37], the criteria for ICU admission and blinding of outcome determination to predictor variables were often not reported, leading to an unclear risk of bias.

Within the analysis domain, all eleven studies had insufficient outcome events per variable (EPV) (<20 EPV for model development studies and <100 for model validation studies) leading to a high risk of bias. Furthermore, no study reported on model calibration and only two studies [37,38] appropriately handled and reported on missing data. Prognostic predictive models were particularly at risk of inadequately accounting for, or reporting on, censored patients who were still hospitalised without the outcome (e.g., ICU admission) at the end of the study period. Only one study appropriately accounted for censored data by combining deep learning techniques with traditional Cox regression [36].

### 3.3. Adherence to Reporting Standards

The modified TRIPOD checklist comprised 25 terms, including 17 terms for reporting of methods and eight terms for results. Figure 2 describes the adherence of studies to reporting standards, as assessed by the modified TRIPOD checklist. Studies reported on a median of 48% (IQR: 48–59%) of relevant TRIPOD items, with 10 of 25 TRIPOD items having 50% adherence or less. Additionally, the following eight TRIPOD items had 25% adherence or less: reporting on treatments administered to study participants (item 5c),

blinding of outcome and predictor assessment (items 6b and 7b), study size determination (item 8), reporting on characteristics of study participants, including proportions of participants with missing data (item 13b), reporting of unadjusted associations between predictors and outcomes in multivariable logistic regression models (item 14b), explanation of how to use the prediction model (item 15b), and calibration and method of calibration (adjusted item 16b).

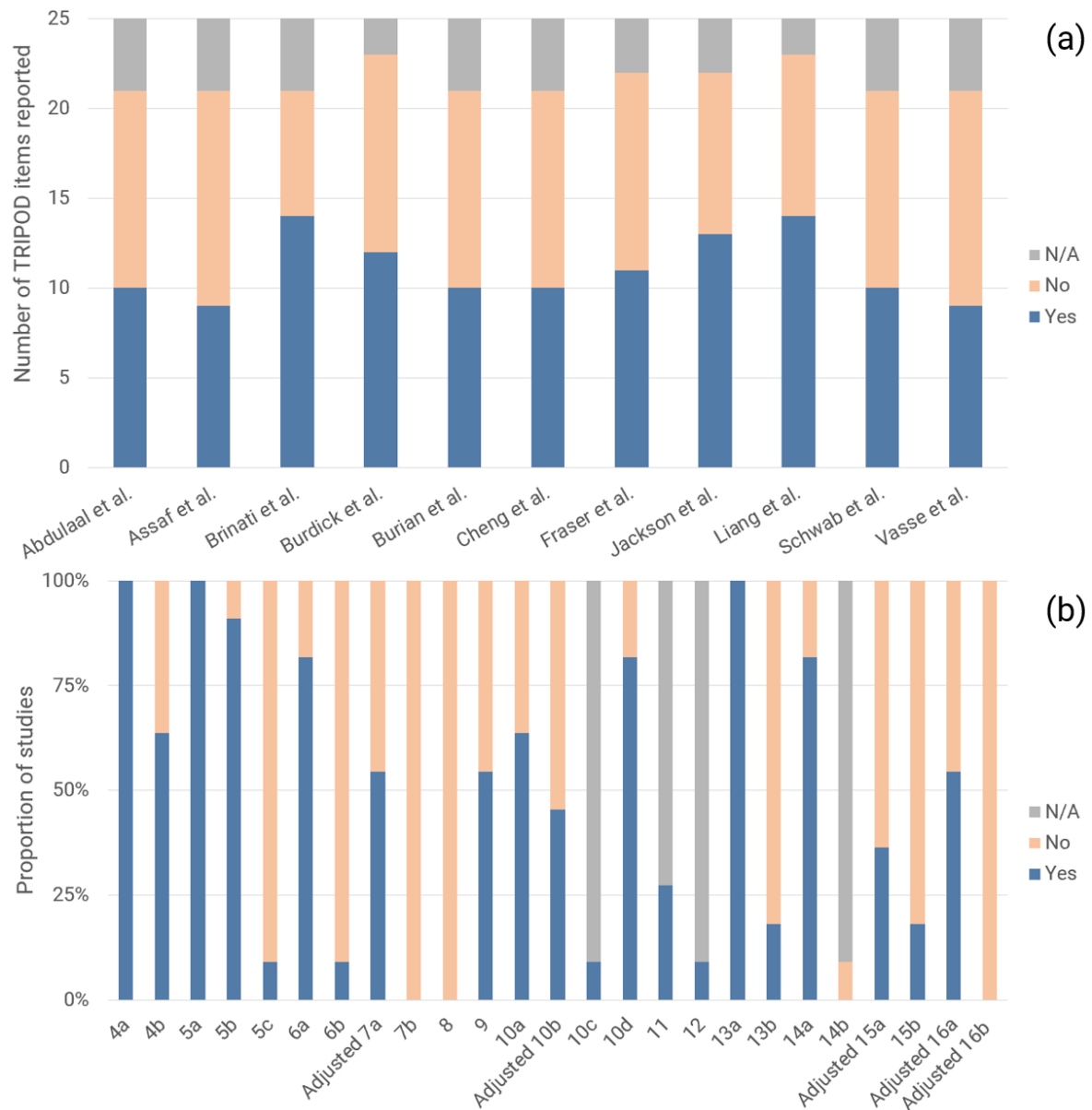


Figure 2. (a) Number of TRIPOD items reported per study and (b) Proportion of studies reporting on each TRIPOD item.

### 3.4. Diagnosis

Three studies investigated diagnostic AI predictive models; two studies developed models to predict the outcome of COVID-19 status at admission to the ED. Only one study was externally validated: Vasse, et al. [40] developed a decision tree based on cellular population data using random forest for feature selection (accuracy = 60.5%). Brinati, et al.'s [38] random forest model (C-index = 0.84, accuracy = 82%) and three-way random forest model (accuracy = 86%) achieved better performance, but were validated



using weaker  $k$ -fold cross-validation. Both studies included leucocyte or a leucocyte sub-population count as a predictor in their final model.

The third study [39] developed a decision tree for determining COVID-19 infection status in the ICU based on plasma inflammatory analyte features selected by a random forest classifier. On five-fold cross-validation, this classifier achieved an accuracy of 98%.

### 3.5. Prognosis

Most studies on prognostic AI predictive models (9/10, 90%) predicted ICU admission, mechanical ventilation, or a similar composite outcome of severe or critical illness. Collectively, such studies reported C-indices between 0.79 and 0.98. Liang, et al.'s [36] deep learning survival Cox model had the largest training cohort of 1590 patients and achieved a C-index of 0.890, 0.852, and 0.967 when externally validated on cohorts of 801, 305, and 73 patients from Wuhan, Hubei, and Guangzhou, respectively. Schwab et al.'s [37] support vector machine achieved a superior C-index of 0.98 on a weaker internal validation and a smaller sample size for testing model performance.

The artificial neural network trained by Abdulaal, et al. [30] using data collected at ED admission (C-index = 0.901) was the only prognostic AI model developed to predict in-hospital mortality in COVID-19 patients.

Apart from predictive modelling, Durhan et al. [41] and Mushtaq et al. [42] evaluated the prognostic utility of two separate deep learning-based software that determine the normal lung proportion and total lung involvement, respectively. Scores obtained from each software achieved a C-index of 0.944 and 0.77 for predicting ICU admission, respectively. While multivariate predictive models were not developed, both studies were subject to similar issues in development and reporting, including ambiguous criteria for ICU admission, inappropriate handling of missing data using complete-case analysis, and lack of reporting on treatments received by participants and on blinding of the outcome.

### 3.6. Other Applications

Apart from diagnostic and prognostic applications, Belciug, et al. [43] utilised an artificial immune system algorithm, a type of evolutionary AI algorithm, to optimise a queueing model for simulating hospital bed allocation in the ICU. The final model, intended as a tool for hospital managers, proposes an optimal admission rate and number of beds while balancing the costs associated with increasing capacity and refusing patients. The model was applied to ICU data published by the Ministry of Health of Italy and estimated a minimum rejection rate of 3.4% and 1.7% of patients requiring ICU admission from 13 March 2020 to 23 March 2020 (average daily volume of 200 patients) and 23 March 2020 to 30 March 2020 (average daily volume of 63 patients), respectively. However, these estimates were not validated.

## 4. Discussion

Our study is the first systematic review of AI applications for COVID-19 in intensive care and emergency settings. Applications were largely limited to diagnostic and prognostic predictive modelling, with only one study investigating a separate application of simulating ICU bed occupancy for resource optimisation. Due to high risk of bias, inadequate validation, or poor adherence to reporting standards in all reviewed studies, we have found no AI application for COVID-19 ready for clinical deployment in intensive care or emergency settings.

Among the reviewed articles, we found a limited range of AI applications being studied within intensive care and emergency settings. An exploratory review identified early detection and diagnosis, resource management of hospital beds or healthcare workers, and automatic monitoring and prognostication as possible applications of AI for the COVID-19 pandemic [20]. However, current applications within the reviewed articles mainly comprised prognostic models for critical illness or diagnostic models to predict COVID-19 status, none of which are ready for clinical use. Only one preliminary study

by Belciug et al. [43], which lacked validation, investigated allocative simulation and resource optimisation in the ICU, while no study investigated automatic monitoring or prognostication of COVID-19 patients. Belciug, et al. studied ICU resource optimisation employing queueing theory, a mathematical field of study, and artificial immune systems, an evolutionary AI algorithm that is uncommonly utilised in medical research. Unfamiliarity and the absence of general adoption of these methods within the medical community may contribute to the paucity of studies exploring less common but potentially impactful AI applications. As highlighted in previous literature [22,44], robust interdisciplinary collaboration and communication will be crucial in stimulating broader applications of AI for COVID-19 in intensive care and emergency settings, as well as the in medical literature at large.

Assessment of AI predictive models also revealed significant deficiencies in model development, validation, and reporting. These findings corroborate with earlier systematic reviews on predictive models for COVID-19 [22] and in intensive care settings [24]. Studies developing AI models should adhere to the TRIPOD reporting guidelines [25], PROBAST [26], or, ideally, recent AI-specific guidelines. These include the guidelines for transparency, reproducibility, ethics, and effectiveness (TREE) [45], Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) [46], and Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) [47]. While the above guidelines provide comprehensive explanations and elaborations, we emphasise hereinafter several common problematic areas within the reviewed studies and recommendations for future studies.

The most common source of bias was an inadequate sample size, which was found in all studies. A low sample size introduces the risk of over-fitting and model optimism. A benchmark for the development of logistic regression models is 20 EPV [4,26,48], while models using AI algorithms like random forest, support vector machines, and neural networks may require up to 200 EPV to account for model optimism [49]; a minimum of 100 EPV is recommended for validation studies [26]. Missing data also contributed significantly to bias; only two studies appropriately handled and reported on missing data. Ideally, the proportion of missing data for each variable should be reported [25] and multiple imputation should be used to avoid bias from inappropriate exclusion of participants with missing data (i.e., complete-case analysis) [50–53]. However, if complete-case analysis is used, authors should provide a comparative analysis of model performance with and without excluded participants to facilitate the judgement of bias from exclusion. For prognostic studies, studies often failed to appropriately account for censored patients (e.g., neither discharged nor admitted to the ICU). Censored patients should be handled using a time-to-event analysis such as Cox regression; inappropriate exclusion of these patients may lead to a skewed dataset that includes fewer patients without the outcome, introducing bias into the model [26]. For diagnostic studies, bias was often introduced by using the reverse transcription-polymerase chain reaction (RT-PCR) test as the ground truth or gold-standard for COVID-19 diagnosis, despite potentially poor sensitivity [54]. We recommend repeat RT-PCR testing to minimise the likelihood of false-negative tests in both diagnostic model development and validation studies.

Several key areas for improvements in reporting were identified in our study, including treatments received by participants, blinding, and study size determination. In particular, no study reported on calibration, a crucial yet often unevaluated measure of model performance [55]. We recommend assessing calibration using the calibration hierarchy described by Van Calster, et al. [56] instead of the commonly used Hosmer-Lemeshow test [57]. This avoids artificial stratification of patients into risk groups and other limitations associated with the Hosmer-Lemeshow test [55].

Studies should also strive to validate their data using stricter validation techniques. Studies with smaller sample sizes should utilise re-sampling techniques, such as bootstrapping or *k*-fold cross-validation. Studies with larger sample sizes should use a non-random split of data (e.g., by location or time) or perform external validation on independent data,

for example, from a different study [25,26,58]. Validation using the same data for model development is inappropriate as it only provides apparent model performance. Similarly, validation using a random split of data, such as a ‘train-test’ split, has lower power than re-sampling techniques [25,59] and should be avoided.

In addition to the limitations in quality and reporting of AI applications, the narrow scope of applications being investigated naturally leads to fewer AI applications eventually being suitable for clinical use. While AI has been practically applied for the identification of candidate drugs for drug repurposing [60] and contact tracing [21], its application and utility for COVID-19 in clinical settings have been insignificant to date. Several studies have employed AI techniques for the detection and classification of COVID-19 images [23], however, none have been validated as a clinical diagnostic adjunct in the ED. Factors that may contribute to this lack of clinical validation include the high risk of bias within existing models [22,61], limited applicability of radiographic images for discriminating between multiple differential diagnoses, and the high prevalence of asymptomatic radiographs in patients who present soon after the onset of symptoms [62,63]. Notwithstanding the high risk of bias and poor reporting of the reviewed AI models, AI algorithms tend to produce uninterpretable “black box” predictive models, which may lead to decreased acceptability of both diagnostic and prognostic AI applications amongst clinicians and hospital administrators. Some studies [35,39,40] have attempted to overcome this by using AI techniques for feature selection and presenting the final model as a decision tree or scoring system with clearly defined input variables. However, such simplifications of AI models curtail performance and limit the utility of the final model.

The above barriers to the validation and integration of AI in clinical settings may preclude significant contribution of AI to combatting the COVID-19 pandemic in intensive care units and emergency departments in the near future. However, improvements in the development, validation, and reporting of AI applications will be critical in advancing the applicability and acceptance of these systems in clinical settings in later phases of the COVID-19 pandemic and in future global health crises. Encouragingly, leading journals such as the *Lancet* family of journals have committed to enforcing AI-specific guidelines such as CONSORT-AI and SPIRIT-AI for submissions with an AI intervention [64]. However, concerted effort is needed from the entire research community, including journals, editors, and authors, to normalise the use of these guidelines and checklists. Such changes will encourage improved development, reporting, and eventual clinical uptake of future AI applications.

### *Limitations*

The results from our systematic review should be considered along with the following limitations. Firstly, our search excluded non-peer-reviewed articles, such as abstracts, posters, conference proceedings, and papers from preprint servers like bioRxiv and medRxiv, which may neglect the most recent literature but ensures a baseline quality of included studies. Secondly, we may have missed some relevant articles despite using a comprehensive search strategy due to publication in journals not indexed in the searched databases and variations in terminology used to describe AI algorithms and intensive care and emergency settings. We may also have missed AI applications that were deployed without publication in scientific literature; in particular, given the intense media attention and the pressure to deploy solutions quickly, AI solutions developed by governments and industry are more likely to be published in mass media formats rather than scientific journals. Thirdly, assessment according to PROBAST and, to a lesser extent, TRIPOD reporting guidelines still rely on a degree of subjectivity, despite comprehensive explanations and elaborations. Hence, other reviewers may arrive at slightly differing results. Lastly, the unprecedented volume of research on COVID-19 has resulted in a rapidly evolving body of literature. Hence, our findings are merely descriptive of the current situation, which may change with welcome improvements and additions to the medical literature.

## 5. Conclusions

Despite widespread interest in novel technologies for the COVID-19 pandemic, our systematic review of the literature reveals that current AI applications were limited in both the range of applications and clinical applicability. Several significant issues in the development, validation, and reporting of AI applications undermine safe and effective implementation of these systems within intensive care units or emergency departments. The integration of new AI-specific reporting guidelines like CONSORT-AI and SPIRIT-AI into research and publication processes will be a vital step in creating future AI applications that are clinically acceptable in the current pandemic, future pandemics, and within the wider medical field. We also emphasise the importance of closer interdisciplinary collaboration between AI experts and clinicians.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijerph18094749/s1>, supplementary File S1: PRISMA checklist (file format: .docx); supplementary File S2: Search strategy (file format: .docx); supplementary File S3: Modified TRIPOD checklist (file format: .docx).

**Author Contributions:** Conceptualization, N.L.; methodology, M.L.C. and N.L.; formal analysis, M.L.C. and N.L.; investigation, M.L.C., M.E.H.O., F.J.S., Z.Z., S.L.L., A.F.W.H. and N.L.; writing—original draft preparation, M.L.C. and N.L.; writing—review and editing, M.L.C., M.E.H.O., F.J.S., Z.Z., S.L.L., A.F.W.H. and N.L.; supervision, N.L.; funding acquisition, N.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Duke-NUS Signature Research Programme funded by the Ministry of Health, Singapore. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Careno, L.; Costantini, E.; Greco, M.; Barra, F.L.; Rendiniello, V.; Mainetti, M.; Bui, R.; Zanella, A.; Grasselli, G.; Lagioia, M.; et al. Hospital surge capacity in a tertiary emergency referral centre during the COVID-19 outbreak in Italy. *Anaesthesia* **2020**, *75*, 928–934. [[CrossRef](#)]
2. Grasselli, G.; Pesenti, A.; Cecconi, M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy. *JAMA* **2020**, *323*, 1545–1546. [[CrossRef](#)] [[PubMed](#)]
3. Phua, J.; Weng, L.; Ling, L.; Egi, M.; Lim, C.-M.; Divatia, J.V.; Shrestha, B.R.; Arabi, Y.M.; Ng, J.; Gomersall, C.D.; et al. Intensive care management of coronavirus disease 2019 (COVID-19): Challenges and recommendations. *Lancet Respir. Med.* **2020**, *8*, 506–517. [[CrossRef](#)]
4. Ogundimu, E.O.; Altman, D.G.; Collins, G.S. Adequate sample size for developing prediction models is not simply related to events per variable. *J. Clin. Epidemiol.* **2016**, *76*, 175–182. [[CrossRef](#)] [[PubMed](#)]
5. Mojoli, F.; Mongodi, S.; Orlando, A.; Arisi, E.; Pozzi, M.; Civardi, L.; Tavazzi, G.; Baldanti, F.; Bruno, R.; Iotti, G.A. Our recommendations for acute management of COVID-19. *Crit. Care* **2020**, *24*, 207. [[CrossRef](#)]
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
7. Mitchell, T.M. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997; p. xvii. 414p.
8. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial intelligence in radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510. [[CrossRef](#)]
9. Du-Harpur, X.; Watt, F.M.; Luscombe, N.M.; Lynch, M.D. What is AI? Applications of artificial intelligence to dermatology. *Br. J. Dermatol.* **2020**, *183*, 423–430. [[CrossRef](#)]
10. Ting, D.S.W.; Pasquale, L.R.; Peng, L.; Campbell, J.P.; Lee, A.Y.; Raman, R.; Tan, G.S.W.; Schmetterer, L.; Keane, P.A.; Wong, T.Y. Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **2019**, *103*, 167–175. [[CrossRef](#)]
11. Blanch, L.; Sales, B.; Montanya, J.; Lucangelo, U.; Garcia-Esquirol, O.; Villagra, A.; Chacon, E.; Estruga, A.; Borelli, M.; Burgueño, M.J.; et al. Validation of the Better Care® system to detect ineffective efforts during expiration in mechanically ventilated patients: A pilot study. *Intensive Care Med.* **2012**, *38*, 772–780. [[CrossRef](#)]

12. Chen, C.-W.; Lin, W.-C.; Hsu, C.-H.; Cheng, K.-S.; Lo, C.-S. Detecting ineffective triggering in the expiratory phase in mechanically ventilated patients based on airway flow and pressure deflection: Feasibility of using a computer algorithm. *Crit. Care Med.* **2008**, *36*, 455–461. [[CrossRef](#)] [[PubMed](#)]
13. Clifton, D.A.; Wong, D.; Clifton, L.; Wilson, S.; Way, R.; Pullinger, R.; Tarassenko, L. A Large-Scale Clinical Validation of an Integrated Monitoring System in the Emergency Department. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 835–842. [[CrossRef](#)]
14. Curtis, D.W.; Pino, E.J.; Bailey, J.M.; Shih, E.I.; Waterman, J.; Vinterbo, S.A.; Stair, T.O.; Gutttag, J.V.; Greenes, R.A.; Ohno-Machado, L. SMART—An Integrated Wireless System for Monitoring Unattended Patients. *J. Am. Med. Inform. Assoc.* **2008**, *15*, 44–53. [[CrossRef](#)]
15. Escobar, G.J.; Turk, B.J.; Ragins, A.; Ha, J.; Hoberman, B.; LeVine, S.M.; Balleca, M.A.; Liu, V.; Kipnis, P. Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. *J. Hosp. Med.* **2016**, *11*, S18–S24. [[CrossRef](#)]
16. Jones, S.S.; Evans, R.S. An agent based simulation tool for scheduling emergency department physicians. *AMIA Annu. Symp. Proc.* **2008**, *2008*, 338–342.
17. Lin, A.X.; Ho, A.F.W.; Cheong, K.H.; Li, Z.; Cai, W.; Chee, M.L.; Ng, Y.Y.; Xiao, X.; Ong, M.E.H. Leveraging Machine Learning Techniques and Engineering of Multi-Nature Features for National Daily Regional Ambulance Demand Prediction. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4179. [[CrossRef](#)]
18. Sun, Y.; Heng, B.H.; Seow, Y.T.; Seow, E. Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emerg. Med.* **2009**, *9*. [[CrossRef](#)]
19. Liu, N.; Zhang, Z.; Wah Ho, A.F.; Ong, M.E.H. Artificial intelligence in emergency medicine. *J. Emerg. Crit. Care Med.* **2018**, *2*, 82. [[CrossRef](#)]
20. Vaishya, R.; Javaid, M.; Khan, I.H.; Haleem, A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2020**, *14*, 337–339. [[CrossRef](#)]
21. Lalmuanawma, S.; Hussain, J.; Chhakchhuak, L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals* **2020**, *139*, 110059. [[CrossRef](#)]
22. Wynants, L.; Van Calster, B.; Collins, G.S.; Riley, R.D.; Heinze, G.; Schuit, E.; Bonten, M.M.J.; Dahly, D.L.; Damen, J.A.A.; Debray, T.P.A.; et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* **2020**, *369*, m1328. [[CrossRef](#)] [[PubMed](#)]
23. Albahri, O.S.; Zaidan, A.A.; Albahri, A.S.; Zaidan, B.B.; Abdulkareem, K.H.; Al-qaysi, Z.T.; Alamoodi, A.H.; Aleesa, A.M.; Chyad, M.A.; Alesa, R.M.; et al. Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects. *J. Infect. Public Health* **2020**, *13*, 1381–1396. [[CrossRef](#)] [[PubMed](#)]
24. Shillan, D.; Sterne, J.A.C.; Champneys, A.; Gibbison, B. Use of machine learning to analyse routinely collected intensive care unit data: A systematic review. *Crit. Care* **2019**, *23*. [[CrossRef](#)] [[PubMed](#)]
25. Moons, K.G.M.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.A.; Macaskill, P.; Steyerberg, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [[CrossRef](#)]
26. Moons, K.G.M.; Wolff, R.F.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann. Intern. Med.* **2019**, *170*, W1–W33. [[CrossRef](#)]
27. Collins, G.S.; Moons, K.G.M. Reporting of artificial intelligence prediction models. *Lancet* **2019**, *393*, 1577–1579. [[CrossRef](#)]
28. Wang, W.; Kiik, M.; Peek, N.; Curcin, V.; Marshall, I.J.; Rudd, A.G.; Wang, Y.; Douiri, A.; Wolfe, C.D.; Bray, B. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE* **2020**, *15*, e0234722. [[CrossRef](#)]
29. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ* **2015**, *350*, g7594. [[CrossRef](#)] [[PubMed](#)]
30. Abdulaal, A.; Patel, A.; Charani, E.; Denny, S.; Mughal, N.; Moore, L. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. *J. Med. Internet Res.* **2020**, *22*, e20259. [[CrossRef](#)]
31. Assaf, D.; Gutman, Y.a.; Neuman, Y.; Segal, G.; Amit, S.; Gefen-Halevi, S.; Shilo, N.; Epstein, A.; Mor-Cohen, R.; Biber, A.; et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern. Emerg. Med.* **2020**. [[CrossRef](#)]
32. Burdick, H.; Lam, C.; Mataraso, S.; Siefkas, A.; Braden, G.; Dellinger, R.P.; McCoy, A.; Vincent, J.L.; Green-Saxena, A.; Barnes, G.; et al. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. *Comput. Biol. Med.* **2020**, *124*, 103949. [[CrossRef](#)] [[PubMed](#)]
33. Burian, E.; Jungmann, F.; Kaissis, G.A.; Lohöfer, F.K.; Spinner, C.D.; Lahmer, T.; Treiber, M.; Dommasch, M.; Schneider, G.; Geisler, F.; et al. Intensive Care Risk Estimation in COVID-19 Pneumonia Based on Clinical and Imaging Parameters: Experiences from the Munich Cohort. *J. Clin. Med.* **2020**, *9*, 1514. [[CrossRef](#)]
34. Cheng, F.-Y.; Joshi, H.; Tandon, P.; Freeman, R.; Reich, D.L.; Mazumdar, M.; Kohli-Seth, R.; Levin, M.A.; Timsina, P.; Kia, A. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. *J. Clin. Med.* **2020**, *9*, 1668. [[CrossRef](#)]
35. Jackson, B.R.; Gold, J.A.W.; Natarajan, P.; Rossow, J.; Neblett Fanfair, R.; da Silva, J.; Wong, K.K.; Browning, S.D.; Bamrah Morris, S.; Rogers-Brown, J.; et al. Predictors at admission of mechanical ventilation and death in an observational cohort of adults hospitalized with COVID-19. *Clin. Infect. Dis.* **2020**, ciaa1459. [[CrossRef](#)] [[PubMed](#)]

36. Liang, W.; Yao, J.; Chen, A.; Lv, Q.; Zanin, M.; Liu, J.; Wong, S.; Li, Y.; Lu, J.; Liang, H.; et al. Early triage of critically ill COVID-19 patients using deep learning. *Nat. Commun.* **2020**, *11*, 3543. [[CrossRef](#)] [[PubMed](#)]
37. Schwab, P.; DuMont Schütte, A.; Dietz, B.; Bauer, S. Clinical Predictive Models for COVID-19: Systematic Study. *J. Med. Internet Res.* **2020**, *22*, e21439. [[CrossRef](#)]
38. Brinati, D.; Campagner, A.; Ferrari, D.; Locatelli, M.; Banfi, G.; Cabitza, F. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J. Med. Syst.* **2020**, *44*, 135. [[CrossRef](#)] [[PubMed](#)]
39. Fraser, D.D.; Cepinskas, G.; Slessarev, M.; Martin, C.; Daley, M.; Miller, M.R.; O’Gorman, D.B.; Gill, S.E.; Patterson, E.K.; Dos Santos, C.C. Inflammation Profiling of Critically Ill Coronavirus Disease 2019 Patients. *Crit. Care Explor.* **2020**, *2*, e0144. [[CrossRef](#)] [[PubMed](#)]
40. Vasse, M.; Ballester, M.C.; Ayaka, D.; Sukhachev, D.; Delcominette, F.; Habarou, F.; Jolly, E.; Sukhacheva, E.; Pascreau, T.; Farfour, É. Interest of the cellular population data analysis as an aid in the early diagnosis of SARS-CoV-2 infection. *Int. J. Lab. Hematol.* **2020**, *43*, 116–122. [[CrossRef](#)] [[PubMed](#)]
41. Durhan, G.; Ardali Duzgun, S.; Basaran Demirkazik, F.; Irmak, I.; Idilman, I.; Gulsun Akpınar, M.; Akpınar, E.; Ocal, S.; Telli, G.; Topeli, A.; et al. Visual and software-based quantitative chest CT assessment of COVID-19: Correlation with clinical findings. *Diagn. Interv. Radiol.* **2020**, *26*, 557–564. [[CrossRef](#)]
42. Mushtaq, J.; Pennella, R.; Lavalley, S.; Colarieti, A.; Steidler, S.; Martinenghi, C.M.A.; Palumbo, D.; Esposito, A.; Rovere-Querini, P.; Tresoldi, M.; et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: Analysis of 697 Italian patients. *Eur. Radiol.* **2020**, *31*, 1770–1779. [[CrossRef](#)] [[PubMed](#)]
43. Belciug, S.; Bejinariu, S.I.; Costin, H. An Artificial Immune System Approach for a Multi-compartment Queuing Model for Improving Medical Resources and Inpatient Bed Occupancy in Pandemics. *Adv. Electr. Comput. Eng.* **2020**, *20*, 23–30. [[CrossRef](#)]
44. Liu, N.; Chee, M.L.; Niu, C.; Pek, P.P.; Siddiqui, F.J.; Ansah, J.P.; Matchar, D.B.; Lam, S.S.W.; Abdullah, H.R.; Chan, A.; et al. Coronavirus disease 2019 (COVID-19): An evidence map of medical literature. *BMC Med. Res. Methodol.* **2020**, *20*, 177. [[CrossRef](#)]
45. Vollmer, S.; Mateen, B.A.; Bohner, G.; Király, F.J.; Ghani, R.; Jonsson, P.; Cumbers, S.; Jonas, A.; McAllister, K.S.L.; Myles, P.; et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* **2020**, *368*, l6927. [[CrossRef](#)]
46. Liu, X.; Cruz Rivera, S.; Moher, D.; Calvert, M.J.; Denniston, A.K.; Ashrafian, H.; Beam, A.L.; Chan, A.-W.; Collins, G.S.; Deeks, A.D.J.; et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Lancet Digit. Health* **2020**, *2*, e537–e548. [[CrossRef](#)]
47. Cruz Rivera, S.; Liu, X.; Chan, A.-W.; Denniston, A.K.; Calvert, M.J.; Ashrafian, H.; Beam, A.L.; Collins, G.S.; Darzi, A.; Deeks, J.J.; et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *Lancet Digit. Health* **2020**, *2*, e549–e560. [[CrossRef](#)]
48. van Smeden, M.; de Groot, J.A.H.; Moons, K.G.M.; Collins, G.S.; Altman, D.G.; Eijkemans, M.J.C.; Reitsma, J.B. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med. Res. Methodol.* **2016**, *16*, 163. [[CrossRef](#)]
49. van der Ploeg, T.; Austin, P.C.; Steyerberg, E.W. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Med. Res. Methodol.* **2014**, *14*, 137. [[CrossRef](#)] [[PubMed](#)]
50. Janssen, K.J.M.; Donders, A.R.T.; Harrell, F.E.; Vergouwe, Y.; Chen, Q.; Grobbee, D.E.; Moons, K.G.M. Missing covariate data in medical research: To impute is better than to ignore. *J. Clin. Epidemiol.* **2010**, *63*, 721–727. [[CrossRef](#)] [[PubMed](#)]
51. Schafer, J.L. Multiple imputation: A primer. *Stat. Methods Med. Res.* **2016**, *8*, 3–15. [[CrossRef](#)] [[PubMed](#)]
52. White, I.R.; Royston, P.; Wood, A.M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **2011**, *30*, 377–399. [[CrossRef](#)]
53. Kang, H. The prevention and handling of the missing data. *Korean J. Anesthesiol.* **2013**, *64*, 402–406. [[CrossRef](#)]
54. Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* **2020**, *296*, E115–E117. [[CrossRef](#)] [[PubMed](#)]
55. Van Calster, B.; McLernon, D.J.; van Smeden, M.; Wynants, L.; Steyerberg, E.W. Calibration: The Achilles heel of predictive analytics. *BMC Med.* **2019**, *17*, 230. [[CrossRef](#)]
56. Van Calster, B.; Nieboer, D.; Vergouwe, Y.; De Cock, B.; Pencina, M.J.; Steyerberg, E.W. A calibration hierarchy for risk models was defined: From utopia to empirical data. *J. Clin. Epidemiol.* **2016**, *74*, 167–176. [[CrossRef](#)] [[PubMed](#)]
57. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*; Wiley-Interscience Publication: Hoboken, NJ, USA, 2000.
58. Steyerberg, E.W.; Bleeker, S.E.; Moll, H.A.; Grobbee, D.E.; Moons, K.G.M. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J. Clin. Epidemiol.* **2003**, *56*, 441–447. [[CrossRef](#)]
59. Steyerberg, E.W.; Harrell, F.E.; Borsboom, G.J.J.M.; Eijkemans, M.J.C.; Vergouwe, Y.; Habbema, J.D.F. Internal validation of predictive models. *J. Clin. Epidemiol.* **2001**, *54*, 774–781. [[CrossRef](#)]
60. Zhou, Y.; Wang, F.; Tang, J.; Nussinov, R.; Cheng, F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit. Health* **2020**, *2*, e667–e676. [[CrossRef](#)]
61. Tartaglione, E.; Barbano, C.A.; Berzovini, C.; Calandri, M.; Grangetto, M. Unveiling COVID-19 from CHEST X-ray with Deep Learning: A Hurdles Race with Small Data. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6933. [[CrossRef](#)] [[PubMed](#)]
62. Laghi, A. Cautions about radiologic diagnosis of COVID-19 infection driven by artificial intelligence. *Lancet Digit. Health* **2020**, *2*, e225. [[CrossRef](#)]

- 
63. Vancheri, S.G.; Savietto, G.; Ballati, F.; Maggi, A.; Canino, C.; Bortolotto, C.; Valentini, A.; Dore, R.; Stella, G.M.; Corsico, A.G.; et al. Radiographic findings in 240 patients with COVID-19 pneumonia: Time-dependence after the onset of symptoms. *Eur. Radiol.* **2020**, *30*, 6161–6169. [[CrossRef](#)] [[PubMed](#)]
  64. The Lancet Digital, H. Guiding better design and reporting of AI-intervention trials. *Lancet Digit. Health* **2020**, *2*, e493. [[CrossRef](#)]