Early COVID-19 respiratory risk stratification using machine learning

Molly J Douglas [iD] ,[1,2] Brian W Bell,[2] Adrienne Kinney,[2] Sarah A Pungitore,[2] Brian P Toner[2]

ABSTRACT

Background COVID-19 has strained healthcare systems globally. In this and future pandemics, providers with limited critical care experience must distinguish between moderately ill patients and those who will require aggressive care, particularly endotracheal intubation. We sought to develop a machine learning-informed Early COVID-19 Respiratory Risk Stratification (ECoRRS) score to assist in triage, by providing a prediction of intubation within the next 48 hours based on objective clinical parameters.

Methods Electronic health record data from 3447 COVID-19 hospitalizations, 20.7% including intubation, were extracted. 80% of these records were used as the derivation cohort. The validation cohort consisted of 20% of the total 3447 records. Multiple randomizations of the training and testing split were used to calculate confidence intervals. Data were binned into 4-hour blocks and labeled as cases of intubation or no intubation within the specified time frame. A LASSO (least absolute shrinkage and selection operator) regression model was tuned for sensitivity and sparsity.

Results Six highly predictive parameters were identified, the most significant being fraction of inspired oxygen. The model achieved an area under the receiver operating characteristic curve of 0.789 (95% CI 0.785 to 0.812). At 90% sensitivity, the negative predictive value was 0.997.

Discussion The ECoRRS score enables non-specialists to identify patients with COVID-19 at risk of intubation within 48 hours with minimal undertriage and enables health systems to forecast new COVID-19 ventilator needs up to 48 hours in advance.

Level of evidence IV.

INTRODUCTION

The COVID-19 global pandemic has caused unprecedented levels of population illness and healthcare resource utilization.[1–4] Infection with the causative agent of COVID-19, SARS-CoV-2, can range from asymptomatic[5 6] to life-threatening,[4 7 8] and illness requiring mechanical ventilation carries a high mortality rate of 25% to 60%.[7 9 10]

The combination of heavy illness burden and finite resources has made triage a necessity in many health systems, with a particular strain on intensive care units (ICUs).[1 3 11] Patients with acute respiratory failure may require endotracheal intubation and placement on a ventilator for respiratory support, interventions which are only performed in an ICU setting. Appropriate triage can reduce unnecessary ICU admissions and promote allocation of resources to the sickest patients. Factors

WHAT IS ALREADY KNOWN ABOUT THIS TOPIC?

⇒ The COVID-19 pandemic has strained healthcare resources and highlighted the importance of appropriate triage to allocate resources most efficiently.

WHAT THIS STUDY ADDS

⇒ This retrospective modeling study derives a six-variable model for predicting the risk of respiratory failure requiring intubation, in any 48-hour period, for patients with COVID-19, with an area under the receiver operating characteristic curve of 0.8.

HOW MIGHT THIS STUDY AFFECT RESEARCH, PRACTICE OR POLICY?

⇒ This streamlined model allows non-experts to assist in accurate triage to an appropriate level of care and can aid in system-level planning for bed and staffing needs.

shown to be associated with severe COVID-19 include advanced age,[12 13] cardiovascular disease, chronic kidney disease, diabetes, and laboratory findings such as lymphopenia, thrombocytopenia, and elevated inflammatory markers.[14–18]

Machine learning has been used to further the understanding of COVID-19, including for disease diagnosis[19–26] and transmission.[27–30] Further, an April 2020 systematic review by Wynants et al[31] discussed 50 published models for predicting disease progression or severity, but recommended none for clinical practice due methodological limitations including small sample sizes, inadequate training versus testing cohorts, or other factors leading to high risk of bias or limited external validity. A January 2021 review of artificial intelligence (AI) applications for COVID-19 by Tayarani et al[19] reviewed 14 additional studies of machine learning for predicting COVID-19 severity and found promise in works using demographics, laboratory values, and other electronic health record (EHR) data. Online calculators have been published with some studies.[15 16]

However, there remains a lack of standardization on how to predict an individual's disease trajectory and risk of severe illness. Thus, assessing the relative weight of risk factors in any particular patient's case has remained largely a provider-level task. Our goal in this work is to develop a tool to aid in risk assessment for progression to severe disease. Specifically, we aimed to analyze demographic and clinical data with statistical and machine learning techniques, and to develop a prediction score, usable

at the bedside by non-experts, to stratify the risk of progression to intubation within the next 48 hours for patients hospitalized with COVID-19.

METHODS
Methods and results are reported in accordance with the 2015 statement for Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.[32]

Data source
De-identified patient-level data were provided via a hospital-affiliated clinical data warehouse. Patients testing positive for SARS-CoV-2 at three academic medical centers in Arizona between January and April 2020 were included. Extracted variables included age, sex, vital signs, laboratory values (including blood counts, electrolytes, blood gas results, and inflammatory markers), oxygen requirements, and timing of intubation.

Patient comorbidities were extracted to describe the study cohort. However, knowledge of comorbidities is dependent on prior interaction with the healthcare system and on patient reporting or availability of medical records. This information may be unavailable at the urgent point of care. Accordingly, comorbidity information was omitted from model training to build a score robust to the incomplete data that may be available in times of health system crisis. Further, data on self-identification of race and ethnicity were not reliably available within the electronic medical record (EMR), so race and ethnicity were not considered in modeling.

Data preprocessing and missing data
Data were reformatted into 4-hour time blocks (rows). A 4-hour interval was chosen to match the frequency of routine vital sign checks in non-ICU units, representing the highest data sampling rate that was likely to be available across the population. Vital signs were then summarized as mean, minimum, and maximum for each block, as well as the initial value recorded on presentation for each patient. Laboratory values, measured less frequently, were represented as current and initial values. Respiratory support other than intubation was quantified by fraction of inspired oxygen (FiO_2) and oxygen delivery device (ie, nasal cannula, face mask, high-flow humidified cannula, etc). Where necessary, FiO_2 was estimated as 0.21 (room air) plus an additional 0.04 for every 1 L/min increase in oxygen flow rate.[33] Each "row" (4-hour block) was labeled with whether the patient required intubation within the subsequent 48 hours, as well as the number of hours from the end of that time block until the time of their intubation. Where values were missing, the last measured vital signs were carried forward for up to 12 hours and the laboratory values for up to 72 hours. Otherwise, missing fields were left blank. Rows with greater than 85% missing values were excluded. Parameters were excluded from modeling if they were populated in fewer than 15% of rows. This left 67 parameters for use in model training, including the initial and summary values as separate model inputs. Bivariate comparisons between the intubated and non-intubated groups were done using the χ^2 test for categorical data and the Mann-Whitney U test for continuous data. A complete list of the parameters initially considered in modeling, prior to elimination of those with low prevalence in the data set, is available in the online supplemental information. Finally, the data were randomly split into 80% training and 20% testing sets.

Modeling
The primary outcome used in model development was whether or not the patient was intubated within 48 hours of the end of
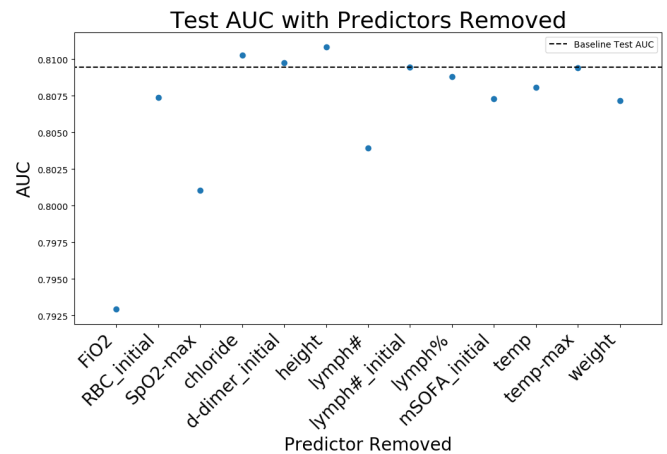

Figure 1 Impact of the top 13 parameters on LASSO model performance during training. Model AUC is plotted after dropping each of the top 13 parameters in turn. A lower postelimination AUC indicates the feature is more important in the model. Parameters yielding minimal reduction (<0.002) or an increase in AUC on elimination were removed from the final model. The dotted line, "baseline test AUC", shows the AUC of the model with all 13 parameters included. AUC, area under the receiver operating characteristic curve; FiO_2, fraction of inspired oxygen; LASSO, least absolute shrinkage and selection operator; lymph#, lymphocyte count; mSOFA_initial, initial modified Sequential Organ Failure Assessment score; RBC_initial, initial red blood cell count; SpO$_2$-max, maximum oxygen saturation for the 4-hour block; temp, temperature; weight, body weight.

each 4-hour time block. A patient's physiological state during each time block was considered as a separate model input, such that each "row" formed an independent training example. Model performance was assessed by the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and qualitatively for ease of application in clinical practice.

LASSO regression
A least absolute shrinkage and selection operator (LASSO) regularized linear regression model was trained.[34] The regularization parameter α had little impact on AUC, but affected the number of non-zero weights (sparsity) and specificity of the model. We noted a sharp drop in specificity as α approached 1, so $\alpha=0.1$ was selected to minimize the number of non-zero weights without sacrificing specificity. This resulted in the inclusion of 10 to 15 predictors, depending on the training and testing data split. Feature importance was then explored by rerunning the model across 100 randomizations of the training and testing data split. Thirteen parameters were used in >50% of model runs, and these were then used in an elimination algorithm where model performance was tested after dropping each parameter in turn (figure 1). Features with minimal (<0.002) reduction in AUC or with high potential for clinical redundancy (such as current temperature and maximum temperature) were removed, leaving only seven predictors: fraction of inspired oxygen (FiO_2), initial red blood cell count (RBC_initial), maximum oxygen saturation for the 4-hour block (SpO$_2$-max), lymphocyte count (lymph#), initial modified Sequential Organ Failure Assessment score (mSOFA_initial), current temperature (temp), and body weight (weight). LASSO was run again with just these seven predictors, and

all predictors except mSOFA_initial had non-zero coefficient values, resulting in a model with just six predictors.

Model performance was then assessed on the testing cohort. The CIs for LASSO performance were bootstrapped using the empirical bootstrap,[35] where the testing set was resampled with replacement 1000 times, and 95% confidence bands were calculated and plotted using the simultaneous joint confidence regions method.[36]

XGBoost

An eXtreme Gradient Boosting (XGBoost) model was trained.[37] Model tuning initially focused on sensitivity and sparsity. Bracketing algorithms were used to select the optimal values for scale_pos_weight (to more heavily weight cases of intubation, given the preponderance of negative examples in the data set), maximum tree depth (to optimize model complexity), as well as the regularization parameter γ. Given the goal of a bedside-usable prediction score, we initially focused on building a single-tree model. However, this yielded performance inferior to LASSO regression, with an AUC of 0.74, sensitivity of 0.88, and specificity of 0.60, so a more complex model using 100 trees was tuned. Tuning this model for maximum sensitivity rather than sparsity (γ=0) yielded a model with improved performance, as described in the Results section. Feature importance was explored by gain in model performance.

RESULTS
Cohort

There were 3447 patient encounters meeting the inclusion criteria, of which 20.7% required intubation. The baseline cohort characteristics regarding comorbidities and all parameters used in model training are presented in table 1. After

Table 1 Cohort initial characteristics

Characteristics	Total (N=3447)	Not intubated (n=2733)	Intubated (n=714)	P value
Demographics				
Age	49.4	47.0	58.4	0.01
Percent female	51.5	54.3	40.6	<0.01
Weight (kg)	89.3	89.2	89.6	0.37
Body mass index	32.8	32.6	33.6	<0.01
Comorbidities (not used in modeling)				
Chronic Obstructive Pulmonary Disease (%)	7.9	6.1	14.8	<0.01
Asthma (%)	13.3	13.3	13.3	<0.01
Heart failure (%)	11.3	7.2	26.9	<0.01
Diabetes (%)	34.7	29.4	54.9	<0.01
Labs and vital signs				
Bilirubin (mg/dL)	0.65	0.64	0.69	<0.01
BUN (mg/dL)	24.33	23.60	27.14	0.40
BUN:creatinine ratio	23.85	23.55	25.01	0.02
Calcium (mg/dL)	8.68	8.73	8.49	<0.01
Chloride (mEq/L)	101.38	101.42	101.25	0.01
D-dimer (ng/mL)	188.57	173.81	245.09	<0.01
Fraction of inspired oxygen (%)	58.96	59.05	58.63	<0.01
Glasgow Coma Scale	13.39	13.68	12.29	<0.01
Glomerular filtration rate (mL/min)	82.09	83.27	77.59	<0.01
Glucose (mg/dL)	146.36	143.33	157.98	0.35
Hematocrit (%)	35.86	35.67	36.61	0.01
Hemoglobin (g/dL)	11.83	11.77	12.06	0.03
Lymphocyte count (10^9/L)	1.29	1.32	1.17	<0.01
Lymphocytes (%)	16.40	17.15	13.52	<0.01
Mean corpuscular hemoglobin (g/dL)	29.26	29.26	29.24	<0.01
Mean corpuscular hemoglobin concentration (g/dL)	32.53	32.55	32.46	0.46
Mean corpuscular volume (fL)	89.88	89.84	90.01	<0.01
Mean platelet volume (fL)	10.52	10.51	10.58	<0.02
mSOFA score	3.62	3.18	5.30	0.28
Oxygen saturation (%)	93.85	93.93	93.54	0.16
Platelet count (10^9/L)	257.96	257.30	260.48	<0.01
Red blood cell count (10^{12}/L)	4.07	4.06	4.09	0.35
Red cell distribution width (%)	14.76	14.80	14.62	0.02
Red cell distribution width, SD (fL)	47.86	47.95	47.50	<0.01
Temperature (°F)	98.42	98.40	98.50	<0.01
White cell count (thousand/μL)	89.24	89.15	89.57	0.37

The table shows the distribution of comorbidities and mean initial physiological metrics used in model training for the overall cohort, intubated patients and non-intubated patients.
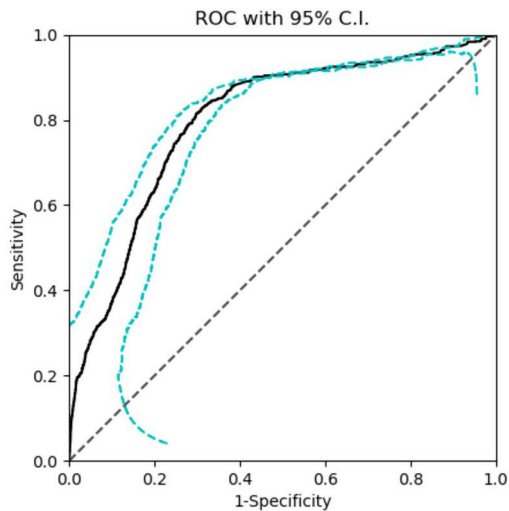BUN, blood urea nitrogen; mSOFA, modified Sequential Organ Failure Assessment.

Figure 2 ROC for the final LASSO model (ECoRRS score). The model predicts intubation within the subsequent 48 hours based on six clinical parameters. The AUC is 0.789 (95% CI 0.785 to 0.812). 95% confidence bands (dotted curves) are shown, calculated via the simultaneous joint confidence regions method. AUC, area under the receiver operating characteristic curve; ECoRRS, Early COVID-19 Respiratory Risk Stratification; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic curve.
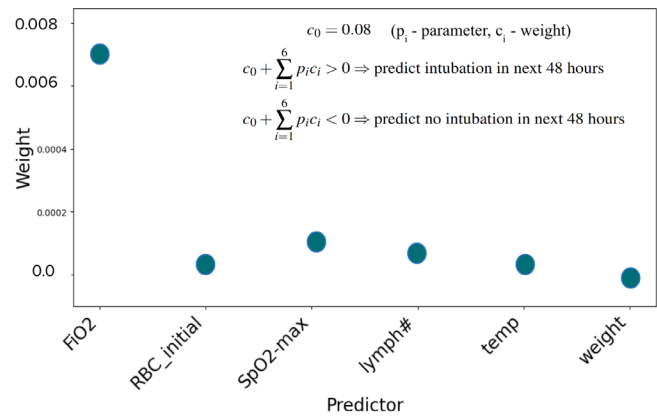

Figure 3 Parameter weights for the final LASSO model (ECoRRS score). The six parameters included in the final LASSO model are shown versus their model weights: FiO_2, RBC_initial, SpO_2-max, lymph#, temp, and weight. The model is applied by summing the value of each parameter multiplied by its coefficient and adding the constant 0.08. Positive values predict intubation within the next 48 hours, and negative values predict no intubation within the next 48 hours. FiO_2 was the most significant predictor of intubation. ECoRRS, Early COVID-19 Respiratory Risk Stratification; FiO_2, fraction of inspired oxygen; LASSO, least absolute shrinkage and selection operator; lymph#, current lymphocyte count; RBC_initial, initial red blood cell count at admission; SpO_2-max, maximum oxygen saturation for the 4-hour block; temp, current temperature; weight, body weight.

data preprocessing as discussed in the Methods section, the average missing data rate was 57% across the 4-hour time blocks, with an SD of 30%. We considered all patients who did not have a documented intubation to be in the non-intubated cohort, so there were no unclassified patients with respect to intubation.

LASSO regression

LASSO modeling, optimized for sensitivity and sparsity (ie, minimization of the number of inputs required), yielded an AUC of 0.798, with 95% CI of 0.785 to 0.812 (figure 2). At the 90% sensitivity operating point, we observed a specificity of 61.7% (95% CI 0.524 to 0.710), NPV of 0.997 (95% CI 0.996 to 0.998), and PPV of 0.040 (95% CI 0.033 to 0.047). The six parameters included in the final LASSO model were FiO_2, RBC_initial, SpO_2-max, current lymph#, current temperature (temp), and body weight (weight). The relative weights of each predictor are shown in figure 3. FiO_2 was the most significant predictor, followed by maximum oxygen saturation (SpO_2). The score is calculated by summing the value of each predictor multiplied by its coefficient and adding the constant (C0). If necessary, FiO_2 is estimated as 0.21 (room air) plus an additional 0.04 for each 1 L/min increase in oxygen flow rate.[33] Positive values predict intubation within the next 48 hours, and negative values predict no intubation within the next 48 hours. The greater the magnitude of the score, the greater the certainty of the prediction.

XGBoost

XGBoost classification tree modeling, optimized for sensitivity and trained on all parameters in the data set, yielded an AUC of 0.86, with a sensitivity of 0.99 at a specificity of 0.74. The NPV was 0.999 and the PPV was 0.082. Of the parameters, FiO_2 was consistently the most important by gain in model performance. The final model used 100 unique trees which combine to produce the prediction. A subsection of one of these tree diagrams is shown in figure 4.

Early COVID-19 Respiratory Risk Stratification prediction score

Both models were highly unlikely to undertriage patients, with NPV of 99.7% (LASSO) and 99.9% (XGBoost). The XGBoost model, however, achieved approximately double the PPV of LASSO and thus is less likely to overtriage patients (ie, indicate a need for intubation when the patient will not be intubated within the specified time frame). Given its complexity, the XGBoost model would require clinicians to enter a large number of variables into a specialized software program to see a prediction; this presents a significant barrier to rapid deployment for emergency triage. In contrast, the LASSO model, with only six parameters, can be used by any practitioner with a simple calculator or spreadsheet program. Thus, we present the LASSO model as the Early COVID-19 Respiratory Risk Stratification (ECoRRS) score. The coefficients and constant to calculate the ECoRRS score are shown in table 2. Positive results predict the need for intubation within 48 hours, and negative results predict no intubation within that time frame. The greater the magnitude of the score, the greater the certainty of the prediction.

DISCUSSION

We analyzed EHR data with two methods, LASSO regularized linear regression and XGBoost classification trees, to predict intubation within the next 48 hours for patients hospitalized with COVID-19. Both models achieved high sensitivity and very low rates of undertriage. XGBoost performed as well or better on all metrics compared with LASSO. However, given the marked simplicity and sparsity of LASSO relative to XGBoost, the LASSO model, which uses six objective inputs, is presented as the ECoRRS score.

Figure 4 Representative portion of a single tree from the XGBoost model. The final model contains 100 unique trees which combine to yield the model prediction. Terminal node ("leaf") values represent the log odds of the probability of intubation. To arrive at the predicted probability, the values of the appropriate leaves of each tree in the model are summed and transformed into a probability using the logistic function. XGBoost, eXtreme Gradient Boosting; HCT, hematocrit.

The ECoRRS score can be used to predict intubation and forecast resource utilization up to 48 hours in advance, which has implications for both individual patient care and for system-wide planning and staffing. The score tolerates overtriage to maximize sensitivity, identifying a subpopulation "at risk" of intubation. At the system level, however, hospitals can multiply the number of patients scoring positive on ECoRRS by the model's PPV and arrive at a relatively precise estimate of the number of inpatients likely to newly require a ventilator within the next 48 hours. This can facilitate timely redistribution of staff and resources to the areas of greatest need.

With regard to individual patient care, our framework relies on objective measurements and not patient history or comorbidities, which may be unavailable at the urgent point of care. Additionally, relying on objective measures, rather than subjective assessments by healthcare providers, supports the utility of ECoRRS as a triage tool for use by personnel with minimal healthcare training when systems are overburdened.

Multiple other investigators have sought to develop predictive algorithms for COVID-19 disease severity. Notably, Marcos and colleagues[16] developed an open-source online calculator using just nine variables to classify patients as high or low risk for severe disease, using a methodology similar to that presented here. Our model differs in that it provides prediction of intubation specifically within a 48-hour window and does not rely on knowledge of comorbidities to predict disease trajectory.

This study has multiple limitations. First, the indications for intubation were not protocolized and the decision to intubate was at the treating clinician's discretion. Thus, differences in individual practice may have impacted the study's results. Further, COVID-19 treatment has evolved since our data collection period (January–April 2020). Prone positioning, which has historically been used as an adjunct for intubated patients with severe acute respiratory distress syndrome,[38 39] came into practice to improve oxygenation in non-intubated patients with COVID-19. Proning increased in popularity during our study period, but data on the precise rate and intensity of proning in our cohort were not available. Studies have shown that prone positioning improves oxygenation and possibly reduces mortality in COVID-19, but it is not clearly associated with a reduced need for intubation.[40 41] As the most powerful predictor of need for intubation in our cohort was FiO_2, it is likely that the benefits of proning would be reflected in FiO_2 requirements, allowing the score to remain useful with increased utilization of prone positioning.

Additionally, remdesivir was introduced for COVID-19 under emergency use authorization in May 2020 and full US Food and Drug Administration approval followed in October 2020.[42 43] However, subsequent studies have shown minimal impact of this drug on disease trajectory,[44] and we suspect remdesivir's introduction to have little impact on the ECoRRS score's generalizability. Convalescent plasma was also introduced in Spring 2020,[45 46] with significant hopes for modifying disease progression, although large trials subsequently found this treatment too was ineffective.[47 48] In contrast, glucocorticoids in patients requiring supplemental oxygen became standard of care during our study period, after the RECOVERY trial.[49] The impact of this major therapeutic is likely captured only in the latter half of our cohort.

Further, our data source is linked to both strengths and significant limitations. With assistance from a hospital-affiliated clinical data warehouse, we extracted real-world EHR data. Such data are notoriously challenging and often include high rates of missing or incorrect values.[50] Our average missing data rate of 57% is similar to that reported in previously reported studies, including an evaluation of blood pressure documentation in the EHR which was found to vary in missing rate from 0.1% to 52%.[50] These missing data may have led to bias in our conclusions and model performance. However, it also may reflect incomplete information that healthcare workers operate with on a regular basis.

Finally, our 3447 patients were from three academic hospitals located within the same state. Validation studies in a wider multicenter cohort are needed to better assess the external validity of the ECoRRS score. The authors plan to undertake this using data from geographically diverse and non-academic hospitals within the same health network, which spans 6 states and 30 facilities.

The contrast of the user-friendliness of the LASSO model versus the accuracy of the XGBoost model highlights an active challenge in healthcare machine learning and informatics. Although numerous algorithms have been developed for healthcare, few have been deployed in the clinical setting, leading some to question the hopes for AI in medicine.[51–53] Although EHR systems remain closed environments, the use of novel algorithms will require clinicians to manually enter data into a secondary

Table 2 Coefficients and constant to calculate the ECoRRS score

Predictor	Coefficient
FiO_2 (as fraction)	7.060×10^{-4}
Initial red blood cell count (million/μL)	1.950×10^{-5}
SpO_2-max (%)	1.045×10^{-4}
Current lymphocyte count (thousand/μL)	6.765×10^{-5}
Temperature (°C)	2.955×10^{-5}
Body weight (kg)	-1.117×10^{-5}
Constant (C0)	0.08

Based on the LASSO model, the score is calculated by summing the value of each predictor multiplied by its coefficient and adding the constant (C0). If necessary, FiO_2 is estimated as 0.21 (room air) plus an additional 0.04 for each 1 L/min increase in oxygen flow rate.[33] Positive values predict intubation within the next 48 hours, and negative values predict no intubation within the next 48 hours. The greater the magnitude of the score, the greater the certainty of the prediction. ECoRRS, Early COVID-19 Respiratory Risk Stratification; FiO_2, fraction of inspired oxygen; SpO_2-max, maximum oxygen saturation for the 4-hour block.

system or calculator, which creates a substantial barrier to algorithm deployment and also to building the infrastructure for ongoing model evaluation with new populations. A future with enhanced collaboration between EHR developers, researchers, and regulatory organizations[54] could facilitate more comprehensive model training, testing, and validation. Such collaboration could also allow algorithms processing large numbers of data inputs, such as our XGBoost model, to find utility in clinical practice.

CONCLUSION

The ECoRRS score enables non-specialists to identify patients with COVID-19 at risk of intubation within 48 hours with minimal undertriage and enables health systems to forecast new COVID-19 ventilator needs up to 48 hours in advance.

ORCID iD
Molly J Douglas http://orcid.org/0000-0002-2567-3274

REFERENCES

1 Lee CCM, Thampi S, Lewin B, Lim TJD, Rippin B, Wong WH, Agrawal RV. Battling COVID-19: critical care and peri-operative healthcare resource management strategies in a tertiary academic medical centre in Singapore. *Anaesthesia* 2020;75:861–71.

2 Grasselli G, Pesenti A, Cecconi M, Giacomo G, Antonio P, Maurizio C. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *JAMA* 2020;323:1545–6.

3 Miller IF, Becker AD, Grenfell BT, Metcalf CJE, Becker Alexander D, Jessica E. Disease and healthcare burden of COVID-19 in the United States. *Nat Med* 2020;26:1212–7.

4 Wu Z, McGoogan JM, Zunyou W, McGoogan Jennifer M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020;323:1239–42.

5 Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med* 2020;173:362–7.

6 Kasper MR, Geibe JR, Sears CL, Riegodedios AJ, Luse T, Von Thun AM, McGinnis MB, Olson N, Houskamp D, Fenequito R, *et al*. An outbreak of Covid-19 on an aircraft carrier. *N Engl J Med* 2020;383:2417–26.

7 Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, Tobin KA, Cerfolio RJ, Francois F, Horwitz LI, Petrilli Christopher M, Jones Simon A, Jie Y. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020;369:m1966.

8 Cates J, Lucero-Obusan C, Dahl RM, Schirmer P, Garg S, Oda G, Hall AJ, Langley G, Havers FP, Holodniy M, Jordan C, Cynthia L-O, Dahl Rebecca M, *et al*. Risk for In-Hospital Complications Associated with COVID-19 and Influenza - Veterans Health Administration, United States, October 1, 2018-May 31, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:1528–34.

9 Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, Safiya R, Hirsch Jamie S, Mangala N, *et al*. the Northwell COVID-19 Research Consortium. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the new York City area. *JAMA* 2020;323:2052–9.

10 Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, Holden KA, Read JM, Dondelinger F, Carson G, Docherty Annemarie B, Green Christopher A, *et al*. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985.

11 McMahon DE, Peters GA, Ivers LC, Freeman EE. Global resource shortages during COVID-19: bad news for low-income countries. *PLoS Negl Trop Dis* 2020;14:e0008412.

12 Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PGT, Fu H, *et al*. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 2020;20:669–77.

13 Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Williamson Elizabeth J, Krishnan B, *et al*. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.

14 Xiang B, Cong F, Yu Z, Bai S, Liu Z, Chen Q, Xu Y, Xia T, Gong S, Xie X, *et al*. Predicting COVID-19 malignant progression with AI techniques. *medRxiv* 2020:2020.03.20.20037325.

15 Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, Wong S, Li Y, Lu J, Liang H, *et al*. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun* 2020;11:3543.

16 Marcos M, Belhassen-García M, Sánchez-Puente A, Sampedro-Gomez J, Azibeiro R, Dorado-Díaz P-I, Marcano-Millán E, García-Vidal C, Moreiro-Barroso M-T, Cubino-Bóveda N, *et al*. Development of a severity of disease score and classification model by machine learning for hospitalized COVID-19 patients. *PLoS One* 2021;16:e0240200.

17 Nemati M, Ansary J, Nemati N. Machine-Learning approaches in COVID-19 survival analysis and Discharge-Time likelihood prediction using clinical data. *Patterns* 2020;1:100074.

18 Fernanda S, Satchiko H-SN, Batistados SE, Maciel SC, Ludovico GD. Predicting the disease outcome in COVID-19 pos- itive patients through machine learning: a retrospective cohort study with Brazilian data. *medRxiv* 2020:2020.06.26.20140764.

19 Tayarani N M-H. Applications of artificial intelligence in Battling against covid-19: a literature review. *Chaos Solitons Fractals* 2021;142:110338.

20 Manjurul AM, Alam Tasfiq E, Theodore T, Pedro H. Deep MLP-CNN model using Mixed-Data to distinguish between COVID-19 and Non-COVID-19. *Patients Symmetry* 2020;12:1526.

21 Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT, Abd EM, Hosny Khalid M, Ahmad S, Darwish Mohamed M, Songfeng L, Sahlol Ahmed T. New machine learning method for image-based diagnosis of COVID-19. *PLoS One* 2020;15:e0235187.

22 André B, João M, Thiago D, Alexan F. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *medRxiv* 2020:2020.04.04.20052092.

23 Zargari KA, Morteza H, Ali SS. COVID-Classifier: an automated ma- chine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images. *medRxiv* 2020.

24 Abed MM, Hameed AK, Begonya G-Z, Mostafa S, Maashi M, Al-Waisy AS, Subhi M, Mutlag AA, Le D-N. A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of COVID-19 based on X-ray images. *Computers, Materials and Continua* 2021;66:3290–10.

25 Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F. A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdiscip Sci* 2021;13:103–17.

26 Yazeed Z, Shira D-R, Noam S. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digital Medicine* 2021;4:1–5.

27 Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J, Vespignani A, Santillana M, Dianbo L, Leonardo C. Real-Time forecasting of the COVID-19 outbreak in Chinese provinces: machine learning approach using novel digital data and estimates from mechanistic models. *J Med Internet Res* 2020;22:e20285.

28 Gergo P, Imre F, Amir M, Pedram G, Richard G. COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics* 2020;8:890.

29 Kumar SL, Kumar JS. A gradient boosting machine learning approach in mod- eling the impact of temperature and humidity on the transmission rate of COVID-19 in India. *Appl Intell* 2020.

30 Clement JC, Vijayakumar P, Sriharipriya KC, Nan- dakumar R, 2021. A survey on mathematical, machine learning and deep learning models for COVID- 19 transmission and diagnosis IEEE reviews in biomedical engineering. Conference Name: IEEE Reviews in Biomedical Engineering 1.

31 Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Bonten MMJ, Dahly DL, Damen JAA, Debray TPA, Laure W, Ben VC, *et al*. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;369:m1328.

32 Collins GS, Reitsma JB, Altman DG, Moons KGM, Reitsma Johannes B, Altman Douglas G, Moons Karel GM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.

33 Wettstein RB, Shelledy DC, Peters JI, Wettstein Richard B, Jay P I. Delivered oxygen concentrations using low-flow and high-flow nasal cannulas. *Respir Care* 2005;50:604–9.

34 Robert T. Regression shrinkage and selection via the LASSO. *Journal of the Royal statistical Society Series B* 1996;58:267–88.

35 Bradley E. Computers and the theory of statistics: thinking the Unthinkable Siam review. 1979;21:460–80.

36 Sofus M, Foster P, 2004. Confidence bands for ROC curves. Methods and an Empirical Study Proceedings of the First Workshop on ROC Analysis in AI.

37 Tianqi C, Carlos G. *XGBoost proceedings of the 22nd ACM SIGKDD international Conference on knowledge discovery and data mining*, 2016.

38 Guérin C, Reignier J, Richard J-C, Beuret P, Gacouin A, Boulain T, Mercier E, Badet M, Mercat A, Baudin O, Claude G, Jean R, Jean-Christophe R, *et al*. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med* 2013;368:2159–68.

39 Beitler JR, Shaefi S, Montesi SB, Devlin A, Loring SH, Talmor D, Malhotra A. Prone positioning reduces mortality from acute respiratory distress syndrome in the low tidal volume era: a meta-analysis. *Intensive Care Med* 2014;40:332–41.

40 Weatherald J, Solverson K, Zuege DJ, Loroff N, Fiest KM, Parhar KKS. Awake prone positioning for COVID-19 hypoxemic respiratory failure: a rapid review. *J Crit Care* 2021;61:63–70.

41 Brigitta F, Alexandria P, Rupert P, Zudin P. Prone position for non- intubated spontaneously breathing patients with hypoxic respiratory failure: a systematic review and meta-analysis. *British Journal of Anaesthesia* 2021.

42 Commissioner Office of the. *Coronavirus (COVID-19) update: FDA issues emergency use Autho- rization for potential COVID-19 treatment*: FDA, 2020.

43 Commissioner Office of the. *FDA Approves first treatment for COVID-19*: FDA, 2020.

44 Consortium WHO Solidarity Trial. Repurposed antiviral drugs for Covid-19 interim who Soli- darity trial results. *N Engl J Med* 2020.

45 Wang Y, Zhang L, Sang L, Ye F, Ruan S, Zhong B, Song T, Alshukairi AN, Chen R, Zhang Z, *et al*. Kinetics of viral load and antibody response in relation to COVID-19 severity. *J Clin Invest* 2020;130:5235–44.

46 Casadevall A, Joyner MJ, Pirofski L-A. SARS-CoV-2 viral load and antibody responses: the case for convalescent plasma therapy. *J Clin Invest* 2020;130:5112–4.

47 Anup A, Aparna M, Gunjan K, Pranab C, Tarun B, Pankaj M. Convalescent plasma in the management of moderate covid-19 in adults in India: open label phase II multicentre randomised controlled trial (PLACID trial). *BMJ* 2020;371:m4232.

48 Pathak EB, Pathak Elizabeth B. Convalescent plasma is ineffective for covid-19. *BMJ* 2020;371:m4072.

49 RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, Staplin N, Brightling C, Ustianowski A, *et al*. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med* 2021;384:693–704.

50 Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev* 2010;67:503–27.

51 Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–43.

52 Trishan P, Heather M, Anthony CL. The inconvenient truth about AI in healthcare. *npj Digit Med* 2019;2:1–3.

53 Cosgriff CV, Stone DJ, Weissman G, Pirracchio R, Celi LA. The clinical artificial intelligence department: a prerequisite for success. *BMJ Health Care Inform* 2020;27:e100183.

54 Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med* 2018;15:e1002689.

55 Perez F, Granger BE. IPython: a system for interactive scientific computing. *Comput Sci Eng* 2007;9:21–9.

56 Fabian P, Gaël V, Alexandre G, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al*. Scikit-learn: machine learning in python. *Journal of Machine Learning Research* 2011;12:2825–30.

57 Eric J, Travis O, Pearu P. *{SciPy}: Open source scientific tools for {Python}*, 2001.

58 Wes M. *Data structures for statistical computing in python proceedings of the 9th python in science conference. (SCIPY 2010)*, 2010:6.

59 Wes M. *Pandas: a foundational python library for data analysis and statistics python high performance science computer*, 2011.