

RESEARCH

Open Access



Novel criteria to classify ARDS severity using a machine learning approach

Mohammed Sayed^{1*}, David Riaño^{1*} and Jesús Villar^{2,3,4*} 

Abstract

Background: Usually, arterial oxygenation in patients with the acute respiratory distress syndrome (ARDS) improves substantially by increasing the level of positive end-expiratory pressure (PEEP). Herein, we are proposing a novel variable $[PaO_2/(FiO_2 \times PEEP)]$ or P/FP_E for $PEEP \geq 5$ to address Berlin's definition gap for ARDS severity by using machine learning (ML) approaches.

Methods: We examined P/FP_E values delimiting the boundaries of mild, moderate, and severe ARDS. We applied ML to predict ARDS severity after onset over time by comparing current Berlin PaO_2/FiO_2 criteria with P/FP_E under three different scenarios. We extracted clinical data from the first 3 ICU days after ARDS onset ($N = 2738, 1519, \text{ and } 1341$ patients, respectively) from MIMIC-III database according to Berlin criteria for severity. Then, we used the multicenter database eICU (2014–2015) and extracted data from the first 3 ICU days after ARDS onset ($N = 5153, 2981, \text{ and } 2326$ patients, respectively). Disease progression in each database was tracked along those 3 ICU days to assess ARDS severity. Three robust ML classification techniques were implemented using Python 3.7 (LightGBM, RF, and XGBoost) for predicting ARDS severity over time.

Results: P/FP_E ratio outperformed PaO_2/FiO_2 ratio in all ML models for predicting ARDS severity after onset over time (MIMIC-III: AUC 0.711–0.788 and CORR 0.376–0.566; eICU: AUC 0.734–0.873 and CORR 0.511–0.745).

Conclusions: The novel P/FP_E ratio to assess ARDS severity after onset over time is markedly better than current PaO_2/FiO_2 criteria. The use of P/FP_E could help to manage ARDS patients with a more precise therapeutic regimen for each ARDS category of severity.

Keywords: Intensive care units, Acute respiratory distress syndrome, Lung severity, Machine learning, Prediction models

Background

Acute respiratory distress syndrome (ARDS) is an acute and intense inflammatory disease process of the lungs with an associated high mortality rate of about 40% in

non-COVID-19 ARDS patients [1, 2]. ARDS is a highly heterogeneous syndrome without a specific diagnostic test [3–5]. According to the LUNG-SAFE study, ARDS is unrecognized in more than half of patients at the time of fulfillment of ARDS criteria [1]. The current “Berlin definition” is under controversy [5–8]. The previous American-European Consensus Conference (AECC) [9] and the Berlin definitions are predominantly based on the value of the PaO_2/FiO_2 ratio at the time of ARDS onset [10].

A working definition of ARDS is essentially required for clinical trials, epidemiologic studies, and biological studies. Moreover, a definition of ARDS is required for

*Correspondence: mgamal.sayed@urv.cat; david.riano@urv.cat; jesus.villar54@gmail.com

¹ Banzai Research Group On Artificial Intelligence, Department of Computer Engineering, Universitat Rovira i Virgili, Av Paisos Catalans 26, 43007 Tarragona, Spain

³ Multidisciplinary Organ Dysfunction Evaluation Research Network, Research Unit, Hospital Universitario Dr Negrín, Barranco de la Ballena s/n, 4th Floor -South Wing, 35019 Las Palmas de Gran Canaria, Spain
Full list of author information is available at the end of the article



clinicians to initiate treatments that would improve clinical outcomes [11], although stratification of ARDS—as defined by Berlin criteria—has been shown not very useful for assessing lung severity [8, 12]. The empirical $\text{PaO}_2/\text{FiO}_2$ cut-offs for “severity” of 100, 200, and 300 mmHg are arbitrary and poorly validated [13]. A recently published Reevaluation of Systemic Early Neuromuscular Blockade (ROSE) trial emphasized the variability of these $\text{PaO}_2/\text{FiO}_2$ cut-offs as the investigators did not enroll patients based on the $\text{PaO}_2/\text{FiO}_2$ at the time of ARDS onset, but based on a $\text{PaO}_2/\text{FiO}_2 < 150$ mmHg within the first 48-h after ARDS diagnosis [14, 15]. The $\text{PaO}_2/\text{FiO}_2$ ratio strongly depends on ventilator settings, including positive end-expiratory pressure (PEEP), inspiratory/expiratory time (I:E) ratio, and FiO_2 , and the requirement of a minimum PEEP of 5 cmH₂O did not substantially improve Berlin prediction compared to AECC [13, 16]. Besides, Berlin definition does not account for the non-linear relationship of PaO_2 and FiO_2 [17] and has a limited predictive accuracy in recent trials [18–21].

Assessment of severity in ARDS remains a challenge. The relation between oxygenation and prognosis in ARDS varies among published reports [20]. For example, the current mild ARDS category may not be significantly associated with 28-day mortality [22–24]. However, although stratification of severity based on Berlin criteria may be helpful to identify severe ARDS patients, it may have less significance to differentiate between mild and moderate ARDS [20]. A recent study identified two different subgroups of moderate ARDS using a 150 mmHg $\text{PaO}_2/\text{FiO}_2$ threshold and may represent a more homogeneous distribution of ARDS patients across subgroups of severity [25–27]. Whether ARDS outcome relates to severity of respiratory failure [28], a higher severity is a risk factor for prolonged mechanical ventilation [19]. Since $\text{PaO}_2/\text{FiO}_2$ does not account for PEEP in its calculation, reported $\text{PaO}_2/\text{FiO}_2$ provides a sense of ARDS severity without knowledge of applied PEEP levels.

The main goal of this study was proposing a novel variable [$\text{PaO}_2/(\text{FiO}_2 \times \text{PEEP})$] or P/FP_E for $\text{PEEP} \geq 5$ cmH₂O that, together with corresponding thresholds, could serve as an improved criterion to assess ARDS severity. The thresholds are 60 to 40 mmHg/cmH₂O for mild, 40 to 20 for moderate, and less than 20 for severe. This new criterion adequately addressed Berlin’s definition gap in computing ARDS severity by including PEEP in the new oxygenation ratio. Increasing the PEEP level with the same FiO_2 yields different PaO_2 and SpO_2 [29]. Thus, including PEEP in calculating the degree of oxygenation severity could be better than current Berlin definition. We examined this hypothesis by applying machine learning (ML) approaches for predicting ARDS severity over time.

Methods

Study design and patient population

Two clinical databases were used for evaluation. Data of the first 3 ICU days (considering day 1 for representative data within the first 24 h after ARDS onset, day 2 for data within 24–48 h after onset, and day 3 for data within 48–72 h after onset) ($N = 2738, 1519, \text{ and } 1341$ patients, respectively) were extracted from a single-center database MIMIC-III (MetaVision, 2008–2012) [30]. The median length of an ICU stay (LOS) of all selected ARDS patients in MIMIC-III was 11.29 days (Q1–Q3: 7.85–17.54). Similarly, data of the first 3 ICU days after ARDS onset ($N = 5153, 2981, \text{ and } 2326$ patients, respectively) were extracted from a multicenter database eICU (2014–2015) [31]. The median length of an ICU LOS of all selected ARDS patients in eICU was 11.72 days (6.92–18.84). All selected patients from both databases fulfilled the Berlin criteria for ARDS and were stratified into mild, moderate or severe ARDS [6] and received mechanical ventilation (MV) for > 48 h [32, 33]. Disease progression of ARDS in each database was tracked along those 3 ICU days to assess lung severity. Patients younger than 18 years were excluded.

Data extraction

Clinical data of ARDS patients were extracted from both databases (MIMIC-III and eICU) using Python 3.7, an interpreted, interactive, object-oriented, open-source programming language. The selection of clinical variables was based on previous studies [1, 19, 34–37].

MIMIC-III

MIMIC-III is a large, publicly available database including de-identified health-related data of approximately 60,000 admissions of ICU patients [30]. The input variables include baseline demographic information (age); hemodynamic parameters including mean, maximum and minimum heart rate (HR); ventilator parameters including mean, maximum and minimum respiratory rate (RR), SpO_2 , and PEEP. These predictors on the third ICU day after assessing lung severity, including their description (mean and 95% CI), are presented in Table 1, and Additional file 1: Tables S1 and S2. The main target variable was ARDS severity (where 0 = mild, 1 = moderate, and 2 = severe). ICU mortality (Fig. 1, Additional file 1: Figs. S1 and S2) and duration of MV were also obtained (Additional file 1: Table S3).

eICU

eICU is a multicenter and publicly available ICU database with high level of detail in the data about more than 200,000 ICU admissions [31]. Input variables included: baseline demographic information (age); ventilator

Table 1 Input variables and their descriptive statistics in MIMIC-III at 72-h

	Mild	Moderate	Severe	All
A. ARDS patients	506 (37.73%)	678 (50.56%)	157 (11.71%)	1341 (100%)
B. Descriptive feature—means and 95% CI				
Age	61.77 [60.37, 63.17]	60.61 [59.42, 61.79]	60.24 [57.42, 63.07]	61.01 [60.14, 61.87]
PEEP	7.41 [7.11, 7.71]	9.40 [9.06, 9.75]	11.68 [10.83, 12.52]	8.92 [8.68, 9.16]
Heart Rate_Mean	92 [90, 94]	92 [91, 94]	96 [93, 99]	93 [92, 94]
Respiratory Rate_Mean	21 [20, 21]	21 [21, 22]	22 [21, 23]	21 [21, 22]
Heart Rate_Max	114 [112, 116]	114 [112, 116]	120 [116, 124]	115 [113, 116]
Heart Rate_Min	75 [74, 77]	76 [75, 78]	78 [75, 81]	76 [75, 77]
Respiratory Rate_Max	30 [29, 31]	30 [29, 31]	32 [31, 34]	30 [30, 31]
Respiratory Rate_Min	13 [13, 14]	13 [13, 14]	13 [13, 14]	13 [13, 14]
SpO ₂ _Mean	97 [97, 98]	96 [96, 97]	96 [95, 96]	97 [96, 97]
SpO ₂ _Max	100 [100, 101]	100 [99, 100]	100 [99, 100]	100 [99, 100]
SpO ₂ _Min	90 [89, 90]	88 [87, 89]	85 [83, 87]	88 [88, 89]

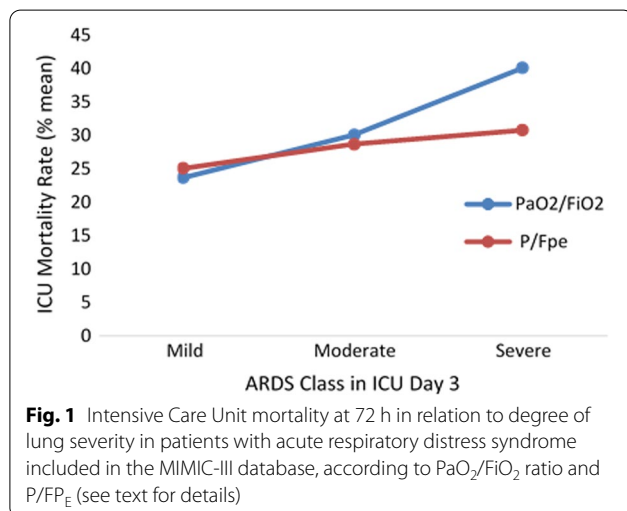


Fig. 1 Intensive Care Unit mortality at 72 h in relation to degree of lung severity in patients with acute respiratory distress syndrome included in the MIMIC-III database, according to PaO₂/FiO₂ ratio and P/F_pe (see text for details)

parameters including PEEP; blood gas parameters including FiO₂, PaO₂, and PaCO₂ (Table 2, Additional file 1: Tables S4 and S5). The main target variable was ARDS severity (where 0 = mild, 1 = moderate, and 2 = severe). ICU mortality (Fig. 2, Additional file 1: Figs. S3 and S4)

and duration of MV were also obtained (Additional file 1: Table S6).

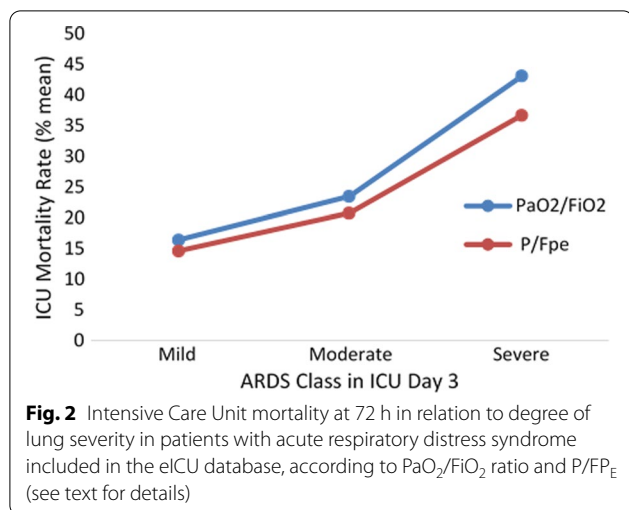
Experimental methods

Before starting our analysis, the thresholds of the P/F_pe index (with PEEP ≥ 5) were experimentally tuned. We computed the minimum and maximum P/F_pe values of the patients in the two databases, which were 2 and 60 mmHg/cmH₂O, respectively. Then, several cut-offs were studied in order to determine the ones that could be more accurate in the stratification of ARDS severity. For this purpose, we tested round values (to be easily remembered by intensivists) in the range 2–60 and analyzed P/F_pe index of the ARDS severity groups obtained. The partition showing a better separation of the ARDS severity groups obtained was achieved in this study for the following thresholds (with PEEP ≥ 5): 60–40 for mild, 40–20 for moderate, and < 20 for severe.

Our study is based only on ML analysis and not on the conventional statistical hypothesis testing analysis. In general, ML is an exploratory process and a current application of artificial intelligence to generate predictive models. Using this technology, there is not a

Table 2 Input variables and their descriptive statistics in eICU at 72-h

	Mild	Moderate	Severe	All
A. ARDS patients	872 (37.49%)	1025 (44.07%)	429 (18.44%)	2326 (100%)
B. Descriptive feature—means and 95% CI				
Age	64.77 [63.78, 65.76]	62.73 [61.83, 63.64]	59.97 [58.67, 61.28]	62.99 [62.39, 63.59]
PEEP	5.95 [5.80, 6.09]	7.16 [6.99, 7.34]	10.09 [9.72, 10.46]	7.25 [7.12, 7.38]
FiO ₂	0.40 [0.39, 0.41]	0.50 [0.49, 0.51]	0.81 [0.79, 0.83]	0.52 [0.51, 0.53]
PaO ₂	98.89 [97.17, 100.62]	80.52 [79.25, 81.78]	74.81 [72.83, 76.79]	86.36 [85.34, 87.37]
PaCO ₂	39.93 [39.33, 40.53]	42.38 [41.72, 43.04]	44.23 [43.13, 45.33]	41.80 [41.38, 42.23]



one-model-fits-all solution. Precisely, there is no ML method that reaches the highest accuracy for all domains, datasets, or problem types [38]. The optimal model differs from one problem to another based on the characteristics of variables and observations. Our aim was to implement ML models capable of predicting ARDS severity over time to compare the PaO₂/FiO₂ ratio—as mandated by the current Berlin criteria for ARDS—with the proposed new P/FP_E ratio according to the following three scenarios: (1) Scenario I: predicting ARDS severity in the 3rd ICU day using information captured in the 1st ICU day; (2) Scenario II: predicting ARDS severity in the 3rd ICU day using information captured in the 2nd ICU day; (3) Scenario III: predicting ARDS severity in the 3rd ICU day using information captured in the 1st and 2nd ICU days.

We implemented three robust supervised ML algorithms using Python 3.7. The ML algorithms were Light Gradient Boosting Machine (LightGBM) [39], Random Forest (RF) [40], and eXtreme Gradient Boosting (XGBoost) [41]. Grid search was used to identify the optimal values for their input parameters. The quality of the prediction models was computed based on a tenfold cross-validation approach. AUC and CORR (correlation between the predicted and actual values of severity level) were used to assess model performance in predicting ARDS severity as a categorical prediction. To provide a meaning to the findings, we used the classification of performance suggested by Hosmer and Lemeshow [42]: “excellent” if AUC ≥ 0.9; “good” if AUC is between 0.8 and 0.9; “fair” if AUC is between 0.7 and 0.8; “poor” if AUC is between 0.6 and 0.7; and “very poor” if AUC is below 0.6. For CORR, we used the interpretation suggested by Mukaka [43] who proposed “very high” for CORR ≥ 0.9 (positive correlation) or CORR ≤ -0.9 (negative

correlation); “high” if CORR is between 0.7 and 0.9 (positive) or -0.9 and -0.7 (negative); “moderate” if CORR is between 0.5 and 0.7 (positive) or -0.7 and -0.5 (negative); “low” if CORR is between 0.3 and 0.5 (positive) or -0.5 and -0.3 (negative), and “negligible” otherwise.

Results

The findings of the three classification ML methods for the three predictive scenarios in the two databases are presented in Tables 3 and 4. Table 3 shows the quality of ML predictions for MIMIC-III, confronting the results obtained for PaO₂/FiO₂ (Table 3(a)) with those obtained for P/FP_E (Table 3(b)). Table 4 shows the same comparative results in patients from the eICU database.

For MIMIC-III, the best ML severity predictive model on the third ICU day was obtained by scenario II and by P/FP_E with an AUC = 0.788 and CORR = 0.566, using LightGBM algorithm. When PaO₂/FiO₂ is used, AUC = 0.635 and CORR = 0.19, but these performances were obtained with different algorithms. In qualitative terms, P/FP_E ratio improves PaO₂/FiO₂ ratio from “poor” to “fair” AUC, and from “negligible” to “moderate” CORR.

For the eICU database, the results were slightly better. The best ML severity predictive model was also observed for scenario II. This finding confirms that the best approach to predict ARDS severity on the third ICU day is to consider the condition of the patient in the second ICU day after ARDS onset, rather than the first ICU day or both. For eICU data, the best AUC and CORR values are 0.873 and 0.745 for P/FP_E; and 0.863 and 0.725 for PaO₂/FiO₂. These results are qualified as a “good” predictive accuracy and a “high” correlation.

In general, P/FP_E ratio has a better behavior in the prediction of ARDS severity than PaO₂/FiO₂ ratio in terms of AUC and CORR. Whereas PaO₂/FiO₂ obtained up to 0.635 AUC and up to 0.19 CORR in MIMIC-III, the use of P/FP_E reached 0.788 AUC and 0.566 CORR. This represents increments of +0.153 AUC and +0.376 CORR and shows the advantages of using the P/FP_E ratio.

Discussion

In this large study, we propose a novel variable or formula (P/FP_E) and corresponding thresholds for classifying ARDS severity. We investigated several ML methods to generate severity predictive models in almost 8,000 patients with ARDS over time after ARDS diagnosis. Our findings confirmed that the best approach to predict ARDS severity on the third ICU day is to consider the condition of the patient in the second ICU day after ARDS onset, rather than during the first ICU day as mandated by Berlin criteria.

Table 3 Quality of the third ICU day severity predictive ML models for MIMIC-III

Algorithm	AUC, mean \pm SD	CORR, mean \pm SD
(a) PaO₂/FiO₂ results		
<i>Scenario I: Predicting ARDS Severity in the 3rd ICU day using the data in 1st ICU day</i>		
XGBoost	0.616 \pm 0.039	0.190 \pm 0.068
RF	0.622 \pm 0.048	0.173 \pm 0.089
LightGBM	0.612 \pm 0.039	0.138 \pm 0.084
<i>*Scenario II: Predicting ARDS Severity in the 3rd ICU day using the data in 2nd ICU day</i>		
XGBoost	0.621 \pm 0.023	0.147 \pm 0.121
*RF	0.635 \pm 0.020	0.139 \pm 0.094
LightGBM	0.622 \pm 0.025	0.126 \pm 0.120
<i>Scenario III: Predicting ARDS Severity in the 3rd ICU day using the data in 1st & 2nd ICU days</i>		
XGBoost	0.619 \pm 0.030	0.150 \pm 0.106
RF	0.627 \pm 0.022	0.177 \pm 0.108
LightGBM	0.618 \pm 0.022	0.086 \pm 0.101
(b) P/FP_E results		
<i>Scenario I: Predicting ARDS Severity in the 3rd ICU day using the data in 1st ICU day</i>		
XGBoost	0.711 \pm 0.029	0.385 \pm 0.064
RF	0.712 \pm 0.027	0.408 \pm 0.060
LightGBM	0.716 \pm 0.029	0.376 \pm 0.073
<i>*Scenario II: Predicting ARDS Severity in the 3rd ICU day using the data in 2nd ICU day</i>		
XGBoost	0.785 \pm 0.025	0.514 \pm 0.053
RF	0.787 \pm 0.023	0.546 \pm 0.061
*LightGBM	0.788 \pm 0.020	0.566 \pm 0.044
<i>Scenario III: Predicting ARDS Severity in the 3rd ICU day using the data in 1st & 2nd ICU days</i>		
XGBoost	0.782 \pm 0.025	0.548 \pm 0.049
RF	0.780 \pm 0.023	0.538 \pm 0.065
LightGBM	0.785 \pm 0.021	0.511 \pm 0.055

* Identifies the optimal scenario and ML model

For the MIMIC-III database, predictive models using the P/FP_E ratio attained outstanding improvements in terms of AUC (15% improvement) and CORR (37.6% improvement), when compared to the previous PaO₂/FiO₂ models. For the eICU database, models based on P/FP_E also outperformed PaO₂/FiO₂ predictions, with 14.8% and 2% improvements of AUC and CORR, respectively. The difference in terms of the accuracy between the two databases is remarkable regarding CORR. This is due to the fact that eICU is a multicenter ICU database with high granularity data (i.e., high level of detail in the data) for over 200,000 admissions to ICUs. By contrast, MIMIC-III is a single-center ICU database for approximately 60,000 admissions of ICU patients. Therefore, in all extracted data of the three ICU days, the number of extracted patients from eICU was greater than the number of extracted patients from MIMIC-III. Consequently, this would lead to better ML results in terms of CORR for the eICU database. Overall, the novel P/FP_E ratio outperformed the PaO₂/FiO₂ ratio in all ML applied models and showed that predictions based on the patient condition

in the second day after onset are better than predictions based on the first 24 h (7.2–13.8% AUC and 1.5–22% CORR improvements), followed by the predictions based on both the first and the second day conditions (0.1–0.3% AUC and 0.18–14% CORR improvements).

In contrast to our study, most recent studies developed ML approaches to predict the risk of ARDS in critically ill patients prior to ARDS onset [36, 44, 45], based on single-center databases [36, 45] and using one single ML algorithm [36, 44]. Consequently, their findings have serious limitations for the generalizability in the context of assessing the prediction of ARDS outcome.

This large study proposes a novel criterion to reclassify ARDS patients in terms of severity by using ML methods on an extensive amount of data from two large datasets of critically ill patients. The relatively good accuracy of P/FP_E (when compared to PaO₂/FiO₂) in stratifying ARDS patients could allow to overcome the major clinical drawbacks of the current Berlin definition. Also, this study is implementing ML models for predicting severity over time after ARDS onset. Critically ill patients are an

Table 4 Quality of the third ICU day severity predictive ML models for eICU

Algorithm	AUC, mean \pm SD	CORR, mean \pm SD
(a) PaO₂/FiO₂ results		
<i>Scenario I: Predicting ARDS Severity in the 3rd ICU day using the data in 1st ICU day</i>		
XGBoost	0.712 \pm 0.032	0.398 \pm 0.061
RF	0.714 \pm 0.030	0.393 \pm 0.059
LightGBM	0.713 \pm 0.028	0.373 \pm 0.069
<i>*Scenario II: Predicting ARDS Severity in the 3rd ICU day using the data in 2nd ICU day</i>		
*XGBoost	0.863 \pm 0.016	0.725 \pm 0.028
RF	0.863 \pm 0.016	0.700 \pm 0.040
LightGBM	0.860 \pm 0.014	0.714 \pm 0.028
<i>Scenario III: Predicting ARDS Severity in the 3rd ICU day using the data in 1st & 2nd ICU days</i>		
XGBoost	0.860 \pm 0.015	0.717 \pm 0.025
RF	0.854 \pm 0.017	0.693 \pm 0.038
LightGBM	0.857 \pm 0.014	0.713 \pm 0.027
(b) P/FP_E results		
<i>Scenario I: Predicting ARDS Severity in the 3rd ICU day using the data in 1st ICU day</i>		
XGBoost	0.735 \pm 0.034	0.525 \pm 0.056
RF	0.735 \pm 0.034	0.514 \pm 0.057
LightGBM	0.734 \pm 0.034	0.511 \pm 0.053
<i>*Scenario II: Predicting ARDS Severity in the 3rd ICU day using the data in 2nd ICU day</i>		
*XGBoost	0.873 \pm 0.022	0.745 \pm 0.033
RF	0.868 \pm 0.016	0.739 \pm 0.039
LightGBM	0.869 \pm 0.023	0.728 \pm 0.043
<i>Scenario III: Predicting ARDS Severity in the 3rd ICU day using the data in 1st & 2nd ICU days</i>		
XGBoost	0.872 \pm 0.020	0.725 \pm 0.040
RF	0.860 \pm 0.015	0.731 \pm 0.038
LightGBM	0.871 \pm 0.022	0.717 \pm 0.040

ideal population for clinical database investigations using machine learning algorithms because while the data from ICUs are extensive, the value of many diagnostic and therapeutic interventions remains largely unproven [46].

ARDS is considered one of the major reasons of ICU admission, and it is associated with a high hospital mortality [1]. Despite its high mortality rate and high rates of ICU utilization, ARDS remains critically misdiagnosed and globally under-diagnosed in the ICU settings [1]. Furthermore, increasing ARDS severity is associated with increased mortality rate [6]. The PaO₂/FiO₂ ratio categorizes ARDS patients according to the severity of their oxygenation deficit without considering the level of applied PEEP in the assessment of lung severity. The PaO₂/FiO₂ ratio does not appropriately show the severity of ARDS for PEEP \geq 5. However, the application of PEEP plays a significant role in improving oxygenation. It is well established that changes in PEEP alter the PaO₂/FiO₂ in lung-injured patients [29]. Attempting to predict lung severity and patient outcomes based solely in PaO₂/FiO₂ on this basis is inherent flawed. Thus, the stratification of ARDS patients as proposed by the Berlin criteria

is useless for assessing severity of lung injury and could be of no benefit for enrolling patients into therapeutic clinical trials. The P/FP_E for PEEP \geq 5 appropriately addressed Berlin's definition gap in computing ARDS severity by including PEEP in the novel ratio. Clearly, our study showed that P/FP_E thresholds improved prediction of ARDS severity. This can lead to important medical implications by accurately anticipate specific treatment for each ARDS category, which could eventually decrease ARDS mortality. In other words, P/FP_E can represent a good solution for the clinical assessment of ARDS severity and as a guidance for treatment of ARDS.

Our study has several strengths. First, we have analyzed a large population of ARDS patients within their first three ICU days after onset. Second, we have described and validated our findings using both a large single-center database (MIMIC-III) and a large multi-center database (eICU). Third, we have investigated several ML predictive models for ARDS severity over time after ARDS onset. We believe that our approach is generalizable across other ARDS populations. However, we acknowledge some limitations to our study. First, our

work is based on a retrospective analysis of data whose results concerning P/FP_E benefits should be confirmed in further prospective studies. Second, our analysis is concerned with the evolution and stratification of patients in their third ICU day after ARDS onset. Although the first 72-h are essential in the management and progression of ARDS patients, our study lacks the assessment of a long-term outcome (e.g., ICU mortality, 60-day mortality). Third, further longitudinal studies on complete evolution of ARDS patients could help to find out new evidence(s) on the management of ARDS since our ML results achieved outstanding improvements compared to the current state, with “fair” to “good” predictions of ARDS severity [42]. Forth, one could argue that extracorporeal membrane oxygenation (ECMO) is not considered in this study. ECMO is a clinical outcome and can only temporarily sustain severe ARDS patients to bridge periods of time when oxygenation through the lungs cannot be achieved via MV. Moreover, ECMO is a constrained resource that is not available in all ICUs. Hence, for the purpose of our study, we only considered patients receiving MV for >48 h [32, 33]. Fifth, regarding the potential consequences of using the new ratio at the bedside, further studies are needed to examine whether it could help for clinical decision making and guiding therapy. Our study opens a possibility to better define ARDS severity, as a new research area for patient care improvement.

Conclusions

This large study proposes a novel criterion based on the P/FP_E formula to assess ARDS severity using ML, which is significantly better than the current Berlin criteria using baseline PaO_2/FiO_2 . Clinically, applying the proposed new criteria for ARDS severity enables clinical care physicians to assess lung severity by involving PEEP information. Moreover, being able to better adjust the severity profiles of ARDS patients will potentially improve the selection of more adequate therapeutic regimens for each ARDS category, which could contribute to reduce ARDS mortality. However, additional studies are required in order to confirm this. In both databases (MIMIC-III and eICU) and either in Berlin or P/FP_E scenario II (assessment of oxygenation deficit after 24 h of ARDS diagnosis and routine ICU treatment) was the best severity predictive scenario. From a ML perspective, P/FP_E outperformed PaO_2/FiO_2 in all ML models predicting ARDS severity after onset over time in all scenarios either in MIMIC-III or eICU. Accordingly, this study can serve as an example of how ML is a worth-considering technology to gain new insights in the development of ARDS predictive models which could contribute to improve ICU resource allocation and mortality reduction.

Abbreviations

AECC: American–European Consensus Criteria; ARDS: Acute respiratory distress syndrome; AUC: Area under the receiver operating characteristic curve; CI: Confidence interval; CORR: Correlation between the predicted and actual values of the target variable; ECMO: Extracorporeal membrane oxygenation; eICU: EICU Collaborative Research Database; FiO_2 : Fraction of the oxygen in the inspired air; HR: Heart rate; LightGBM: Light Gradient Boosting Machine; ICU: Intensive care unit; ML: Machine learning; MIMIC-III: Medical Information Mart for Intensive Care Database; MV: Mechanical ventilation; $PaCO_2$: Partial pressure of arterial CO_2 ; PaO_2 : Arterial oxygen tension; PaO_2/FiO_2 or P/F : Ratio of partial pressure of arterial O_2 to fraction of inspired O_2 ; P/FP_E : New ARDS severity criteria; PEEP or P_E : Positive end-expiratory pressure; RF: Random forest; ROSE: Reevaluation of Systemic Early Neuromuscular Blockade; RR: Respiratory rate; SD: Standard deviation; SpO_2 : Oxygen saturation; XGBoost: Extreme Gradient Boosting.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13054-021-03566-w>.

Additional file 1. Predictors at 24-h and 48-h and other clinical outcomes, and their descriptive statistics in MIMIC-III and eICU.

Authors' contributions

M. Sayed had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. M. Sayed, D. Riaño and J. Villar participated in the research question. M. Sayed was responsible for developing the software and obtaining the results. M. Sayed and D. Riaño performed the analysis of the results. M. Sayed, D. Riaño and J. Villar were responsible for drafting the first and subsequent versions of the manuscript. J. Villar provided critical appraisal during data analysis. All authors read and approved the final manuscript.

Funding

No funding was received for the access, development and analysis of the data. M. Sayed and D. Riaño are funded by the Spanish Ministry of Science and Innovation, Madrid, Spain (PID2019-105789RB-I00). J. Villar is funded by Instituto de Salud Carlos III, Madrid, Spain (#CB06/06/1088; #PI16/00049; #PI19/00141), the European Regional Development's Funds (FEDER), and the Asociación Científica Pulmón y Ventilación Mecánica.

Availability of data and materials

By request to M. Sayed and D. Riaño.

Code availability

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable. The datasets used for the analysis in this study are publicly available, and the datasets for the analysis are de-identified.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest in relation to this manuscript.

Author details

¹ Banzai Research Group On Artificial Intelligence, Department of Computer Engineering, Universitat Rovira I Virgili, Av Paisos Catalans 26, 43007 Tarragona, Spain. ² Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain. ³ Multidisciplinary Organ Dysfunction Evaluation Research Network, Research Unit, Hospital Universitario Dr Negrín, Barranco de la Ballena s/n, 4th Floor -South Wing, 35019 Las Palmas de

Gran Canaria, Spain. ⁴ Keenan Research Center for Biomedical Science at the Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada.

Received: 14 January 2021 Accepted: 5 April 2021

Published online: 20 April 2021

References

- Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in Intensive Care Units in 50 countries. *JAMA*. 2016;315(8):788–800. <https://doi.org/10.1001/jama.2016.0291>.
- Thompson BT, Chambers RC, Liu KD. Acute Respiratory Distress Syndrome. *N Engl J Med*. 2017;377(19):1904–5. <https://doi.org/10.1056/nejmc1711824>.
- Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med*. 2014;2(8):611–20. [https://doi.org/10.1016/S2213-2600\(14\)70097-9](https://doi.org/10.1016/S2213-2600(14)70097-9).
- Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS. Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med*. 2018;44(11):1859–69. <https://doi.org/10.1007/s00134-018-5378-3>.
- Del Sorbo L, Ranieri VM, Ferguson ND. The Berlin definition met our needs: yes. *Intensive Care Med*. 2016;42(5):643–7. <https://doi.org/10.1007/s00134-016-4286-7>.
- ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, et al. Acute respiratory distress syndrome: the Berlin definition. *JAMA*. 2012;307(23):2526–33. <https://doi.org/10.1001/jama.2012.5669>.
- 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 Med*. 2012;38(10):1573–82. <https://doi.org/10.1007/s00134-012-2682-1>.
- Villar J, Perez-Mendez L, Kacmarek RM. The Berlin definition met our needs: no. *Intensive Care Med*. 2016;42(5):648–50. <https://doi.org/10.1007/s00134-016-4242-6>.
- Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. Report of the American-European consensus conference on ARDS: definitions, mechanisms, relevant outcomes and clinical trial coordination. *Intensive Care Med*. 1994;20(3):225–32. <https://doi.org/10.1007/bf01704707>.
- Costa EL, Amato MB. The new definition for acute lung injury and acute respiratory distress syndrome: is there room for improvement? *Curr Opin Crit Care*. 2013;19(1):16–23. <https://doi.org/10.1097/MCC.0b013e32835c50b1>.
- Thompson BT, Matthay MA. The Berlin definition of ARDS versus pathological evidence of diffuse alveolar damage. *Am J Respir Crit Care Med*. 2013;187(7):675–7. <https://doi.org/10.1164/rccm.201302-0385ed>.
- Villar J, Blanco J, del Campo R, Andaluz-Ojeda D, Díaz-Domínguez FJ, Muriel A, et al. Assessment of PaO₂/FI_O₂ for stratification of patients with moderate and severe acute respiratory distress syndrome. *BMJ Open*. 2015;5(3):e006812. <https://doi.org/10.1136/bmjopen-2014-006812>.
- Huber W, Findeisen M, Lahmer T, Herner A, Rasch S, Mayr U, et al. Prediction of outcome in patients with ARDS: A prospective cohort study comparing ARDS-definitions and other ARDS-associated parameters, ratios and scores at intubation and over time. *PLoS ONE*. 2020;15(5):e0232720. <https://doi.org/10.1371/journal.pone.0232720>.
- Moss M, Huang DT, Brower RG, Ferguson ND, Ginde AA, Gong MN, et al. Early neuromuscular blockade in the acute respiratory distress syndrome. *N Engl J Med*. 2019;380(21):1997–2008. <https://doi.org/10.1056/NEJMoA1901686>.
- Slutsky AS, Villar J. Early paralytic agents for ARDS? Yes, no, and sometimes. *N Engl J Med*. 2019;380(21):2061–3. <https://doi.org/10.1056/NEJMe1905627>.
- Phillips CR. The Berlin definition: real change or the emperor's new clothes? *Crit Care*. 2013;17(4):174. <https://doi.org/10.1186/cc12761>.
- Allardet-Servent J, Forel JM, Roch A, Guerville C, Chiche L, Castanier M, et al. FI_O₂ and acute respiratory distress syndrome definition during lung protective ventilation. *Crit Care Med*. 2009;37(1):202–07, e4–6. <https://doi.org/https://doi.org/10.1097/CCM.0b013e31819261db>
- Balzer F, Menk M, Ziegler J, Pille C, Wernecke KD, Spies C, et al. Predictors of survival in critically ill patients with acute respiratory distress syndrome (ARDS): an observational study. *BMC Anesthesiol*. 2016;16(1):108. <https://doi.org/10.1186/s12871-016-0272-4>.
- Dai Q, Wang S, Liu R, Wang H, Zheng J, Yu K. Risk factors for outcomes of acute respiratory distress syndrome patients: a retrospective study. *J Thorac Dis*. 2019;11(3):673–85. <https://doi.org/10.21037/jtd.2019.02.84>.
- Kamo T, Tasaka S, Suzuki T, Asakura T, Suzuki S, Yagi K, et al. Prognostic values of the Berlin definition criteria, blood lactate level, and fibro-proliferative changes on high-resolution computed tomography in ARDS patients. *BMC Pulm Med*. 2019;19(1):37. <https://doi.org/10.1186/s12890-019-0803-0>.
- Lai CC, Sung MI, Liu HH, Chen CM, Chiang SR, Liu WL, 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)*. 2016;95(14):e3333. <https://doi.org/10.1097/MD.0000000000003333>.
- Hernu R, Wallet F, Thiollière F, Martin O, Richard JC, Schmitt Z, et al. An attempt to validate the modification of the American-European consensus definition of acute lung injury/acute respiratory distress syndrome by the Berlin definition in a university hospital. *Intensive Care Med*. 2013;39(12):2161–70. <https://doi.org/10.1007/s00134-013-3122-6>.
- Pirracchio R, Gropper MA. Heterogeneity in Intensive Care: low severity does not mean low risk! *Anesthesiology*. 2019;130(2):190–1. <https://doi.org/10.1097/ALN.0000000000002537>.
- Pham T, Serpa Neto A, Pelosi P, Laffey JG, De Haro C, Lorente JA, et al. Outcomes of patients presenting with mild acute respiratory distress syndrome: insights from the LUNG SAFE study. *Anesthesiology*. 2019;130(2):263–83. <https://doi.org/10.1097/ALN.0000000000002508>.
- Maiolo G, Collino F, Vasques F, Rapetti F, Tonetti T, Romitti F, et al. Re-classifying acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 2018;197(12):1586–95. <https://doi.org/10.1164/rccm.201709-1804OC>.
- Bourenne J, Carvelli J, Papazian L. Evolving definition of acute respiratory distress syndrome. *J Thorac Dis*. 2019;11(Suppl 3):S390–3. <https://doi.org/10.21037/jtd.2018.12.24>.
- Villar J, Fernández RL, Ambrós A, Parra L, Blanco J, Domínguez-Berrot AM, et al. A clinical classification of the acute respiratory distress syndrome for predicting outcome and guiding medical therapy. *Crit Care Med*. 2015;43(2):346–53. <https://doi.org/10.1097/CCM.0000000000000703>.
- Ferring M, Vincent JL. Is outcome from ARDS related to the severity of respiratory failure? *Eur Respir J*. 1997;10(6):1297–300. <https://doi.org/10.1183/09031936.97.10061297>.
- Villar J, Pérez-Méndez L, López J, Belda J, Blanco J, Saralegui I, et al. An early PEEP/FIO₂ trial identifies different degrees of lung injury in patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 2007;176(8):795–804. <https://doi.org/10.1164/rccm.200610-1534OC>.
- Physionet.org, 'MIMIC-III Critical Care Database'. [Online]. Available: <https://mimic.physionet.org/about/mimic>. Accessed 2 July 2020.
- Physionet.org, 'eICU Collaborative Research Database'. [Online]. Available: <https://eicu-crd.mit.edu/about/eicu/>. Accessed 19 Oct 2020.
- Jia X, Malhotra A, Saeed M, Mark RG, Talmor D. Risk factors for ARDS in patients receiving mechanical ventilation for >48 h. *Chest*. 2008;133(4):853–61. <https://doi.org/10.1378/chest.07-1121>.
- Mahmoud O. Mechanical power is associated with increased mortality and worsened oxygenation in ARDS. *Chest*. 2020;158(4):A679.
- Monchi M, Bellenfant F, Cariou A, Joly LM, Thebert D, Laurent I, et al. Early predictive factors of survival in the acute respiratory distress syndrome. A multivariate analysis. *Am J Respir Crit Care Med*. 1998;158(4):1076–81. <https://doi.org/10.1164/ajrccm.158.4.9802009>.
- Pintado MC, de Pablo R, Trascasa M, Milicica JM, Rogero S, Daguerre M, et al. Individualized PEEP setting in subjects with ARDS: a randomized controlled pilot study. *Respir Care*. 2013;58(9):1416–23. <https://doi.org/10.4187/respcare.02068>.
- 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). *J Crit Care*. 2020;60:96–102. <https://doi.org/10.1016/j.jccr.2020.07.019>.
- Flaatten H, Gjerde S, Guttormsen AB, Haugen O, Hoivik T, Onarheim H, et al. Outcome after acute respiratory failure is more dependent on dysfunction in other vital organs than on the severity of the respiratory failure. *Crit Care*. 2003;7(4):R72. <https://doi.org/10.1186/cc2331>.

38. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J Clin Epidemiol*. 2013;66(4):398–407. <https://doi.org/10.1016/j.jclinepi.2012.11.008>.
39. Dehua W, Yang Z, Yi Z. LightGBM: An effective miRNA classification method in breast cancer patients. In: Proceedings of the 2017 international conference on computational biology and bioinformatics (ICCB 2017). ACM, New York, NY, USA, 2017;7–11. <https://doi.org/10.1145/3155077.3155079>
40. Boulesteix AL, Janitza S, Kruppa J, Konig IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Int Rev Data Min Knowl Discov*. 2012;2(6):493–507. <https://doi.org/10.1002/widm.1072>.
41. Chen T, Carlos G. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (KDD '16). Association for Computing Machinery, New York, NY, USA. 2016;785–94. <https://doi.org/10.1145/2939672.2939785>
42. Hosmer W, Lemeshow JR. Applied logistic regression. New York: Wiley; 2004.
43. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24(3):69–71.
44. Ding XF, Li JB, Liang HY, Wang ZY, Jiao TT, 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(1):326. <https://doi.org/10.1186/s12967-019-2075-0>.
45. Zeiberg D, Prahad T, Nallamothu BK, Iwashyna TJ, Wiens J, Sjoding MW. Machine learning for patient risk stratification for acute respiratory distress syndrome. *PLoS ONE*. 2019;14(3):e0214465. <https://doi.org/10.1371/journal.pone.0214465>.
46. Rush B, Stone DJ, Celi LA. From big data to artificial intelligence: harnessing data routinely collected in the process of care. *Crit Care Med*. 2018;46(2):345–6. <https://doi.org/10.1097/CCM.0000000000002892>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

