



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

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



Special Communication

Learning from past respiratory failure patients to triage COVID-19 patient ventilator needs: A multi-institutional study

Harris Carmichael^{a,b,1}, Jean Coquet^{a,1}, Ran Sun^{a,1}, Shengtian Sang^{a,1}, Danielle Groat^f, Steven M. Asch^{a,c}, Joseph Bledsoe^{b,h}, Ithan D. Peltan^{f,g}, Jason R. Jacobs^f, Tina Hernandez-Boussard^{a,d,e,*}

^a Department of Medicine, Stanford University, Stanford, CA, United States

^b Healthcare Delivery Institute, Intermountain Healthcare, Murray, UT, United States

^c Center for Innovation to Implementation, VA Palo Alto Medical Center, United States

^d Department of Biomedical Data Science, Stanford University, Stanford, CA, United States

^e Department of Surgery, Stanford University, Stanford, CA, United States

^f Division of Pulmonary and Critical Care Medicine, Department of Medicine, Intermountain Medical Center, Murray, UT, United States

^g Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Utah, Salt Lake City, UT, United States

^h Department of Emergency Medicine, Stanford University, Stanford, CA, United States



ARTICLE INFO

Keywords

COVID-19

Invasive mechanical ventilation

Machine learning

Triage tool

Respiratory failure

ABSTRACT

Background: Unlike well-established diseases that base clinical care on randomized trials, past experiences, and training, prognosis in COVID-19 relies on a weaker foundation. Knowledge from other respiratory failure diseases may inform clinical decisions in this novel disease. The objective was to predict 48-hour invasive mechanical ventilation (IMV) within 48 h in patients hospitalized with COVID-19 using COVID-like diseases (CLD).

Methods: This retrospective multicenter study trained machine learning (ML) models on patients hospitalized with CLD to predict IMV within 48 h in COVID-19 patients. CLD patients were identified using diagnosis codes for bacterial pneumonia, viral pneumonia, influenza, unspecified pneumonia and acute respiratory distress syndrome (ARDS), 2008–2019. A total of 16 cohorts were constructed, including any combinations of the four diseases plus an exploratory ARDS cohort, to determine the most appropriate cohort to use. Candidate predictors included demographic and clinical parameters that were previously associated with poor COVID-19 outcomes. Model development included the implementation of logistic regression and three ensemble tree-based algorithms: decision tree, AdaBoost, and XGBoost. Models were validated in hospitalized COVID-19 patients at two healthcare systems, March 2020–July 2020. ML models were trained on CLD patients at Stanford Hospital Alliance (SHA). Models were validated on hospitalized COVID-19 patients at both SHA and Intermountain Healthcare.

Results: CLD training data were obtained from SHA ($n = 14,030$), and validation data included 444 adult COVID-19 hospitalized patients from SHA ($n = 185$) and Intermountain ($n = 259$). XGBoost was the top-performing ML model, and among the 16 CLD training cohorts, the best model achieved an area under curve (AUC) of 0.883 in the validation set. In COVID-19 patients, the prediction models exhibited moderate discrimination performance, with the best models achieving an AUC of 0.77 at SHA and 0.65 at Intermountain. The model trained on all pneumonia and influenza cohorts had the best overall performance (SHA: positive predictive value (PPV) 0.29, negative predictive value (NPV) 0.97, positive likelihood ratio (PLR) 10.7; Intermountain: PPV, 0.23, NPV 0.97, PLR 10.3). We identified important factors associated with IMV that are not traditionally considered for respiratory diseases.

Abbreviations: ALT, alanine aminotransferase; ARDS, acute respiratory distress syndrome; AST, aspartate aminotransferase; AUC, area under the receiver operator characteristic curve; BUN, blood urea nitrogen; CRP, c-reactive protein; IMV, invasive mechanical ventilation; LDH, lactate dehydrogenase; NPV, negative predictive value; PPV, positive predictive value; SMOTE, synthetic minority oversampling technique; SVM, support vector machine; WBC, white blood cell.

* Corresponding author at: Stanford School of Medicine, 1265 Welch Road, #245, Stanford University, Stanford, CA 94306, United States.

E-mail address: boussard@stanford.edu (T. Hernandez-Boussard).

¹ Equal contributors.

<https://doi.org/10.1016/j.jbi.2021.103802>

Received 9 October 2020; Received in revised form 27 April 2021; Accepted 3 May 2021

Available online 27 May 2021

1532-0464/© 2021 Elsevier Inc. This article is made available under the Elsevier license (<http://www.elsevier.com/open-access/userlicense/1.0/>).

Conclusions: The performance of prediction models derived from CLD for 48-hour IMV in patients hospitalized with COVID-19 demonstrate high specificity and can be used as a triage tool at point of care. Novel predictors of IMV identified in COVID-19 are often overlooked in clinical practice. Lessons learned from our approach may assist other research institutes seeking to build artificial intelligence technologies for novel or rare diseases with limited data for training and validation.

1. Introduction

SARS-CoV-2 has infected over 30 million people worldwide and its disease syndrome, COVID-19, is responsible for more than 940,000 deaths as of September 17, 2020 [1]. With limited disease-specific clinical experience or data, the medical community is substituting evidence and clinical experience from similar diseases to guide treatment and monitoring practices in COVID-19 [2]. Although experience is accumulating, it remains difficult to accurately identify those patients that need close attention and distinguish those that can safely be monitored in lower acuity settings.

In the early and middle stages of COVID-19, patients present with respiratory symptoms that are clinically indistinguishable from hundreds of other upper respiratory infections in terms of symptoms and severity [3–5]. Critical differences appear later in the disease course when patients present in severe hypoxemia and can rapidly deteriorate late in the symptomatic timeline, requiring advanced oxygenation support. An important unanswered question in this pandemic is how to determine which patients will need advanced respiratory support and which will not. The ability to accurately predict invasive mechanical ventilation (IMV) in COVID-19 would allow health systems to appropriately utilize limited resources to more closely monitor high risk patients and could be used to target enrollment for trials of therapies intended to prevent IMV.

Barriers to accurate, unbiased prediction of COVID-19 disease trajectory include lack of readily available patient data to train and test models in the early outbreak of the disease and no public datasets available to test and validate models outside of a single healthcare setting. In this study, we hypothesized that machine learning (ML) prediction models trained on patients with respiratory infections similar to COVID-19 could accurately identify hospitalized COVID-19 patients who are at risk of IMV within 48 h. In addition, we share lessons learned from the process of implementing this ML framework and the importance of finding the most representative data at the beginning of the pandemic crisis. The framework can be generalized to benefit any healthcare facility facing limited disease-specific data, as was the case with COVID-19.

2. Material and methods

2.1. Study design

This multicenter retrospective study included patients from two hospitals belonging to the Stanford Healthcare Alliance (SHA) in California using Epic medical record systems and 22 Intermountain Healthcare hospitals in Utah and Idaho using Cerner medical record systems. We evaluated model performance in predicting IMV among COVID-19 hospitalized patients using models developed in COVID-like disease (CLD) patients. The framework of the study is described elsewhere [6]. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRI-POD) checklist was followed [7].

2.2. Patient cohorts

COVID-like disease patients: The CLD cohort identified from SHA included patients hospitalized between January 1, 2008 and June 30, 2019 with an acute respiratory infection (unspecified pneumonia,

bacterial pneumonia, viral pneumonia, or influenza) or ARDS on the basis of international classification of diseases, ninth and tenth revision (ICD-9-CM and ICD-10-CM) diagnosis codes (eTable 1). Fig. 1a presents the cohort selection process of the CLD patients. Patients were excluded if they were younger than 18 years old or had Do Not Intubate (DNI) status. We then divided patients into five disease groups based on ICD code using the following hierarchy: ARDS > influenza > viral pneumonia > bacterial pneumonia > unspecified pneumonia. Patients with no laboratory test results during their hospitalizations due to immediate intubation after admission or other unknown reasons were excluded from the training set.

COVID-19 patients: At SHA and Intermountain, hospitalized adult patients with COVID-19 were identified with a positive SARS-CoV-2 polymerase chain reaction result or, at SHA, a diagnosis code for COVID-19 (eTable 1 in the Supplement). Patients included were admitted from March 9 to July 25, 2020 at SHA and from March 16 to May 24, 2020 at Intermountain. Fig. 1b describes selection of COVID-19 cohorts from the two institutions.

2.3. Data collection

Patient demographic and clinical information were captured from electronic health records (EHRs), including baseline demographics, comorbidities, co-infections, symptoms, and laboratory results during hospitalization (Table 1). These variables were identified from the literature that were associated with COVID-19 disease severity [8–11]. Relevant comorbidities were extracted up to three years prior to hospitalization, including cardiovascular disease, diabetes, cancer, hypertension, chronic respiratory disease, respiratory failure, kidney disease, Alzheimer's disease, and cirrhosis. Symptoms from 15 days before admission were also extracted for each patient at SHA and symptoms from time of COVID testing were obtained from all Intermountain patients. We identified patients' coinfection with other respiratory pathogens using both ICD codes and the results of respiratory pathogen panels.

2.4. Outcome

The study outcome was the risk of IMV within 48 h. Daily risk scores were generated for each patient based on the most recent data available, providing a rolling window of risk scores throughout the patient's hospital stay. CPT codes for endotracheal intubation (31500) or ventilator management (94002–94005) were used to identify IMV at SHA; respiratory documentation abstracted from the electronic data warehouse was used to identify IMV initiation at Intermountain. A sensitivity analysis was performed for prediction of IMV at 24 and 72 h. The non-IMV group included patients who were discharged alive and died without IMV.

2.5. Machine learning (ML) algorithms

Cohort Generation. We explored the performance of multiple models trained on 16 training cohorts including 15 cohorts with different combinations of influenza, bacterial pneumonia, viral pneumonia and unspecified pneumonia plus a separate exploratory cohort of patients who developed ARDS during their hospital stay. Initially, many consensus-driven treatment recommendations for COVID were modeled on ARDS algorithms [12]. As ARDS is a syndrome with many possible

etiologies, including non-infectious etiologies, it was modeled separately. Details of cohort generation have been previously been published [6].

Model Development. We developed a rolling window of predictions; each hospital day, a risk score for IMV in the upcoming 48 h was derived for each patient using the most recent laboratory results that were available up until two weeks, with one value per variable. To handle the data imbalance challenge, undersampling in conjunction with an oversampling strategy was used to balance samples between groups. Random undersampling was applied to trim the number of patients in the majority class while oversampling strategies were performed to generate synthetic samples of the minority class. Oversampling approaches considered included synthetic minority oversampling technique (SMOTE), borderline SMOTE, and SVM SMOTE. For each of the 16 training cohorts, 70% of data were randomly selected as training set and 30% were used as a validation set. Several machine learning models that have been frequently applied to predict similar outcomes and with good

interpretability were selected for model training, including logistic regression and decision tree, AdaBoost and XGBoost. We used the three decision tree based algorithms because they have been previously applied to predict clinical events in patients with respiratory diseases based on EHR data [13–15].

To determine the best hyperparameters for each ML algorithm, a grid search and 10-fold cross-validation were used on the training data, and hyperparameters of the model with the best AUC were selected. Using the selected hyperparameters, we tested the model in the remaining unseen 30% of the data. The training procedure stopped once the performance on the validation dataset did not improve after 20 training iterations. Sixteen ML models were derived based on different combinations of the five CLD cohorts. The model training processes are presented in eFigure 1.

Model Performance. The performance of the trained models were quantified using the area under the receiver operator characteristic curve (AUC). We further validated the models on the COVID-19 cohorts.

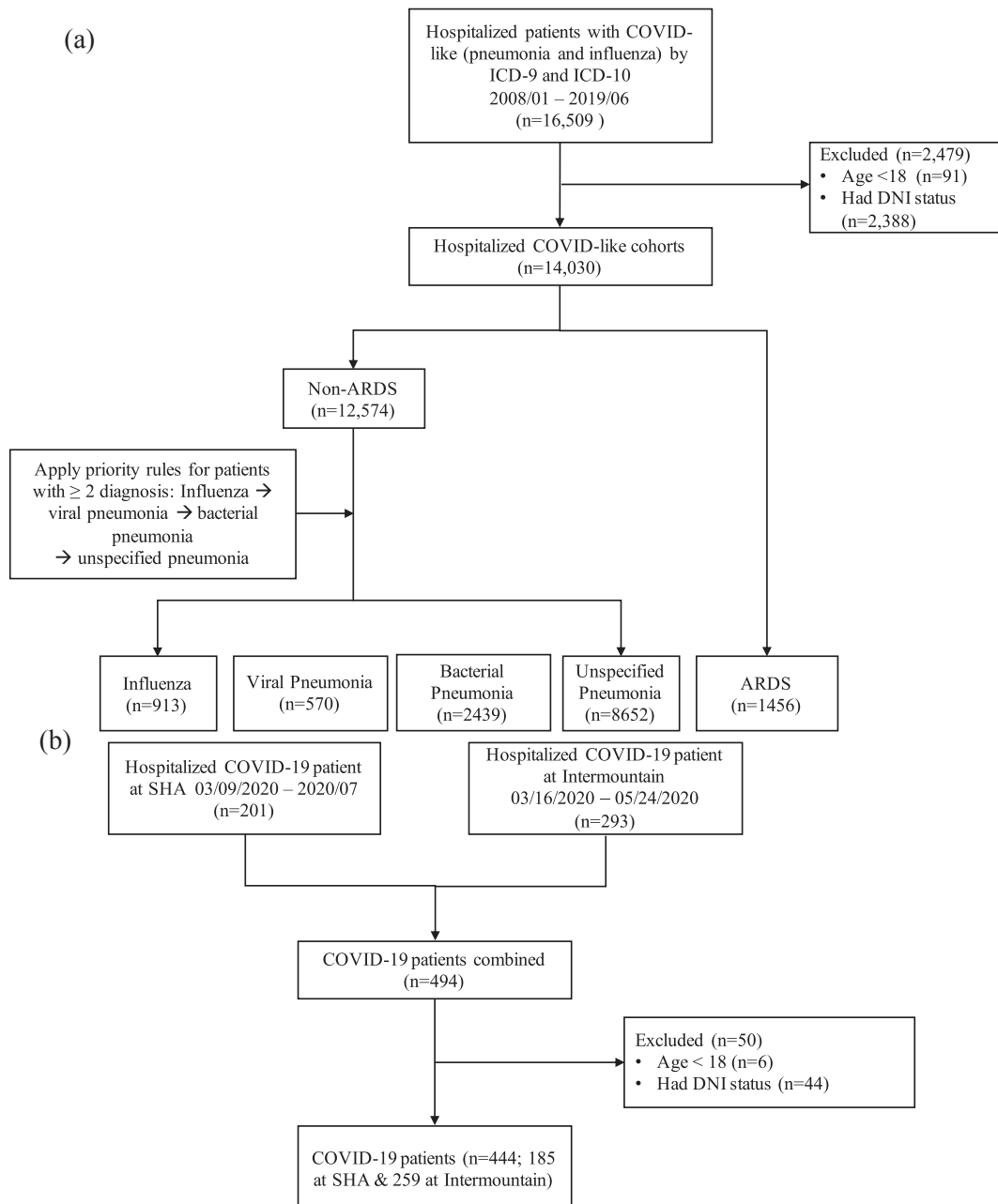


Fig. 1. Cohort selection diagram for (A) COVID-like patients at Stanford Hospital Alliance (SHA) and (B) COVID-19 patients at SHA and Intermountain Healthcare.

Table 1
Features included in model development.

Demographics and clinical characteristics	Laboratory findings	Co-infection*
<ul style="list-style-type: none"> ● Age at admission, years ● Gender ● Ever smoke (all life before IP) ● Comorbidity present (3 years before admission) <ul style="list-style-type: none"> ○ Cancer ○ Chronic Respiratory Disease ○ Cardiovascular Disease ○ Hypertension ○ Type 2 Diabetes ○ Respiratory Failure ○ Kidney Disease ○ Alzheimer's Disease ○ Cirrhosis ● Symptoms* <ul style="list-style-type: none"> ○ Cough ○ Dyspnea ○ Tachypnea, respiratory rate > 20 ○ Hypoxemia, oxygen saturation ≤ 90% ○ Rhinorrhea ○ Nose congestion ○ Fever, temperature > 37C/98.6F ○ Sputum ○ Pharyngitis (sore throat) ○ Headache ○ Fatigue ○ Conjunctivitis ○ Diarrhea ○ Anosmia ○ Myalgias 	<ul style="list-style-type: none"> ● White blood cell count, K/uL ● Lymphocyte count, K/uL ● ALT, U/L ● Creatinine, mg/dL ● Lactate dehydrogenase, U/L ● Creatine kinase, U/L ● High-sensitivity cardiac troponin I, ng/mL ● D-dimer, ng/mL ● Prothrombin time, seconds ● Serum ferritin, ng/mL ● IL-6, pg/mL ● Procalcitonin, ng/mL ● Lactate dehydrogenase, U/L ● Platelet count, K/uL ● C-reactive protein, mg/dL ● Total bilirubin, mg/dL ● Blood urea nitrogen, mg/dL ● Albumin, g/dL ● O2 Saturation, mm/Hg ● Fraction of inspired oxygen, FiO2, % ● Aspartate Aminotransferase, U/L ● Sodium (Na+), mmol/L ● Potassium (K+), mmol/L 	<ul style="list-style-type: none"> ● Adenovirus ● Chlamydia pneumoniae ● Coronavirus ● Influenza A ● Influenza B ● Metapneumovirus ● Mycoplasma pneumoniae ● Parainfluenza 1 ● Parainfluenza 2 ● Parainfluenza 3 ● Parainfluenza 4 ● Rhinovirus/Enterovirus ● Respiratory syncytial virus (RSV)

* within 15 days before admission.

Additional evaluation criteria considered included accuracy (ACC), positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, and positive likelihood ratio (PLR). A threshold of 0.5 was selected and values greater than 0.5 indicated a positive prediction (i.e. IMV would occur in the next 48 h). We further calculated Brier scores to determine the accuracy of the models' probabilistic predictions on the two testing sites. SHapley Additive exPlanations (SHAP) values were calculated for model interpretation [16].

2.6. Sensitivity analysis

Sensitivity analyses measured the prediction model's performance of IMV during alternative follow-up windows (24 h and 72 h).

2.7. Statistical analysis

We used descriptive statistics to compare the demographic and clinical characteristics of the CLD and COVID-19 combined cohorts from SHA and Intermountain. We further compared patient symptoms and laboratory test results between IMV and non-IMV COVID-19 patients. Independent *t*-test or Mann-Whitney *U* test were used for continuous features, and Pearson Chi-square test or Fisher's exact test was applied for categorical features wherever appropriate. Statistical significance for primary analysis was set at *p*-value < 0.05. All tests were 2-tailed. ML models were implemented with Python scikit-learn and XGBoost packages and analyses were performed in R version 3.5.2.

3. Results

We included a total of 14,030 patients with CLDs and 444 patients in the COVID-19 cohort (Table 2). Patients who had no laboratory results (*n* = 2,695, 19.2%) were excluded. COVID-19 patients were significantly younger (53.4 ± 17.2 vs. 63.7 ± 18.9 , *p* < 0.001) with fewer patients with comorbidities (47.7% vs. 64.9%, *p* < 0.001) than the CLD patients. Minorities were disproportionately represented among COVID-19 patients, who were also less likely to have public insurance (56.5% vs 63.9%, *p* < 0.001) or have a smoking history (21.4% vs 40.4%, *p* < 0.001).

Among patients with CLDs, 26.9% received IMV beginning an average of 4.1 ± 7.6 days after hospital admission, and in the COVID-19 cohort 13.5% received IMV beginning 2.9 ± 3.2 days from admission (*p* = 0.001). COVID-19 patients requiring IMV had lower platelets and lymphocyte counts, higher lactate dehydrogenase (LDH) and c-reactive protein (CRP), and higher fraction of inspired oxygen (Table 3).

3.1. Model performance

Across all patients, we retrieved a total of 79,233 patient days, of which 2841 were positive intubation days. The XGBoost model outperformed other ML models in the training data and was used for all subsequent analyses; the performance metrics for the XGBoost model in the training data are presented in eTable 3. The model that included all pneumonia and influenza cohorts, without ARDS, performed best across all metrics at both validation sites. Table 4 present the performance of the best performing model at SHA and Intermountain. In the COVID-19 validation cohorts, the AUC was 0.77 at SHA and 0.65, at Intermountain (Fig. 2), and the statistics as a triage tool for impending IMV remained robust with excellent specificity and PLR (Table 4). While other models had higher AUCs, the model containing the four pneumonia and influenza cohorts had better overall performance metrics, such as PPV, NPV and ACC (eTables 4 and 5). The sensitivities were 0.15 at SHA and 0.10 at Intermountain, while specificity was 0.99 at both sites. The Brier score was 0.047 at SHA and 0.033 at Intermountain. Only 1.8% and 1.2% of total patient days were predicted as positive at SHA and IM respectively. The PPV in COVID-19 patients was 0.29 at SHA and 0.23 at Intermountain while NPV was 0.97 at both institutions. PLR were 10.7 and 10.0 at SHA and Intermountain, respectively.

3.2. Algorithm variable importance

The top ten most important variables for the IMV prediction were: D-dimer, serum ferritin, LDH, fraction of inspired oxygen, total bilirubin, SPO2, CRP, prothrombin time, sodium (Na+), and creatine kinase. The full ranked list of variables is available in eFigure 2 in the Supplement.

3.3. Sensitivity analysis

We used the trained model that predicted IMV within 48 h to test different prediction lead times, specifically IMV at 24 and 72 h. IMV prediction at 24 h had higher AUCs, yet lower PPVs compared to the 48-hour model at both validation sites. Predicting IMV at 72 h produced AUCs lower than the 48-hour models yet higher PPVs, with the highest PPV reaching 0.381 in the SHC validation set. (eTable 6)

4. Discussion

Faced with a novel disease and high - even overwhelming - patient volumes, clinicians and hospitals need to understand how well tools and care patterns designed for apparently similar respiratory syndromes can function for COVID-19 care. We leveraged data from other respiratory syndrome patients (COVID-19 like patients) to train and test a ML model to predict IMV within 48 h and measured its performance when applied to hospitalized COVID-19 patients from two independent healthcare

Table 2
Baseline characteristics of patients hospitalized in the COVID-Like Disease (CLD) cohort and the COVID-19 cohort.

Variables	CLD	COVID-19			p-value*
		SHA	Intermountain	Total	
Total, N	14,030	185	259	444	
Age (years), Mean ± SD	63.7 ± 18.9	53.9 ± 17.8	53.0 ± 16.5	53.4 ± 17.2	<0.001
Gender					
Male	7673 (54.7)	92 (49.7)	145 (56.0)	237 (53.4)	0.582
Race, N (%)					
White	7990 (56.9)	42 (22.7)	179 (69.1)	221 (49.8)	<0.001
Asian	2213 (15.8)	24 (13.0)	<10	26 (5.9)	
Black	867 (6.2)	<10	<10	15 (3.4)	
Other	2612 (18.6)	97 (52.4)	72 (27.8)	169 (38.1)	
Ethnicity					
Hispanic	2023 (14.4)	94 (50.8)	98 (37.8)	192 (43.2)	<0.001
Insurance, N(%)					
Private	3309 (23.6)	38 (20.5)	123 (47.5)	161 (36.3)	< 0.001
Public	8976 (63.9)	130 (70.2)	121 (46.7)	251 (56.5)	
Other	333 (2.4)	17 (9.2)	14 (5.4)	31 (7.0)	
Ever smoke, N (%)	5675 (40.4)	55 (29.7)	40 (15.4)	95 (21.4)	< 0.001
One or more comorbidity, N(%)	9108 (64.9)	64 (34.6)	148 (57.1)	212 (47.7)	<0.001
IMV, N (%)	3775 (26.9)	25 (13.5)	35 (13.5)	60 (13.5)	<0.001
Time to IMV (days), Mean ± SD	4.1 ± 7.6	3.2 ± 4.3	2.7 ± 2.2	2.9 ± 3.2	0.001
Inpatient mortality, N (%)	1925 (13.7)	<10†	11 (4.3)	–	<0.001

Note. *p value was calculated comparing patients with CLDs to the COVID-19 combined cohort; Percentages may not add up to 100 due to missing values.

Table 3
Clinical characteristics of hospitalized COVID-19 patients, stratified by Invasive Mechanical Ventilation (IMV).

Variable	non-IMV			IMV			p-value
	SHA	Intermountain	Combined	SHA	Intermountain	Combined	
Patients (N)	137	224	361	25	35	60	
Age	57.8 ± 15.2	53.3 ± 16.6	55.0 ± 16.1	56.8 ± 20.0	53.1 ± 16.4	54.6 ± 18.0	0.861
Gender, male	67 (48.9)	122 (54.46)	189 (52.3)	11(44.0)	23 (65.7)	34 (56.7)	0.535
Ever smoke	31 (22.6)	33 (14.73)	64 (17.7)	12(48.0)	<10	19 (31.7)	0.012
1 + comorbidities	41 (29.9)	124 (55.36)	165 (45.7)	10(40.0)	24 (68.57)	35 (58.3)	0.070
Co-Infections (Yes/No)	10 (7.3)	<10	11 (3.0)	<10	0(0.0)	<10	0.435
Symptoms at admission							
Cough	105 (76.6)	182 (81.3)	287 (79.5)	17 (68.0)	28 (80.0)	45 (75.0)	0.429
Dyspnea	104 (75.9)	188 (83.9)	292 (80.0)	20 (80.0)	32 (91.4)	52 (86.7)	0.284
Fever	97 (70.8)	133 (59.4)	230 (63.7)	16 (64.0)	27 (77.1)	43 (71.7)	0.232
Fatigue	62 (45.3)	67 (29.9)	129 (35.7)	10 (40.0)	13 (37.1)	23 (38.3)	0.698
Myalgias	52 (38.0)	83 (37.1)	135 (37.4)	10 (40.0)	<10	19 (31.7)	0.394
Headache	28 (20.4)	46 (20.5)	74 (20.5)	<10	<10	11 (18.3)	0.699
Diarrhea	24 (17.5)	79 (35.3)	103 (28.5)	<10	16 (45.7)	23 (38.3)	0.125
Tachypnea	50 (36.5)	33 (13.5)	83 (23.0)	11 (44.0)	<10	19 (31.7)	0.146
Pharyngitis	22 (16.1)	66 (29.5)	88 (24.4)	<10	13 (37.1)	20 (33.3)	0.141
Hypoxemia	34 (24.8)	25 (10.3)	59 (16.3)	15 (60.0)	<10	23 (38.3)	<0.001
Sputum	25 (18.2)	44 (19.6)	69 (19.1)	<10	<10	11 (18.3)	0.887
Anosmia	20 (14.6)	26 (11.6)	46 (12.7)	<10	<10	<10	0.816
Rhinorrhea	<10	60 (26.8)	64 (17.7)	<10	13 (37.1)	14 (23.3)	0.301
Nose congestion	<10	0	<10	<10	0 (0.0)	<10	0.599
Laboratory Values, Mean ± SD							
White blood cell count, K/uL	7.0 ± 2.5	6.6 ± 2.5	6.7 ± 2.5	8.3 ± 4.1	6.7 ± 5.8	7.4 ± 5.1	0.095
Lymphocyte count, K/uL	1.3 ± 0.7	1.3 ± 0.8	1.3 ± 0.8	0.9 ± 0.4	0.9 ± 0.5	0.9 ± 0.4	<0.001
Platelet count, K/uL	262.7 ± 97.8	243.5 ± 93.1	250.8 ± 94.9	197.0 ± 70.8	192.2 ± 96.6	194.4 ± 85.7	<0.001
Alanine aminotransferase, U/L	50.4 ± 41.2	54.3 ± 67.9	52.7 ± 58.8	35.7 ± 23.7	51.6 ± 55.3	44.2 ± 43.7	0.308
Aspartate aminotransferase, U/L	54.1 ± 44.6	57.4 ± 51.7	56.1 ± 49.1	43.9 ± 18.6	73.2 ± 81.6	59.6 ± 61.2	0.634
Total bilirubin, mg/dL	0.6 ± 1.8	0.6 ± 0.3	0.6 ± 1.1	0.6 ± 0.2	0.7 (0.4)	0.6 ± 0.3	0.841
Albumin, g/dL	3.5 ± 0.5	3.5 ± 0.4	3.5 ± 0.5	3.3 ± 0.6	3.6 ± 0.5	3.5 ± 0.5	0.453
Blood urea nitrogen, mg/dL	15.8 ± 11.9	14.3 ± 8.9	14.8 ± 10.1	22.4 ± 17.8	18.3 ± 11.5	20.2 ± 14.7	<0.001
Troponin I, ng/mL	0.08 ± 0.1	0.45 ± 4.1	0.42 ± 3.89	0.12 ± 0.17	0.03 ± 0.1	0.05 ± 0.1	0.671
D-dimer, ng/mL	1233.4 ± 699.1	1398 ± 2018	1374.7 ± 1893.1	1285.5 ± 890.9	1308 ± 716	1303.8 ± 743.5	0.883
Lactate dehydrogenase, LDH, U/L	363.7 ± 113.7	359.9 ± 132.6	361.6 ± 124.6	460.7 ± 150.2	540.9 ± 230.4	499.4 ± 193.0	<0.001
Serum ferritin, ng/mL	896.6 ± 905.6	851.8 ± 1074	870.3 ± 1007.9	1289.5 ± 896.6	1315 ± 1122	1300.1 ± 994.6	0.031
Procalcitonin, ng/mL	0.2 ± 0.3	0.4 ± 1.4	0.3 ± 1.1	0.6 ± 0.7	0.6 ± 1.4	0.6 ± 1.1	0.112
C-reactive protein, mg/dL	8.3 ± 6.0	9.0 ± 6.8	8.8 ± 6.6	17.9 ± 11.7	13.5 ± 9.1	15.6 ± 10.4	<0.001
Creatine kinase, U/L	188.0 ± 226.9	481.2 ± 1660	322.6 ± 1136.2	219.1 ± 258.2	192.0 ± 229.0	209.6 ± 248.9	0.660
Sodium (Na+), mmol/L	136.7 ± 3.2	137.0 ± 2.9	136.9 ± 3.0	137.7 ± 3.0	135.7 ± 3.5	136.6 ± 3.2	0.546
Potassium (K+), mmol/L	4.0 ± 0.3	3.8 ± 0.4	3.9 ± 0.3	4.0 ± 0.7	3.8 ± 0.5	3.9 ± 0.6	0.884
Creatinine, mg/dL	1.0 ± 1.3	1.0 ± 1.2	1.0 ± 1.2	1.3 ± 1.8	1.0 ± 0.4	1.1 ± 1.3	0.363
Prothrombin time, (s)	14.0 ± 2.4	15.1 ± 4.5	14.3 ± 3.2	14.4 ± 1.7	15.8 ± 6.2	14.8 ± 3.4	0.545
Oxygen saturation, SpO2, %	95.7 ± 1.9	93.9 ± 2.2	94.6 ± 2.1	95.6 ± 2.3	92.9 ± 2.8	94.2 ± 2.6	0.189
Fraction of inspired oxygen, %	0.29 ± 0.1	0.56 ± 0.3	0.4 ± 0.2	0.6 ± 0.2	0.6 ± 0.3	0.6 ± 0.2	<0.001

Table 4Performance of the best performing model^e at Stanford Hospital Alliance (SHA) Intermountain Healthcare for COVID-like disease and COVID-19 hospitalized patients.

Validation datasets	AUC ^a	ACC ^a	PPV ^b	Sensitivity	F-score	Specificity	NPV ^c	PLR ^d
COVID-like disease patients	0.871	0.829	0.731	0.491	0.588	0.940	0.849	8.183
COVID-19 patients at SHA	0.772	0.957	0.286	0.150	0.197	0.986	0.970	10.714
COVID-19 patients at Intermountain	0.648	0.965	0.235	0.100	0.140	0.99	0.974	9.999

^a AUC: Area under the ROC Curve; ^aACC: accuracy; ^bPPV: positive predictive value; ^cNPV: negative predictive value. ^dPLR: positive likelihood ratio; ^eselected model with all four diseases: viral pneumonia, bacterial pneumonia, unspecified pneumonia, and influenza.

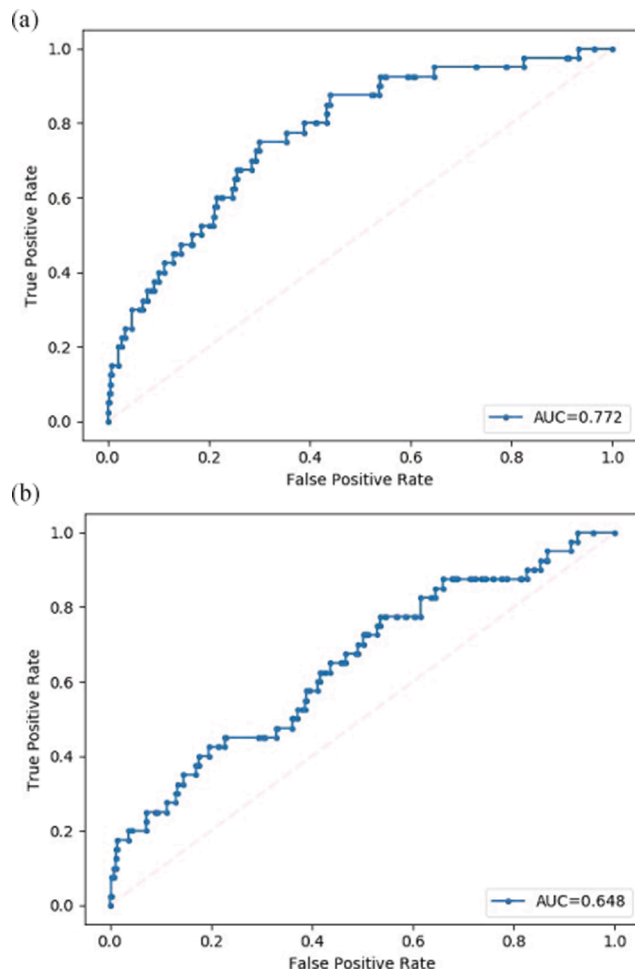


Fig. 2. Receiver Operating Characteristic Curves for trained Model on COVID-19 patients at (A) Stanford Hospital Alliance and (B) Intermountain Healthcare.

centers with different EHR systems. Our framework leverages the abundance of retrospective respiratory failure patients to provide a testing bed for discovery and innovation. Based on our model's excellent clinical performance in PLR, the best utility would be in highlighting those patients at highest risk of IMV within 48 h out of all hospitalized COVID patients with a moderate risk of respiratory failure. Under anticipated surges of hospitalized patients based on local community incidence, this tool may help clinicians and hospital systems with limited specialized resources focus limited expertise on COVID-19 patients at highest risk of deterioration. Given the relatively low rates of positive results from the model (13.5%) there is a permissive tolerance of false positives that may receive closer monitoring but not require intubation. Excellent NPV and specificity allow confidence in standard of care monitoring for those patients with negative results from the model. Such a tool could also help enrich therapeutic trials with patients more likely to benefit from particular therapies, thereby improving study power and efficiency. We believe lessons learned from the COVID

experience can serve others interested in designing and implementing prognostic models in novel or rare diseases, where training data may be limited.

The model performed well at SHA but less optimal in Intermountain Differences from the two facilities were significant and include, but are not limited to, EMR systems, EHR source data definitions and data acquisition systems, geographic area covered, patient demographics (Tables 2 and 3), provider practice habits, equipment, and dedicated tertiary training facility vs collection of community hospitals and tertiary referral center. This variation may explain some of the differences in AUC but also make it likely that these results are robust against even greater variation than already observed and reported in this manuscript. Importantly, while AUC's did vary, the tool continued to maintain utility at the included sites. The AUC and F-score derived in both the training and validation data sets are comparable to other validated inpatient prediction models such as APACHE II and IV [17] and current COVID prediction models [18]. However, many predictive models for COVID-19 report only model accuracy and fail to report the additional information needed to guide clinical decisions, such as sensitivity and specificity. In order for any predictive model to be useful, the clinical utility of the model must be a priority and an understanding of how such a model can be used at point of care is essential. In this study the top performing model was based on performance across all metrics, not only the AUC. We therefore selected the model trained on the cohort that included all pneumonia and influenza patients, which did not have the highest AUC in the validation sets but had higher PPVs and PLR. In addition, our study builds on the existing literature by predicting IMV within a 48-hour rolling window, updating individual patient inputs daily, and was externally validated using data from a multihospital health system employing a different EHR system.

Hospitals experiencing a surge of COVID-19 patients will have limited experienced hospital clinicians or infectious disease providers, which emphasizes the need for a tool to help triage patients and identify those most likely to clinically deteriorate. ML algorithms, such as our model, are ideal triage tools that may sort patients for higher priority to be reviewed by these experts. In consonance with emerging evidence on the association of inflammation and COVID-19 outcomes [19], our model identified inflammatory and coagulation markers (e.g., D-dimer, ferritin, LDH, CRP, etc.) as features more predictive than factors such as age, degree of hypoxic respiratory failure, and other markers of organ failure. The fact that our model leverages potentially underappreciated laboratory features, not included in many prior predictive models, for hospitalized respiratory syndromes indicates that these results could identify impending decline in patients otherwise overlooked. Additionally, excellent PLR—positive results dramatically increase probability of IMV—at both institutions and indicate our model can appropriately identify patients for close monitoring of imminent IMV. Although PPV was low, this largely reflects the low daily intubation rate (1.2% at our external validation site, Intermountain). A “positive” risk score nevertheless indicated a substantial increase in individual risk as assessed by the PLR. Lastly, the rolling daily risk assessments are also an important feature of our predictive model and is particularly important due to the rapid deterioration observed in COVID-19 patients. This model would be well suited for automatic screening to target COVID patients for closer monitoring on admission and during daily rounds.

Despite the disease's novelty, the literature is already flooded with

prediction models for COVID-19, including risk, diagnostic, and prognostic models [14,18,20–22]. Our results are similar to previous work attempting to predict clinical COVID-19 patient clinical deterioration [23]. While some of our predictive variables overlap with previous models, our model identifies emerging inflammatory markers previously overlooked. Given the pressing issues of the pandemic, some of these models were under-developed and potentially biased, resulting in retraction of reports [24]. Our framework and transparency in reporting of the details of the training population, model architecture, and hyperparameter tuning follows emerging standards for reporting ML models in healthcare [7,25]. We have made our algorithms and code publicly available, and hope others will test and build on our work.

This study has limitations that may affect the interpretation of results. First, there was a relatively small number of COVID-19 patients receiving IMV in our population. This could negatively impact the stability of model's performance estimates. Second, we selected features that were associated with COVID-19 outcomes in existing literature. As our knowledge of the disease progresses, additional features are emerging that may also contribute to the risk of IMV, which were not included in the model. However, this study lays the framework for additional studies to validate emerging risk predictors. Finally, clinical practice variation, differences in institutional protocols, case volumes, and availability of clinical resources could all affect the predictive accuracy of our model. COVID-19 patient surges may substantially alter the threshold for initiating IMV, which may limit the utility of our model in regions with extreme patient volumes. As part of a future investigation, an interesting strategy would be to identify the best threshold for both institutes using a subset of the COVID-19 cohort to calibrate a local threshold. Due to the small number of patients with COVID in some cohort groups, our study was limited in the scope of machine learning methods to train the model. We selected several machine learning models that have been successfully applied in other studies and with good interpretability in terms of feature significance, which is essential when implementing such predictive models in clinical settings. However, with the increase in the number of COVID-19 patients and the possibility of performing a crossover study, a future investigation would be to apply several deep learning methods, such as recurrent neural network (RNN), which would allow us to take into consideration the temporal sequence between hospitalization days. These deep learning methods may have the potential to outperform our current model in terms of accuracy and should be considered in future studies with larger sample sizes

Integrating prospective and retrospective outcomes from regions with varied patient volumes and demographics would strengthen our model and we encourage the reporting of this data to improve our models.

5. Conclusion

Medical knowledge often progresses by analogizing from pathophysiologically related but distinct disease entities. To address the lack of COVID specific data, we used the abundant historical data from related conditions to address prognostication for a novel epidemic infection. The commonality between COVID-19 and other common respiratory failure diseases can help provide guidance to providers treating COVID-19 patients during this novel pandemic where clinical experience and evidence remain sparse. When validated on COVID-19 patients from two healthcare systems, our best-performing model demonstrated the capacity to meaningfully inform the clinical suspicion that a COVID-19 patient will need IMV. Future efforts to increase model sensitivity and further bolster positive predictive value will improve the model's clinical utility.

5.1. Lessons learned

In the beginning of the pandemic, many institutes were challenged

with a limited number of cases for model training and limited knowledge about the novel COVID-19 disease. The quick surge of COVID-19 patients overflows health facilities and a decision support tool was in great need for triaging patients. Many rushed to produce models with limited training data [20] that led to potentially biased models [24]. This work proposes an alternative approach that addresses the needs of many institutions to support decision making in the void of robust data sets. Such a framework could support the rapid dissemination of prognostic models at the next pandemic or for rare diseases.

In the haste of developing AI-based COVID decision tools, most published models have fallen short of providing safe and effective guidance [24]. In a continuously updated systematic review of COVID-19 prediction models, most models lack transparency and are at high risk of bias [20]. External validation is not common and performance drops dramatically when these models are tested on a different dataset, which can make these models irrelevant for deployment at the point of care outside the training data site [26]. The model reported in this study also suffers from generalizability to some extent, however there are important lessons that can be learned from this endeavor. Data extraction procedures vary greatly across systems. To ensure generalizability, one must empirically validate each feature used in the model, including missingness, distribution, time captured, and units captured. For example, oxygen flow was an essential feature of this model, yet captured very differently at the two sites. For one site, conversion of values for FiO2 were performed; the values recorded included either the liters of oxygen per minutes delivered or the actual FiO2 values, we converted the oxygen delivered into FiO2 when necessary. While the other site has this feature recorded cleaner and more complete. In this work, one institute created a panel to obtain many features for each COVID patient hospitalized. In contrast, features at the other hospital had to be scraped from flowcharts. Generalizability was also affected by the differences in patient populations, where there were important differences in race, ethnicity, smoking status, and pre-existing conditions, which might also be a reflection of data capture. However, given that we are one of the few COVID models that have actually performed cross-site validation, the model is likely more generalizable than most. This work demonstrated the feasibility of cross-site validation in a very short period of time and suggests that clinical decision tools need to prove robustness against a wide range of generalizability challenges.

In addition to model development, we provide guidance on the implementation of our framework and multi-institute validation. First, it is important to identify 'like-cohorts' that share outcome characteristics and clinical manifestations with the disease of interest. These like-cohorts must be common with sufficient sample sizes for training, testing, and validating the ML models. Second, a system should be in place to extract cohorts and variables from the EHRs and, importantly, identify new variables that might be unique to the disease of interest and stored in flowsheets or as unstructured data in clinical narrative text. Third, at the onset of the pandemic, COVID-19 treatment was unsure and guidelines did not exist, large practice variation was seen across the globe and needs to be considered when implementing clinical decision support tools. Therefore, cross-validation across care settings is necessary. During cross-validation, every variable used in the model must be thoroughly investigated, including data capture, storage, missingness and validity. Finally, geographical variation must be considered, as the outbreak of COVID-19 in New York City is greatly different from outbreaks in other cities across the globe, again highlighting the need for cross validation.

The lessons learned for clinical model application were that models trained on data of a "patient-like me" cohort (i.e., COVID-like diseases) can have good accuracy and present clinical utility by identifying in-patients at high risk of decomposition. Importantly, we also learned that laboratory factors associated with inflammation (i.e., D-dimer, ferritin, LDH, CRP), that have not been previously appreciated as predictors for patient deterioration in respiratory illnesses, were highly influential in our model. The framework provides an opportunity to quickly identify

unsuspected risk factors associated with disease outcomes at the onset of future pandemics.

6. Funding xxx

IDP was supported by a career development award (K23GM129661) from the National Institute of General Medical Sciences.

7. Ethics approval and consent to participate

This study received approval from the Institutional Review Board (IRB) of both participating healthcare systems.

8. Consent for publication

Not applicable.

9. Availability of data and materials

All code and trained ML models are publicly available via GitHub as well as a de-identified sample dataset [URL to be published with manuscript].

10. Authors' contributions

THB takes responsibility for the integrity of the data and the accuracy of the data analysis.

- Concept and design: THB
- Statistical analysis: JC, RS, and HLC
- Collection of data: JC, RS, DG, JJ, IDP
- Interpretation of data: All authors
- Drafting of the manuscript: HLC, JC, RS, SS, THB
- Critical revision of the manuscript: All authors
- Administrative, technical, or material support: THB, IDP
- Study supervision: THB, IDP

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ithan Peltan has received research support to their institution from Janssen Pharmaceuticals, Immunexpress Inc., and Regeneron Pharmaceuticals. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jbi.2021.103802>.

References

- [1] COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). <https://coronavirus.jhu.edu/map.html> (accessed July 28, 2020).
- [2] S. Boccia, W. Ricciardi, J.P.A. Ioannidis, What other countries can learn from Italy during the COVID-19 pandemic, *JAMA - J. Am. Med. Assoc.* 180 (2020) 927–928, <https://doi.org/10.1001/jamainternmed.2020.1447>.
- [3] M.J. Cummings, M.R. Baldwin, D. Abrams, S.D. Jacobson, B.J. Meyer, E. M. Balough, J.G. Aaron, J. Claassen, L.R.E. Rabbani, J. Hastie, B.R. Hochman, J. Salazar-Schicchi, N.H. Yip, D. Brodie, M.R. O'Donnell, Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study, *Lancet.* 395 (2020) 1763–1770, [https://doi.org/10.1016/S0140-6736\(20\)31189-2](https://doi.org/10.1016/S0140-6736(20)31189-2).
- [4] J. Zhu, P. Ji, J. Pang, Z. Zhong, H. Li, C. He, J. Zhang, C. Zhao, Clinical characteristics of 3,062 COVID-19 patients: a meta-analysis, *J. Med. Virol.* (2020) 1–13, <https://doi.org/10.1002/jmv.25884>.
- [5] B. Spellberg, M. Haddix, R. Lee, S. Butler-Wu, P. Holtom, H. Yee, P. Gounder, Community prevalence of SARS-CoV-2 among patients with influenzalike illnesses presenting to a Los Angeles Medical Center in March 2020, *Jama.* (2020) 1–2, <https://doi.org/10.1001/jama.2020.4958>.
- [6] S. Sang, R. Sun, J. Coquet, H. Carmichael, T. Seto, T. Hernandez-Boussard, Learning from past respiratory infections to predict COVID-19 outcomes: retrospective study. *J. Med. Int. Res.* 23(2) e23026. <https://doi.org/10.2196/23026>.
- [7] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement, *Circulation.* 131 (2015) 211–219, <https://doi.org/10.1161/CIRCULATIONAHA.114.014508>.
- [8] W. Liang, H. Liang, L. Ou, B. Chen, A. Chen, C. Li, Y. Li, W. Guan, L. Sang, J. Lu, Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19, *JAMA Intern. Med.* (2020).
- [9] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, B. Cao, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, *Lancet.* 6736 (2020) 1–9, [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3).
- [10] Z. Wu, J.M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention, *JAMA - J. Am. Med. Assoc.* 2020 (2019) 3–6, <https://doi.org/10.1001/jama.2020.2648>.
- [11] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet* 395 (2020) 507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
- [12] L. Gattinoni, S. Coppola, M. Cressoni, M. Busana, D. Chiumello, Covid-19 does not lead to a "typical" acute respiratory distress syndrome, *Am. J. Respir. Crit. Care Med.* (2020) 1–5, <https://doi.org/10.1164/rccm.202003-0817le>.
- [13] P. Yang, T. Wu, M. Yu, F. Chen, C. Wang, J. Yuan, J. Xu, G. Zhang, A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters, *PLoS One.* 15 (2020) 1–21, <https://doi.org/10.1371/journal.pone.0226962>.
- [14] A. Vaid, S. Somani, A.J. Russak, Machine Learning to predict mortality and critical events in COVID-19 positive New York City patients, *MedRxiv.* (2020).
- [15] F. Ma, L. Yu, L. Ye, D.D. Yao, W. Zhuang, Length-of-stay prediction for pediatric patients with respiratory diseases using decision tree methods, *IEEE J. Biomed. Heal. Inform.* (2020).
- [16] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Nips.* 16 (2017) 426–430.
- [17] Y.E. Varghese, M.S. Kalaiselvan, M.K. Renuka, A.S. Arunkumar, Comparison of acute physiology and chronic health evaluation II (APACHE II) and acute physiology and chronic health evaluation IV (APACHE IV) severity of illness scoring systems, in a multidisciplinary ICU, *J. Anaesthesiol. Clin. Pharmacol.* 33 (2017) 248–253, <https://doi.org/10.4103/0970-9185.209741>.
- [18] L. Jehi, X. Ji, A. Milinovich, S. Erzurum, A. Merlino, S. Gordon, J.B. Young, M. W. Kattan, Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19, *PLoS One.* 15 (2020), e0237419, <https://doi.org/10.1371/journal.pone.0237419>.
- [19] P. Horby, W.S. Lim, J.R. Emberson, M. Mafham, J.L. Bell, L. Linsell, N. Staplin, Dexamethasone in hospitalized patients with covid-19 — preliminary report, *N. Engl. J. Med.* (2020) 1–11, <https://doi.org/10.1056/nejmoa2021436>.
- [20] L. Wynants, B. Van Calster, G.S. Collins, R.D. Riley, G. Heinze, E. Schuit, M.M.J. Bonten, J.A.A. Damen, T.P.A. Debray, M. De Vos, P. Dhiman, M.C. Haller, M.O. Harhay, L. Henckaerts, N. Kreuzberger, A. Lohmann, K. Luijckx, J. Ma, C.L. Andaur Navarro, J.B. Reitsma, J.C. Sergeant, C. Shi, N. Skoetz, L.J.M. Smits, K.I.E. Snell, M. Sperrin, R. Spijker, E.W. Steyerberg, T. Takada, S.M.J. Van Kuijk, F.S. Van Royen, C. Wallisch, L. Hooft, K.G.M. Moons, M. Van Smeden, Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal, *BMJ.* 369 (2020). <https://doi.org/10.1136/bmj.m1328>.
- [21] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, An interpretable mortality prediction model for COVID-19 patients, *Nat. Mach. Intell.* (2020) 1–6.
- [22] Y. Shang, T. Liu, Y. Wei, J. Li, L. Shao, M. Liu, Y. Zhang, Z. Zhao, H. Xu, Z. Peng, F. Zhou, X. Wang, Scoring systems for predicting mortality for severe patients with COVID-19, *EclinicalMedicine.* 24 (2020) 1–11, <https://doi.org/10.1016/j.eclinm.2020.100426>.
- [23] H. Burdick, C. Lam, S. Mataraso, A. Siefkas, G. Braden, R.P. Dellinger, A. McCoy, J. L. Vincent, A. Green-Saxena, G. Barnes, J. Hoffman, J. Calvert, E. Pellegrini, R. Das, Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial, *Comput. Biol. Med.* 124 (2020), 103949, <https://doi.org/10.1016/j.combiomed.2020.103949>.
- [24] E. Rösli, B. Rice, T. Hernandez-boussard, Bias at warp speed: How AI may contribute to the disparities gap in the time of COVID-19, *J. Am. Med. Informatics Assoc.* (2020), <https://doi.org/10.1093/jamia/ocaa210>.
- [25] T. Hernandez-Boussard, S. Bozkurt, J.P.A. Ioannidis, N.H. Shah, MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care, *J. Am. Med. Informatics Assoc.* 00 (2020) 1–5, <https://doi.org/10.1093/jamia/ocaa088>.
- [26] M. Barish, S. Bolourani, L.F. Lau, S. Shah, T.P. Zanos, External validation demonstrates limited clinical utility of the interpretable mortality prediction model for patients with COVID-19, *Nature Mach. Intell.* 3 (1) (2021) 25–27, <https://doi.org/10.1038/s42256-020-0180-7>.