scientific reports

Check for updates

OPEN

Early predictive values of clinical assessments for ARDS mortality: a machine-learning approach

Ning Ding[1,2,5], Tanmay Nath[3,5], Mahendra Damarla[4], Li Gao[1]✉ & Paul M. Hassoun[4]✉

Acute respiratory distress syndrome (ARDS) is a devastating critical care syndrome with significant morbidity and mortality. The objective of this study was to evaluate the predictive values of dynamic clinical indices by developing machine-learning (ML) models for early and accurate clinical assessment of the disease prognosis of ARDS. We conducted a retrospective observational study by applying dynamic clinical data collected in the ARDSNet FACTT Trial (n = 1000) to ML-based algorithms for predicting mortality. In order to compare the significance of clinical features dynamically, we further applied the random forest (RF) model to nine selected clinical parameters acquired at baseline and day 3 independently. An RF model trained using clinical data collected at day 3 showed improved performance and prognostication efficacy (area under the curve [AUC]: 0.84, 95% CI: 0.78–0.89) compared to baseline with an AUC value of 0.72 (95% CI: 0.65–0.78). Mean airway pressure (MAP), bicarbonate, age, platelet count, albumin, heart rate, and glucose were the most significant clinical indicators associated with mortality at day 3. Thus, clinical features collected early (day 3) improved performance of integrative ML models with better prognostication for mortality. Among these, MAP represented the most important feature for ARDS patients' early risk stratification.

Keywords Ards; Machine-Learning; Mortality; Mean Airway Pressure

Acute respiratory distress syndrome (ARDS) is a devastating critical care syndrome affecting over 200,000 patients in the U.S. each year, and an important cause of respiratory failure with significant morbidity and mortality[1]. At this time, identifying specific risk factors most associated with ARDS-related death are needed for early intervention in patients at risk.

In recent years, research has focused on identifying important clinical features and biomarkers, as well as their combined predictive capabilities for death in ARDS patients[2]. Among these, ventilation parameters such as positive end-expiratory pressure (PEEP) and plateau pressure demonstrated predictive importance for mortality in patients with ARDS[3]. Mean airway pressure (MAP) is a key component of the oxygenation index, which has also been associated with mortality in multiple studies of outcomes in both adult and pediatric respiratory failure[4–6]. Comparing to PEEP and plateau pressure, there is a paucity of research on MAP's dynamic change and early prognostic importance in ARDS.

Unsupervised learning methods have been increasingly used in complex clinical syndromes such as ARDS to address the issue of clinical and biological heterogeneity[7,8]. In contrast, current supervised machine learning (ML) models allow analysis of a variety of collected variables and development of an ARDS-specific mortality prediction system. We hypothesize that ARDS is characterized by unique clinical features, and these crucial clinical features can be applied to ML algorithms for accurate prediction of worst outcomes. To investigate this hypothesis, we employed an integrative approach, which incorporates clinical data collected in the ARDSNet FACTT Trial[9] and ML modeling for ARDS prognostication, followed by comparisons of the accuracy of different ML models and determining the importance of clinical features, especially the ventilation parameters. Based on the prognostic value of the best-performing ML model, we further determined the importance of crucial clinical features in prioritizing patients for early intervene that can potentially reduce the mortality rate for ARDS.

[1]Division of Allergy and Clinical Immunology, Johns Hopkins University School of Medicine, 5501 Hopkins Bayview Circle, Baltimore, MD 21224-6821, USA. [2]Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China. [3]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA. [4]Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, 1830 East Monument St, Baltimore, MD 21287, USA. [5]These authors contributed equally: Ning Ding and Tanmay Nath. ✉email: lgao2@jhmi.edu; phassou1@jhmi.edu

Methods

Patient population

We performed a ML predictive modeling using data obtained from patients enrolled in the ARDSNet FACTT randomized clinical trial. The design of the FACTT study has been described previously[9,10]. Briefly, FACTT trial enrolled 1,000 patients with ARDS between 2000 and 2005. The trial randomized subjects in a two-by-two factorial design; one arm compared conservative versus liberal fluid management[9], whereas the other arm compared monitoring patients with ARDS with a pulmonary artery versus central venous catheter[10]. Patients were followed for 60 days or until discharge home with unassisted breathing. The primary outcome was mortality at 60 days before discharge home[9–11]. This study was approved by the Johns Hopkins University Institutional Review Board and all patients gave informed consent at the time of enrollment. All procedures were followed in accordance with the ethical standards (details provided in Supplementary Methods).

Clinical data

When evaluating clinical conditions of ARDS, clinical data from the baseline and early stages of the disease (such as days 3) were frequently utilized[12–14]. Fewer studies, nevertheless, assessed the prognostic significance of information gathered beyond baseline to identify at risk patients early on for worse outcomes. In this study, we assessed baseline clinical characteristics and data gathered early on day 3. Clinical predictors of 60-day mortality were chosen from 29 parameters collected in the FACTT trial using a backward stepwise selection scheme (*see* Supplementary Methods); we also considered the importance of selected parameters as indicated by published studies[2,11]. Backward elimination approach is a well-documented logistic regression method for variable selection and previously has been employed for evaluating risk factors associated with prognosis of ARDS[15]. The 29 clinical parameters were in 5 major categories including (1) baseline characteristics; (2) vital signs and circulatory; (3) respiratory and ventilatory; (4) blood and coagulation; and (5) metabolism and renal. We also excluded the following clinical parameters: (1) parameters closely related to fluid strategy; and (2) parameters with data missing rate greater than 30% (tidal volume, bilirubin). Finally, nine predictors (age, sex, pneumonia as cause of ARDS, heart rate, mean airway pressure, glucose, albumin, platelet count, bicarbonate) were selected and utilized for the development of ML models.

Prediction model building and evaluation

The primary aim of our study was to establish a ML model for predicting 60-day mortality. As summarized in Fig. 1, five steps were performed including variables selection, applying analysis strategy, building ML models and model comparison as well as model application. First, we employed non-imputed datasets by dropping observations with missing data. We also imputed missing data using an iterative multivariate imputation technique[16] for two datasets (day 0 and day 3). Second, the entire dataset was randomly divided into 70% training and 30% testing for all ML classifiers. The training set was used to build the ensemble model, while the testing set was used to evaluate the predictive performance of the model. Third, we employed six typically used supervised ML classification algorithms (Random Forest [RF], XGBoost, Support vector machine [SVM], Logistic regression [LR], Multi-layer perceptron [MLP], and Stacking Classifier [SC] models) for classifying survivors and non-survivors. Next, we performed a fivefold cross validation on the training dataset on each model to evaluate their cross-validation performance. Finally, the different models were evaluated and compared using area under the receiver operating characteristic (ROC) curve (AUC) and confusion matrix (Supplementary figure S1), which can be summarized using precision, sensitivity and F1 score (which is a weighted average of precision and sensitivity). We plotted the calibration curve of the models for predicting survival utilizing the sigmoid regressor based on Platt's logistic model. See the Supplementary Methods for further details.

Our analysis was conducted in python version 3.6 (https://www.python.org) using the library Scikit Learn[17].

Statistical analyses

Continuous variables were expressed as mean ± SE or median (interquartile range), as appropriate, while categorical variables were presented as numbers (percentage). Qualitative and quantitative differences between subgroups were analyzed by chi-square test for categorical parameters and Student's t test or Mann–Whitney's test for continuous parameters, as appropriate. Missing values were imputed using multivariate imputation. A general linear model with repeated measures was utilized to evaluate the trend over time of MAP at baseline, days 1 through 3. The group-by-time interaction term was tested first. If significant, between-group (survivor and non-survivor) differences at each time point were examined. Then within-group changes over time (trend) were tested in both survivor and non-survivor groups independently, with Bonferroni correction applied. Two-tailed *P* values less than 0.05 were considered statistically significant. Statistical analyses were performed with SPSS statistical software (version 22, IBM® SPSS Inc., Chicago, IL, USA).

Ethics approval

This study was approved by the Johns Hopkins University Institutional Review Board (approval number: NA_00034898) and all procedures were followed in accordance with the ethical standards.

Results

Baseline patient characteristics comparison

Table 1 shows baseline patient characteristics between survivors and non-survivors at day 60. After excluding observations with missing data, there were 700 patients (survivors = 505, non-survivors = 195) at day 0 and 593 patients (survivors = 453, non-survivors = 140) at day 3 for non-imputed datasets. The non-survivors were older

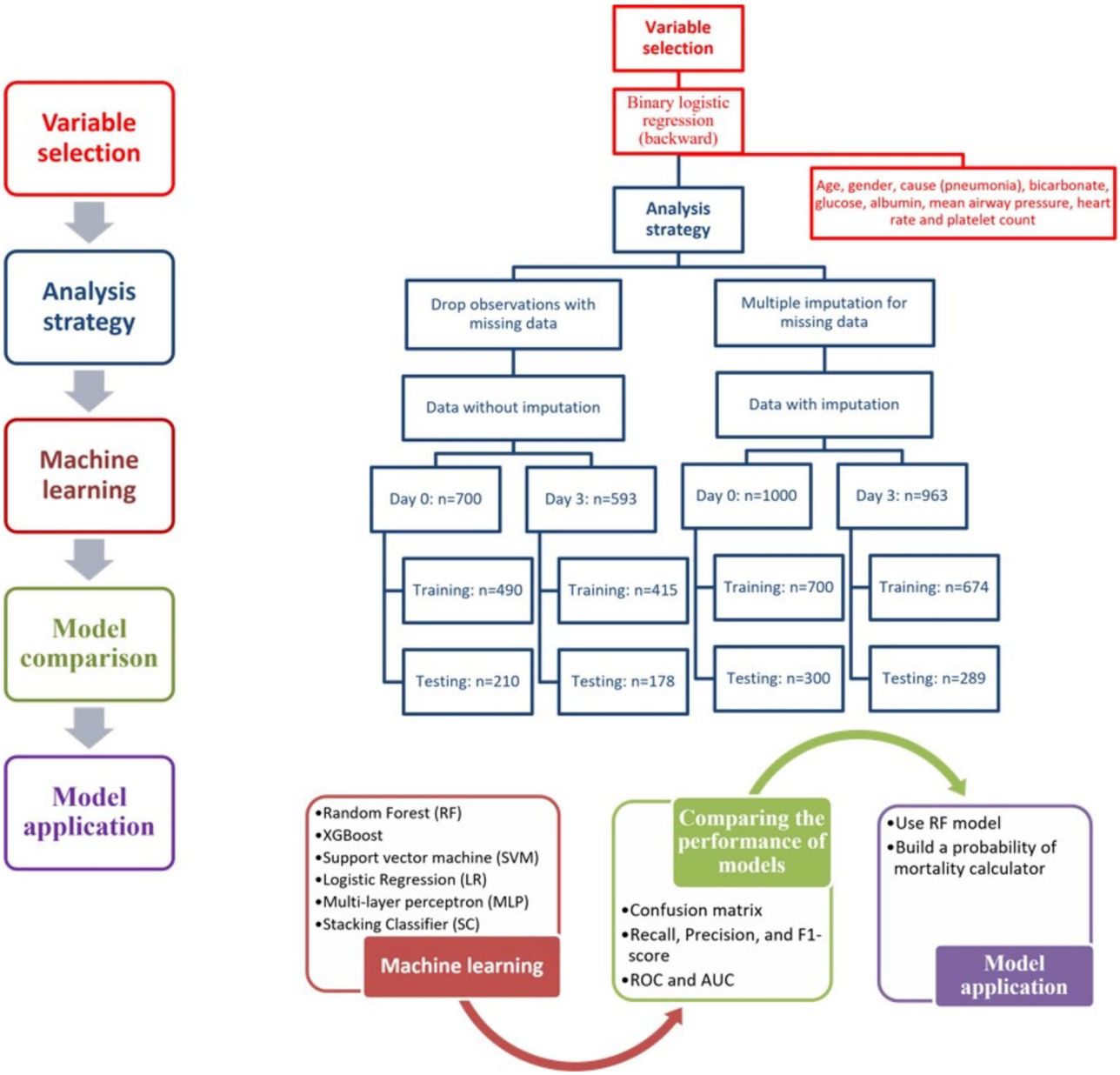

Figure 1. Overview of the analysis plan.

and had high percentage of pneumonia and sepsis but lower percentage of trauma compared to survivors. The non-survivors had significantly higher respiratory rate, FiO_2, serum potassium, BUN and creatinine, lower body temperature, mean arterial pressure, serum albumin, bicarbonate and arterial pH at either day 0 or day 3. At day 3, the non-survivors had higher PEEP, peak pressure, plateau pressure, MAP, chloride and blood glucose but lower sodium, hemoglobin and platelet count compared to survivors.

Variable selection

Multivariate binary logistic regression (backward elimination) was applied to explore the risk factors for 60-day mortality. Twenty-nine variables were entered into the analysis. Supplementary table S1 showed nine predictors (age, sex, pneumonia, heart rate, MAP, glucose, albumin, platelet and bicarbonate) were independently associated with 60-day mortality, which were then included in the development of ML models. Of note, we randomly split the non-imputed datasets from day 0 and day 3 into training and testing groups and compared patient characteristics, there were no significant differences for all nine parameters (Supplementary table S2). The other two well established logistic regression methods for variable selection (Enter and Forward selection) yielded comparable results; however, the backward elimination method included albumin, for which there was previously research indicating a possible link to ARDS mortality. Similar to findings reported in prior research[18–21], we found that patients with ARDS were more likely to die due to increased heart rate, blood glucose, and reduced platelet counts and albumin. Moreover, we observed that acid–base balance was a significant predictor of death, which

Clinical parameters	Day 0 (n = 700)			Day 3 (n = 593)		
	Survivor (n = 505)	Non-survivor (n = 195)	*P* value	Survivor (n = 453)	Non-survivor (n = 140)	*P* value
Age (yr) §	47.51 ± 0.66	55.86 ± 1.28	*< 0.001*	47.41 ± 0.7	55.8 ± 1.55	*< 0.001*
Sex: Male, n (%) §	256 (50.7)	111 (56.9)	0.139	229 (50.6)	78 (55.7)	0.285
Cause of ARDS, n (%)						
Pneumonia §	317 (62.8)	141 (72.3)	*0.017*	284 (62.7)	103 (73.6)	*0.018*
Sepsis	200 (39.6)	99 (50.8)	*0.007*	178 (39.3)	69 (49.3)	*0.036*
Aspiration	106 (21)	38 (19.5)	0.659	98 (21.6)	30 (21.4)	0.959
Trauma	47 (9.3)	7 (3.6)	*0.011*	44 (9.7)	5 (3.6)	*0.022*
Multiple transfusion	13 (2.6)	7 (3.6)	0.470	13 (2.9)	6 (4.3)	0.406
Other	48 (9.5)	12 (6.2)	0.156	41 (9.1)	9 (6.4)	0.329
Fluid strategy, n (%)			0.561			0.487
Liberal	244 (48.3)	99 (50.8)		221 (48.8)	73 (52.1)	
Conservative	261 (51.7)	96 (49.2)		232 (51.2)	67 (47.9)	
Vital signs and circulatory						
Respiratory rate (bpm)	24.72 ± 0.34	27.34 ± 0.59	*< 0.001*	26.58 ± 0.39	29.8 ± 0.65	*< 0.001*
Heart rate (bpm) §	100.44 ± 0.93	104.63 ± 1.57	*0.019*	94.25 ± 0.94	97.34 ± 1.82	0.116
Body temperature (℃)	37.58 ± 0.04	37.3 ± 0.09	*0.002*	37.41 ± 0.04	37.22 ± 0.08	*0.028*
Mean arterial pressure (mm Hg)	77.45 ± 0.62	74.14 ± 0.97	*0.005*	83.88 ± 0.72	78.08 ± 1.2	*< 0.001*
Respiratory						
FiO_2	0.63 ± 0.01	0.67 ± 0.02	*0.013*	0.49 ± 0.01	0.59 ± 0.02	*< 0.001*
PaO_2 (mm Hg)	89.86 ± 1.88	89.83 ± 3.19	0.093	82.86 ± 1.61	82.88 ± 2.53	0.995
$PaCO_2$ (mm Hg)	40.31 ± 0.44	39.46 ± 0.85	0.338	44.47 ± 0.62	44.38 ± 1.04	0.942
P/F ratio (mm Hg)	153.34 ± 3.15	148.06 ± 5.73	0.397	183.24 ± 4.38	162.98 ± 8.59	*0.024*
SpO_2	94.65 ± 0.24	94.1 ± 0.34	0.225	94.53 ± 0.18	93.58 ± 0.52	0.091
Mechanical ventilation						
PEEP (cm H_2O)	9.38 ± 0.18	9.86 ± 0.3	0.163	7.56 ± 0.17	9.95 ± 0.37	*< 0.001*
Peak (cm H_2O)	32.17 ± 0.41	33.06 ± 0.74	0.272	29.94 ± 0.51	32.66 ± 0.89	*0.006*
Plateau pressure (cm H_2O)	25.9 ± 0.37	26.77 ± 0.66	0.236	23.31 ± 0.37	26.65 ± 0.86	*< 0.001*
Mean airway pressure (cm H_2O) §	15.23 ± 0.26	16.62 ± 0.48	*0.007*	13.21 ± 0.26	17.08 ± 0.84	*< 0.001*
Blood and coagulation						
Potassium (mmol/L)	3.96 ± 0.03	4.13 ± 0.05	*0.005*	3.81 ± 0.03	3.97 ± 0.05	*0.006*
Sodium (mmol/L)	138.91 ± 0.23	138.81 ± 0.43	0.828	141.86 ± 0.27	140.45 ± 0.6	*0.032*
Chlorine (mmol/L)	107.65 ± 0.3	108.12 ± 0.52	0.408	105.94 ± 0.36	107.8 ± 0.67	*0.012*
Blood glucose (mg/dl) §	136.09 ± 2.47	145.95 ± 7.27	0.200	136.91 ± 2.45	157.42 ± 6.19	*0.002*
Hemoglobin (g/L)	10.49 ± 0.09	10.21 ± 0.13	0.082	10.07 ± 0.08	9.75 ± 0.13	*0.041*
Platelet count (10^9/L) §	198.83 ± 5.31	228.75 ± 46.89	0.527	208.4 ± 6.32	145.17 ± 10.67	*< 0.001*
Total protein	5.04 ± 0.05	4.9 ± 0.08	0.138	5.26 ± 0.05	4.77 ± 0.08	*< 0.001*
Albumin (g/dl) §	2.24 ± 0.03	2.07 ± 0.05	*0.003*	2.19 ± 0.03	1.91 ± 0.05	*< 0.001*
Metabolism and renal						
Bicarbonate (mmol/L) §	22.52 ± 0.22	20.82 ± 0.37	*< 0.001*	27.58 ± 0.28	23.62 ± 0.53	*< 0.001*
Arterial pH	7.36 ± 0.004	7.34 ± 0.008	*0.004*	7.41 ± 0.004	7.34 ± 0.009	*< 0.001*
BUN (mg/dl)	20.73 ± 0.74	30.8 ± 1.45	*< 0.001*	25.79 ± 0.89	36.03 ± 2	*< 0.001*
Creatinine (mg/dl)	1.18 ± 0.04	1.58 ± 0.07	*< 0.001*	1.22 ± 0.05	1.77 ± 0.11	*< 0.001*

Table 1. Patient characteristics at baseline (day 0) and day 3 comparing survivors to non-survivors at day 60. Independent-Samples T test was used for continuous variables normally distributed (presented as mean ± standard error), and Mann–Whitney test was used for non-normally distributed data (presented as median (interquartile range)). Chi-square test was used for categorical variables (presented as n (%)). Significant *P* values ($P < 0.05$) were italicized. *ARDS* acute respiratory distress syndrome, *APACHE* Acute Physiologic Assessment and Chronic Health Evaluation, *PaO_2* partial pressure of oxygen, *$PaCO_2$* partial pressure of carbon dioxide, *P/F* PaO_2/FiO_2, *SpO_2* pulse oxygen saturation, *PEEP* positive end expiratory pressure, *peak* peak inspiratory pressure, *BUN* blood urea nitrogen. § Parameters selected for machine learning.

is consistent with findings from previous studies[22]. MAP was widely used for the ventilatory management in critical illness, including ARDS[23]. Given its emerging significance in ARDS prognostication, in the following section, we further evaluated the association between elevated MAP and organ dysfunction among non-survivors.

Elevated MAP was associated with ARDS mortality

We compared the repeated assessments of MAP between survivors and non-survivors throughout the first three days following the onset of ARDS (Supplementary Results and figure S2). The interaction term between groups (survivors and non-survivors) and time indicated significant interaction effects: ($F = 11.67$, $P < 0.001$), thus we further examined the effects of group status and time separately. For between-group differences at each time point, there was no significant difference between survivors and non-survivors at day 0 ($P = 0.597$). But from day 1 to day 3, survivors' MAP was much lower than non-survivors (day 1: $P < 0.01$, days 2 and 3: $P < 0.001$). Furthermore, we assessed changes of MAP over time. Over the course of the three-day follow-up, the trend of MAP among survivors decreased progressively, with day 3 seeing the lowest values ($P < 0.001$). In contrast, the trend of MAP in non-survivors did not change over time ($P = 0.926$). We further dichotomized ARDS patients into MAP_{high} (n = 357) and MAP_{low} (n = 353) subgroups based on their MAP levels using a median split (Supplementary table S3). At day 3, patients in the MAP_{high} group showed worse respiratory (P/F ratio) and renal function metrics (higher BUN and creatinine, $P < 0.001$) as well as raised ventilation parameters. In contrast, we did not observe worse renal function in the MAP_{high} group at baseline. Furthermore, comparing 235 survivors and 122 non-survivors in the MAP_{high} subgroup at day 3, we found that the bicarbonate and arterial pH levels in the non-survivor group were significantly lower (below normal ranges) than in the survivors group (22.50 ± 0.58 vs. 27.53 ± 0.39, 7.31 ± 0.009 vs. 7.39 ± 0.005, $P < 0.001$), indicating metabolic acidosis. On the other hand, at baseline, no discernible difference was found. Consequently, at day 3, ARDS patients with higher MAP levels were more likely to experience metabolic dysfunction and organ damage, which may have contributed to the high death rate (38.1% vs. 19.2%, $P < 0.001$).

Assessment of AUC values of six prediction models in the testing set

The performance of each predictive model was assessed in the testing sets based on its receiver operating characteristic (ROC) curve, judged by its area under the ROC curve (AUC), and the 95% confidence interval (CI) for each AUC value.

Imputed data

First, we applied six ML classification algorithms (RF, XGBoost, LR, SVM, MLP and SC) to the imputed testing sets. Then, we obtained the average AUC and 95% CI for each model to evaluate the performance (Fig. 2). At baseline (n = 300, panel A), the AUC values and 95% CI were 0.64 (0.56–0.72), 0.64 (0.56–0.72), 0.72 (0.65–0.78), 0.57 (0.49–0.65), 0.7 (0.63–0.77) and 0.64 (0.57–0.72), respectively. Only the LR classifier obtained a satisfactory AUC value of 0.72 (above 0.70). At day 3 (n = 289, panel **B**), all 6 classifiers achieved satisfactory AUC values above 0.70. The AUC values were 0.84 (0.78–0.89), 0.82 (0.77–0.88), 0.80 (0.74–0.86), 0.77 (0.7–0.83), 0.79 (0.73–0.85) and 0.84 (0.78–0.89), respectively. Overall, we observed enhanced performance for all 6 models at day 3 with the RF and SC classifiers having the highest AUC value (0.84), followed by XGBoost (0.82).

Non-imputed data

At baseline (n = 210, panel **C**), the AUC values were 0.68, 0.65, 0.69, 0.55, 0.62 and 0.67, respectively. The LR classifier obtained the highest AUC value of 0.69 (95% CI: 0.61–0.78). At day 3 (n = 178, panel D), all 6 classifiers achieved satisfactory AUC values above 0.70. The AUC values and 95% CI were 0.82 (0.75–0.89), 0.79 (0.71–0.88), 0.79 (0.71–0.87), 0.80 (0.73–0.87), 0.83 (0.77–0.9) and 0.83 (0.76–0. 9), respectively. Similar to the imputed data, non-imputed data at day 3 exhibited enhanced performance for all 6 models, with the MLP and SC classifiers having the highest AUC value (0.83), followed by RF (0.82). Intriguingly, the accuracy of the prediction values (ROC-AUCs) appears to increase from approximately 0.7 at day 0 to above 0.8 at day 3 in the prediction models.

Comparing the performance of six prediction models

ML classifiers demonstrated high accuracy as indicated by their prediction values (ROC-AUCs) utilizing data from the FACTT trial. To further demonstrate good discrimination of prediction models, we compared the comprehensive performance of 6 classifiers (Supplementary figure S3) utilizing imputed data at day 0 (panel A) and day 3 (panel B) in the testing sets by computing a confusion matrix for each classifier. As shown in Table 2 and Supplementary table S4B, the six models presented varying performances as indicated by the efficacy metrics generated from confusion matrix: precision, sensitivity and F1 score. The RF and LR classifiers achieved the best F1 score of 0.86 for predicting survivors at baseline. In contrast, the MLP classifier achieved the best F1 score of 0.45 for predicting non-survivors. At day 3, both the RF and SC classifiers achieved the best F1 scores of 0.88 for predicting survivors whereas the scores for non-survivors were 0.51 and 0.53, respectively. Additionally, the efficacy of the six classifiers in the training sets are shown in Table 2 and Supplementary table S4A. The RF classifier, at day 3, also demonstrated the highest F1 scores of 0.94 and 0.77 for survivors and non-survivors, respectively. These results indicated that RF classifier consistently exhibited better prognostic values for classifying survivors and non-survivors at either day 0 or day 3.

To test the calibration of the model, we ultimately drew the calibration curves of the models at baseline and 3 days after ARDS onset. At day 3, most of the models under-predicted the true probabilities with the prediction/observation points distributed above the 45° (dashed accuracy-equals-confidence) line (Supplementary figure S4B). In contrast, at baseline models were over-confident until about 0.6 and then under-predicted around

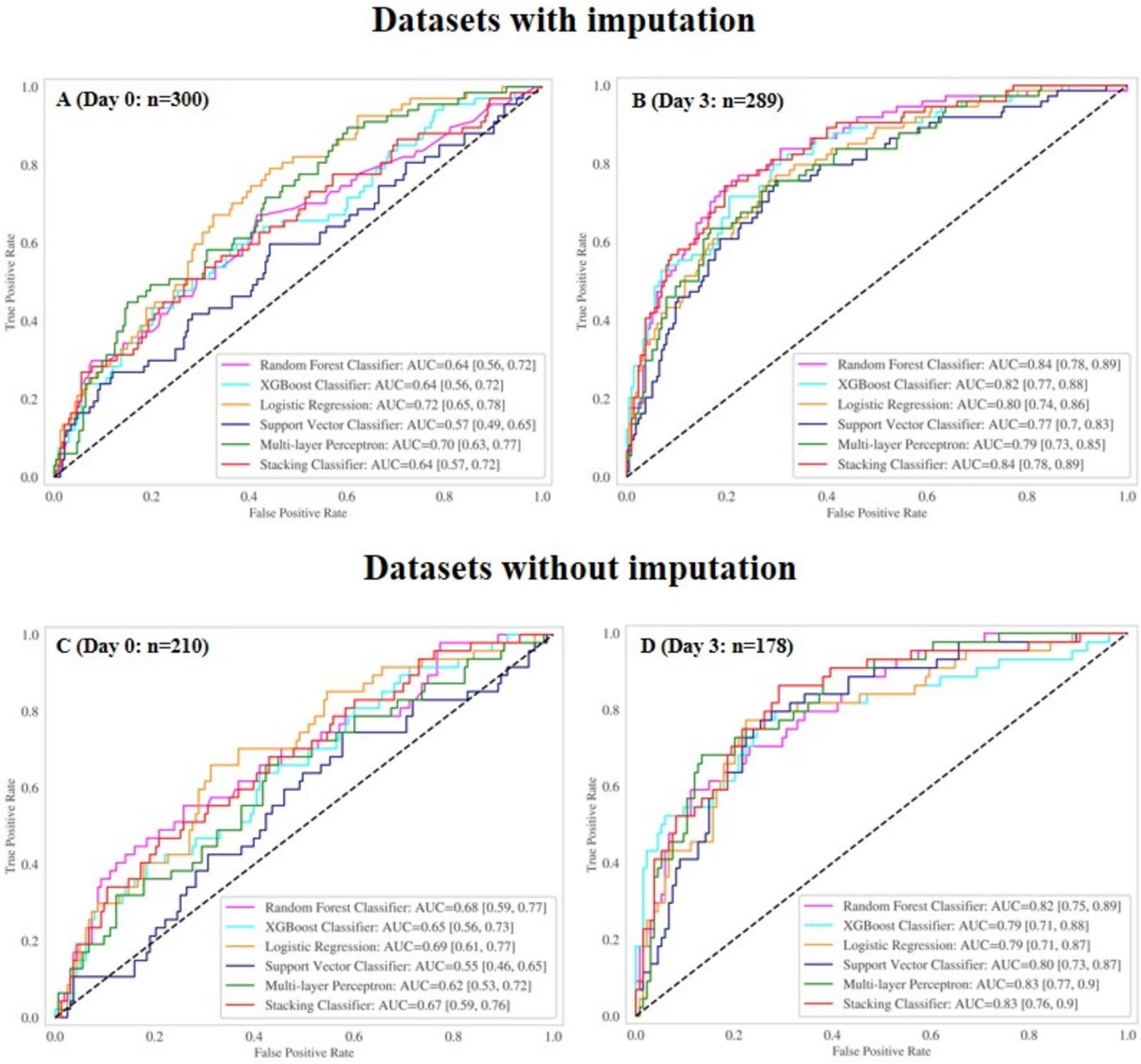

Figure 2. ROC curves of six ML classifiers for predicting 60-day mortality in ARDS patients. Panel (**A**): data collected from Day 0 in the testing dataset (n = 300 with imputation); Panel (**B**): data collected from Day 3 in the testing dataset (n = 289 with imputation). Definition of abbreviations: ROC = receiver operating characteristics, AUC = area under the curve.

0.8 (Supplementary figure S4A). In addition, the results showed that the RF model performed relatively better in each dataset.

Feature importance of the RF classifier in the testing set

We determined the importance of quantitative clinical features according to the estimated probability of 60-day mortality from the RF classifier, which was the best-performing model in the testing sets at day 3 with the highest AUC value of 0.84 when applied to imputed data. Feature importance is the average of reduction in impurity index over all trees when a particular feature is used at split point[17]; we measured feature rankings by using the 'gini impurity or mean decrease in impurity' metric in RF (*see* Supplementary Methods). Figure 3 illustrated the outcomes of relative feature importance for each single attribute. The relative ranked top seven features (from high to low) in the RF predictor at baseline (panel A) were age, platelet count, bicarbonate, MAP, heart rate, glucose, and pneumonia. The relative ranked top features were changed at day 3 (panel B) as the following: MAP, bicarbonate, age, platelet count, albumin, heart rate and glucose. Of note, MAP, one of the ventilator-related features, ranked as the most important mortality risk predictor at day 3, possibly reflecting ARDS progression. We observed similar trends in data without imputation. Age, bicarbonate and MAP were among the top three

Datasets	Models	Predictive values for survivors (Day 0 = 498, Day 3 = 516)			Predictive values for non-survivors (Day 0 = 202, Day 3 = 158)		
		Precision (95% CI)	Sensitivity (95% CI)	F1 score (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	F1 score (95% CI)
A. In the training set							
Day 0 (n = 700)	RF	0.85 (0.83, 0.88)	0.98 (0.98, 1.0)	0.91 (0.91, 0.93)	0.96 (0.93, 0.99)	0.59 (0.54, 0.66)	0.73 (0.69, 0.78)
	XGBoost	0.87 (0.86, 0.9)	0.98 (0.98, 0.99)	0.92 (0.92, 0.94)	0.95 (0.92, 0.98)	0.66 (0.61, 0.72)	0.78 (0.74, 0.82)
	SVM	0.94 (0.93, 0.97)	0.98 (0.98, 1.0)	0.96 (0.96, 0.98)	0.97 (0.95, 0.99)	0.87 (0.83, 0.91)	0.91 (0.89, 0.94)
Day 3 (n = 674)	RF	0.90 (0.88, 0.92)	0.99 (0.98, 1.0)	0.94 (0.93, 0.96)	0.95 (0.92, 0.98)	0.65 (0.6, 0.72)	0.77 (0.73, 0.82)
	SVM	0.89 (0.87, 0.92)	0.97 (0.97, 0.99)	0.93 (0.92, 0.95)	0.89 (0.84, 0.94)	0.62 (0.57, 0.69)	0.73 (0.69, 0.78)
	SC	0.89 (0.87, 0.92)	0.98 (0.98, 1.0)	0.93 (0.93, 0.95)	0.94 (0.9, 0.98)	0.62 (0.56, 0.68)	(0.7, 0.8)
Datasets	**Models**	**Predictive values for survivors (n, Day 0 = 233, Day 3 = 215)**			**Predictive values for non-survivors (n, Day 0 = 67, Day 3 = 74)**		
		Precision (95% CI)	**Sensitivity (95% CI)**	**F1 score (95% CI)**	**Precision (95% CI)**	**Sensitivity (95% CI)**	**F1 score (95% CI)**
B. In the testing set							
Day 0 (n = 300)	RF	0.81 (0.78, 0.85)	0.91 (0.88, 0.94)	0.86 (0.83, 0.89)	0.47 (0.34, 0.61)	0.28 (0.19, 0.38)	0.35 (0.25, 0.45)
	LR	0.80 (0.77, 0.85)	0.93 (0.9, 0.95)	0.86 (0.84, 0.89)	0.48 (0.34, 0.63)	0.23 (0.16, 0.33)	0.32 (0.22, 0.42)
	MLP	0.84 (0.81, 0.89)	0.8 (0.76, 0.85)	0.82 (0.79, 0.85)	0.41 (0.33, 0.51)	0.49 (0.39, 0.59)	0.45 (0.36, 0.53)
Day 3 (n = 289)	RF	0.82 (0.78, 0.86)	0.95 (0.93, 0.98)	0.88 (0.86, 0.91)	0.76 (0.65, 0.88)	0.39 (0.3, 0.49)	0.51 (0.42, 0.61)
	XGBoost	0.82 (0.78, 0.86)	0.93 (0.91, 0.97)	0.87 (0.85, 0.9)	0.69 (0.58, 0.82)	0.40 (0.31, 0.5)	0.51 (0.42, 0.6)
	SC	0.82 (0.78, 0.86)	0.95 (0.93, 0.98)	0.88 (0.86, 0.91)	0.76 (0.65, 0.88)	0.40 (0.31, 0.5)	0.53 (0.43, 0.62)

Table 2. The efficacy (precision, sensitivity and F1 score) of top three machine-learning classifiers in the training (A) and testing (B) sets (using imputed data). Missing data was imputed using an iterative multivariate imputation technique. *RF* random forest, *LR* logistic regression, *SVM* support vector machine, *MLP* multi-layer perceptron, *SC* stacking classifier. 95% CI = 95% confidence interval.

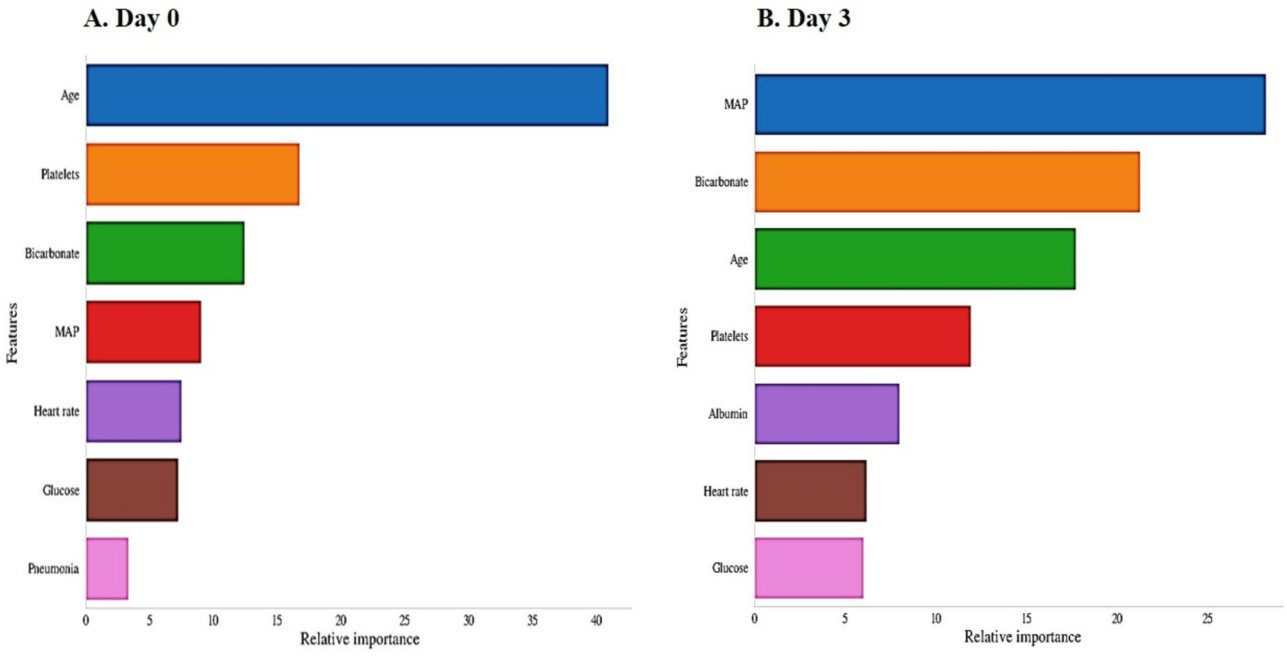

Figure 3. Importance of clinical features in the training dataset according to the estimated probability of 60-day mortality in Random Forest models. Normalized values for each single attribute (in a range of 0–100, the most important predictor variable was assigned with the value of 100) were illustrated as relative ranked features (from high to low). The training datasets from day 0 (n = 700, Panel (**A**)) and day 3 (n = 674, Panel (**B**)) with imputation were applied.

important features demonstrating improved predictive performance for risk of mortality at day 3 (Supplementary figure S5). As a result, ML modeling offered more proof that MAP is crucial for early ARDS prognostication.

Discussion

ARDS is an inflammatory syndrome characterized by acute respiratory failure due to non-cardiogenic pulmonary edema and hypoxemia, and is associated with significant morbidity and mortality[1]. Despite advances in the understanding of the pathophysiology of ARDS, mortality rates remain high – ranging between 16.1% and 45.4% in recent reports[24]. There are no known effective pharmacological interventions for ARDS, proven interventions are limited to supportive care. Leveraging cutting-edge analytical workflows (*e.g.*, ML and artificial intelligence approaches) in the analysis of large datasets from the inpatient setting is likely to identify critical risk factors most associated with poor outcomes of ARDS and facilitate the development of strategies for early intervention in patients at the highest risk. In the current study, we have applied a mathematical modelling approach based on state-of-the-art ML algorithms to identify the most discriminative clinical features of ARDS-associated mortality. We compared efficacy of mortality prediction models derived from the ARDSNet FACTT trial datasets (day 0 vs. day 3), and the ML approach revealed key clinical features representing multi-organ dysfunction for ARDS mortality prediction.

There is an increasing trend of utilizing modern ML algorithms to develop and evaluate highly accurate classifier models for both early diagnosis and mortality prediction in ARDS [25,26]. However, most of the studies only examined data collected at baseline despite the fact that predictive values tend to be improved at early days beyond ARDS onset (*e.g.*, at day 3). We applied six commonly adopted ML classifiers to data collected at two early time points – day 0 and day 3, for predicting ARDS mortality. The area under the receiver operating characteristic curve, confusion matrix, precision, sensitivity and F1 score (which is the harmonic mean of the precision and sensitivity) were used to evaluate and compare the comprehensive performance of model types. As expected, the accuracy of the prediction values (*e.g.*, AUC values) increased from approximately 0.7 at day 0 to above 0.8 at day 3 in the prediction models. The RF classifier consistently demonstrated outstanding performance at both days, and obtained the highest AUC value of 0.84 at day 3 (Fig. 2B). The SC classifier which also displayed outstanding performance at day 3, is an ensemble ML algorithm that learns how to best combine the predictions from multiple well-performing ML models to obtain better predictive performance[27]. With regard to RF, it is a powerful ensemble-based ML classifier made up of multiple decision trees[28], and it demonstrated high predictive ability for prognostication in the data from the FACTT trial. During the training process, it randomly samples the training dataset with replacement (also known as bootstrapping) to build a decision tree. Additionally, it considers random subsets of features to split the nodes. The final predictions of the RF are made by averaging the predictions of each individual tree. Additionally, we analyzed the importance of clinical features in predicting 60-day mortality within the RF model. Our results suggested seven clinical parameters (MAP, bicarbonate, age, platelet count, albumin, heart rate and glucose) were the most important features at day 3 for ARDS mortality risk in the FACTT trial. The illustration of clinical feature importance (Fig. 3) may give physicians an intuitive understanding of the key features within the RF model, which provided the highest precision and sensitivity.

To account for the dynamic process of disease progression, efforts have been made previously to compare the performance of mortality risk factors utilizing data collected after ARDS onset. Bone et al.[29] analyzed the dynamic change of PaO_2/ FiO_2 ratio during the first 7 days of ARDS onset and found survivors had a significantly higher PaO_2/ FiO_2 ratio at day 1–7 than non-survivors, even though the two group of patients had a similar PaO_2/ FiO_2 ratio at day 0. Lai et al.[30] found that using clinical data from 1-day after ARDS onset could predict outcomes better than using data collected at baseline. Go et al.[12] examined the change of oxygenation index (OI) over the first seven days of ARDS and found that failure to improve OI at day 7 was associated with higher mortality. In this study, we also found evidence that variables measured after ARDS onset (*e.g.*, day 3) have better predictive performance than those at baseline. In the field of critical care medicine, increased accuracy in predicting mortality may have a major impact on various aspects of patient care, *i.e.* improved prognostication to allow more accurate patient stratification for clinical trials and help inform family discussions. Thus, utilizing clinical indices collected early after ARDS onset may improve the performance of mortality prediction in ARDS.

In this study, we found that MAP was the most significant clinical characteristic for predicting 60-day death, along with other crucial clinical features. Previous studies have reported the value of PEEP, plateau pressure or tidal volume in predicting mortality in ARDS patients[31]. In contrast, fewer studies have evaluated the predictive value of MAP. Recently, Sahetya SK *et al*[6] reported the prognostic value of MAP at baseline (within 24 h of mechanical ventilation). However, they did not assess whether dynamic data may provide enhanced predictive value. In this study (Supplementary figure S2), despite the fact that there was no significant difference in MAP between survivors and non-survivors at baseline, non-survivors showed a tendency toward significantly higher MAP than survivors by days one to three ($P<0.001$), suggesting that MAP represents the most important predictor for ARDS mortality. While driving pressure and plateau pressure reflect lung stress, MAP provides a more complete estimation of lung disease severity, respiratory compliance, and need for respiratory support than driving pressure or plateau pressure alone. MAP will increase if airway resistance increases, compliance of the lung or chest wall decreases, or dead space and work of breathing increase[23]. Plateau pressure represents the stiffness in the respiratory system, which predicts mortality in patients with ARDS[3]. MAP correlates directly with plateau pressure but also varies with minute ventilation, which could reflect dead space or acidosis. Furthermore, MAP took part in the entire respiratory cycle, contained more information of mechanical ventilation. In our study, we confirmed: (1) ARDS patients in the MAP_{high} subgroup had increased risk of metabolic dysfunction and organ injuries associated with high mortality rate ($P<0.001$, Supplementary table S3); and (2) utilizing mechanical ventilation parameters such as MAP collected at day 3 after treatment could provide better prognostic value for ARDS mortality.

The strengths of our study include utilizing robust clinical parameters collected in a large multi-center study – the FACTT trial, and implementing the state-of-the art ML workflows. However, our study has some limitations. First, we generated the ML models from secondary analyses of previously conducted randomized

controlled trial; these models must be evaluated in observational cohorts prospectively before they can be generalized to the ARDS population and used in the clinical setting. Second, given that the proposed ML method is data-driven, repeating the whole procedure in data collected from early days beyond day 3 (*e.g.*, day 7) may reveal various performances. Third, a combination of clinical predictors and biomarkers may enhance the performance of mortality predicting models in ARDS[2,32]. Biological markers of cell-specific injury, acute inflammation, and altered coagulation were correlate with mortality in multicenter clinical trials in ARDS[33–36]. Further, novel biomarkers discovered by a systems biology multi- "omics" approach may hold the promise to establish predictive or prognostic stratification methods and ultimately helps to develop more tailored therapeutics for ARDS patients[37,38].

Of note, we have considered an internal validation strategy in our study design. We utilized a common approach that is to split the single FACTT dataset into two parts: a training cohort, and a separate testing/validation cohort that is not used in developing the model itself. The superior performance of ML classifiers in the training sets was confirmed in the testing sets. The decision rules developed within the RF model at day 3 can predict the mortality rates of patients in advance with more than 80% accuracy. Given the novelty of our findings and its potential for translation into practice, the model was further developed into a web ARDS mortality 'calculator' (https://mortality-predictor.streamlit.app/). This exploratory online tool uses the RF classifier, which is already trained on data from day 3 of the FACTT trial to compute the prediction and returns a numerical value for ARDS mortality (Supplementary figure S6).

In conclusion, utilizing a large ARDS dataset, we developed ML-based models for risk stratification in critically ill ARDS patients and identified MAP as the most important clinical predictor for mortality. Future prospective research is warranted to validate the proposed models and to translate the advantages of ML models into improved patient outcomes through early intervene.

Data availability

The datasets used and/or analyzed during the current study are available from http://www.ardsnet.org/.

Received: 17 November 2023; Accepted: 26 July 2024
Published online: 01 August 2024

References

1. Rubenfeld, G. D. & Herridge, M. S. Epidemiology and outcomes of acute lung injury. *Chest* **131**, 554–562 (2007).
2. Ware, L. B. *et al.* Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury. *Chest* **137**, 288–296 (2010).
3. Hager, D. N., Krishnan, J. A., Hayden, D. L., Brower, R. G. & Network, A. C. T. Tidal volume reduction in patients with acute lung injury when plateau pressures are not high. *Am. J. Respiratory Crit. Care Med.* **172**, 1241–1245 (2005).
4. Gajic, O. *et al.* Prediction of death and prolonged mechanical ventilation in acute lung injury. *Critical Care* **11**, R53 (2007).
5. Balzer, F. *et al.* Predictors of survival in critically ill patients with acute respiratory distress syndrome (ARDS): An observational study. *BMC Anesthesiology* **16**, 108 (2016).
6. Sahetya, S. K. *et al.* Mean Airway Pressure As a Predictor of 90-Day Mortality in Mechanically Ventilated Patients. *Crit. Care Med.* **48**, 688–695 (2020).
7. Sayed, M., Riano, D. & Villar, J. Novel criteria to classify ARDS severity using a machine learning approach. *Crit. Care* **25**, 150 (2021).
8. Rehm, G. B. *et al.* Use of Machine Learning to Screen for Acute Respiratory Distress Syndrome Using Raw Ventilator Waveform Data. *Crit. Care Explor.* **3**, e0313 (2021).
9. National Heart, L. *et al.* Comparison of two fluid-management strategies in acute lung injury. *N. Engl. J. Med.* **354**, 2564–2575 (2006).
10. National Heart, L. *et al.* Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. *N. Engl. J. Med.* **354**, 2213–2224 (2006).
11. Wiedemann, H. P. A perspective on the fluids and catheters treatment trial (FACTT). Fluid restriction is superior in acute lung injury and ARDS. *Cleve Clin. J. Med.* **75**, 42–48 (2008).
12. Go, L. *et al.* Failure to Improve the Oxygenation Index Is a Useful Predictor of Therapy Failure in Acute Respiratory Distress Syndrome Clinical Trials. *Crit. Care Med.* **44**, e40-44 (2016).
13. Sapru, A. *et al.* Plasma soluble thrombomodulin levels are associated with mortality in the acute respiratory distress syndrome. *Intens. Care Med.* **41**, 470–478 (2015).
14. Delucchi, K. *et al.* Stability of ARDS subphenotypes over time in two randomised controlled trials. *Thorax* **73**, 439–445 (2018).
15. Shen, Y. *et al.* Interaction between low tidal volume ventilation strategy and severity of acute respiratory distress syndrome: A retrospective cohort study. *Critical Care* **23**, 254 (2019).
16. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work?. *International J. Methods Psychiatric Res.* **20**, 40–49 (2011).
17. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
18. Kilinc, G. & Atasoy, A. A. Evaluation of Patients Treated in Intensive Care Due to COVID-19: A Retrospective Study. *Infection Chemother.* **54**, 328–339 (2022).
19. Alvarado, M., Campos-Campos, L., Guerrero-Romero, F. & Simental-Mendia, L. E. The Triglycerides and Glucose Index Is an Independent Risk Factor for Acute Respiratory Distress Syndrome in Patients with COVID-19. *Metab. Syndrome Related Disord.* https://doi.org/10.1089/met.2023.0247 (2024).
20. Wu, S. *et al.* Factors Associated with Mortality Among Severe Omicron Patients for COVID-19. *Infection Drug Resist.* **17**, 1309–1319 (2024).
21. Kumar, M. *et al.* Hypoalbuminemia: incidence and its impact on acute respiratory distress syndrome and 28-day outcome in trauma patients. *European J. Trauma Emergency Surg.* **49**, 2305–2314 (2023).
22. Sinha, P. *et al.* Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome phenotypes: A secondary analysis of randomised controlled trials. *Lancet. Respir. Med.* **8**, 247–257 (2020).
23. Marini, J. J. & Ravenscraft, S. A. Mean airway pressure: physiologic determinants and clinical importance–Part 1: Physiologic determinants and measurements. *Crit. Care Med.* **20**, 1461–1472 (1992).
24. Bellani, G. *et al.* Noninvasive Ventilation of Patients with Acute Respiratory Distress Syndrome. Insights from the LUNG SAFE Study. *Am. J. Respir. Crit. Care Med.* **195**, 67–77 (2017).

25. Le, S. *et al.* Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). *J. Crit. Care* **60**, 96–102 (2020).
26. Sinha, P., Churpek, M. M. & Calfee, C. S. Machine Learning Classifier Models Can Identify Acute Respiratory Distress Syndrome Phenotypes Using Readily Available Clinical Data. *Am. J. Respire. Acrit. Care Med.* **202**, 996–1004 (2020).
27. Smyth, P. & Wolpert, D. Linearly combining density estimators via stacking. *Mach. Learn.* **36**, 59–83 (1999).
28. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324 (2001).
29. Bone, R. C. *et al.* An early test of survival in patients with the adult respiratory distress syndrome. The PaO2/FIo2 ratio and its differential response to conventional therapy. Prostaglandin E1 Study Group. *Chest* **96**, 849–851 (1989).
30. Lai, C. C. *et al.* The Ratio of Partial Pressure Arterial Oxygen and Fraction of Inspired Oxygen 1 Day After Acute Respiratory Distress Syndrome Onset Can Predict the Outcomes of Involving Patients. *Medicine (Baltimore)* **95**, e3333 (2016).
31. Amato, M. B. *et al.* Driving pressure and survival in the acute respiratory distress syndrome. *N. England J. Med.* **372**, 747–755 (2015).
32. Zhao, Z. *et al.* External validation of a biomarker and clinical prediction model for hospital mortality in acute respiratory distress syndrome. *Intens. Care med.* **43**, 1123–1131 (2017).
33. Eisner, M. D. *et al.* Plasma surfactant protein levels and clinical outcomes in patients with acute lung injury. *Thorax* **58**, 983–988 (2003).
34. Ware, L. B., Fang, X. & Matthay, M. A. Protein C and thrombomodulin in human acute lung injury. *Am. J. Physiol. Lung. Cell. Mol. Physiol.* **285**, L514-521 (2003).
35. Ware, L. B., Eisner, M. D., Thompson, B. T., Parsons, P. E. & Matthay, M. A. Significance of von Willebrand factor in septic and nonseptic patients with acute lung injury. *Am. J. Respir. Crit. Care Med.* **170**, 766–772 (2004).
36. Parsons, P. E., Matthay, M. A., Ware, L. B., Eisner, M. D. & National Heart, L. B. I. A. R. D. S. C. T. N. Elevated plasma levels of soluble TNF receptors are associated with morbidity and mortality in patients with acute lung injury. *Am. J. Physiol. Lung. Cell. Mol. Physiol.* **288**, L426-431 (2005).
37. Liao, S. Y. *et al.* Identification of early and intermediate biomarkers for ARDS mortality by multi-omic approaches. *Sci. Rep.* **11**, 18874 (2021).
38. Zheng, F. *et al.* Novel biomarkers for acute respiratory distress syndrome: Genetics, epigenetics and transcriptomics. *Biomark. Med.* **16**, 217–231 (2022).

Acknowledgements

This manuscript was prepared using FACTT Research Materials obtained from the NHLBI. The authors thank the FACTT participants and participating physicians, investigators and staff for making this research possible. More information about the study is at http://www.ardsnet.org/.

Author contributions

The project was conceived and planned by M.D., L.G. and P.H. Statistical analysis was performed by N.D., T.N. and L.G. The manuscript was written by N.D., T. N., L.G., M.D. and P.H. All authors read and approved the final manuscript.

Funding

This work was supported by National Institutes of Health (NIH) grant number R21HL145216 (to L.G. and P.H.).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-68653-8.

Correspondence and requests for materials should be addressed to L.G. or P.M.H.