Prediction models for diagnosis and prognosis of covid-19: living systematic review and critical appraisal

Prediction models for diagnosis and prognosis of covid-19: living systematic review and critical appraisal

Laure Wynants(http://orcid.org/0000-0002-3037-122X),1 2 assistant professor, Ben Van Calster(http://orcid.org/0000-0003-1613-7450),2 3 associate professor, Gary S Collins(http://orcid.org/0000-0002-2772-2316),4 5 professor, Richard D Riley,6 professor, Georg Heinze(http://orcid.org/0000-0003-1147-8491),7 associate professor, Ewoud Schuit(http://orcid.org/0000-0002-9548-3214),8 9 assistant professor, Marc M J Bonten(http://orcid.org/0000-0002-9095-9201),8 10 professor, Darren L Dahly (http://orcid.org/0000-0003-0110-324X),11 12 principal statistician, Johanna A Damen (http://orcid.org/0000-0001-7401-4593),8 9 assistant professor, Thomas P A Debray(http://orcid.org/0000-0002-1790-2719),8 9 assistant professor, Valentijn M T de Jong (http://orcid.org/0000-0001-9921-3468),8 9 assistant professor, Maarten De Vos(http://orcid.org/0000-0002-3482-5145),2 13 associate professor, Paula Dhiman (http://orcid.org/0000-0002-0989-0623),4 5 research fellow, Maria C Haller(http://orcid.org/0000-0003-4635-6291),7 14 medical doctor, Michael O Harhay(http://orcid.org/0000-0002-0553-674X),15 16 assistant professor, Liesbet Henckaerts(http://orcid.org/0000-0003-2880-1400),17 18 assistant professor, Pauline Heus (http://orcid.org/0000-0002-6886-4652),8 9 assistant professor, Michael Kammer (http://orcid.org/0000-0003-4821-9928),7 19 research associate, Nina Kreuzberger(http://orcid.org/0000-0001-8922-0488),20 research associate, Anna Lohmann(http://orcid.org/0000-0002-4004-4265),21 researcher in training, Kim Luijken(http://orcid.org/0000-0001-5192-8368),21 doctoral candidate, Jie Ma(http://orcid.org/0000-0002-3900-1903),5 medical statistician, Glen P Martin (http://orcid.org/0000-0002-3410-9472),22 lecturer, David J McLernon (http://orcid.org/0000-0001-8905-2429),23 senior research fellow, Constanza L Andaur Navarro (http://orcid.org/0000-0002-7745-2887),8 9 doctoral student, Johannes B Reitsma,8 9 associate professor, Jamie C Sergeant(http://orcid.org/0000-0002-9000-4413),24 25 senior lecturer, Chunhu Shi(http://orcid.org/0000-0003-0151-0451),26 research associate, Nicole Skoetz(http://orcid.org/0000-0003-4744-6192),19 medical doctor, Luc J M Smits(http://orcid.org/0000-0003-0785-1345),1 professor, Kym I E Snell(http://orcid.org/0000-0001-9373-6591),6 lecturer, Matthew Sperrin(http://orcid.org/0000-0002-5351-9960),27 senior lecturer, René Spijker(http://orcid.org/0000-0003-4445-1201),8 9 28 information specialist, Ewout W Steyerberg(http://orcid.org/0000-0002-7787-0122),3 professor, Toshihiko Takada(http://orcid.org/0000-0002-8032-6224),8 assistant professor, Ioanna Tzoulaki,29 30 assistant professor, Sander M J van Kuijk,31 research fellow, Bas van Bussel (http://orcid.org/0000-0003-1621-7848), 1, 32 medical doctor, Iwan C C van der Horst (https://orcid.org/0000-0003-3891-8522), 32 professor, Florien S van Royen(http://orcid.org/0000-0002-6785-214X),8 research fellow, Jan Y Verbakel (https://orcid.org/0000-0002-7166-7211),33 34 assistant professor, Christine Wallisch(http://orcid.org/0000-0003-3943-6234),7 35 36 research fellow, Jack Wilkinson (http://orcid.org/0000-0003-3513-4677),22 research fellow, Robert Wolff (http://orcid.org/0000-0003-3270-7791),37 medical doctor, Lotty Hooft (http://orcid.org/0000-0002-7950-2980),8 9 associate professor, Karel G M Moons,8 9 professor, Maarten van Smeden(http://orcid.org/0000-0002-5529-1541),8 assistant professor

[1]Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Peter Debyeplein 1, 6229 HA Maastricht, Netherlands

[2]Department of Development and Regeneration, KU Leuven, Leuven, Belgium

[3]Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, Netherlands

[4]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Musculoskeletal Sciences, University of Oxford, Oxford, UK

[5]NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[6]Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Keele, UK

[7]Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[8]Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands

[9]Cochrane Netherlands, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands

[10]Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, Netherlands

[11]HRB Clinical Research Facility, Cork, Ireland

[12]School of Public Health, University College Cork, Cork, Ireland

[13]Department of Electrical Engineering, ESAT Stadius, KU Leuven, Leuven, Belgium

[14]Ordensklinikum Linz, Hospital Elisabethinen, Department of Nephrology, Linz, Austria

[15]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[16]Palliative and Advanced Illness Research Center and Division of Pulmonary and Critical Care Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[17]Department of Microbiology, Immunology and Transplantation, KU Leuven-University of Leuven, Leuven, Belgium

[18]Department of General Internal Medicine, KU Leuven-University Hospitals Leuven, Leuven, Belgium

[19]Department of Nephrology, Medical University of Vienna, Vienna, Austria

[20]Evidence-Based Oncology, Department I of Internal Medicine and Centre for Integrated Oncology Aachen Bonn Cologne Dusseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[21]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, Netherlands

[22]Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

[23]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

[24]Centre for Biostatistics, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[25]Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[26]Division of Nursing, Midwifery and Social Work, School of Health Sciences, University of Manchester, Manchester, UK

[27]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

[28]Amsterdam UMC, University of Amsterdam, Amsterdam Public Health, Medical Library, Netherlands

[29]Department of Epidemiology and Biostatistics, Imperial College London School of Public Health, London, UK

[30]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

[31]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, Maastricht, Netherlands

[32]Department of Intensive Care, Maastricht University Medical Center+, Maastricht University, Maastricht, The Netherlands

[33]EPI-Centre, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[34]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

[35]Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

[36]Berlin Institute of Health, Berlin, Germany

[37]Kleijnen Systematic Reviews, York, UK

Correspondence to: L Wynants laure.wynants@maastrichtuniversity.nl

Abstract

Objective To review and appraise the validity and usefulness of published and preprint reports of prediction models for diagnosing coronavirus disease 2019 (covid-19) in patients with suspected infection, for prognosis of patients with covid-19, and for detecting people in the general population at increased risk of covid-19 infection or being admitted to hospital with the disease.

Design Living systematic review and critical appraisal by the COVID-PRECISE (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse sEttings) group.

Data sources PubMed and Embase through Ovid, up to 1 July 2020, supplemented with arXiv, medRxiv, and bioRxiv up to 5 May.

Study selection Studies that developed or validated a multivariable covid-19 related prediction model.

Data extraction At least two authors independently extracted data using the CHARMS (critical appraisal and data extraction for systematic reviews of prediction modelling studies) checklist; risk of bias was assessed using PROBAST (prediction model risk of bias assessment tool).

Results 37 421 titles were screened, and 169 studies describing 232 prediction models were included. The review identified seven models for identifying people at risk in the general population; 118 diagnostic models for detecting covid-19 (75 were based on medical imaging, 10 to diagnose disease severity); and 107 prognostic models for predicting mortality risk, progression to severe disease, intensive care unit admission, ventilation, intubation, or length of hospital stay. The most frequent types of predictors included in the covid-19 prediction models are vital signs, age, comorbidities, and image features. Flu-like symptoms are frequently predictive in diagnostic models, while sex, C reactive protein, and lymphocyte counts are frequent prognostic factors. Reported C index estimates ranged from 0.71 to 0.99 in prediction models for the general population, from 0.65 to more than 0.99 in diagnostic models, and from 0.54 to 0.99 in prognostic models. All models were rated at high or unclear risk of bias, mostly because of non-representative selection of control patients, exclusion of patients who had not experienced the event of interest by the end of the study, high risk of model overfitting, and unclear reporting. Thirty five percent did not include a description of the target population or care setting, and only 5% was externally validated using a calibration plot.

Conclusion Prediction models for covid-19 are quickly entering the academic literature to support medical decision making at a time when they are urgently needed. This review indicates that proposed models are poorly reported, at high risk of bias, and their reported performance is probably optimistic. However, we have identified two promising models that should be validated in multiple cohorts, preferably through collaborative efforts and data sharing to allow an investigation of heterogeneity in performance. Details on all reviewed models are publically available at www.covprecise.org. Methodological guidance should be followed because unreliable predictions could cause more harm than benefit in guiding clinical decisions. Finally, authors should adhere to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) reporting guideline.

Systematic review registration Protocol https://osf.io/ehc47/, registration https://osf.io/wy245.

Readers' note This article is a living systematic review that will be updated to reflect emerging evidence. Updates may occur for up to two years from the date of original publication. This version is update 3 of the original article published on 7 April 2020 (*BMJ* 2020;369:m1328). Previous updates can be found as data supplements (https://www.bmj.com/content/369/bmj.m1328/related#datasupp) and on https://www.covprecise.org/.

What is already known on this topic

The sharp recent increase in coronavirus disease 2019 (covid-19) incidence has put a strain on healthcare systems worldwide; an urgent need exists for efficient early detection of covid-19 in the general population, for diagnosis of covid-19 in patients with suspected disease, and for prognosis of covid-19 in patients with confirmed disease

Viral nucleic acid testing and chest computed tomography imaging are standard methods for diagnosing covid-19, but are time consuming

Earlier reports suggest that elderly patients, patients with comorbidities (chronic obstructive pulmonary disease, cardiovascular disease, hypertension), and patients presenting with dyspnoea are vulnerable to more severe morbidity and mortality after infection

What this study adds

Seven models identified patients at risk in the general population (using proxy outcomes for covid-19)

Thirty three diagnostic models were identified for detecting covid-19, in addition to 75 diagnostic models based on medical images,10 diagnostic models for severity classification, and 107 prognostic models for predicting, among others, mortality risk, progression to severe disease

Proposed models are poorly reported and at high risk of bias, raising concern that their predictions could be unreliable when applied in daily practice

Two prediction models, one for diagnosis and one for prognosis, were identified as being of higher quality than others and efforts should be made to validate these in other datasets

Introduction

The novel coronavirus disease 2019 (covid-19) presents an important and urgent threat to global health. Since the outbreak in early December 2019 in the Hubei province of the People's Republic of China, the number of patients confirmed to have the disease has exceeded 47 million as the disease spread globally, and the number of people infected is probably much higher. More than 1.2 million people have died from covid-19 (up to 3 November 2020).[1] Despite public health responses aimed at containing the disease and delaying the spread, several countries have been confronted with a critical care crisis, and more countries could follow.[2-4] Outbreaks lead to important increases in the demand for hospital beds and shortage of medical equipment, while medical staff themselves also get infected. Several regions have had or are experiencing second waves, and despite improvements in testing and tracing, several regions are again facing the limits of their test capacity, hospital resources and healthcare staff.

To mitigate the burden on the healthcare system, while also providing the best possible care for patients, efficient diagnosis and information on the prognosis of the disease are needed. Prediction models that combine several variables or features to estimate the risk of people being infected or experiencing a poor outcome from the infection could assist medical staff in triaging patients when allocating limited healthcare resources. Models ranging from rule based scoring systems to advanced machine learning models (deep learning) have been proposed and published in response to a call to share relevant covid-19 research findings rapidly and openly to inform the public health response and help save lives.[5]

We aimed to systematically review and critically appraise all currently available prediction models for covid-19, in particular models to predict the risk of covid-19 infection or being admitted to hospital with the disease, models to predict the presence of covid-19 in patients with suspected infection, and models to predict the prognosis or course of infection in patients with covid-19. We included model development and external validation studies. This living systematic review, with periodic updates, is being conducted by the international COVID-PRECISE (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse sEttings: www.covprecise.org) group in collaboration with the Cochrane Prognosis Methods Group.

Methods

We searched the publicly available, continuously updated publication list of the covid-19 living systematic review.[6] We validated whether the list is fit for purpose (online

supplementary material) and further supplemented it with studies on covid-19 retrieved from arXiv. The online supplementary material presents the search strings. We included studies if they developed or validated a multivariable model or scoring system, based on individual participant level data, to predict any covid-19 related outcome. These models included three types of prediction models: diagnostic models for predicting the presence or severity of covid-19 in patients with suspected infection; prognostic models for predicting the course of infection in patients with covid-19; and prediction models to identify people in the general population at risk of covid-19 infection or at risk of being admitted to hospital with the disease.

We searched the database repeatedly up to 1 July 2020 (supplementary table 1). As of the third update (search date 1 July), we only include peer reviewed articles (indexed in PubMed and Embase through Ovid). Pre-prints (from bioRxiv, medRxiv, and arXiv) that were already included in previous updates of the systematic review remain included in the analysis. Reassessment takes place after publication of a preprint in a peer reviewed journal. No restrictions were made on the setting (eg, inpatients, outpatients, or general population), prediction horizon (how far ahead the model predicts), included predictors, or outcomes. Epidemiological studies that aimed to model disease transmission or fatality rates, diagnostic test accuracy, and predictor finding studies were excluded. Starting with the second update, retrieved records were initially screened by a text analysis tool developed using artificial intelligence to prioritise sensitivity (supplementary material). Titles, abstracts, and full texts were screened for eligibility in duplicate by independent reviewers (pairs from LW, BVC, MvS) using EPPI-Reviewer,[13] and discrepancies were resolved through discussion.

Data extraction of included articles was done by two independent reviewers (from LW, BVC, GSC, TPAD, MCH, GH, KGMM, RDR, ES, LJMS, EWS, KIES, CW, AL, JM, TT, JAAD, KL, JBR, LH, CS, MS, MCH, NS, NK, SMJvK, JCS, PD, CLAN, RW, GPM, IT, JYV, DLD, JW, FSvR, PH, VMTdJ, BvB, ICCvdH, DJM, MK and MvS). Reviewers used a standardised data extraction form based on the CHARMS (critical appraisal and data extraction for systematic reviews of prediction modelling studies) checklist[14] and PROBAST (prediction model risk of bias assessment tool; www.probast.org) for assessing the reported prediction models.[15] We sought to extract each model's predictive performance by using whatever measures were presented. These measures included any summaries of discrimination (the extent to which predicted risks discriminate between participants with and without the outcome), and calibration (the extent to which predicted risks correspond to observed risks) as recommended in the TRIPOD (transparent reporting of a multivariable

prediction model for individual prognosis or diagnosis; www.tripod-statement.org) statement.[16] Discrimination is often quantified by the C index (C index=1 if the model discriminates perfectly; C index=0.5 if discrimination is no better than chance). Calibration is often quantified by the calibration intercept (which is zero when the risks are not systematically overestimated or underestimated) and calibration slope (which is one if the predicted risks are not too extreme or too moderate).[17] We focused on performance statistics as estimated from the strongest available form of validation (in order of strength: external (evaluation in an independent database), internal (bootstrap validation, cross validation, random training test splits, temporal splits), apparent (evaluation by using exactly the same data used for development)). Any discrepancies in data extraction were discussed between reviewers, and remaining conflicts were resolved by LW and/or MvS. The online supplementary material provides details on data extraction. Some studies investigated multiple models and some models were investigated in multiple studies (i.e., in external validation studies). The unit of analysis was a model within a study, unless stated otherwise. We considered aspects of PRISMA (preferred reporting items for systematic reviews and meta-analyses)[18] and TRIPOD[16] in reporting our study. Details on all reviewed studies and prediction models are publically available at www.covprecise.org.

Patient and public involvement

It was not possible to involve patients or the public in the design, conduct, or reporting of our research. The study protocol and preliminary results are publicly available on https://osf.io/ehc47/, medRxiv and www.covprecise.org.

Results

We retrieved 37 412 titles through our systematic search (of which 23 203 were included in the present update; supplementary table 1, fig 1). We included a further nine studies that were publicly available but were not detected by our search. Of 37 421 titles, 444 studies were retained for abstract and full text screening (of which 169 are included in the present update). One hundred sixty nine studies describing 232 prediction models met the inclusion criteria (of which 62 studies and 87 models added since the present update, supplementary table 1).[7-12 19-119 165-226] These studies were selected for data extraction and critical appraisal. The unit of analysis is the model within a study: of these 232 models, 208 were unique, newly developed models for covid-19. The remaining 24 analyses were external validations of existing models (in a study other than the model development study). Some models were validated more than once (in different studies, as described below). Many models are publicly available (Box 1).

A database with the description of each model and its risk of bias assessment can be found on

www.covprecise.org.

Fig 1 PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart of study inclusions and exclusions
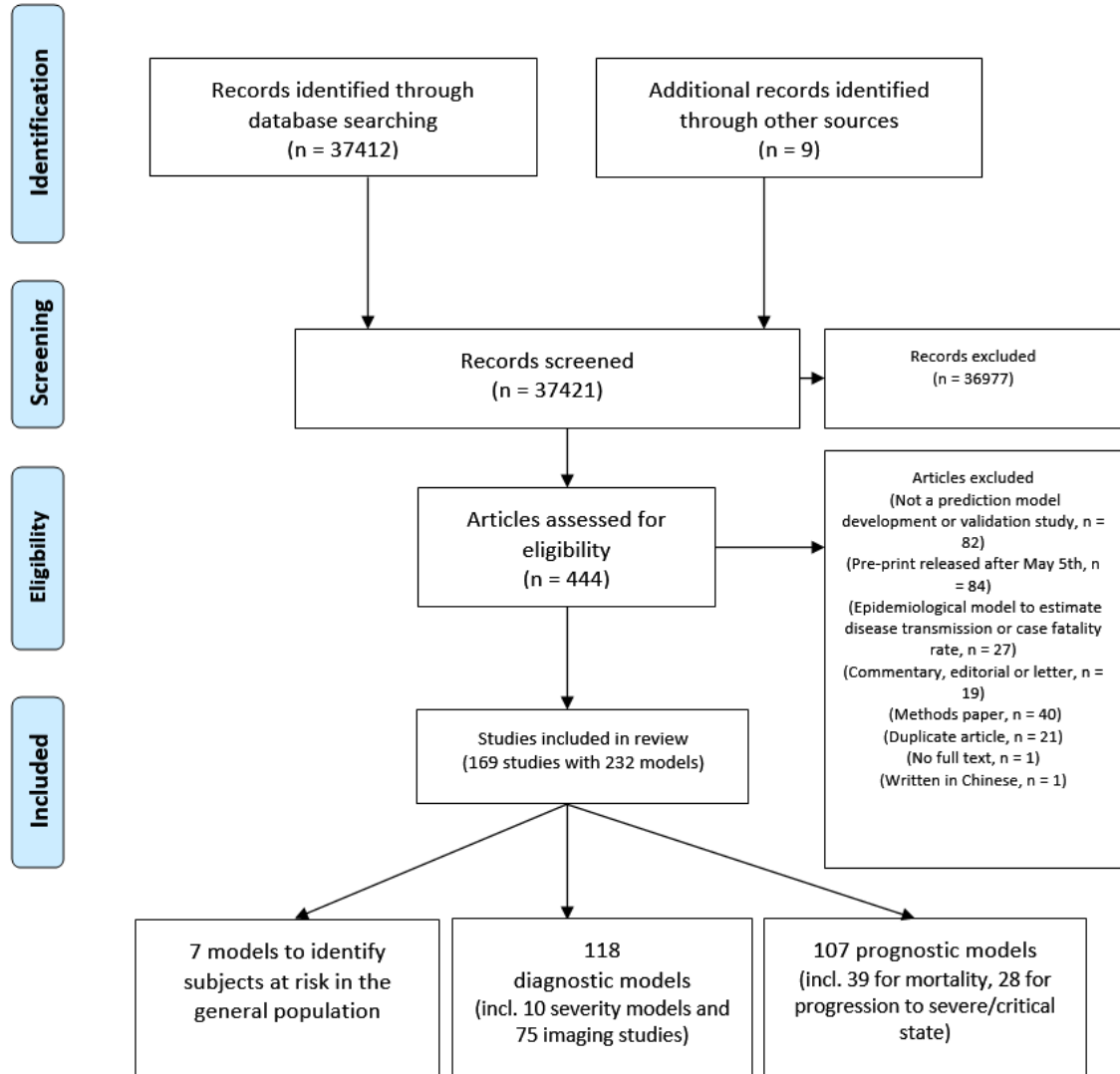
Primary datasets

One hundred seventy four (75%) models used data from a single country (Table 1). Forty two (18%) models used international data and for sixteen models (7%) it was unclear how many (and which) countries contributed data. Two models (1%) used simulated data and twelve models (5%) used proxy data to estimate covid-19 related risks (eg, Medicare claims data from 2015 to 2016). Most models were intended for use in confirmed covid-19 cases (47%) and a hospital setting (51%). The average patient age varied from 39 to 71 years, and the proportion of men varied from 35% to 75%, although this information was often not reported. One study developed a prediction model for use in paediatric patients[25].

Table 1. Characteristics of reviewed models

	Number of models* (%) or median (IQR)
Country **	
Single country data	174 (75%)
China	97 (42%)
Italy	23 (10%)
United States of America	17 (7%)
South-Korea	10 (4%)
France	5 (2%)
Singapore	4 (2%)
Turkey	4 (2%)
Brazil	3 (1%)
Spain	2 (1%)
United Kingdom	2 (1%)
Other single country	8 (3%)
International (combined) data	42 (18%)
Unknown origin of data	16 (7%)
Used proxy (non-covid-19) data	12 (5%)
Used simulated data	2 (1%)
Target setting	
Hospitalized patients	119 (51%)
Patient at triage centre/fever clinic	12 (5%)
Patients in general practice	3 (1%)
Other	23 (10%)
Unclear	75 (32%)
Target population	
Confirmed covid-19	108 (47%)
Suspected covid-19	84 (36%)
Other	13 (6%)
Unclear	27 (12%)
Type of model	
Predict risks of covid-19 in the general population	7 (3%)

Diagnostic (covid-19 vs. not covid-19)	33 (14%)
Diagnostic classification of covid-19 severity	10 (4%)
Diagnostic, imaging data only	75 (32%)
Prognostic	107 (46%)
Study type	
Developed in reviewed study	50 (22%)
Developed and internally validated in reviewed study	113 (49%)
Developed and externally validated in reviewed study	45 (19%)
Externally validated in reviewed study	24 (10%)
Sample size (development)	338 (134 to 707)
Number of events (development)	69 (37 to 160)
Sample size (external validation)	189 (76 to 312)
Number of events (external validation)	40 (24 to 122)

*The analysis unit is a model within a study. Some studies investigated multiple models and some models were investigated in multiple studies (i.e., in external validation studies). ** A study that uses development data from one country and validation data from another is classified as international.

Based on the studies that reported study dates, data were collected from December 2019 to June 2020. Some centres provided data to multiple studies and several studies used open Github[120] or Kaggle[121] data repositories (version or date of access often unspecified), and so it was unclear how much these datasets overlapped across our identified studies.

Among the diagnostic model studies, the reported prevalence of covid-19 varied between 7% and 71% (if a cross sectional or cohort design was used). Because 75 diagnostic studies used either case-control sampling or an unclear method of data collection, the prevalence in these diagnostic studies might not be representative of their target population.

Among the studies that developed prognostic models to predict mortality risk in people with confirmed or suspected infection, the percentage of deaths varied between 1% and 52%. This wide variation is partly because of substantial sampling bias caused by studies excluding participants who still had the disease at the end of the study period (that is, they had neither recovered nor died). Additionally, length of follow-up varied between studies (but was often not reported), and there likely is local and temporal variation in how people were diagnosed as having covid-19 or were admitted to the hospital (and therefore recruited for the studies).

Box 1: Availability of models in format for use in clinical practice

Two hundred and eight unique models were developed in the included studies. Thirty of these (14%) models were presented as a model equation including intercept and regression coefficients. Eight models (4%) were only partially presented (e.g., intercept or baseline hazard were missing). The remaining did not provide the underlying model equation. Seventy two models (35%) are available as a tool for use in clinical practice (in addition to or instead of a published equation). Twenty seven models were presented as a web calculator (13%), 12 as a sum score (6%), 11 as a nomogram (5%), 8 as a software object (4%), 5 as a decision tree or set of predictions for subgroups (2%), 3 as a chart score (1%), and 6 in other usable formats (3%).

All these presentation formats make predictions readily available for use in the clinic. However, because all models were at high or uncertain risk of bias, we do not recommend their routine use before they are externally validated, ideally by independent investigators.

Models to predict risk of covid-19 in the general population

We identified seven models that predicted risk of covid-19 in the general population. Three models from one study used hospital admission for non-tuberculosis pneumonia, influenza, acute bronchitis, or upper respiratory tract infections as proxy outcomes in a dataset without any patients with covid-19.[8] Among the predictors were age, sex, previous hospital admission, comorbidities, and social determinants of health. The study reported C indices of 0.73, 0.81, and 0.81. A fourth model used deep learning on thermal videos from the faces of people wearing facemasks to determine abnormal breathing (not covid related) with a reported sensitivity of 80%.[90] A fifth model uses demographics, symptoms and contact history in a mobile app to assist general practitioners in collecting data and to risk-stratify patients. It was contrasted with two further models including additional blood values and blood values plus computed tomography (CT) images. The authors reported a C index of 0.71 with demographics only, which rose to 0.97 and 0.99 as blood values and imaging characteristics were added.[194] Calibration was not assessed in any of the general population models.

Diagnostic models to detect covid-19 in patients with suspected infection

We identified 33 multivariable models to diagnose covid-19. Most models targeted patients with suspected covid-19. Reported C index values ranged between 0.65 and 0.99. Calibration was assessed for seven models using calibration plots (including two at external validation), with mixed results. The most frequently included predictors (at least ten times) were vital signs (eg, temperature, heart rate, respiratory rate, oxygen saturation, blood pressure), flu-like signs and symptoms (eg, shiver, fatigue), age, electrolytes, image features (eg, pneumonia signs on CT scan), contact with covid-19 confirmed cases, lymphocyte count,

neutrophil count, cough or sputum, sex, leukocytes, liver enzymes and red cell distribution width.

Ten studies aimed to diagnose severe disease in patients with covid-19: nine in adults with reported C indices between value of 0.80 and 0.99, and one in children that reported perfect classification of severe disease.[25] Calibration was not assessed in any of the models. Predictors of severe covid-19 used more than once were comorbidities, liver enzymes, C reactive protein, imaging features, lymphocyte count and neutrophil count.

Seventy seven prediction models were proposed to support the diagnosis of covid-19 or covid-19 pneumonia (and some also to monitor progression) based on images. Most studies used CT images or chest radiographs. Others used spectrograms of cough sounds[53] and lung ultrasound.[73] The predictive performance varied considerably, with reported C index values ranging from 0.70 to more than 0.99. Only one model based on imaging was evaluated using a calibration plot, and it appeared well calibrated at external validation [227].

Prognostic models for patients with diagnosis of covid-19

We identified 107 prognostic models for patients with a diagnosis of covid-19. The intended use of these models (that is, when to use them, and for whom) was often not clearly described. Prediction horizons varied between one and 37 days, but were often unspecified.

Of these models, 39 estimated mortality risk and 28 aimed to predict progression to a severe or critical disease. The remaining studies used other outcomes (single or as part of a composite) including recovery, length of hospital stay, intensive care unit admission, intubation, (duration of) mechanical ventilation, acute respiratory distress syndrome, cardiac injury and thrombotic complication. One study used data from 2015 to 2019 to predict mortality and prolonged assisted mechanical ventilation (as a non-covid-19 proxy outcome).[113]

The most frequently used categories of prognostic factors (for any outcome, included at least 20 times) included age, comorbidities, vital signs, image features, sex, lymphocyte count, and C reactive protein.

Studies that predicted mortality reported C indices between 0.68 and 0.98. Four studies also presented calibration plots (including at external validation), all indicating miscalibration[7 67 116] or showing plots for integer scores without clearly explaining how these were translated into predicted risks.[186] The studies that developed models to predict progression to a severe or critical disease reported C indices between 0.58 and 0.99. Five of

these models also were evaluated using calibration plots, two of them at external validation. Even though calibration appeared good, plots were constructed in an unclear way.[83 119]

Reported C indices for other outcomes varied between 0.54 (ICU admission) and 0.99 (severe symptoms 3 days after admission), and five models had calibration plots (of which two at external validation), with mixed results.

Risk of bias

All models were at high (226 or 97%) or unclear (6 or 3%) risk of bias according to assessment with PROBAST, which suggests that their predictive performance when used in practice is probably lower than that reported (Fig 2). Therefore, we have cause for concern that the predictions of the proposed models are unreliable when used in other people. Fig 2 and Box 2 gives details on common causes for risk of bias for each type of model.

Fig 2 PROBAST (prediction model risk of bias assessment tool) risk of bias for all included models combined (n=232) and broken down per type of model
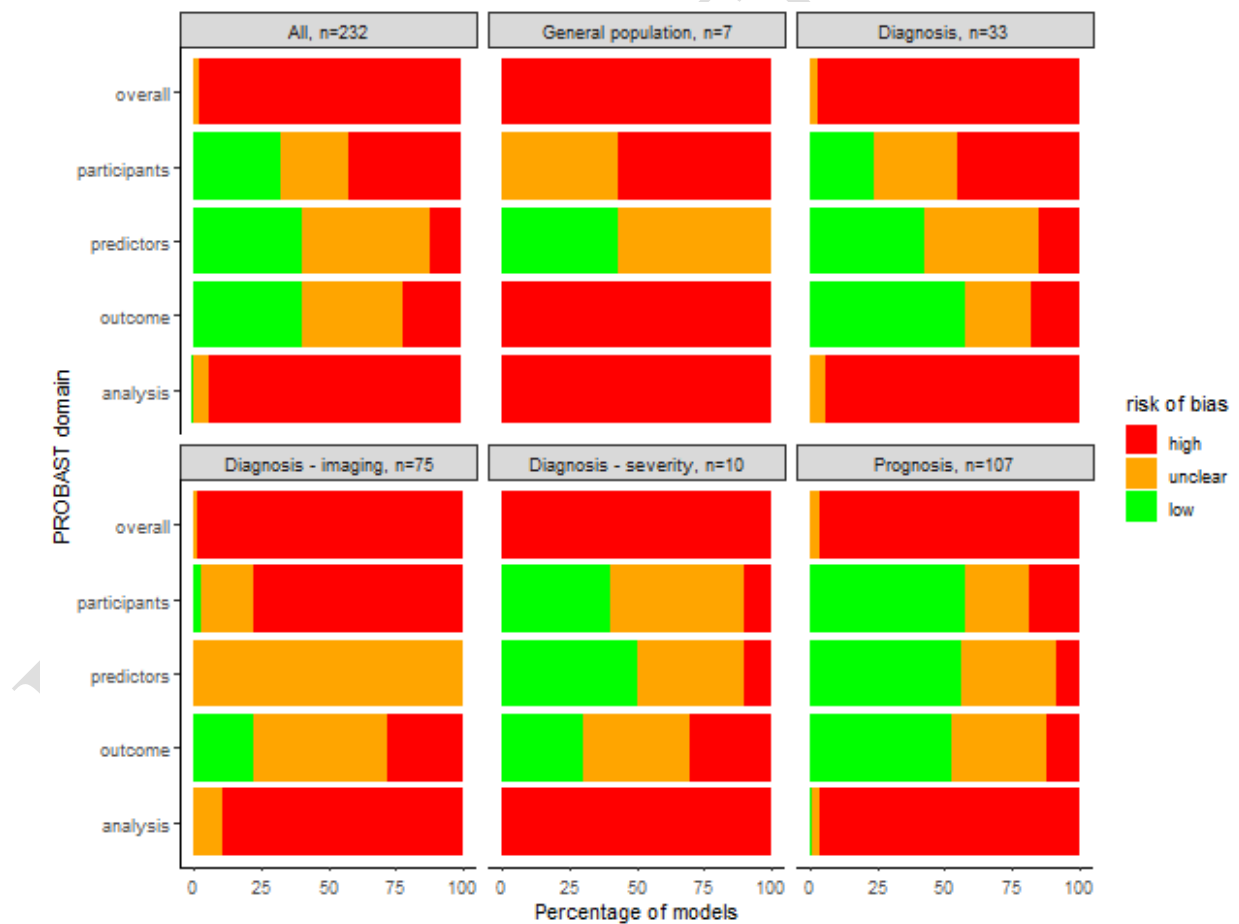
Box 2: Common causes of risk of bias in the reported prediction models
Models to predict coronavirus disease 2019 (covid-19) risk in general population
All of these models had unclear or high risk of bias for the participant, outcome, and analysis domain. All were based on proxy outcomes to predict covid-19 related risks, such as presence of or hospital admission due to severe respiratory disease, in the absence of data of patients with covid-19.[8 90 194]

Diagnostic models
Thirty percent used inappropriate data sources (eg, due to a non-nested case-control design), 27% used inappropriate in- or exclusion criteria such that the study data was not representative of the target population, and 24% selected controls that were not representative of the target population for a diagnostic model (eg, controls for a screening model had viral pneumonia). Other frequent problems were dichotomization of predictors (27%), and tests used to determine the outcome (24%) or predictor definitions or measurement procedures (21%) that varied between participants.

Diagnostic models based for severity classification
Twenty percent of models used predictor data that was assessed while the severity (the outcome) was known. Other concerns include non-standard or lack of a prespecified outcome definition (20%), predictor measurements (e.g., fever) being part of the outcome definition (20%) and outcomes being assessed with knowledge of predictor measurements (20%).

Diagnostic models based on medical imaging
Generally, studies did not clearly report which patients had imaging during clinical routine. Seventy three percent used an inappropriate or unclear study design to collect data (eg, a non-nested case-control). It was often unclear (52%) whether the selection of controls was made from the target population (that is, patients with suspected covid-19). Outcome definitions were often not defined or determined in the same way in all participants (24%). Diagnostic model studies that used medical images as predictors were all scored as unclear on the predictor domain. These publications often lacked clear information on the preprocessing steps (e.g., cropping of images). Moreover, complex machine learning algorithms transform images into predictors in a complex way, which makes it challenging to fully apply the PROBAST predictors section for such imaging studies. However, a more favourable assessment of the predictor domain does not lead to better overall judgement regarding risk of bias for the included models. Careful description of model specification and subsequent estimation were frequently lacking, challenging the transparency and reproducibility of the models. Studies used different deep learning architectures, some were established and others specifically designed, without benchmarking the used architecture against others.

Prognostic models
Dichotomization of predictors was a frequent concern (21%). Other problems include inappropriate in- or exclusions of study participants (18%). Study participants were often excluded because they did not develop the outcome at the end of the study period but were still in follow-up (that is, they were in hospital but had not recovered or died), yielding a selected study sample (11%). Additionally, many models (15%) did not account for censoring or competing risks.

Ninety eight models (42%) had a high risk of bias for the participants domain, which indicates that the participants enrolled in the studies might not be representative of the models' targeted populations. Unclear reporting on the inclusion of participants led to an unclear risk of bias assessment in 58 models (25%), and 76 (33%) had a low risk of bias for the participants domain. Fifteen models (6%) had a high risk of bias for the predictor domain, which indicates that predictors were not available at the models' intended time of use, not clearly defined, or influenced by the outcome measurement. One hundred and thirty five (58%) models were rated unclear and 82 (35%) rated at low risk of bias for the predictor domain. Most studies used outcomes that are easy to assess (eg, death, presence of covid-19 by laboratory confirmation), and hence 95 (41%) were rated at low risk of bias. Nonetheless, there was cause for concern about bias induced by the outcome measurement in 50 models (22%), for example due to the use of subjective or proxy outcomes (e.g., non-covid-19 severe respiratory infections). Eighty seven models (38%) had an unclear risk of bias due to opaque or ambiguous reporting.

Two hundred and eighteen (94%) models were at high risk of bias for the analysis domain. The reporting was insufficiently clear to assess risk of bias in the analysis in thirteen studies (6%). Only one model had a low risk of bias for the analysis domain (<1%). Twenty nine (13%) models had low risk of bias on all domains except analysis, indicating adequate data collection and study design but issues that could have been avoided by conducting a better statistical analysis. Many studies had small to modest sample sizes (see table 1), which led to an increased risk of overfitting, particularly if complex modelling strategies were used. In addition, 50 models (22%) were neither internally nor externally validated. Performance statistics calculated on the development data from these models are likely optimistic. Calibration was only assessed for 23 models using calibration plots (10%), of which 12 on external validation data.

Two models are worthy of note, that were generally of good quality, were built on large datstes, and rated low risk of bias on most domains but had an overall rating of unclear risk of bias, due to a unclear details on only one signalling question within the analysis domain, see Table 2 for a summary. Jehi and colleagues presented findings from developing a diagnostic model, however, there was substantial missing data and it remains unclear whether the use of median imputation influenced results, and there are unexplained discrepancies between the online calculator, nomogram, and published logistic regression model [184]. Hence, the calculator should not be used without further validation. Knight and colleagues developed a prognostic model for in-hospital mortality, however, continuous predictors were

dichotomized, which reduces granularity of predicted risks (even though the model had a C index comparable with that of a generalized additive model) [186]. The model was also converted into an sum score but it was unclear how the scores were translated to the predicted mortality risks that were used to evaluate calibration.

Table 2 Prediction models with unclear risk of bias overall and large development samples.

Study; setting; and outcome	Model	Sample size: total No of participants (No with outcome)*	Predictive performance		Overall risk of bias using PROBAST
			Strongest type of validation reported	**Performance (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))**	
Diagnostic models					
Jehi et al [184]; data from U.S.A., patients with suspected covid-19; covid-19 diagnosis	Jehi model	Development: 11672 (818) External validation: 2295 (290)	External validation, same country, new centres and later period	C index 0.84 (0.82 to 0.86)	Unclear
Prognostic models					
Knight et al [186]; data from U.K.; suspected or confirmed symptomatic inpatients; in-hospital mortality	4C Mortality Score	Development: 35463 (11426) Temporal validation: 22361 (6729)	Temporal validation	C index 0.77 (0.76 to 0.77)	Unclear

NPV=negative predictive value; PPV=positive predictive value; PROBAST=prediction model risk of bias assessment tool; * a large dataset according to PROBAST is at least 10 events per candidate variable (EPV) for model development and at least 100 events for validation. In case EPV could not be extracted or calculated from the study report, 100 events for model development was the lower bound to be included in this table.

External validation

Forty six models were developed and externally validated in the same study (in an independent dataset, excluding random training test splits and temporal splits). In addition, there were 24 external validations of models developed for covid-19 or before the covid-19 pandemic in separate studies. Unfortunately, none of the external validations were scored as low risk of bias, 3 were rated as unclear risk of bias, and 67 as high risk of bias. One common concern is that datasets used for the external validation were likely not representative of the target population (eg, not recruiting consecutive patients, using an inappropriate study design, using unrepresentative controls, excluding patients still in follow-up). Consequently, predictive performance could differ if the models are applied in the targeted population. Moreover, only 21% had 100 or more events for external validation, which is the recommended minumum.[137] [138] Only 16% of external validations presented a calibration plot.

Table 3 shows the results of external validations that had at most an unclear risk of bias and at least 100 events in the external validation set. The model by Jehi et al has been discussed above [184]. Luo and colleagues performed a validation of the CURB-65 score, originally developed to predict mortality of community-acquired pneumonia, to assess its abilty to predict in-hospital mortality in patients with confirmed covid-19. This was conducted in a large retrospective cohort of patients admitted to two Chinese designated hospitals to treat patients with SARS-CoV-2 pneumonia [198]. It was unclear whether all consecutive patients were included (although this is likely given the retrospective design), there was no calibration plot because the score gives an integer as output rather tgan estimates risks, and the score uses dichotomized predictors. Overall, the external validation by Luo et al was performed well. It should be noted that studies that validated CURB-65 in covid-19 patients obtained C indexes of 0.58, 0.74, 0.75, 0.84 and 0.88 [173] [191] [198] [207] [228]. These observed differences may be due to differences in risk of bias (all except Luo et al were high risk of bias), heterogeneity in study populations (South Korea, China, Turkey, and U.S.A.), outcome definitions (progression to severe covid-19 vs. mortality), and sampling variability (number of events 36, 55, 131, 201, unclear).

Table 3 External validations with unclear risk of bias and large validation samples.

Study; setting; and outcome	Model	Sample size: total No of participants for model validation set (No with outcome) *	Predictive performance		Overall risk of bias using PROBAST
			Type of validation	Performance* (C index, sensitivity (%), specificity (%), PPV/NPV (%), calibration slope, other (95% CI, if reported))	
Diagnostic models					
Jehi et al [184]; data from U.S.A., patients with suspected covid-19; covid-19 diagnosis	Jehi model	Development: 11672 (818) External validation: 2295 (290)	External validation, same country, new centres and later period	C index 0.84 (0.82 to 0.86)	Unclear
Prognostic models					
Luo et al[198]; data from China, inpatients with confirmed covid-19; in-hospital mortality	CURB-65	1018 (201)	Independent external validation	C index 0.84 (0.82 to 0 .93)	Unclear

PROBAST=prediction model risk of bias assessment tool; CURB-65= Confusion, Urea, Respiratory rate, Blood pressure plus age at least 65 years NPV=negative predictive value; PPV=positive predictive value; * a large dataset according to PROBAST is at least 10 events per candidate predictor for model development and at least 100 events for validation.

Discussion

In this systematic review of prediction models related to the covid-19 pandemic, we identified and critically appraised 232 models described in 169 studies. These prediction models can be divided into three categories: models for the general population to predict the risk of having covid-19 or being admitted to hospital for covid-19; models to support the diagnosis of covid-19 in patients with suspected infection; and models to support the prognostication of patients with covid-19. All models reported moderate to excellent predictive performance, but all were appraised to have high or uncertain risk of bias owing to a combination of poor reporting and poor methodological conduct for participant selection, predictor description, and statistical methods used. Models were developed on data from different countries, but the majority used data from a single country. Often, the available sample sizes and number of events for the outcomes of interest were limited. This is a well-known problem when building prediction models and increases the risk of overfitting the model.[139] A high risk of bias implies that the performance of these models in new samples will probably be worse than that reported by the researchers. Therefore, the estimated C indices, often close to 1 and indicating near perfect discrimination, are probably optimistic. The majority of studies developed new models specifically for covid-19, but only 45 carried out an external validation, and calibration was rarely assessed. We cannot yet recommend any of the identified prediction models for widespread use in clinical practice, although a few models originated from studies that were of better quality. We suggest that these models should be further validated in other data sets, and ideally by independent investigators[184 186].

Challenges and opportunities

The main aim of prediction models is to support medical decision making in individual patients. Therefore, it is vital to identify a target setting in which predictions serve a clinical need (eg emergency department, intensive care unit, general practice, symptom monitoring app in the general population), and a representative dataset from that setting (preferably comprising consecutive patients) on which the prediction model can be developed and validated. This clinical setting and patient characteristics should be described in detail (including timing within the disease course, the severity of disease at the moment of prediction, and the comorbidity), so that readers and clinicians are able to understand if the proposed model could be suited for their population. Unfortunately, the studies included in our systematic review often lacked an adequate description of the target setting and study population, which leaves users of these models in doubt about the models' applicability.

Although we recognise that the earlier studies were done under severe time constraints, we recommend that any studies currently in preprint and all future studies should adhere to the TRIPOD reporting guideline[16] to improve the description of their study population and guide their modelling choices. TRIPOD translations (eg, in Chinese and Japanese) are also available at https://www.tripod-statement.org.

A better description of the study population could also help us understand the observed variability in the reported outcomes across studies, such as covid-19 related mortality and covid-19 prevalence. The variability in mortality could be related to differences in included patients (eg age, comorbidities) and interventions for covid-19. The variability in prevalence could in part be reflective of different diagnostic standards across studies. Note that the majority of diagnostic models use viral nucleic acid test results as the gold standard, which may have unacceptable false negative rates.

Covid-19 prediction will often not present as a simple binary classification task. Complexities in the data should be handled appropriately. For example, a prediction horizon should be specified for prognostic outcomes (eg, 30 day mortality). If study participants have neither recovered nor died within that time period, their data should not be excluded from analysis, which some reviewed studies have done. Instead, an appropriate time to event analysis should be considered to allow for administrative censoring.[17] Censoring for other reasons, for instance because of quick recovery and loss to follow-up of patients who are no longer at risk of death from covid-19, could necessitate analysis in a competing risk framework.[140]

We reviewed 75 studies that used only medical images to diagnose covid-19, covid-19 related pneumonia, or to assist in segmentation of lung images, the majority using advanced machine learning methodology. The predictive performance measures showed a high to almost perfect ability to identify covid-19, although these models and their evaluations also had a high risk of bias, notably because of poor reporting and an artificial mix of patients with and without covid-19. Currently, none of these models are recommended to be used in clinical practice. An independent systematic review and critical appraisal (using PROBAST[16]) of machine learning models for covid-19 using chest radiographs and CT scans came to the same conclusions, even though they focused on models that met a minimum requirement of study quality based on specialised quality metrics for the assessment of radiomics and deep-learning-based diagnostic models in radiology.[229]

A prediction model applied in a new healthcare setting or country often produces predictions that are miscalibrated[141] and might need to be updated before it can safely be

applied in that new setting.[17] This requires data from patients with covid-19 to be available from that system. Instead of developing and updating predictions in their local setting, individual participant data from multiple countries and healthcare systems might allow better understanding of the generalisability and implementation of prediction models across different settings and populations. This approach could greatly improve the applicability and robustness of prediction models in routine care.[142-146]

The evidence base for the development and validation of prediction models related to covid-19 will continue to increase over the coming months. To leverage the full potential of these evolutions, international and interdisciplinary collaboration in terms of data acquisition, model building and validation is crucial.

Study limitations

With new publications on covid-19 related prediction models rapidly entering the medical literature, this systematic review cannot be viewed as an up-to-date list of all currently available covid-19 related prediction models. Also, 80 of the studies we reviewed were only available as preprints. These studies might improve after peer review, when they enter the official medical literature; we will reassess these peer reviewed publications in future updates. We also found other prediction models that are currently being used in clinical practice without scientific publications,[159] and web risk calculators launched for use while the scientific manuscript is still under review (and unavailable on request). These unpublished models naturally fall outside the scope of this review of the literature.[160] As we have argued extensively elsewhere,[161] transparent reporting that enables validation by independent researchers is key for predictive analytics, and clinical guidelines should only recommend publicly available and verifiable algorithms.

Implications for practice

All reviewed prediction models were found to have an unclear or high risk of bias, and evidence from independent external validations of the newly developed models is still scarce. However, the urgency of diagnostic and prognostic models to assist in quick and efficient triage of patients in the covid-19 pandemic might encourage clinicians and policymakers to prematurely implement prediction models without sufficient documentation and validation. Earlier studies have shown that models were of limited use in the context of a pandemic,[162] and they could even cause more harm than good.[163] Therefore, we cannot recommend any model for use in practice at this point.

The current oversupply of insufficiently validated models is not useful for clinical practice. Future studies should focus on validating, comparing, improving, and updating promising available prediction models.[17] The models by Knight and colleagues [186] and Jehi and colleagues [184] are good candidates for validation studies in other data. We urge Jehi and colleagues to make all model equations available for independent validation. [184] Such external validations should assess not only discrimination, but also calibration and clinical utility (net benefit),[141 146 163] in large datasets [137 138] collected using an appropriate study design. In addition, these models' transportability to other countries or settings remains to be investigated. Owing to differences between healthcare systems (eg, Chinese and European) and over time in when patients are admitted to and discharged from hospital, as well as the testing criteria for patients with suspected covid-19, we anticipate most existing models will be miscalibrated, but researchers could attempt to update and adjust the model to the local setting.

Most reviewed models used data from a hospital setting, but there is a paucity of models for primary care and the general population. Additional research is needed, including validation of any recently proposed models not yet included in the current update of the living review (eg Clift et al[230]). The models reviewed to date predicted the covid-19 diagnosis or assess the risk of mortality or deterioration, whereas long-term morbidity and functional outcomes remain understudied and may become a target outcome of interest in future studies developing prediction models.[231 232]

When creating a new prediction model, we recommend building on previous literature and expert opinion to select predictors, rather than selecting predictors in a purely data driven way.[17] This is especially important for datasets with limited sample size.[164] Frequently used predictors included in multiple models identified by our review are vital signs, age, comorbidities, and image features, and these should be considered when appropriate. Flu-like symptoms should be considered in diagnostic models, and sex, C reactive protein, and lymphocyte counts could be considered as prognostic factors.

By pointing to the most important methodological challenges and issues in design and reporting of the currently available models, we hope to have provided a useful starting point for further studies, which should preferably validate and update existing ones. This living systematic review has been conducted in collaboration with the Cochrane Prognosis Methods Group. We will update this review and appraisal continuously to provide up-to-date information for healthcare decision makers and professionals as more international research emerges over time.

Conclusion

Several diagnostic and prognostic models for covid-19 are currently available and they all report moderate to excellent discrimination. However, these models are all at high or unclear risk of bias, mainly because of model overfitting, inappropriate model evaluation (e.g. calibration ignored), use of inappropriate data sources and unclear reporting. Therefore, their performance estimates are probably optimistic and not representative for the target population. The COVID-PRECISE group does not recommend any of the current prediction models to be used in practice, but two models originated from higher quality studies and should be (independently) validated in other datasets. For details of the reviewed models see www.covprecise.org. Future studies aimed at developing and validating diagnostic or prognostic models for covid-19 should explicitly address the concerns raised. Sharing data and expertise for the validation and updating of covid-19 related prediction models is urgently needed.

We thank the authors who made their work available by posting it on public registries or sharing it confidentially. A preprint version of the study is publicly available on medRxiv.

The lead authors affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Dissemination to participants and related patient and public communities: The study protocol is available online at https://osf.io/ehc47/.

References

1   Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020:S1473-3099(20)30120-1. doi:10.1016/S1473-3099(20)30120-1. PubMed

2   Arabi YM, Murthy S, Webb S. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med 2020. doi:10.1007/s00134-020-05955-1. PubMed

3   Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA 2020. doi:10.1001/jama.2020.4031. PubMed

4   Xie J, Tong Z, Guan X, Du B, Qiu H, Slutsky AS. Critical care crisis and some recommendations during the COVID-19 epidemic in China. Intensive Care Med 2020. doi:10.1007/s00134-020-05979-7. PubMed

5   Wellcome Trust. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak 2020. https://wellcome.ac.uk/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-covid-19-outbreak.

6   Institute of Social and Preventive Medicine. Living evidence on COVID-19 2020. https://ispmbern.github.io/covid-19/living-review/index.html.

7   Xie J, Hungerford D, Chen H, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. medRxiv [Preprint] 2020. doi:10.1101/2020.03.28.20045997

8   DeCaprio D, Gartner J, Burgess T, et al. Building a COVID-19 vulnerability index. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200307347D.

9   Bai X, Fang C, Zhou Y, et al. Predicting COVID-19 malignant progression with AI techniques. medRxiv [Preprint] 2020. doi:10.1101/2020.03.20.20037325

10   Feng C, Huang Z, Wang L, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected covid-19 pneumonia in fever clinics. medRxiv [Preprint] 2020. doi:10.1101/2020.03.19.20039099

11   Jin C, Chen W, Cao Y, et al. Development and evaluation of an AI system for covid-19 diagnosis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.20.20039834

12   Meng Z, Wang M, Song H, et al. Development and utilization of an intelligent application for aiding COVID-19 diagnosis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.18.20035816

13   Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4.0: software for research synthesis [program]. EPPI-Centre Software. London: Social Science Research Unit, Institute of Education, University of London, 2010.

14   Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11:e1001744. doi:10.1371/journal.pmed.1001744. PubMed

15   Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1-33. doi:10.7326/M18-1377. PubMed

16   Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73. doi:10.7326/M14-0698. PubMed

17   Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer US, 2019 doi:10.1007/978-3-030-16399-0.

18   Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6:e1000100. doi:10.1371/journal.pmed.1000100. PubMed

19   Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv [Preprint] 2020. doi:10.1101/2020.02.24.20027268

20   Lu J, Hu S, Fan R, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. medRxiv [Preprint] 2020. doi:10.1101/2020.02.20.20025510

21   Qi X, Jiang Z, YU Q, et al. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv [Preprint] 2020. doi:10.1101/2020.02.29.20029603

22   Yan L, Zhang H-T, Xiao Y, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv [Preprint] 2020. doi:10.1101/2020.02.27.20028027

23   Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS One 2020;15:e0230548. doi:10.1371/journal.pone.0230548. PubMed

24   Song Y, Zheng S, Li L, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with CT images. medRxiv [Preprint] 2020. doi:10.1101/2020.02.23.20026930

25   Yu H, Shao J, Guo Y, et al. Data-driven discovery of clinical routes for severity detection in covid-19 pediatric cases. medRxiv [Preprint] 2020. doi:10.1101/2020.03.09.20032219

26   Gozes O, Frid-Adar M, Greenspan H, et al. Rapid AI development cycle for the coronavirus (covid-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200305037G

27   Chen J, Wu L, Zhang J, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv [Preprint] 2020. doi:10.1101/2020.02.25.20021568

28   Xu X, Jiang X, Ma C, et al. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200209334X

29   Shan F, Gao Y, Wang J, et al. Lung infection quantification of covid-19 in CT images with deep learning. arXiv e-prints 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200304655S

30   Wang S, Kang B, Ma J, et al. A deep learning algorithm using CT images to screen for corona virus disease (covid-19). medRxiv [Preprint] 2020. doi:10.1101/2020.02.14.20023028

31   Song C-Y, Xu J, He J-Q, et al. COVID-19 early warning score: a multi-parameter screening tool to identify highly suspected patients. medRxiv [Preprint] 2020. doi:10.1101/2020.03.05.20031906

32   Barstugan M, Ozkaya U, Ozturk S. Coronavirus (COVID-19) classification using CT images by machine learning methods. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200309424B

33   Jin S, Wang B, Xu H, et al. AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks. medRxiv [Preprint] 2020. doi:10.1101/2020.03.19.20039354

34   Li L, Qin L, Xu Z, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest CT. Radiology 2020:200905. doi:10.1148/radiol.2020200905. PubMed

35   Lopez-Rincon A, Tonda A, Mendoza-Maldonado L, et al. Accurate identification of SARS-CoV-2 from viral genome sequences using deep learning. bioRxiv [Preprint] 2020. doi:10.1101/2020.03.13.990242

36   Shi F, Xia L, Shan F, et al. Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200309860S

37   Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 2020;24:108. doi:10.1186/s13054-020-2833-7. PubMed

38   Zheng C, Deng X, Fu Q, et al. Deep learning-based detection for covid-19 from chest CT using weak label. medRxiv [Preprint] 2020. doi:10.1101/2020.03.12.20027185

39   Chowdhury MEH, Rahman T, Khandakar A, et al. Can AI help in screening Viral and COVID-19 pneumonia? arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200313145C.

40   Sun Y, Koh V, Marimuthu K, et al. Epidemiological and clinical predictors of covid-19. Clin Infect Dis 2020;ciaa322. doi:10.1093/cid/ciaa322. PubMed

41   Martin A, Nateqi J, Gruarin S, et al. An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. bioRxiv [Preprint] 2020. doi:10.1101/2020.03.25.008805

42   Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for covid-19 diagnostic and prognostic analysis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.24.20042317

43   Wang Z, Weng J, Li Z, et al. Development and validation of a diagnostic nomogram to predict covid-19 pneumonia. medRxiv [Preprint] 2020. doi:10.1101/2020.04.03.20052068

44   Sarkar J, Chakrabarti P. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with covid-19. medRxiv [Preprint] 2020. doi:10.1101/2020.03.25.20043331

45   Wu J, Zhang P, Zhang L, et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv [Preprint] 2020. doi:10.1101/2020.04.02.20051136

46   Zhou Y, Yang Z, Guo Y, et al. A new predictor of disease severity in patients with covid-19 in Wuhan, China. medRxiv [Preprint] 2020. doi:10.1101/2020.03.24.20042119

47   Abbas A, Abdelsamea M, Gaber M. Classification of covid-19 in chest x-ray images using DeTraC deep convolutional neural network. medRxiv [Preprint] 2020. doi:10.1101/2020.03.30.20047456

48   Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine, 2020. doi:10.1007/s13246-020-00865-4.

49   Bukhari SUK, Bukhari SSK, Syed A, et al. The diagnostic evaluation of Convolutional Neural Network (CNN) for the assessment of chest X-ray of patients infected with COVID-19. medRxiv [Preprint] 2020. doi:10.1101/2020.03.26.20044610

50   Chaganti S, Balachandran A, Chabin G, et al. Quantification of tomographic patterns associated with covid-19 from chest CT. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200401279C.

51   Fu M, Yi S-L, Zeng Y, et al. Deep learning-based recognizing covid-19 and other common infectious diseases of the lung by chest CT scan images. medRxiv [Preprint] 2020. doi:10.1101/2020.03.28.20046045

52   Gozes O, Frid-Adar M, Sagie N, et al. Coronavirus detection and analysis on chest CT with deep learning. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200402640G.

53   Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for covid-19 from cough samples via an app. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200401275I.

54   Li K, Fang Y, Li W, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). Eur Radiol 2020; doi:10.1007/s00330-020-06817-6. PubMed

55   Li X, Li C, Zhu D. COVID-MobileXpert: on-device covid-19 screening using snapshots of chest x-ray. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200403042L.

56   Hassanien AE, Mahdy LN, Ezzat KA, et al. Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine. medRxiv [Preprint] 2020. doi:10.1101/2020.03.30.20047787

57   Tang Z, Zhao W, Xie X, et al. Severity assessment of coronavirus disease 2019 (covid-19) using quantitative features from chest CT images. arXiv e-prints [Preprint] 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200311988T.

58   Zhang J, Xie Y, Li Y, et al. COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection. arXiv e-prints 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200312338Z.

59   Zhou M, Chen Y, Wang D, et al. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. medRxiv [Preprint] 2020. doi:10.1101/2020.03.24.20043117

60   Huang H, Cai S, Li Y, et al. Prognostic factors for COVID-19 pneumonia progression to severe symptom based on the earlier clinical features: a retrospective analysis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.28.20045989

61   Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with covid-19 using artificial intelligence to help medical decision-making. medRxiv [Preprint] 2020. doi:10.1101/2020.03.30.20047308

62   Zeng L, Li J, Liao M, et al. Risk assessment of progression to severe conditions for patients with COVID-19 pneumonia: a single-center retrospective study. medRxiv [Preprint] 2020. doi:10.1101/2020.03.25.20043166

63   Al-Najjar H, Al-Rousan N. A classifier prediction model to predict the status of coronavirus covid-19 patients in South Korea. Eur Rev Med Pharmacol Sci 2020;24:3400-3. PubMed doi:10.26355/eurrev_202003_20709

64   Angelov P, Soares E. Explainable-by-design approach for covid-19 classification via CT-scan. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20078584.

65   Arpan M, Surya K, Harish R, et al. CovidAID: COVID-19 Detection Using Chest X-Ray. ArXiv e-prints [Preprint] 2020

66   Bai HX, Wang R, Xiong Z, et al. AI augmentation of radiologist performance in distinguishing covid-19 from pneumonia of other etiology on chest CT. Radiology 2020;201491. doi:10.1148/radiol.2020201491. PubMed

67   Barda N, Riesel D, Akriv A, et al. Performing risk stratification for COVID-19 when individual level data is not available, the experience of a large healthcare organization. medRxiv [Preprint] 2020 doi:10.1101/2020.04.23.20076976.

68   Bassi PRAS, Attux R. A deep convolutional neural network for covid-19 detection using chest x-rays. ArXiv e-prints [Preprint] 2020

69   Batista AFdM. Miraglia JL, Donato THR, et al. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv [Preprint] 2020 doi:10.1101/2020.04.04.20052092.

70   Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, et al. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. J Clin Endocrinol Metab 2020 doi: 10.1210/clinem/dgaa346

71   Benchoufi M, Bokobza J, Anthony C, et al. Lung injury in patients with or suspected COVID-19 : a comparison between lung ultrasound and chest CT-scanner severity assessments, an observational study. MedRxiv [Preprint] 2020 doi:10.1101/2020.04.24.20069633.

72   Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiol Med 2020;125:509-13. doi:10.1007/s11547-020-01200-3. PubMed

73   Born J, Brandle G, Cossio M, et al. POCOVID-Net: Automatic detection of covid-19 from a new lung ultrasound imaging dataset (POCUS). ArXiv e-prints [Preprint] 2020.

74   Brinati D, Campagner A, Ferrari D, et al. Detection of covid-19 infection from routine blood exams with machine learning: a feasibility study. medRxiv [Preprint] 2020. doi:10.1101/2020.04.22.20075143.

75   Carr E, Bendayan R, O'Gallagher K, et al. Supplementing the National Early Warning Score (NEWS2) for anticipating early deterioration among patients with covid-19 infection. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20078006

76   Castiglioni I, Ippolito D, Interlenghi M, et al. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy. medRxiv [Preprint] 2020. doi:10.1101/2020.04.08.20040907.

77   Chassagnon G, Vakalopoulou M, Battistella E, et al. AI-driven CT-based quantification, staging and short-term outcome prediction of covid-19 pneumonia. medRxiv [Preprint] 2020. doi:10.1101/2020.04.17.20069187.

78   Chen X, Tang Y, Mo Y, et al. A diagnostic model for coronavirus disease 2019 (COVID-19) based on radiological semantic and clinical features: a multi-center study. Eur Radiol 2020. doi:10.1007/s00330-020-06829-2. PubMed

79   Colombi D, Bodini FC, Petrini M, et al. Well-aerated lung on admitting chest CT to predict adverse outcome in covid-19 pneumonia. Radiology 2020;201433. doi:10.1148/radiol.2020201433. PubMed

80   Das A, Mishra S, Gopalan SS. Predicting community mortality risk due to CoVID-19 using machine learning and development of a prediction tool. medRxiv [Preprint] 2020. doi:10.1101/2020.04.27.20081794.

81   Diaz-Quijano FA, Silva JMNd, Ganem F, et al. A model to predict SARS-CoV-2 infection based on the first three-month surveillance data in Brazil. medRxiv [Preprint] 2020. doi:10.1101/2020.04.05.20047944.

82   Guiot J, Vaidyanathan A, Deprez L, et al. Development and validation of an automated radiomic CT signature for detecting covid-19. medRxiv [Preprint] 2020. doi:10.1101/2020.04.28.20082966.

83   Guo Y, Liu Y, Lu J, et al. Development and validation of an early warning score (EWAS) for predicting clinical deterioration in patients with coronavirus disease 2019. medRxiv [Preprint] 2020. doi:10.1101/2020.04.17.20064691

84   Hu C, Liu Z, Jiang Y, et al. Early prediction of mortality risk among severe covid-19 patients using machine learning. medRxiv [Preprint] 2020. doi:10.1101/2020.04.13.20064329.

85   Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critical ill patients with novel coronavirus disease. Acad Emerg Med 2020;27:461-8. doi:10.1111/acem.13992. PubMed

86   Hu R, Ruan G, Xiang S, et al. Automated diagnosis of covid-19 using deep learning and data augmentation on chest CT. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20078998.

87   Islam MT, Fleischer JW. Distinguishing L and H phenotypes of covid-19 using a single x-ray image. medRxiv [Preprint] 2020. doi:10.1101/2020.04.27.20081984.

88   Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with covid-19 pneumonia: the CALL score. Clin Infect Dis 2020;ciaa414. doi:10.1093/cid/ciaa414. PubMed

89   Jiang X, Coffee M, Bari A, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Computers. Materials & Continua 2020;63:537-51 doi:10.32604/cmc.2020.010691.

90   Jiang Z, Hu M, Fan L, et al. Combining visible light and infrared imaging for efficient detection of respiratory infections such as covid-19 on portable device. ArXiv e-prints [Preprint] 2020.

91   Kana GEB, Kana ZMG, Kana DAF, et al. A web-based diagnostic tool for covid-19 using machine learning on chest radiographs (CXR). medRxiv [Preprint] 2020. doi:10.1101/2020.04.21.20063263.

92 Rezaul KM, Döhmen T, Rebholz-Schuhmann D, et al. DeepCOVIDExplainer: explainable covid-19 predictions based on chest x-ray images. ArXiv e-prints [Preprint] 2020.

93   Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed 2020;196:105581. doi: 10.1016/j.cmpb.2020.105581

94   Kumar R, Arora R, Bansal V, et al. Accurate prediction of covid-19 using chest x-ray images through deep feature learning model with SMOTE and machine learning classifiers. medRxiv [Preprint] 2020. doi:10.1101/2020.04.13.20063461.

95   Kurstjens S, van der Horst A, Herpers R, et al. Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. medRxiv [Preprint] 2020. doi:10.1101/2020.04.20.20067512

96   Levy TJ, Richardson S, Coppa K, et al. Estimating survival of hospitalized covid-19 patients from admission information. medRxiv [Preprint] 2020. doi:10.1101/2020.04.22.20075416.

97   Li Z, Zhong Z, Li Y, et al. From community acquired pneumonia to covid-19: a deep learning based method for quantitative analysis of covid-19 on thick-section CT scans. medRxiv [Preprint] 2020. doi:10.1101/2020.04.17.20070219.

98   Liu Q, Fang X, Tokuno S, et al. Prediction of the clinical outcome of COVID-19 patients using T lymphocyte subsets with 340 cases from Wuhan, China: a retrospective cohort study and a web visualization tool. medRxiv [Preprint] 2020. doi:10.1101/2020.04.06.20056127.

99   Lyu P, Liu X, Zhang R, Shi L, Gao J. The performance of chest CT in evaluating the clinical severity of COVID-19 pneumonia: identifying critical cases based on CT characteristics. Invest Radiol 2020;55:412-21. doi:10.1097/RLI.0000000000000689. PubMed

100 McRae MP, Simmons GW, Christodoulides NJ, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with COVID-19. Lab Chip 2020 doi: 10.1039/d0lc00373e

101   Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 doi: 10.1038/s41591-020-0931-3

102   Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 doi: 10.1038/s41591-020-0916-2

103   Moutounet-Cartan PGB. Deep convolutional neural networks to diagnose covid-19 and other pneumonia diseases from posteroanterior chest x-rays. ArXiv e-prints [Preprint] 2020

104   Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792. doi:10.1016/j.compbiomed.2020.103792. PubMed

105   Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Inform Med Unlocked 2020;19 doi: 10.1016/j.imu.2020.100360

106   Rehman A, Naz S, Khan A, et al. Improving coronavirus (covid-19) diagnosis using deep transfer learning. medRxiv [Preprint] 2020. doi:10.1101/2020.04.11.20054643.

107   Singh D, Kumar V, Vaishali NA, Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. Eur J Clin Microbiol Infect Dis 2020;39:1379-89. doi:10.1007/s10096-020-03901-z. PubMed

108   Singh K, Valley TS, Tang S, et al. Validating a widely implemented deterioration index model among hospitalized covid-19 patients. medRxiv [Preprint] 2020. doi:10.1101/2020.04.24.20079012.

109   Soares F, Villavicencio A, Anzanello MJ, et al. A novel high specificity COVID-19 screening method based on simple blood exams and artificial intelligence. medRxiv [Preprint] 2020. doi:10.1101/2020.04.10.20061036.

110   Tordjman M, Mekki A, Mali RD, et al. Pre-test probability for SARS-Cov-2-related Infection Score: the PARIS score. medRxiv [Preprint] 2020. doi:10.1101/2020.04.28.20081687.

111   Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses 2020;140:109761. doi:10.1016/j.mehy.2020.109761. PubMed

112   Vaid A, Somani S, Russak AJ, et al. Machine learning to predict mortality and critical events in covid-19 positive New York City patients. medRxiv [Preprint] 2020. doi:10.1101/2020.04.26.20073411.

113   Vazquez Guillamet C, Vazquez Guillamet R, Kramer AA, et al. Toward a covid-19 score-risk assessments and registry. medRxiv [Preprint] 2020. doi:10.1101/2020.04.15.20066860.

114   Wang c, Deng R, Gou L, et al. Preliminary study to identify severe from moderate cases of COVID-19 using NLR&RDW-SD combination parameter. medRxiv [Preprint] 2020. doi:10.1101/2020.04.09.20058594.

115   Wu Y-H, Gao S-H, Mei J, et al. JCS: an explainable covid-19 diagnosis system by joint classification and segmentation. ArXiv e-prints [Preprint] 2020.

116   Zhang H, Shi T, Wu X, et al. Risk prediction for poor outcome and death in hospital in-patients with COVID-19: derivation in Wuhan, China and external validation in London. medRxiv [Preprint] 2020. doi:10.1101/2020.04.28.20082222.

117   Zhao B, Wei Y, Sun W, et al. Distinguish coronavirus disease 2019 patients in general surgery emergency by CIAAD scale: development and validation of a prediction model based on 822 cases in China. medRxiv [Preprint] 2020. doi:10.1101/2020.04.18.20071019.

118   Zhu Z, Cai T, Fan L, et al. Clinical value of immune-inflammatory parameters to assess the severity of coronavirus disease 2019. Int J Infect Dis 2020;95:332-9. doi:10.1016/j.ijid.2020.04.041. PubMed

119   Gong J, Ou J, Qiu X, et al. A tool to early predict severe corona virus disease 2019 (covid-19) : a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020;ciaa443. doi:10.1093/cid/ciaa443. PubMed

120   Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv e-prints [Preprint] 2020. https://github.com/ieee8023/covid-chestxray-dataset.

121   Kaggle. COVID-19 Kaggle community contributions 2020. https://www.kaggle.com/covid-19-contributions.

137   Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016;35:214-26. doi:10.1002/sim.6787. PubMed

138   Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58:475-83. doi:10.1016/j.jclinepi.2004.06.017. PubMed

139   Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441. doi:10.1136/bmj.m441. PubMed

140   Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation 2016;133:601-9. doi:10.1161/CIRCULATIONAHA.115.017719. PubMed

141   Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230. PubMed doi:10.1186/s12916-019-1466-7

142   Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges [correction: BMJ 2019;365:l4379]. BMJ 2016;353:i3140. doi:10.1136/bmj.i3140. PubMed

143   Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG; Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. PLoS Med 2015;12:e1001886. doi:10.1371/journal.pmed.1001886. PubMed

144   Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245-7. doi:10.1016/j.jclinepi.2015.04.005. PubMed

145   Wynants L, Kent DM, Timmerman D, Lundquist CM, Van Calster B. Untapped potential of multicenter studies: a review of cardiovascular risk prediction models revealed inappropriate analyses and wide variation in reporting. Diagn Progn Res 2019;3:6. doi:10.1186/s41512-019-0046-9. PubMed

146   Wynants L, Riley RD, Timmerman D, Van Calster B. Random-effects meta-analysis of the clinical utility of tests and prediction models. Stat Med 2018;37:2034-52. doi:10.1002/sim.7653. PubMed

147   Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395:1054-62. doi:10.1016/S0140-6736(20)30566-3. PubMed

148   Li K, Wu J, Wu F, et al. The clinical and chest CT features associated with severe and critical covid-19 pneumonia. Invest Radiol 2020. doi:10.1097/RLI.0000000000000672. PubMed

149   Li B, Yang J, Zhao F, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol 2020. doi:10.1007/s00392-020-01626-9. PubMed

150   Jain V, Yuan J-M. Systematic review and meta-analysis of predictive symptoms and comorbidities for severe COVID-19 infection. medRxiv [Preprint] 2020. doi:10.1101/2020.03.15.20035360

151   Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al; Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). https://www.lancovid.org. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 2020:101623. doi:10.1016/j.tmaid.2020.101623. PubMed

152   Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. Clin Chim Acta 2020;506:145-8. doi:10.1016/j.cca.2020.03.022. PubMed

153   Zhao X, Zhang B, Li P, et al. Incidence, clinical characteristics and prognostic factor of patients with covid-19: a systematic review and meta-analysis. medRxiv [Preprint] 2020. doi:10.1101/2020.03.17.20037572

154   Johansson MA, Saderi D. Open peer-review platform for COVID-19 preprints. Nature 2020;579:29. PubMed doi:10.1038/d41586-020-00613-4

155   Xu B, Kraemer MU, Gutierrez B, et al. Open access epidemiological data from the COVID-19 outbreak. Lancet Infect Dis 2020. doi:10.1016/s1473-3099(20)30119-5

159   Infervision. Infervision launches hashtag#AI-based hashtag#Covid-19 solution in Europe 2020. https://www.linkedin.com/posts/infervision_ai-covid-medicine-activity-6650772755031613440-TqLJ.

160   Surgisphere Corporation. COVID-19 response center 2020. https://surgisphere.com/covid-19-response-center/.

161   Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? J Am Med Inform Assoc 2019;26:1651-4. doi:10.1093/jamia/ocz130. PubMed

162   Enfield K, Miller R, Rice T, et al. Limited utility of SOFA and APACHE II prediction models for ICU triage in pandemic Influenza. Chest 2011;140:913A. doi:10.1378/chest.1118087.

163   Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making 2015;35:162-9. doi:10.1177/0272989X14547233. PubMed

164   van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. Stat Methods Med Res 2019;28:2455-74. doi:10.1177/0962280218784726. PubMed

165. Apostolopoulos ID, Aznaouridis SI, Tzani MA. Extracting Possibly Representative COVID-19 Biomarkers from X-ray Images with Deep Learning Approach and Image Data Related to Pulmonary Diseases. J Med Biol Eng 2020:01/Aug. doi: 10.1007/s40846-020-00529-4

166. Ardakani AA, Kanafi AR, Acharya UR, et al. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med 2020;121 doi: 10.1016/j.compbiomed.2020.103795

167. Bar S, Lecourtois A, Diouf M, et al. The association of lung ultrasound images with COVID-19 infection in an emergency room cohort. Anaesthesia 2020 doi: 10.1111/anae.15175

168. Bi X, Su Z, Yan H, et al. Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count. Platelets 2020:01/Jun. doi: 10.1080/09537104.2020.1760230

169. Borghesi A, Zigliani A, Golemi S, et al. Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy. Int J Infect Dis 2020 doi: 10.1016/j.ijid.2020.05.021

170. Burian E, Jungmann F, Kaissis GA, et al. Intensive Care Risk Estimation in COVID-19 Pneumonia Based on Clinical and Imaging Parameters: Experiences from the Munich Cohort. Journal of clinical medicine 2020;9(5) doi: 10.3390/jcm9051514

171. Cecconi M, Piovani D, Brunetta E, et al. Early Predictors of Clinical Deterioration in a Cohort of 239 Patients Hospitalized for Covid-19 Infection in Lombardy, Italy. Journal of clinical medicine 2020;9(5) doi: 10.3390/jcm9051548

172. Cheng FY, Joshi H, Tandon P, et al. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. Journal of clinical medicine 2020;9(6) doi: 10.3390/jcm9061668

173. Choi MH, Ahn H, Ryu HS, et al. Clinical Characteristics and Disease Progression in Early-Stage COVID-19 Patients in South Korea. Journal of clinical medicine 2020;9(6) doi: 10.3390/jcm9061959

174. Clemency BM, Varughese R, Scheafer DK, et al. Symptom Criteria for COVID-19 Testing of Heath Care Workers. Acad Emerg Med 2020 doi: 10.1111/acem.14009

175. Dong Y, Zhou H, Li M, et al. A novel simple scoring model for predicting severity of patients with SARS-CoV-2 infection. Transbound Emerg Dis 2020 doi: 10.1111/tbed.13651

176. Elasnaoui K, Chawki Y. Using X-ray Images and Deep Learning for Automated Detection of Coronavirus Disease. J Biomol Struct Dyn 2020:Jan/22. doi: 10.1080/07391102.2020.1767212

177. Fu L, Li Y, Cheng A, et al. A Novel Machine Learning-derived Radiomic Signature of the Whole Lung Differentiates Stable From Progressive COVID-19 Infection: A Retrospective Cohort Study. J Thorac Imaging 2020 doi: 10.1097/RTI.0000000000000544

178. Galloway JB, Norton S, Barker RD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: an observational cohort study. J Infect 2020 doi: 10.1016/j.jinf.2020.05.064

179. Gezer NS, Ergan B, Barış MM, et al. COVID-19 S: A new proposal for diagnosis and structured reporting of COVID-19 on computed tomography imaging. Diagn Interv Radiol 2020 doi: 10.5152/dir.2020.20351

180. Gidari A, De S, G V, et al. Predictive value of National Early Warning Score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection. Infect Dis (Lond) 2020:01/Jul. doi: 10.1080/23744235.2020.1784457

181. Hong Y, Wu X, Qu J, et al. Clinical characteristics of Coronavirus Disease 2019 and development of a prediction model for prolonged hospital length of stay. Ann Transl Med 2020;8(7):443. doi: 10.21037/atm.2020.03.147

182. Huang D, Wang T, Chen Z, et al. A novel risk score to predict diagnosis with Coronavirus Disease 2019 (COVID-19) in suspected patients: A retrospective, multi-center, observational study. J Med Virol 2020 doi: 10.1002/jmv.26143

183. Huang J, Cheng A, Lin S, et al. Individualized prediction nomograms for disease progression in mild COVID-19. J Med Virol 2020 doi: 10.1002/jmv.25969

184. Jehi L, Ji X, Milinovich A, et al. Individualizing risk prediction for positive COVID-19 testing: results from 11,672 patients. Chest 2020 doi: 10.1016/j.chest.2020.05.580

185. Joshi RP, Pejaver V, Hammarlund NE, et al. A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. J Clin Virol 2020;129:104502. doi: 10.1016/j.jcv.2020.104502

186. Knight SR, Ho A, Pius R, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 2020;370:m3339. doi: 10.1136/bmj.m3339

187. Ko H, Chung H, Kim KW, et al. COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image. J Med Internet Res 2020 doi: 10.2196/19569

188. Li Q, Zhang J, Ling Y, et al. A simple algorithm helps early identification of SARS-CoV-2 infection patients with severe progression tendency. Infection 2020 doi: 10.1007/s15010-020-01446-z

189. Li Y, Yang Z, Ai T, et al. Association of 'initial CT' findings with mortality in older patients with coronavirus disease 2019 (COVID-19). Eur Radiol 2020 doi: 10.1007/s00330-020-06969-5

190. Li Z, Zeng B, Lei P, et al. Differentiating pneumonia with and without COVID-19 using chest CT images: from qualitative to quantitative. J Xray Sci Technol 2020 doi: 10.3233/XST-200689

191. Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA internal medicine 2020 doi: 10.1001/jamainternmed.2020.2033

192. Liu F, Zhang Q, Huang C, et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. Theranostics 2020;10(12):5613-22. doi: 10.7150/thno.45985

193. Liu X, Shi S, Xiao J, et al. Prediction of the severity of Corona Virus Disease 2019 and its adverse clinical outcomes. Jpn J Infect Dis 2020 doi: 10.7883/yoken.JJID.2020.194

194. Liu Y, Wang Z, Tian Y, et al. Design and development of COVID-19 risk assessment decision support system for general practitioners. J Med Internet Res 2020 doi: 10.2196/19786

195. Liu YP, Li GM, He J, et al. Combined use of the neutrophil-to-lymphocyte ratio and CRP to predict 7-day disease severity in 84 hospitalized patients with COVID-19 pneumonia: A retrospective cohort study. Ann Transl Med 2020;8(10) doi: 10.21037/atm-20-2372

196. Lorente-Ros A, Monteagudo R, J M, et al. Myocardial injury determination improves risk stratification and predicts mortality in COVID-19 patients. Cardiol J 2020 doi: 10.5603/CJ.a2020.0089

197. Luo L, Luo Z, Jia Y, et al. CT differential diagnosis of COVID-19 and non-COVID-19 in symptomatic suspects: a practical scoring method. BMC Pulm Med 2020;20(1):129. doi: 10.1186/s12890-020-1170-6

198. Luo M, Liu J, Jiang W, et al. IL-6 combined with CD8+ T cell count early predict in-hospital mortality for patients with COVID-19. JCI Insight 2020 doi: 10.1172/jci.insight.139024

199. Luo Y, Yuan X, Xue Y, et al. Using the diagnostic model based on routine laboratory tests to distinguish patients infected with SARS-CoV-2 from those infected with influenza virus. Int J Infect Dis 2020 doi: 10.1016/j.ijid.2020.04.078

200. Matos J, Paparo F, Mussetto I, et al. Evaluation of novel coronavirus disease (COVID-19) using quantitative lung CT and clinical data: prediction of short-term outcome. Eur Radiol Exp 2020;4(1):39. doi: 10.1186/s41747-020-00167-0

201. Mazzaccaro D, Giacomazzi F, Giannetta M, et al. Non-Overt Coagulopathy in Non-ICU Patients with Mild to Moderate COVID-19 Pneumonia. Journal of clinical medicine 2020;9(6) doi: 10.3390/jcm9061781

202. Murphy K, Smits H, Knoops AJG, et al. COVID-19 on the Chest Radiograph: A Multi-Reader Evaluation of an AI System. Radiology 2020:201874. doi: 10.1148/radiol.2020201874

203. Obeid JS, Davis M, Turner M, et al. An AI approach to COVID-19 infection risk assessment in virtual visits: a case report. J Am Med Inform Assoc 2020 doi: 10.1093/jamia/ocaa105

204. Pu J, Leader J, Bandos A, et al. Any unique image biomarkers associated with COVID-19? Eur Radiol 2020 doi: 10.1007/s00330-020-06956-w

205. Rajaraman S, Antani S. Weakly Labeled Data Augmentation for Deep Learning: A Study on COVID-19 Detection in Chest X-Rays. Diagnostics (Basel, Switzerland) 2020;10(6) doi: 10.3390/diagnostics10060358

206. Roland LT, Gurrola JG, Loftus PA, et al. Smell and taste symptom-based predictive model for COVID-19 diagnosis. Int Forum Allergy Rhinol 2020 doi: 10.1002/alr.22602

207. Satici C, Demirkol MA, Altunok ES, et al. Performance of Pneumonia Severity Index and CURB-65 in Predicting 30-day mortality in patients with COVID-19. Int J Infect Dis 2020 doi: 10.1016/j.ijid.2020.06.038

208. Song J, Wang H, Liu Y, et al. End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT. Eur J Nucl Med Mol Imaging 2020 doi: 10.1007/s00259-020-04929-1

209. Sun L, Song F, Shi N, et al. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. J Clin Virol 2020:104431. doi: 10.1016/j.jcv.2020.104431

210. Toraih EA, Elshazli RM, Hussein MH, et al. Association of cardiac biomarkers and comorbidities with increased mortality, severity, and cardiac injury in COVID-19 patients: A meta-regression and Decision tree analysis. J Med Virol 2020 doi: 10.1002/jmv.26166

211. Tuncer T, Dogan S, Ozyurt F. An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based corona detection method using lung X-ray image. Chemometr Intell Lab Syst 2020:104054. doi: 10.1016/j.chemolab.2020.104054

212. Vaid S, Kalantar R, Bhandari M. Deep learning COVID-19 detection bias: accuracy through artificial intelligence. Int Orthop 2020 doi: 10.1007/s00264-020-04609-7

213. Vultaggio A, Vivarelli E, Virgili G, et al. Prompt predicting of early clinical deterioration of moderate-to-severe COVID-19 patients: usefulness of a combined score using IL-6 in a preliminary study. J Allergy Clin Immunol Pract 2020 doi: 10.1016/j.jaip.2020.06.013

214. Wang F, Hou H, Wang T, et al. Establishing a model for predicting the outcome of COVID-19 based on combination of laboratory tests. Travel Med Infect Dis 2020:101782. doi: 10.1016/j.tmaid.2020.101782

215. Wang K, Zuo P, Liu Y, et al. Clinical and laboratory predictors of in-hospital mortality in patients with COVID-19: a cohort study in Wuhan, China. Clin Infect Dis 2020 doi: 10.1093/cid/ciaa538

216. Wang L, Liu Y, Zhang T, et al. Differentiating Between 2019 Novel Coronavirus Pneumonia and Influenza Using a Nonspecific Laboratory Marker-Based Dynamic Nomogram. Open Forum Infect Dis 2020;7(5):ofaa169. doi: 10.1093/ofid/ofaa169

217. Wu S, Du Z, Shen S, et al. Identification and validation of a novel clinical signature to predict the prognosis in confirmed COVID-19 patients. Clin Infect Dis 2020 doi: 10.1093/cid/ciaa793

218. Wu X, Hui H, Niu M, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. Eur J Radiol 2020;128:109041. doi: 10.1016/j.ejrad.2020.109041

219. Yang P, Wang P, Song Y, et al. A retrospective study on the epidemiological characteristics and establishment of early warning system of severe COVID-19 patients. J Med Virol 2020 doi: 10.1002/jmv.26022

220. Yang Y, Shen C, Li J, et al. Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19. J Allergy Clin Immunol 2020 doi: 10.1016/j.jaci.2020.04.027

221. Yu C, Lei Q, Li W, et al. Clinical Characteristics, Associated Factors, and Predicting COVID-19 Mortality Risk: A Retrospective Study in Wuhan, China. Am J Prev Med 2020 doi: 10.1016/j.amepre.2020.05.002

222. Zhang C, Qin L, Li K, et al. A Novel Scoring System for Prediction of Disease Severity in COVID-19. Front Cell Infect Microbiol 2020;10:318. doi: 10.3389/fcimb.2020.00318

223. Zhang K, Liu X, Shen J, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. Cell 2020 doi: 10.1016/j.cell.2020.04.045

224. Zheng QN, Xu MY, Zheng YL, et al. Prediction of the rehabilitation duration and risk management for mild-moderate COVID-19. Disaster Med Public Health Prep 2020:Jan/27. doi: 10.1017/dmp.2020.214

225. Zhou Y, He Y, Yang H, et al. Development and validation a nomogram for predicting the risk of severe COVID-19: A multi-center study in Sichuan, China. PLoS One 2020;15(5):e0233328. doi: 10.1371/journal.pone.0233328

226. Zou X, Li S, Fang M, et al. Acute Physiology and Chronic Health Evaluation II Score as a Predictor of Hospital Mortality in Patients of Coronavirus Disease 2019. Crit Care Med 2020 doi: 10.1097/CCM.0000000000004411

227. Wang S, Zha Y, Li W, et al. A Fully Automatic Deep Learning System for COVID-19 Diagnostic and Prognostic Analysis. Eur Respir J 2020 doi: 10.1183/13993003.00775-2020

228. Levy TJ, Richardson S, Coppa K, et al. Estimating Survival of Hospitalized COVID-19 Patients from Admission Information. MedRxiv 2020 doi: 10.1101/2020.04.22.20075416

229. Roberts M, Driggs D, Thorpe M, et al. Machine learning for COVID-19 detection and prognostication using chest radiographs and CT scans: a systematic methodological review. arXiv preprint arXiv:200806388 2020

230. Clift AK, Coupland CAC, Keogh RH, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ 2020;371:m3731. doi: 10.1136/bmj.m3731

231. Mahase E. Covid-19: What do we know about "long covid"? BMJ 2020;370:m2815. doi: 10.1136/bmj.m2815

232. Klok FA, Boon GJAM, Barco S, et al. The Post-COVID-19 Functional Status scale: a tool to measure functional status over time after COVID-19. The European respiratory journal 2020;56(1):2001494. doi: 10.1183/13993003.01494-2020

Supplementary material

Systematic review and critical appraisal of prediction models for diagnosis and prognosis of covid-19 infection

Contents:

Search and screening

Data extraction

Supplementary Table 1. Overview of search dates, hits and inclusions

Supplementary material: search and screening

Search strings

arXiv: ncov or corona or Wuhan or COVID

Search strings for bioRxiv, medRxiv, PubMed, EMBASE are given on ispmbern.github.io/covid-19/living-review/collectingdata.html

Validation

We validated the publicly available systematic review list to examine whether it is fit for purpose by comparing it to relevant hits from bioRxiv and medRxiv when combining covid-19 search terms (covid-19, sars-cov-2, novel corona, 2019-ncov) with methodological search terms (diagnostic, prognostic, prediction model, machine learning, artificial intelligence, algorithm, score, deep learning, regression) in March 2020. All relevant hits were found on the living systematic review list. We supplemented this list with hits from PubMed by searching for "covid-19" because when we performed our initial search this term was not included in the reported living systematic review search terms for PubMed.

AI for initial screening

As the amount of covid-19 literature increased exponentially over time, a new approach was needed to keep up with the literature. The results from manual screening of title and abstracts followed by full text review for final inclusion for the records retrieved up to 7 April were used to build a classification model for covid-19 prediction model studies, using the model building function within Eppi reviewer. To this end we used the final included studies within the review as relevant studies and the rest of the covid-19 retrieved studies as irrelevant documents. Eppi reviewer uses the standard SGCClassifier in Scikit-learn on word tri-grams. As output, new documents get a percentage (from the predict_proba function) where scores close to 100 indicate a high probability of belonging to the class 'relevant document' and scores close to 0 indicate a low probability of belonging to the class 'relevant document'. The classifier was trained on the first sets of screened articles (from searches up to March 24), and tested and retrained on every following set of screened articles (search dates April 7 and May 5).

Testing on the second set (search date April 7) revealed poor positive predictive value but 100% sensitivity at a cut-off of 20%. The poor positive predictive value may be due to the broad scope of our topic (all prediction models in covid-19), poor reporting in abstracts, and small set of included documents. The model was retrained after adding the screening data of the second set (search date April 7), which added a considerable amount of additional documents. This led to a large increase in (apparent) positive predictive value, at the cost of a lower (apparent) sensitivity, which led us to reduce the cut-off to 10%. The largest proportion of documents had a score between 0%-9%. This set did not contain any of the relevant documents. This version of the classifier with cut-off 10% was used for the initial screening of the results from the search on May 5th, and saved around 80% of the screening burden.

Subsequent updating after manual screening of the results identified by the AI algorithm and testing of the classifier on the third set (search date May 5th) led to a sensitivity of 100% at cut-off 10% and reasonable

positive predictive value (60%). This version of the classifier was used for the initial screening of the results from the search on June 1st, and saved around 96% of the screening burden.

Supplementary material: data extraction

Data were extracted on the following items:

• Population (China, other).

• Intended timing of model use (screening of patients, ICU admission, etc.).

• Setting (inpatients, outpatients, suspected cases).

• Participants: study design, recruitment method, number of centres, inclusion criteria, exclusion criteria, patient age (mean and standard deviation or median and interquartile range), patient sex (n males and percentage).

• Predictors: list of candidate predictors, number of candidate predictors, number of additional degrees of freedom for candidate predictors (e.g. for categorical variables with more than two categories or for modelling continuous variables non-linearly), predictors in the final model, number of predictors in the final model, number of additional degrees of freedom in the final model.

• Outcome: definition of outcome, timing of outcome.

• Analysis:

  o Total number of participants, total number of participants with the outcome.
  o Total number of participants with any missing predictor or outcome values.
  o Method for handling missing data.
  o Method for prediction model development (logistic regression, Cox regression, neural network, tree-based, etc.).
  o Method for selection of predictors during multivariable analysis and the criteria used (e.g., p-value used for selection).
  o Handling of categorical and continuous variables.
  o Method(s) for validation (e.g. apparent, internal or external) and for optimism adjustment.
  o Performance measures (calibration, discrimination, other) resulting from validation.
  o Model presentation (coefficients and confidence intervals, final model, alternative presentation formats of the model such as a web-based tool).

• Standard signalling questions and items to assess risk of bias according to PROBAST (Moons, Wolff, et al.), on four domains:

  o participants
  o predictors
  o outcome
  o analysis

If risk of bias was high in at least one of the subdomains, overall risk of bias was judged high, as per PROBAST guidance (Moons, Wolff, et al.).

Supplementary Table 1. Overview of Search dates, hits and inclusions.

search date	publication version	titles retrieved	full text screening	included papers	included models
13 March 2020	first pre-print (medRxiv) version	1916	/	/	/
24 March 2020	first published version (BMJ)	774	85	27	31
7 April 2020	first update (BMJ)	2213	114	24	35
5 May 2020	second update (BMJ)	9306	76	56	79
1 July 2020	third update (BMJ)	23203	169	62	87