PLOS ONE

eARDS: A multi-center validation of an interpretable machine learning algorithm of early onset Acute Respiratory Distress Syndrome (ARDS) among critically ill adults with COVID-19

Lakshya Singhal[1☯], Yash Garg[1☯], Philip Yang[2☯], Azade Tabaie[1], A. Ian Wong[2], Akram Mohammed[3], Lokesh Chinthala[3], Dipen Kadaria[4], Amik Sodhi[4], Andre L. Holder[2], Annette Esper[2], James M. Blum[1,5], Robert L. Davis[3], Gari D. Clifford[1,6], Greg S. Martin[2], Rishikesan Kamaleswaran[1,6]*

1 Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia, United States of America, **2** Division of Pulmonary, Critical Care, Allergy, and Sleep Medicine, Emory University School of Medicine, Atlanta, Georgia, United States of America, **3** Department of Pediatrics, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **4** Department of Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **5** Department of Anaesthesia, Emory University School of Medicine, Atlanta, Georgia, United States of America, **6** Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States of America

☯ These authors contributed equally to this work.

* rkamaleswaran@gatech.edu

Abstract

We present an interpretable machine learning algorithm called 'eARDS' for predicting ARDS in an ICU population comprising COVID-19 patients, up to 12-hours before satisfying the Berlin clinical criteria. The analysis was conducted on data collected from the Intensive care units (ICU) at Emory Healthcare, Atlanta, GA and University of Tennessee Health Science Center, Memphis, TN and the Cerner® Health Facts Deidentified Database, a multi-site COVID-19 EMR database. The participants in the analysis consisted of adults over 18 years of age. Clinical data from 35,804 patients who developed ARDS and controls were used to generate predictive models that identify risk for ARDS onset up to 12-hours before satisfying the Berlin criteria. We identified salient features from the electronic medical record that predicted respiratory failure among this population. The machine learning algorithm which provided the best performance exhibited AUROC of 0.89 (95% CI = 0.88–0.90), sensitivity of 0.77 (95% CI = 0.75–0.78), specificity 0.85 (95% CI = 085–0.86). Validation performance across two separate health systems (comprising 899 COVID-19 patients) exhibited AUROC of 0.82 (0.81–0.83) and 0.89 (0.87, 0.90). Important features for prediction of ARDS included minimum oxygen saturation (SpO_2), standard deviation of the systolic blood pressure (SBP), O_2 flow, and maximum respiratory rate over an observational window of 16-hours. Analyzing the performance of the model across various cohorts indicates that the model performed best among a younger age group (18–40) (AUROC = 0.93 [0.92–0.94]), compared to an older age group (80+) (AUROC = 0.81 [0.81–0.82]). The model

performance was comparable on both male and female groups, but performed significantly better on the severe ARDS group compared to the mild and moderate groups. The eARDS system demonstrated robust performance for predicting COVID19 patients who developed ARDS at least 12-hours before the Berlin clinical criteria, across two independent health systems.

Introduction

The novel coronavirus-2019 disease (COVID-19) pandemic has led to a disruptive global health crisis with significant morbidity and mortality. It has placed a significant burden on the healthcare system, with about 15–29% of COVID-19 cases requiring hospitalization, and about 17–35% of inpatients requiring critical care [1–4]. The morbidity and mortality among the critically ill patients with COVID-19 is particularly high, especially related to respiratory failure and acute respiratory distress syndrome (ARDS). Some prior studies have suggested that the risk of ARDS in mechanically ventilated patients with COVID-19 ranges 40–100% [5–7], and the mortality in those requiring mechanical ventilation is reported to be as high as 50–97%—higher than the mortality rates from other causes of ARDS, including H1N1 influenza [5, 6, 8].

Although ARDS secondary to COVID-19 may satisfy the Berlin definition of ARDS, some features that appear distinct from "classic" ARDS have also been suggested [9, 10]. Such differences include preservation of the respiratory system compliance despite severe hypoxemia in some patients [9, 11], as well as relatively delayed timing of onset compared to the 7-day period included in the Berlin definition [12]. Based on the differences in respiratory system compliance and hypothesized mechanisms of hypoxemia, some studies have proposed subphenotypes of COVID-19 induced ARDS that may behave differently from "classic" ARDS, such as the high- and low-elastance phenotypes [9, 13]. The differences between the subphenotypes were also corroborated by a study of computed tomographic examinations of the lungs in COVID-19 and non-COVID-19 ARDS patients [14]. Although the validity and the clinical significance of these differences between COVID-19 ARDS and "classic" ARDS is uncertain and debatable [15], they nonetheless highlight the heterogeneity in COVID-19 induced respiratory failure and ARDS. The heterogeneity and potentially distinct features of COVID-19 ARDS, combined with the aforementioned high mortality, present unique challenges for its diagnosis, risk-stratification, and management. These new challenges are also applicable for predictive modeling in ARDS. While several prior studies have utilized machine learning models to identify and/or predict general ARDS [16–19], these models have not been trained or validated on populations containing patients with COVID-19.

The high incidence of and mortality from ARDS in COVID-19 highlights an important need for early prediction and recognition of ARDS in this population. Machine learning models for predicting ARDS that are validated in COVID-19 patients have the potential to improve early identification of patients who are at high risk of disease progression and promote timely implementation of indicated treatments. The objective of this study was to address this need by developing and validating a machine learning model for early prediction of ARDS development and its severity in COVID-19 patients before they satisfy the clinical definition of ARDS. In this paper, we introduce a machine learning model called *eARDS*, which predicts the onset of ARDS in critically ill COVID-19 patients up to 48 hours before meeting the Berlin definition.

Materials and methods

This study was approved by the Institutional Review Board at Emory University, Atlanta GA (#IRB00033069), and The University of Tennessee Health Science Center, Memphis TN (#20-07294-XP).

Description of the datasets

The model was derived from the Cerner Real-World Data, consisting of de-identified information from hospitals within the Cerner environment, and was evaluated on patients who developed ARDS during hospitalization at the Emory Healthcare and the UTHSC-Methodist LeBonheur Healthcare systems for patients with a positive SARS-CoV-2 result by qRT-PCR. The eARDS model was trained using the de-identified Cerner Real-World Data™ (CRD), which consists of both COVID19 and non-COVID19 patients from hospitals across various geographic regions and demographics. Data was captured between January till April 2020, and did not include any patients from the Emory healthcare or UTHSC-MLH system.

For validation of the eARDS, we derived data from 767 COVID-19 patients admitted across 4 hospitals within the Emory Healthcare system, Atlanta, GA, and 132 COVID-19 patients admitted across 5 hospitals within the Methodist LeBonheur Healthcare (MLH) system, Memphis TN. Demographic information, medical comorbidities, vital signs, laboratory data, and other clinical information abstracted from the electronic health records (EHR) (S1 Table) were selected from admission till the onset of ARDS from the ICU. These variables were selected based on the literature pertaining to the prediction of ARDS. We extracted data for patients from February to June 1, 2020.

Selection criteria

All patients above 18 years of age who were admitted to the ICU diagnosed with SARS-CoV-2 with at least 48 hours of data were included in the study. Where we observed multiple encounters for the same patient, we treated each encounter as independent if the admissions were at least 30 days apart, and used the first admission. We used Current Procedural Terminology (CPT) and International Classification of Diseases (ICD-10) to identify mechanical ventilation and oxygen therapy. The onset time of ARDS is measured using the Berlin criteria [20], in which we identify t_{Onset} as when the patient requires positive end-expiratory pressure (PEEP) of at least 5 cmH$_2$O and a ratio of arterial partial pressure oxygen to fraction of inspired oxygen (P/F) ratio $< = 300$, within 1 week of initial oxygen support. We further exclude patients based on the percentage missing data i.e., if the percentage *null* value is greater than 90%. We further segregate ARDS patients by severity, using the worst P/F during an encounter, specifically the severity classes were: mild (P/F ratio $> = 200$ and <300), moderate (P/F ratio $> = 100$ and <200), and severe (P/F ratio < 100).

Missing data and class imbalanced

All the missing values of laboratory and vital sign measurements were filled using last-one carry forward imputation and remaining missing values were imputed by the global median of the associated variable in the training dataset (CRD dataset) and validation dataset (Emory and UTHSC-MLH dataset). Separate binary variables were generated to indicate a positive binary value during their oxygen therapy, mechanical ventilation, vasoactive status and for their comorbidities. Class imbalance was addressed through balanced micro batching, in which we balance ARDS patients with an equal random set of Non-ARDS patients.

Data preprocessing

The data of each patient was segmented and resampled at an evenly sampled 2-hour interval. Their median value replaced the multiple measures available for a single variable within the 2-hour interval. Leaving an interval of 12-hours before ARDS onset as the *prediction window*, we further segment data in 16-hour *observational windows* to extract statistical features. From continuous variables, we extracted features including, minimum, maximum, standard deviation, median, and skewness. For categorical variables, such as the presence of therapies or medication, we generated a binary flag to indicate the presence of the variable at the appropriate time intervals. In this manner, we obtained 148 statistical features that are provided as inputs to our model.

Model development and evaluation

We developed our machine learning models on Python using the Scikit-Learn [21] and the XGBoost package [22]. Data management was performed using the Pandas library [23]. During the course of the machine learning pipeline, we evaluated a number of machine learning methods including, Neural Networks [24], Support Vector Machines [25], Random Forests [26], Logistic Regression [27], and eXtreme Gradient Boosting (XGBoost) [28]. We then selected the XGBoost model due to the robust and superior performance across the internal and external validations. In the derivation of eARDS was performed in two steps, first, a model was trained on 80% of the CRD database (training set) which consisted of ARDS in patients positive for SARS-CoV-2. The remaining 20% were preserved as a hold-out dataset, this process was repeated with a random selection of patients and repeated 10 times with replacement to generate average training performance. Prior to training a model, we used a subset of the training data (30%) for hyperparameter selection using Bayesian optimization.

We then validated the model on retrospective data collected on COVID-19 patients from Emory and UTHSC-MLH datasets. We selected a random 80% of the dataset and repeatedly sampled from this dataset 10 times with replacement to generate confidence intervals of the performance statistics.

Feature importance and model interpretability

A popular recent method for explaining machine learning is by the use of SHapley Additive exPlanations (SHAP) [29], which uses optimal credit allocations among entities to derive their contributions, a game theory centric method for feature importance at the prediction level. We used SHAP to extract prediction level explanations along with mean SHAP values generated across predictions to develop interpretations of important predictors.

Results

Clinical characteristics

Table 1 shows the clinical characteristics of the study population. In the training dataset, 35,804 patients were available, of which 14,097 met inclusion criteria. 1,890 patients (13.4%) met ARDS criteria, and of that 964 were positive for SARS-CoV-2. Among the 12,207 patients who did not meet ARDS criteria, 4,712 were positive for SARS-CoV-2. The median age of ARDS patients was 66 [54, 77] while for Non-ARDS, it was 60 [44, 73], the number of males (%) for ARDS is 1049 (56%) and for Non-ARDS is 5,966 (49%). Statistically significant differences ($P < 0.05$) were found among gender, race, and ethnicity in the training dataset.

In the validation datasets, a total of 767 COVID-19 patients were available from Emory Healthcare, and 611 met the inclusion criteria and were included in the analysis. Of these, 145

Table 1. Characteristics of patients in the datasets.

	Emory		UTHSC-MLH		Cerner Real-World Data	
	ARDS	**NON-ARDS**	**ARDS**	**NON-ARDS**	**ARDS**	**NON-ARDS**
Patients, n	145	466	17	60	1890	12207
COVID-19, n (column %)	145 (100)	466 (100)	17 (100)	60 (100)	964 (51%)	4712 (39%)
Age, mean [IQR]	+67 [58, 77]	+61 [50, 73]	68 [59, 80]	62 [52, 73]	66 [54, 77]	60 [44, 73]
Age Groups, n (column %)						
18 yrs. to 40 yrs.	#10 (7%)	56 (12%)	0 (0%)	1 (2%)	#148 (8%)	2342 (19%)
41 yrs. - 60 yrs.	30 (21%)	155 (33%)	6 (35%)	26 (43%)	536 (28%)	3754 (31%)
61 yrs. - 80 yrs.	75 (51%)	194 (42%)	6 (35%)	24 (40%)	828 (44%)	4188 (34%)
81+ yrs.	30 (21%)	61 (13%)	5 (30%)	9 (15%)	378 (20%)	1923 (16%)
Male, number (column %)	82 (57%)	223 (48%)	8(47%)	30(50%)	#1,049 (56%)	5,966 (49%)
First Day Apache II Score*, mean [IQR]	19 [13, 26]	15 [12, 20]	19 [13, 26]	12 [8, 15]	19 [11, 34]	16 [8, 28]
Mechanical Ventilation, n (column %)	#94 (65%)	34 (7%)	#7 (41%)	5 (8%)	#889 (47%)	755 (6%)
P/F Ratio, mean (IQR)	+215 [158, 328]	341 [207, 341]	+200 [145, 340]	340 [272, 340]	+161 [107, 231]	312 [170, 393]
Race, n (column %)						
African American	94 (65%)	313 (68%)	15 (88%)	49 (82%)	#518 (27%)	3,100 (25%)
Caucasian	40 (28%)	92 (20%)	0 (0%)	7 (12%)	916 (49%)	6,407 (53%)
American Indian	0 (0%)	1 (0%)	0 (0%)	0 (0%)	31 (2%)	226 (2%)
Asian	2 (1%)	9 (2%)	0 (0%)	1 (2%)	59 (3%)	356 (3%)
Mixed Race	1 (0%)	0 (0%)	1 (6%)	1 (2%)	1 (0%)	0 (0%)
Unknown Race	8 (6%)	51 (10%)	1 (6%)	2 (3%)	365 (19%)	2117 (17%)
Ethnicity, n (column %)						
Not Hispanic or Latino	#134 (92%)	386 (83%)	16 (94%)	59 (98%)	#1,323 (70%)	8,461 (69%)
Hispanic or Latino	4 (3%)	22 (5%)	1 (6%)	1 (2%)	330 (17%)	2,518 (21%)
Unknown	7 (5%)	58 (12%)	-	-	237 (13%)	1228 (10%)

*Excludes Chronic Health Points due to lack of data availability.

+statistical significance by Wilcoxon Rank-Sum test (P < 0.05).

#statistical significance by Chi-square test (P < 0.05).

https://doi.org/10.1371/journal.pone.0257056.t001

were ARDS patients and 466 were Non-ARDS patients. There were 132 COVID-19 patients available from the UTHSC-MLH dataset, and 77 patients met inclusion criteria and were included in the analysis. Of these, 17 patients met the ARDS criteria. The average age among patients across both datasets was similar, with the UTHSC-MLH patients being a year older on average in both groups. Comparing ARDS and non-ARDS patients did not show statistical significance differences (P < 0.05) in age, gender, race, or ethnicity. Among the Emory dataset, however, we observed statistically significant differences in age and ethnicity (P < 0.05).

Data preprocessing

There were a number of challenges in the preprocessing of the CRD dataset. The most prominent of which were discrepancies among units in similar measures, for instance, FiO_2 values were present in both fractional and percentage forms at various periods, which required standardization, and temperature appeared in both Celsius and Fahrenheit. We further observed variable contamination, for e.g. heart rate with a unit of beats/min appeared labeled as 'mean arterial pressure', which required explicit unit validation. Some measures were erroneously high in this dataset which were categorized as values greater than 99 percentile value for that variable, these were imputed with the global median.
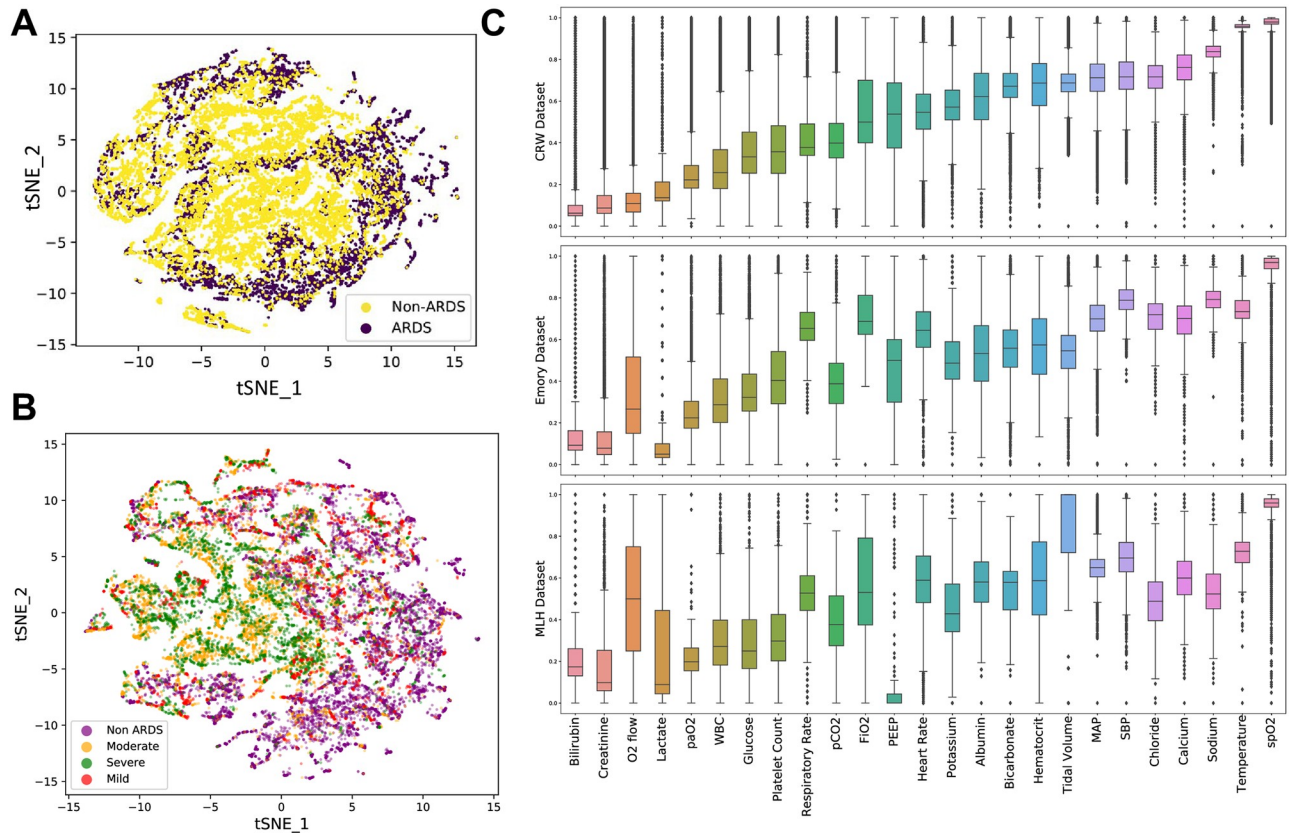
Fig 1. Characteristic analysis of the data. A. Illustrates a t-SNE plot of ARDS vs non-ARDS comparing across all predictors, two localised clusters emerge consisting of ARDS (yellow) patients; B. t-SNE plot of ARDS Severity, classified as mild, moderate and severe, significant convergence can be observed among the moderate and severe groups with non-ARDS and mild groups similarly forming a separate cluster. C. A box-plot of data density is illustrated, with each of the variables compared across the three data sources. Statistical significance is observed in all variables except PaO$_2$ and Chloride.

From validation datasets, we observed some challenges in preprocessing the UTHSC-MLH dataset, particularly as some structured data fields contained text data and/or comments relevant for clinical interpretations. Negative values were observed in fields, which were flagged as erroneous by clinical experts. In the Emory dataset, we observed inconsistencies in some values being present in fractional and percentage forms, which required standardization. Fig 1(A) illustrates the tSNE plot of ARDS vs Non-ARDS patients using the preprocessed data. Significant clustering around the center was observed for the ARDS patients (in yellow), while Non-ARDS (in purple) are clustered around the periphery. Fig 1(B) represents the tSNE plot between Non-ARDS and ARDS severity. A summary of the normalized preprocessed data across each of the three datasets are illustrated in box-plots in Fig 1(C), statistical differences were observed in all variables between the three groups, except for PaO$_2$ and Chloride. S1 Fig illustrates a box-plot figure of missingness, i.e. the percentage of missing values that existed for each variable across the different datasets.

Training and validation model performance

Performance measures derived from evaluating eARDS on the 20% CRD hold-out, and the validation results (UTHSC-MLH and Emory) are summarized in Table 2. As illustrated in Fig 2(A), at 12-hours before ARDS onset, the training model achieved an AUC [95% CI] of 0.89

Table 2. Analysis of model performance over 6-hour and 12-hour prediction horizon.

COVID-19 Positives (6 hours)			
Dataset	CRW	Emory	MLH
AUC	0.90 [0.889, 0.90]	0.85 [0.85, 0.86]	0.88 [0.86, 0.90]
Specificity	0.86 [0.86, 0.86]	0.81 [0.80,0.82]	0.82 [0.80, 0.84]
Sensitivity	0.77 [0.76,0.78]	0.74 [0.73, 0.76]	0.77 [0.74, 0.78]
PPV	0.54 [0.53, 0.54]	0.55 [0.54, 0.59]	0.56 [0.53, 0.59]
COVID-19 Positives (12 Hours)			
Dataset	CRW	Emory	MLH
AUC	0.89 [0.88, 0.90]	0.82 [0.81, 0.83]	0.89 [0.87, 0.90]
Specificity	0.85 [0.85 0.86]	0.80 [0.79, 0.82]	0.81 [0.79, 0.84]
Sensitivity	0.77 [0.75, 0.78]	0.70 [0.69, 0.71]	0.77 [0.74, 0.80]
PPV	0.53 [0.52, 0.54]	0.54 [0.52, 0.55]	0.55 [0.52, 0.59]
All-cause ARDS (12 Hours)			
Dataset	CRW	Emory	MLH
AUC	0.86 [0.85, 0.86]	0.80 [0.79, 0.81]	0.86 [0.84, 0.87]
Specificity	0.79 [0.79, 0.80]	0.77 [0.76, 0.78]	0.79 [0.76, 0.81]
Sensitivity	0.77 [0.76, 0.77]	0.70 [0.69, 0.71]	0.75 [0.72, 0.78]
PPV	0.45 [0.44, 0.45]	0.49 [0.48, 0.51]	0.51 [0.48, 0.55]

[0.88, 0.90] averaged over 10 iterations of bootstrap. The sensitivity, specificity and positive predictive value (PPV), of the hold-out CRD dataset were 0.77 [0.75, 0.78], 0.85 [0.85, 0.86] and 0.53 [0.52, 0.54], respectively.

For the same time period, among the validation datasets, we observed an AUC for Emory and UTHSC-MLH of 0.82 [0.81, 0.83] and 0.89 [0.87, 0.90], respectively, the sensitivity, specificity, and PPV are included in Table 2. At 6-hours before onset, AUC across the CRD (hold-out), Emory and UTHSC-MLH were, 0.90 [0.89, 0.90], 0.85 [0.85, 0.86], and 0.88 [0.86, 0.90] respectively (Table 2).

Fig 2(B) illustrates the temporal performance of the model tuned to different prediction horizons ranging from t_{onset} to 36 hrs prior. The retrospective temporal performance across Emory and UTHSC-MLH was observed to be different, with the Emory dataset suggesting a lower AUC in the hours preceding t_{onset} when compared to UTHSC-MLH which displays a better averaged performance. The aggregated performance on the UTHSC-MLH data was found to be highly variant, as illustrated with a broader 95% CI when compared to the Emory dataset.

Model interpretability

Some features and clinical information showed more significant importance than others for prediction of ARDS. Using mean SHAP values generated from the training data, we generated a ranked list of the top 20 important features (S3 Fig). Among the top 20, the top 8 features contributed more to the generation of an alert for ARDS than the next 12. The top feature for prediction of ARDS was SpO_2 (minimum), followed by SBP (standard deviation), younger age group (18–40), FiO_2 (max), respiratory rate (max), O_2 flow (max and standard deviation) and platelet count (min) in descending order.

Fig 3(A) illustrates an example patient who develops ARDS over a 48 hour period, SpO_2 (minimum), SBP (standard deviation), respiratory rate (maximum), O_2 flow (standard deviation) and heart rate (maximum) cause the probability to increase beyond the alert threshold up to 42 hours before the patient meets the severe ARDS criteria (t = 0). Resuscitative
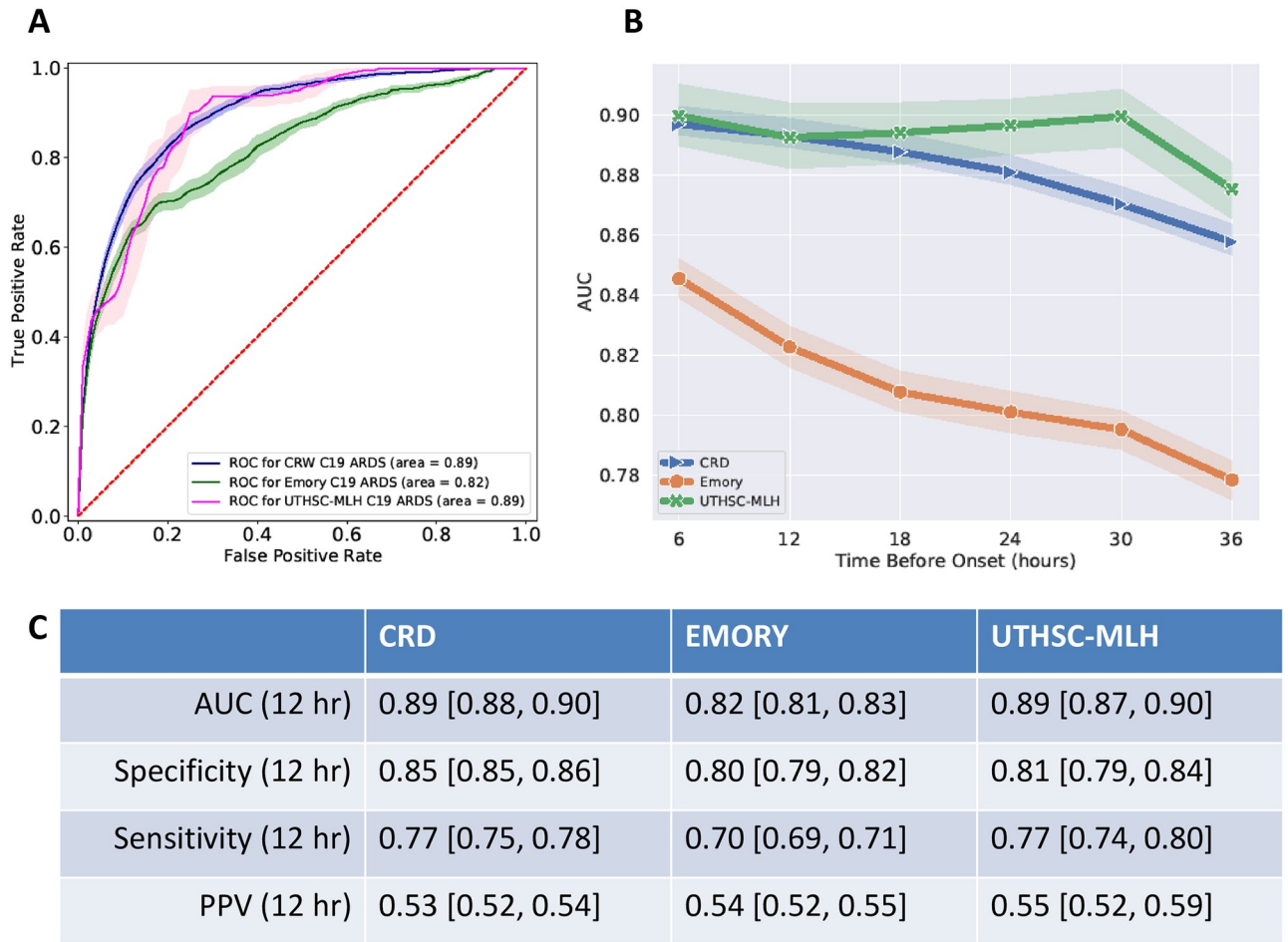
C	CRD	EMORY	UTHSC-MLH
AUC (12 hr)	0.89 [0.88, 0.90]	0.82 [0.81, 0.83]	0.89 [0.87, 0.90]
Specificity (12 hr)	0.85 [0.85, 0.86]	0.80 [0.79, 0.82]	0.81 [0.79, 0.84]
Sensitivity (12 hr)	0.77 [0.75, 0.78]	0.70 [0.69, 0.71]	0.77 [0.74, 0.80]
PPV (12 hr)	0.53 [0.52, 0.54]	0.54 [0.52, 0.55]	0.55 [0.52, 0.59]

Fig 2. Model performance and feature importance. A. The AUROC performance of the eARDS model over hold-out and validation datasets, 95% CI, are illustrated as shaded bands for each model. B. Temporal performance of the model over different prediction horizons ranging from t_{onset} to 36 hours before, MLH is shown to have the highest performance over the time periods, while Emory data has the lowest AUC consistently. C. Depicts the performance measures of the algorithm across the three datasets at 12-hours before ARDS onset.

interventions were observed in the hours leading to t_{Onset}. As illustrated in the figure, a series of interpretable readings, by the way of 'important features' are generated throughout the time period.

Fig 3(B) illustrates the clustered heatmap of the top 20 features among ARDS patients derived using the aggregated SHAP values for each prediction at six hours prior to t_{Onset}. The heatmap is clustered on disease severity, namely, mild, moderate and severe. Patients are enumerated column-wise, and as illustrated in the figure, more severe patients are grouped towards the right side of the heatmap while moderate and less severe are grouped in the left. Among the less severe cluster (left side, Fig 3(B)), minimum values of the SpO_2 and the standard deviation of SBP suppressed the probabilistic value (orange shade indicates suppression) as opposed to the severe patients group (right side, Fig 3(B)). Values of O_2 flow (both maximum and standard deviation) contributed positively to the alert in more severe ARDS. Among the less severe cohort (left side, middle, Fig 3(B)), Age of 18–40 contributes positively to the probabilistic value, in contrast to the more severe cohort. Minimum of PaO_2 and FiO_2 values in the observational window were particularly important among the severe cohort. Beyond the
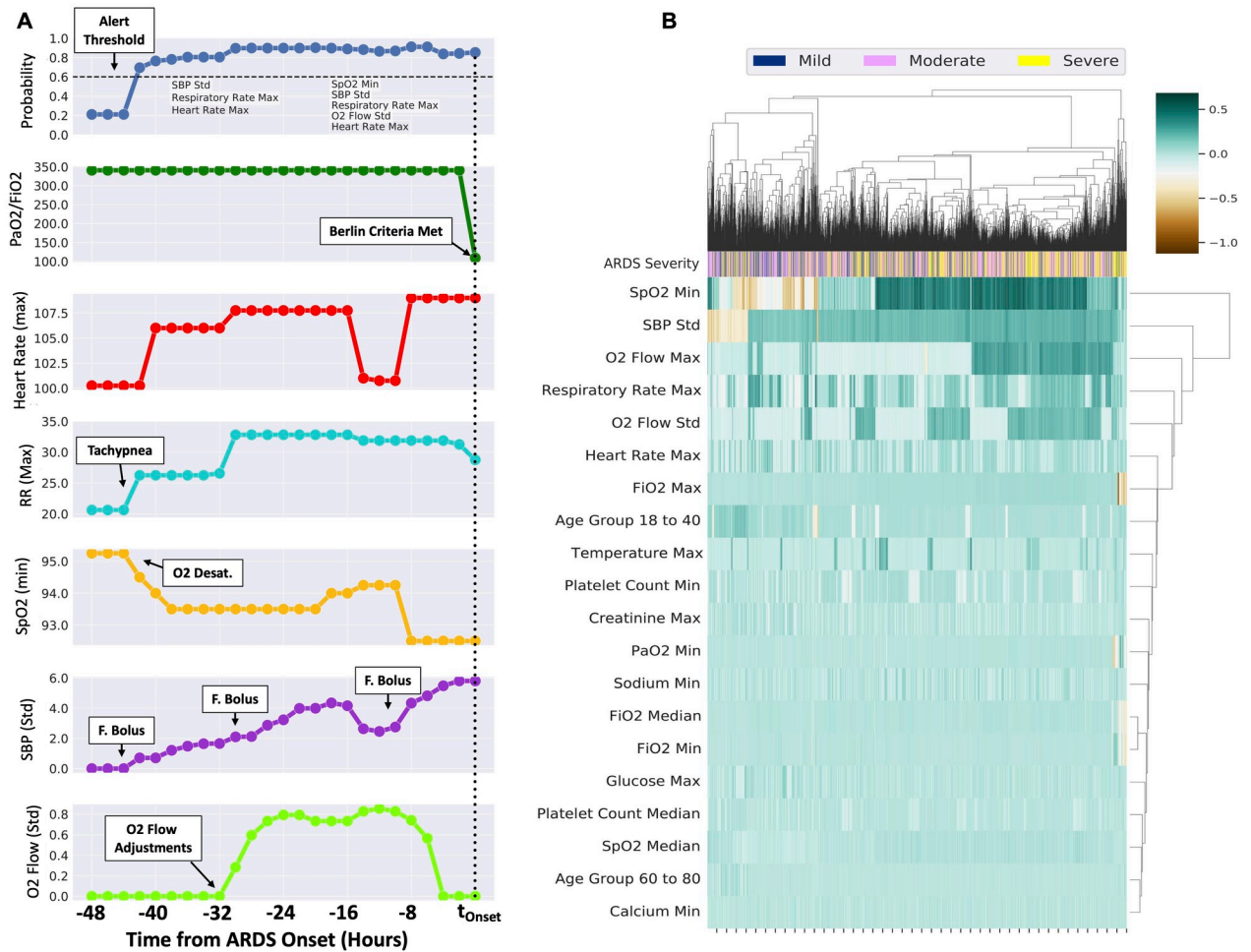
Fig 3. A. Example patient who develops an increased risk for ARDS over 48 hours before meeting the Berlin Criteria, the top panel illustrates probability and the second panel illustrates the PaO_2/FiO_2 ratio, while the third through last panels illustrate the top features identified in the model. Note that P/F ratio was not included in the model input, and illustrated for demonstration purposes. B. Illustrates the cluster heatmap of 20 most significant features derived by aggregating SHAP values for each prediction at six hours prior to t_{onset}. Patients are enumerated column-wise, and clustered on disease severity, namely, mild, moderate and severe. Two groups are observed, the left consisting of the mild and moderate groups, while the right consists more of the severe (yellow) group.

https://doi.org/10.1371/journal.pone.0257056.g003

top 10 features (SpO_2 (*min*) till platelet count (*min*)), the remaining features add incremental value to the overall prediction.

Cohort analysis of performance

Table 3 describes the comparisons among disease severity, which revealed statistical significance (P<0.001) among the classes, with the model performing better in severe ARDS (P/

Table 3. Analysis of model performance on the hold-out dataset against ARDS severity.

ARDS Severity					
Severity*	AUC	Sensitivity	Specificity	PPV	P-Value
Mild	0.83 [0.83, 0.84]	0.75 [0.74, 0.76]	0.78 [0.78, 0.79]	0.11 [0.11, 0.11]	p<0.001
Moderate	0.87 [0.87, 0.88]	0.79 [0.79, 0.80]	0.79 [0.78, 0.79]	0.27 [0.27, 0.27]	
Severe	0.91* [0.90, 0.91]	0.86 [0.85, 0.87]	0.79 [0.79, 0.80]	0.29 [0.28, 0.29]	

*statistical significance (P<0.05) by one-way ANOVA row-wise.

https://doi.org/10.1371/journal.pone.0257056.t003

Table 4. Analysis of model performance on the hold-out CRD dataset across demographics.

	AUC	Sensitivity	Specificity	PPV	P-Value
Gender					
Male	0.89 [0.89, 0.89]	0.85 [0.84, 0.85]	0.76 [0.76, 0.77]	0.46 [0.45, 0.46]	p>0.05
Female	0.89 [0.89, 0.89]	0.77 [0.76, 0.78]	0.84 [0.84, 0.84]	0.49 [0.48, 0.50]	
Age Groups*					
18–40	0.93 [0.92, 0.94]	0.81 [0.79, 0.82]	0.93 [0.93, 0.93]	0.51 [0.50, 0.52]	p<0.001
40–60	0.89 [0.89, 0.90]	0.82 [0.81, 0.82]	0.81 [0.81, 0.82]	0.48 [0.47, 0.49]	
60–80	0.89 [0.89, 0.89]	0.81 [0.80, 0.81]	0.78 [0.78, 0.79]	0.48 [0.48, 0.49]	
80+	0.81 [0.81, 0.82]	0.82 [0.81, 0.83]	0.68 [0.68, 0.69]	0.42 [0.42, 0.43]	
Ethnicity*					
Not Hispanic or Latino	0.88 [0.88, 0.88]	0.80 [0.80, 0.81]	0.79 [0.78, 0.79]	0.46 [0.45, 0.46]	p<0.001
Ethnic group unknown	0.91 [0.90, 0.91]	0.87 [0.86 0.88]	0.79 [0.78, 0.80]	0.47 [0.46, 0.48]	
Hispanic or Latino	0.91 [0.91 0.91]	0.81 [0.80, 0.82]	0.87 [0.86, 0.87]	0.55 [0.54, 0.56]	
Race*					
Black or African American	0.91 [0.90, 0.91]	0.81 [0.80, 0.82]	0.83 [0.82, 0.83]	0.52 [0.51, 0.52]	p<0.001
White	0.87 [0.86, 0.87]	0.80 [0.79, 0.81]	0.78 [0.77, 0.78]	0.44 [0.43, 0.44]	
American Indian or Alaska Native	0.85 [0.83, 0.88]	0.71 [0.66, 0.75]	0.82 [0.79, 0.84]	0.57 [0.52, 0.63]	
Asian or Pacific islander	0.86 [0.84, 0.88]	0.86 [0.84, 0.87]	0.74 [0.73, 0.75]	0.32 [0.31, 0.34]	
Unknown Race	0.92 [0.91, 0.92]	0.88 [0.87, 0.88]	0.82 [0.81, 0.82]	0.50 [0.48, 0.52]	

*statistical significance (P<0.05) by one-way ANOVA row-wise.

F < 100) than in mild ARDS. Table 4 describes the performance of our model when controlled against gender, age, ethnicity, and race in the CRD hold-out dataset at least 12-hours before t_{Onset}. We report the average [95% CI] values of AUC, sensitivity, specificity, and PPV. Performance across age was statistically significant (P<0.001), with eARDS reporting the highest AUC of 0.93 [0.92, 0.94] for the age group of 18–40, patients over 80-years of age had the lowest AUC of 0.81 [0.81, 0.82]. Statistical significance (P<0.001) was also observed in severity, age, ethnicity and race in hold-out data of the training dataset, with the lowest performance on average among American Indian 0.85 [0.83, 0.88] and the highest performance among 'Unknowns' with an average AUC of 0.92 [0.91, 0.92]. A moderate statistical significance (P<0.05) was observed in sex, with females on average having an AUC greater than males.

Comparison to the lung injury prediction score

We compared the performance of the eARDS model to that of the Lung Injury Prediction Score (LIPS) for predicting ARDS (Table 5). The eARDS model performed better than

Table 5. Analysis of the clinical LIPS benchmark across the three datasets.

Prediction using LIPS (COVID-19 positives)						
Dataset	AUC	Sensitivity	Specificity	PPV	LIPS(ARDS)	LIPS(Non-ARDS)
CRW	0.80	0.61	0.84	0.53	2.94 [2.85, 3.04]	0.95 [0.92, 0.99]
Emory	0.79	0.53	0.56	0.15	3.74 [3.39, 4.1]	1.75 [1.64, 1.87]
MLH	0.63	0.12	0.96	0.64	0.88 [0.42, 1.33]	0.42 [0.16, 0.67]
Prediction using LIPS (All-cause)						
Dataset	AUC	Sensitivity	Specificity	PPV	LIPS(ARDS)	LIPS(Non-ARDS)
CRW	0.80	0.63	0.83	0.38	3.00 [2.93, 3.07]	1.01 [0.99, 1.03]
Emory	0.81	0.73	0.81	0.36	3.74 [3.39, 4.09]	1.75 [1.64, 1.87]
MLH	0.61	0.10	0.96	0.60	0.88 [0.42, 1.33]	0.42 [0.16, 0.67]

LIPS for predicting ARDS in COVID-19 patients, with eARDS demonstrating higher AUC across all three datasets. The eARDS model also demonstrated higher AUC than LIPS for predicting ARDS in all patients, with the exception of the Emory dataset in which the AUC was comparable.

Discussion

In this study, we derived and validated a supervised machine learning model called *eARDS* for predicting the onset of ARDS in critically ill COVID-19 patients up to 36 hours before meeting the clinical criteria. In our validation, the eARDS model performed well in predicting ARDS in critically ill COVID-19 patients with an optimal prediction horizon of 12 hours before the onset of ARDS according to the Berlin definition. The high AUC and other performance characteristics of the model demonstrate the utility of the eARDS model in identifying a subset of critically ill COVID-19 patients who were at increased risk of developing ARDS. Common errors, such as missingness and incorrect data points were frequently observed among the laboratory values. These are consistent with findings from the literature [30].

The results of our study have important clinical implications, particularly from the performance of our machine learning model in early prediction of ARDS. The PPV of 0.59 and 0.48 for Emory and UTHSC-MLH validation cohorts, respectively, indicate that 48–59% of patients who were predicted to have ARDS by our model did, in fact, develop ARDS at a later time. Considering that the incidence of ARDS was 13.4% in the overall study population and 17.0% in the COVID-19 population, the PPV of 0.48–0.59 represents a significantly higher incidence of ARDS than baseline in those who were predicted by our model to develop ARDS. Our model also showed better performance in predicting severe ARDS with AUC of 0.91 compared to mild ARDS with AUC of 0.83. These characteristics may allow clinicians to promptly identify a subset of patients who are at high risk of developing ARDS, especially the severe forms of ARDS that would likely require mechanical ventilation and other advanced treatments. This early risk-stratification can inform decisions regarding various interventions, such as the timing of intubation for critically ill COVID-19 patients. While early intubation was not associated with differences in clinical outcomes or mortality in COVID-19 in one single center study, it does appear to correlate with the severity of illness and the rate of progression of disease [31, 32] [references]. Our machine learning model can predict ARDS development well before the actual disease onset, thereby alerting the clinicians of high-risk patients who may soon develop ARDS and prompting an earlier assessment of the need for intubation. In addition, early identification of high-risk patients could allow timely implementation of evidence-based treatments and strategies to prevent further lung injury. Such treatment strategies include low-tidal volume and lung protective ventilation strategies in those already receiving mechanical ventilation [33, 34], conservative fluid management and early utilization of diuretics to optimize fluid balance [35, 36]. Early prediction with our model can also prompt early corticosteroid treatment for those with severe COVID-19 who will require oxygen support or mechanical ventilation, which could mitigate the development of ARDS and improve outcomes [37–39]. Our prediction model can allow additional time for preemptive implementation of these proven strategies to attenuate lung injury in patients with progressively worsening hypoxia.

Prediction of ARDS can also improve compliance with proven ARDS management strategies such as lung-protective ventilation and prone positioning once a diagnosis is made. In a large international cohort of patients with ARDS, the diagnosis of ARDS was missed entirely in 40% of patients, with ARDS recognition ranging from only 51% in mild ARDS to 79% in severe ARDS [40]. The poor rate of recognition may be related to the complexity of the Berlin definition that utilizes both structured data and unstructured elements, such as interpretation

of chest imaging that are not always definitive [18, 41]. Clinician recognition of ARDS was associated with the use of higher PEEP levels and with greater use of prone positioning, neuro-muscular blockade and extracorporeal membrane oxygenation. It has also been reported that patients receiving higher tidal volumes shortly after ARDS onset have a higher mortality [42]. These findings highlight the importance of early prediction and recognition of ARDS in improving the management of ARDS, and the same principles apply to COVID-19 induced ARDS as well. Our present machine learning model and similar models can help clinicians with prediction, early recognition, and prognostication of patients at high risk of developing severe respiratory failure and ARDS. This, in turn, would allow more time for implementing strategies to avoid further lung injury, improve adherence to evidence-based therapies, and guide clinical decisions regarding treatments for severe COVID-19 and ARDS.

The example patient in Fig 3A illustrates the utility of the model in predicting ARDS onset earlier, through analysis and integration of various hemodynamic and oxygenation support variables. The initial prediction for ARDS was driven primarily by a slew of vital sign abnormalities, including the very first signs of tachycardia, tachypnea, hypoxia, and fluctuations in blood pressure. Our model was able to promptly capture the changes in these variables that suggested hemodynamic instability, and generated the initial prediction for ARDS. This alert generation was well in advance of the O_2 flow rate adjustment [10], which can be a surrogate for the clinician's recognition of a deteriorating patient with worsening respiratory failure. Subsequently, the standard deviation of the O_2 flow rate sustained the ARDS prediction probability above the threshold, until the clinical criteria for ARDS was finally met. As demonstrated by this example, our model was able to integrate the earliest signs of clinical deterioration and successfully predict an increased risk of ARDS, predating the clinical suspicion or the actual onset of ARDS by several hours. Therefore, understanding the clinical context through such surrogate information may be a means by which the model recognizes worsening severity of illness.

Prior to the utilization of machine learning in predictive modeling for ARDS, a widely used clinical prediction tool for ARDS was the Lung Injury Prediction Score (LIPS) [43]. However, many of the variables used in LIPS required manual chart abstraction, and the model did not perform well when applied to settings that were different from the original validation study [44]. Machine learning models such as eARDS can automate the analysis of relevant clinical variables, and expedite the prediction of ARDS at an earlier time point than would be feasible with traditional predictive modeling. Furthermore, machine learning models can fit the data more precisely than the traditional models and result in more accurate predictions of ARDS. We demonstrate that eARDS utilized machine learning techniques to successfully analyze a complex combination of structured clinical variables that can be automatically abstracted from the EHR. Consequently, our eARDS model showed better performance for early prediction of ARDS in COVID-19 patients compared to LIPS.

The feature importance heatmaps from our model provide an indication for the most important clinical features for predicting ARDS onset in our model. Not surprisingly, features that are directly related to the patients' respiratory status, including the minimum SpO_2, respiratory rate, and O_2 flow rate, were ranked among the five most important features for predicting ARDS development. The standard deviation of SBP, which is also ranked among the top five most important features for predicting ARDS, may be a surrogate of hemodynamic instability and overall clinical deterioration related to impending respiratory failure. In the clustered heatmap based on disease severity, the maximum and the standard deviation of O_2 flow rates contributed positively to the alert in more severe ARDS, again highlighting that the features associated with the patients' respiratory status were important in predicting ARDS development. The fact that our model placed relative importance on these factors related to vital signs and noninvasive measurements adds strength to our model. Our model could predict

ARDS development without heavy reliance on invasive tests or lab values (e.g. PaO_2), which are more likely to be obtained in patients who are already demonstrating signs of clinical deterioration and/or worsening respiratory failure. This suggests that our model can utilize vital signs and other readily available measurements to predict ARDS development before the onset of overt clinical deterioration, without heavy reliance on potentially biased availability of information.

The performance of our model with regard to age, especially the younger age group of 18–40, is also noteworthy. This may have been attributed to the lower prevalence of ARDS within this cohort, with 6% of the patients 18–40 years of age meeting ARDS criteria, while 20% met criteria within the 81+ years group. Prior literature reported that younger patients are less likely to develop ARDS and tend to suffer less severe illness from COVID-19 [45, 46]. From this, one can anticipate that younger age would not be an important feature in predicting ARDS development, but the age group 18–40 was actually one of the most important features in our model. This finding may be related to the fact that this age group contributed positively to the prediction of mild ARDS in the clustered analysis by disease severity, thus indicating an association between younger age and mild severity of ARDS.

Limitations

There are several limitations to this study. First, we identify a select number of variables from the EMR, driven by a review of existing literature. Expanded coverage of the EMR, including deriving variables from natural language processing of the unstructured data, such as clinical notes and chest imaging studies, may improve the model's performance and specificity. Secondly, we developed our model using a popular state-of-the-art machine learning method. However, we have not demonstrated the performance across more recent deep learning methods. Incorporating such architectures may further improve our model performance. Third, we noted that there was a high degree of missingness of variables within all three datasets; this has frequently been observed when using EMR data [30], arising from the dynamic array of possible care patterns that each patient may receive. In order to address this limitation, we incorporated methods within our pipeline that indicated missingness of a variable at each opportunity, while this effort may provide the model with some context to the nature of missingness, it is still unable to discern whether the missingness was at random or not-at-random. Further studies to this effect, particularly from prospective validation may be necessary to truly discern whether particular variables are missing or were not entered into the EMR. Finally, we developed our model and validated it using only retrospective data. A prospective validation will be essential not only for identifying potential errors and improving the performance of our model, but also to be able to implement it in clinical practice.

In conclusion, we demonstrate that machine learning methods can be applied to predictions of ARDS in patients with COVID-19. We further evaluate the performance of a general ARDS prediction model in critically ill COVID-19 patients and find that our model achieves optimal and statistically significant performance in the severe ARDS group than the mild ARDS group. Further research including the addition of blood-based biomarkers [47, 48], radiographic images, unstructured notes and high-frequency bedside monitoring data streams [49–51] may further improve the performance of the model for bedside clinical decision support.

Supporting information

S1 Fig. Consort diagram of the training dataset (CRD).
(EPS)

S2 Fig. The proportion of missing values for each variable across the three datasets. A higher median value signifies a greater level of missingness.
(PDF)

S3 Fig. Ranking of important features in the eARDS model sorted from highest to lowest importance.
(PDF)

S1 Table. Variables used in machine learning model.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Rishikesan Kamaleswaran.

Formal analysis: Rishikesan Kamaleswaran.

Investigation: Philip Yang, Andre L. Holder.

Methodology: Lakshya Singhal, Yash Garg, Philip Yang, Azade Tabaie, A. Ian Wong, Akram Mohammed, Lokesh Chinthala, Dipen Kadaria, Amik Sodhi, Andre L. Holder, Annette Esper, James M. Blum, Robert L. Davis, Gari D. Clifford, Greg S. Martin, Rishikesan Kamaleswaran.

Supervision: Rishikesan Kamaleswaran.

Validation: Dipen Kadaria, Amik Sodhi, Andre L. Holder, Annette Esper, Greg S. Martin, Rishikesan Kamaleswaran.

Writing – original draft: Philip Yang, Rishikesan Kamaleswaran.

Writing – review & editing: Philip Yang, Dipen Kadaria, Amik Sodhi, Andre L. Holder, Annette Esper, James M. Blum, Robert L. Davis, Gari D. Clifford, Greg S. Martin.

References

1. Joost Wiersinga W, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020; 324: 782–793. https://doi.org/10.1001/jama.2020.12839 PMID: 32648899

2. Myers LC, Parodi SM, Escobar GJ, Liu VX. Characteristics of Hospitalized Adults With COVID-19 in an Integrated Health Care System in California. JAMA. 2020; 323: 2195–2198. https://doi.org/10.1001/jama.2020.7202 PMID: 32329797

3. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, 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. https://doi.org/10.1136/bmj.m1985 PMID: 32444460

4. Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. medRxiv. 2020. https://doi.org/10.1101/2020.04.15.20067157 PMID: 32511638

5. Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. JAMA. 2020. https://doi.org/10.1001/jama.2020.4326 PMID: 32191259

6. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. Covid-19 in Critically Ill Patients in the Seattle Region—Case Series. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2004500 PMID: 32227758

7. Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020; 382: 1708–1720. https://doi.org/10.1056/NEJMoa2002032 PMID: 32109013

8. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020. https://doi.org/10.1001/jama.2020.6775 PMID: 32320003

9. Gattinoni L, Chiumello D, Caironi P, Busana M, Romitti F, Brazzi L, et al. COVID-19 pneumonia: different respiratory treatments for different phenotypes? Intensive Care Med. 2020. https://doi.org/10.1007/s00134-020-06033-2 PMID: 32291463

10. Ferguson ND, Fan E, Camporota L, Antonelli M, Anzueto A, Beale R, et al. The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material. Intensive Care Medicine. 2012. pp. 1573–1582. https://doi.org/10.1007/s00134-012-2682-1 PMID: 22926653

11. Gattinoni L, Coppola S, Cressoni M, Busana M, Rossi S, Chiumello D. COVID-19 Does Not Lead to a "Typical" Acute Respiratory Distress Syndrome. American Journal of Respiratory and Critical Care Medicine. 2020. pp. 1299–1300. https://doi.org/10.1164/rccm.202003-0817le PMID: 32228035

12. Li X, Ma X. Acute respiratory failure in COVID-19: is it "typical" ARDS? Crit Care. 2020; 24: 198. https://doi.org/10.1186/s13054-020-02911-9 PMID: 32375845

13. Rello J, Storti E, Belliato M, Serrano R. Clinical phenotypes of SARS-CoV-2: implications for clinicians and researchers. Eur Respir J. 2020; 55. https://doi.org/10.1183/13993003.01028-2020 PMID: 32341111

14. Chauvelot L, Bitker L, Dhelft F, Mezidi M, Orkisz M, Davila Serrano E, et al. Quantitative-analysis of computed tomography in COVID-19 and non COVID-19 ARDS patients: A case-control study. J Crit Care. 2020; 60: 169–176. https://doi.org/10.1016/j.jcrc.2020.08.006 PMID: 32854088

15. Fan E, Beitler JR, Brochard L, Calfee CS, Ferguson ND, Slutsky AS, et al. COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? Lancet Respir Med 2020; 8:816–821. 2020;8: 816–821. https://doi.org/10.1016/S2213-2600(20)30304-0 PMID: 32645311

16. Ding X-F, Li J-B, Liang H-Y, Wang Z-Y, Jiao T-T, Liu Z, et al. Predictive model for acute respiratory distress syndrome events in ICU patients in China using machine learning algorithms: a secondary analysis of a cohort study. J Transl Med. 2019; 17: 326. https://doi.org/10.1186/s12967-019-2075-0 PMID: 31570096

17. Zeiberg D, Prahlad T, Nallamothu BK, Iwashyna TJ, Wiens J, Sjoding MW. Machine learning for patient risk stratification for acute respiratory distress syndrome. PLoS One. 2019; 14: e0214465. https://doi.org/10.1371/journal.pone.0214465 PMID: 30921400

18. Afshar M, Joyce C, Oakey A, Formanek P, Yang P, Churpek MM, et al. A Computable Phenotype for Acute Respiratory Distress Syndrome Using Natural Language Processing and Machine Learning. AMIA Annu Symp Proc. 2018; 2018: 157–165. PMID: 30815053

19. Le S, Pellegrini E, Green-Saxena A, Summers C, Hoffman J, Calvert J, et al. Supervised Machine Learning for the Early Prediction of Acute Respiratory Distress Syndrome (ARDS). medRxiv. 2020. https://doi.org/10.1016/j.jcrc.2020.07.019 PMID: 32777759

20. Definition Task Force A, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin Definition. JAMA. 2012; 307: 2526–2533. https://doi.org/10.1001/jama.2012.5669 PMID: 22797452

21. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMob Mob Comput Commun. 2015; 19: 29–33.

22. XGBoost Documentation—xgboost 1.5.0-dev documentation. [cited 31 May 2021]. https://xgboost.readthedocs.io/en/latest/

23. pandas documentation—pandas 1.2.4 documentation. [cited 31 May 2021]. https://pandas.pydata.org/docs/

24. den Bakker I. Python Deep Learning Cookbook: Over 75 practical recipes on neural network modeling, reinforcement learning, and transfer learning using Python. Packt Publishing Ltd; 2017.

25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011; 12: 2825–2830.

26. sklearn.ensemble.RandomForestClassifier—scikit-learn 0.24.2 documentation. [cited 31 May 2021]. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

27. sklearn.linear_model.LogisticRegression—scikit-learn 0.24.2 documentation. [cited 31 May 2021]. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

28. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. pp. 785–794.

29. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems 30. Curran Associates, Inc.; 2017. pp. 4765–4774.

30. Ho LV, Ledbetter D, Aczon M, Wetzel R. The Dependence of Machine Learning on Electronic Medical Record Quality. AMIA Annu Symp Proc. 2017; 2017: 883–891. PMID: 29854155

31. Hernandez-Romieu AC, Adelman MW, Hockstein MA, Robichaux CJ, Edwards JA, Fazio JC, et al. Timing of Intubation and Mortality Among Critically Ill Coronavirus Disease 2019 Patients: A Single-Center Cohort Study. Crit Care Med. 2020; 48: e1045–e1053. https://doi.org/10.1097/CCM.0000000000004600 PMID: 32804790

32. Matta A, Chaudhary S, Bryan Lo K, DeJoy R 3rd, Gul F, Torres R, et al. Timing of Intubation and Its Implications on Outcomes in Critically Ill Patients With Coronavirus Disease 2019 Infection. Crit Care Explor. 2020; 2: e0262. https://doi.org/10.1097/CCE.0000000000000262 PMID: 33134950

33. Hager DN, Krishnan JA, Hayden DL, Brower RG, ARDS Clinical Trials Network. Tidal volume reduction in patients with acute lung injury when plateau pressures are not high. Am J Respir Crit Care Med. 2005; 172: 1241–1245. https://doi.org/10.1164/rccm.200501-048CP PMID: 16081547

34. Neto AS, Cardoso SO, Manetta JA, Pereira VGM, Espósito DC, de Oliveira Prado Pasqualucci M, et al. Association Between Use of Lung-Protective Ventilation With Lower Tidal Volumes and Clinical Outcomes Among Patients Without Acute Respiratory Distress Syndrome: A Meta-analysis. JAMA. 2012; 308: 1651–1659. https://doi.org/10.1001/jama.2012.13730 PMID: 23093163

35. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network, Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, et al. Comparison of two fluid-management strategies in acute lung injury. N Engl J Med. 2006; 354: 2564–2575. https://doi.org/10.1056/NEJMoa062200 PMID: 16714767

36. Seitz KP, Caldwell ES, Hough CL. Fluid management in ARDS: an evaluation of current practice and the association between early diuretic use and hospital mortality. J Intensive Care Med. 2020; 8: 78. https://doi.org/10.1186/s40560-020-00496-7 PMID: 33062283

37. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19—Preliminary Report. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2021436 PMID: 32678530

38. Confalonieri M, Urbino R, Potena A, Piattella M, Parigi P, Puccio G, et al. Hydrocortisone Infusion for Severe Community-acquired Pneumonia. American Journal of Respiratory and Critical Care Medicine. 2005. pp. 242–248. https://doi.org/10.1164/rccm.200406-808OC PMID: 15557131

39. Salton F, Confalonieri P, Meduri GU, Santus P, Harari S, Scala R, et al. Prolonged Low-Dose Methylprednisolone in Patients With Severe COVID-19 Pneumonia. Open Forum Infect Dis. 2020; 7: ofaa421. https://doi.org/10.1093/ofid/ofaa421 PMID: 33072814

40. Bellani G, Pham T, Laffey JG. Missed or delayed diagnosis of ARDS: a common and serious problem. Intensive Care Med. 2020; 46: 1180–1183. https://doi.org/10.1007/s00134-020-06035-0 PMID: 32328723

41. Sjoding MW, Hofer TP, Co I, Courey A, Cooke CR, Iwashyna TJ. Interobserver Reliability of the Berlin ARDS Definition and Strategies to Improve the Reliability of ARDS Diagnosis. Chest. 2018; 153: 361–367. https://doi.org/10.1016/j.chest.2017.11.037 PMID: 29248620

42. Needham DM, Yang T, Dinglas VD, Mendez-Tellez PA, Shanholtz C, Sevransky JE, et al. Timing of low tidal volume ventilation and intensive care unit mortality in acute respiratory distress syndrome. A prospective cohort study. Am J Respir Crit Care Med. 2015; 191: 177–185. https://doi.org/10.1164/rccm.201409-1598OC PMID: 25478681

43. Trillo-Alvarez C, Cartin-Ceba R, Kor DJ, Kojicic M, Kashyap R, Thakur S, et al. Acute lung injury prediction score: derivation and validation in a population-based sample. Eur Respir J. 2011; 37: 604–609. https://doi.org/10.1183/09031936.00036810 PMID: 20562130

44. Soto GJ, Kor DJ, Park PK, Hou PC, Kaufman DA, Kim M, et al. Lung Injury Prediction Score in Hospitalized Patients at Risk of Acute Respiratory Distress Syndrome. Crit Care Med. 2016; 44: 2182–2191. https://doi.org/10.1097/CCM.0000000000002001 PMID: 27513358

45. Wu C, Chen X, Cai Y, Xia J 'an, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med. 2020. https://doi.org/10.1001/jamainternmed.2020.0994 PMID: 32167524

46. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA. 2020. https://doi.org/10.1001/jama.2020.5394 PMID: 32250385

47. Akram M, Cui Y, Mas VR, Kamaleswaran R. Differential gene expression analysis reveals novel genes and pathways in pediatric septic shock patients. Sci Rep. 2019; 9: 1–7.

48. Banerjee S, Mohammed A, Wong HR, Palaniyar N, Kamaleswaran R. Machine Learning Identifies Complicated Sepsis Trajectory and Subsequent Mortality Based on 20 Genes in Peripheral Blood Immune Cells at 24 Hours post ICU admission. https://doi.org/10.1101/2020.06.14.150664

49. Kamaleswaran R, Akbilgic O, Hallman MA, West AN, Davis RL, Shah SH. Applying Artificial Intelligence to Identify Physiomarkers Predicting Severe Sepsis in the PICU. Pediatr Crit Care Med. 2018; 19: e495–e503. https://doi.org/10.1097/PCC.0000000000001666 PMID: 30052552

50. Kamaleswaran R, McGregor C, Percival J. Service oriented architecture for the integration of clinical and physiological data for real-time event stream processing. Conf Proc IEEE Eng Med Biol Soc. 2009; 2009: 1667–1670. https://doi.org/10.1109/IEMBS.2009.5333884 PMID: 19964545

51. Mohammed A, Van Wyk F, Chinthala LK, Khojandi A, Davis RL, Coopersmith CM, et al. Temporal Differential Expression of Physiomarkers Predicts Sepsis in Critically Ill Adults. Shock. 2020; Publish Ahead of Print. https://doi.org/10.1097/SHK.0000000000001670 PMID: 32991797