Machine learning predicts lung recruitment in acute respiratory distress syndrome using single lung CT scan

Francesca Pennati[1], Andrea Aliverti[1], Tommaso Pozzi[2], Simone Gattarello[3], Fabio Lombardo[3],
Silvia Coppola[4] and Davide Chiumello[2,4,5]*

Abstract

Background To develop and validate classifier models that could be used to identify patients with a high percentage of potentially recruitable lung from readily available clinical data and from single CT scan quantitative analysis at intensive care unit admission. 221 retrospectively enrolled mechanically ventilated, sedated and paralyzed patients with acute respiratory distress syndrome (ARDS) underwent a PEEP trial at 5 and 15 cmH_2O of PEEP and two lung CT scans performed at 5 and 45 cmH_2O of airway pressure. Lung recruitability was defined at first as percent change in not aerated tissue between 5 and 45 cmH_2O (radiologically defined; recruiters: Δ_{45-5}non-aerated tissue $> 15\%$) and secondly as change in PaO_2 between 5 and 15 cmH_2O (gas exchange-defined; recruiters: $\Delta_{15-5}PaO2 > 24$ mmHg). Four machine learning (ML) algorithms were evaluated as classifiers of radiologically defined and gas exchange-defined lung recruiters using different models including different variables, separately or combined, of lung mechanics, gas exchange and CT data.

Results ML algorithms based on CT scan data at 5 cmH_2O classified radiologically defined lung recruiters with similar AUC as ML based on the combination of lung mechanics, gas exchange and CT data. ML algorithm based on CT scan data classified gas exchange-defined lung recruiters with the highest AUC.

Conclusions ML based on a single CT data at 5 cmH_2O represented an easy-to-apply tool to classify ARDS patients in recruiters and non-recruiters according to both radiologically defined and gas exchange-defined lung recruitment within the first 48 h from the start of mechanical ventilation.

Key points

– Machine learning based on a single CT imaging at the admission in intensive care unit provided a reliable tool to classify ARDS patients in responder and not responder to lung recruitment within the first 48 hours from the start of mechanical ventilation.

Keywords ARDS, Machine learning, Tomography

*Correspondence:
Davide Chiumello
davide.chiumello@unimi.it
Full list of author information is available at the end of the article

Pennati *et al. Annals of Intensive Care*   (2023) 13:60

Page 2 of 11

Introduction

ARDS is typically defined as a non-cardiogenic pulmonary edema characterized by different degree of hypoxemia, alveolar shunt and not aerated lung regions [1]. The commonly suggested lung protective ventilation strategy includes lung recruitment maneuvers to reopen not-aerated lung regions (i.e., collapsed areas) in order to reduce VILI, improve lung oxygenation and CO_2 removal [2]. However, several data showed that lung recruitment, although improving the oxygenation, could also, at the same time, impair the hemodynamics without improving the 28-day mortality [3, 4]. Furthermore, a systematic review and meta-analysis reported that in patients with moderate–severe ARDS, the use of higher PEEP with prolonged lung recruitment was associated with increased risk of death compared to similar PEEP without lung recruitment [5]. Thus, selecting an appropriate ventilatory strategy balancing the levels of PEEP and lung recruitment is therefore critical.

Among the different lung imaging techniques, the CT is the reference method both for a morphological analysis and for an accurate quantitative computation of lung recruitability [6–8]. The measurement of lung potential recruitment is fundamental to establish the therapeutic efficacy of PEEP [8–10]. It has been reported that the amount of lung recruitability ranged from 0 up to 70% of the total lung weight [8, 10]. Moreover, the lung recruitability was poorly predictable, being affected by the distribution of the lung disease, amount of edema, timing of ARDS onset and alteration in respiratory mechanics [9–12].

Recently, several applications of machine learning techniques have been applied in critical care medicine with promising results [13, 14]. Machine learning algorithms have been proposed to classify patients into ARDS subphenotypes using readily available clinical data [15–19]. Various studies demonstrated that machine learning can be used to predict patients who required prolonged mechanical ventilation and also the outcome [14, 15]. In a secondary analysis of a randomized trial applying machine learning, three different ARDS clusters were found, differing in the injury effect of an open lung recruitment strategy and the outcome [20]. Recently, several prediction models of COVID-19 have been also developed with a focus on CT diagnosis and prognosis [21–25].

Thus, we hypothesized to apply machine learning algorithms for the detection of lung recruitment, defined both from radiological and gas exchange data, in ARDS patients. In particular, the aim of the present study was to develop and validate classifier models to identify patients with a high percentage of potentially recruitable lung from readily available clinical data (namely mechanics and gas exchange) and using single CT scan at admission at 5 cmH_2O of PEEP. The secondary aim was to develop models that use a more limited set of available clinical and CT scan data.

Materials and methods

The study is a retrospective analysis of ARDS patients previously enrolled from 2016 to December 2022 and partially included in other already published studies [26]. The study was approved by the Institutional Review board of our hospital (Comitato Etico Interaziendale Milano Area A, protocol number 2016/ST/143 on the 22nd June 2016, entitled "PEEP test") and informed consent was obtained according to the Italian regulations. The study protocol flowchart is shown in Additional file 1: Figure S1.

Study protocol

At Intensive Care Unit (ICU) admission, patients were maintained deeply sedated and paralyzed, ventilated in volume control ventilation, with a tidal volume between 6 and 8 ml/kg of ideal body weight, a respiratory rate to ensure an arterial carbon dioxide partial pressure ($PaCO_2$) between 40 and 50 mmHg; positive end-expiratory pressure (PEEP) and FiO_2 were set by the attending physician to ensure an arterial saturation between 88 and 92%. An esophageal balloon catheter (Smart Cath, Viasys, Palm Springs, USA) was placed in the lower third of the esophagus, as already described before [27].

Patients were enrolled within 48 h from ICU admission and the study protocol was started.

At the beginning of the study, a recruitment maneuver was performed in pressure controlled ventilation at PEEP 5 cmH_2O, with a plateau pressure of 45 cmH_2O, I:E 1:1, respiratory rate of 10 breaths/min for 2 min. Subsequently, the previously applied tidal volume and respiratory rate were resumed and a PEEP trial at 5 and 15 cmH_2O was performed; the FiO_2 was adjusted at the beginning of the trial to ensure an arterial saturation between 88 and 92%. At each PEEP level, after 20 min, end-inspiratory and end-expiratory pauses were performed and arterial and central venous blood gas analysis were obtained; consequently, we recorded partitioned respiratory mechanics, gas exchange and hemodynamics variables. Partitioned respiratory mechanics of lung and chest wall elastance was computed according to the following standard formulas [9]. Gas exchange-defined lung recruitment was assessed as the difference in PaO_2 between 15 and 5 cmH_2O of PEEP.

After the PEEP trial, two whole lung CT scan in static condition at 5 cmH_2O of end-expiratory airway pressure and 45 cmH_2O of end-inspiratory airway pressure were performed.

Pennati *et al. Annals of Intensive Care* (2023) 13:60

Page 3 of 11
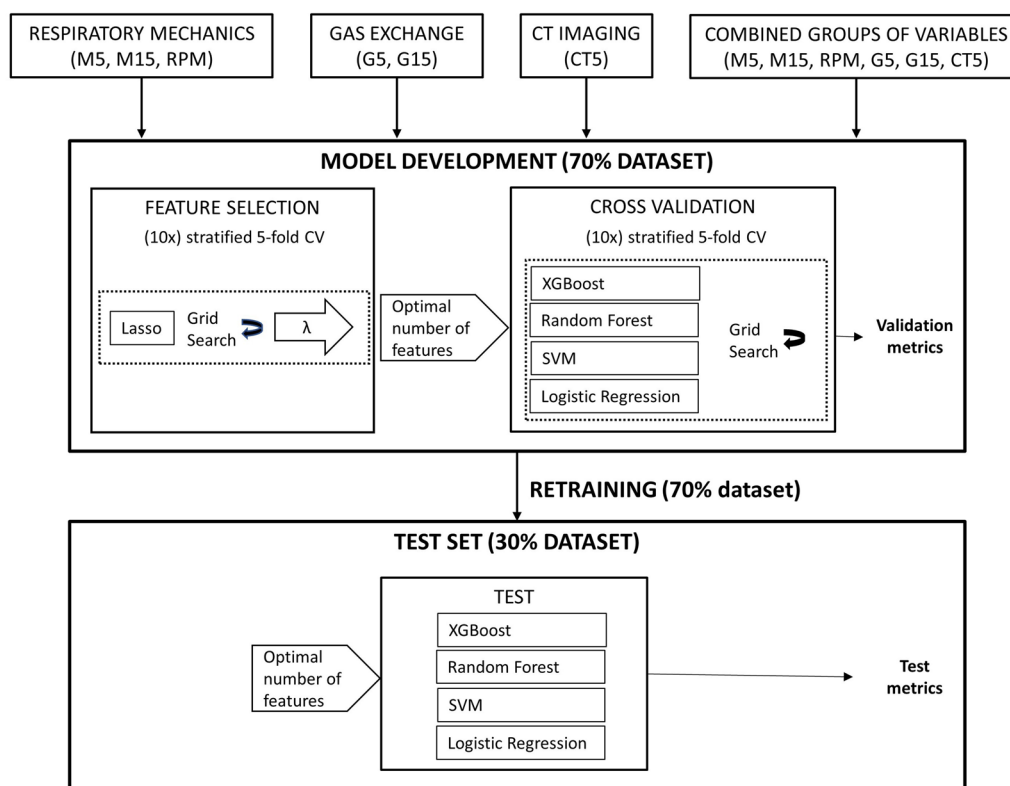
Fig. 1 The machine learning workflow. Input parameters included lung mechanics at PEEP 5 cmH$_2$O (M5), lung mechanics at PEEP 15 cmH$_2$O (M15), respiratory partitioned mechanics (RPM), gas exchange measured at PEEP 5 cmH$_2$O (G5), gas exchange measured at PEEP 15 cmH$_2$O (G15), CT imaging acquired at PEEP 5 cmH$_2$O (CT5). A grid search strategy with a stratified fivefold cross-validation repeated 10 times was performed to optimize algorithms' parameters, for both feature selection and model training. The hold-out test set was used to test the re-trained models in terms of mean area under the receiver operating characteristic curve (AUC), accuracy, sensitivity and specificity

Lung CT quantitative analysis

An integrated approach was used based on manual segmentation of the lung by a dedicated software and subsequently automatically analyzed (Soft-E-Film). The total lung weight, the gas volume, and the amount of the different compartment (not inflated, poor inflated, well inflated and overinflated) were computed [6]. Radiologically defined lung recruitment was assessed as the ratio between the difference in not aerated tissue at 5 cmH$_2$O and 45 cmH$_2$O of airway pressure to the total lung tissue weight at 5 cmH$_2$O of airway pressure.

Statistical analysis

Continuous variables are presented as mean ± standard deviation or median (interquartile range), as appropriate, whereas categorical data are reported as percentages. Clinical data of recruiters and non-recruiters at 5 cmH$_2$O, as well as the differences in respiratory mechanics and gas exchange between 5 cmH$_2$O and 45 cmH$_2$O of airway pressures were compared by the Student's t test or Mann–Whitney rank-sum test, as appropriate. Categorical data were compared by the Chi-square test. Tests were two-sided with significance α level set at less than 0.05.

Machine learning models

Machine learning models were implemented in Python using the Scikit-Learn package [28]. Data management was performed using the Pandas library [29]. The workflow is summarized in Fig. 1.

Outcome The outcome of interest was the prediction of patients with a percentage of potentially recruitable lung greater than the median value of the whole population. Radiologically defined lung recruitability was assessed as the ratio of the change in not aerated lung tissue between 5 cmH$_2$O and 45 cmH$_2$O to the total lung tissue weight at 5 cmH$_2$O at CT scan. Gas exchange-defined lung recruitment was assessed as the difference in PaO$_2$ between 15 and 5 cmH$_2$O of PEEP ($\Delta_{15\text{-}5}$ PaO$_2$). For the two outcomes, the median values for the whole population were 15% and 24 mmHg, respectively. Furthermore, to evaluate the model performance over different lung recruitability thresholds, the analyses on the overall dataset was

Pennati *et al. Annals of Intensive Care*   (2023) 13:60

Page 4 of 11

repeated for all models by assigning classes using cutoffs of 10%, 20% and 30% for radiologically defined lung recruitability and using cut-offs of 20 mmHg, 30 mmHg and 40 mmHg for gas exchange-defined lung recruitability.

Predictor variables   With the rationale of investigating the best measuring conditions in the clinical setting, classifier models were developed using demographic data with the addiction of sparse sets of variables that were grouped according to the variable type: lung mechanics (M), gas exchange (G) and CT imaging data acquired at PEEP 5 cmH$_2$O (CT5). Lung mechanics and gas exchange variable sets were further subdivided according to the measuring condition: lung mechanics at PEEP 5 cmH$_2$O (M5), lung mechanics at PEEP 15 cmH$_2$O (M15), respiratory partitioned mechanics (RPM), gas exchange measured at PEEP 5 cmH$_2$O (G5) and gas exchange measured at PEEP 15 cmH$_2$O (G15). A total of 44 features were used as independent variables for the development of the models (Additional file 1: Figure S2).

Data pre-processing   To ensure the availability of all predictors in models' development, we excluded features with more than 30% missing data. In the remaining features, residual missing data were imputed with the median value of the respective feature. Finally, the data were normalized and scaled to have zero mean and unit variance such that variables with different scales can contribute equally to the analysis. For the purposes of evaluation, we reserved 30% of the dataset, chosen at random, as a hold-out dataset and used the remaining 70% to train, validate, and iterate the predictive models. The hold-out dataset was used to assess the performance of the models on totally unseen data. As imbalanced class distribution can affect model performance, the Synthetic Minority Oversampling Technique (SMOTE) was applied during training to balance the dataset [30].

Feature selection   The purpose of this step is to find the smallest number of relevant and informative features. In the training set, the least absolute shrinkage and selection operator (LASSO) was repeatedly applied, each time with a different random data split, and the features that had been selected in more than 50% of the case were retained. Additional details on feature selection are reported in Additional file 1: Section S3.

Machine learning classification algorithms (classifiers)   Four conventional ML algorithms were implemented to classify recruiters, considering their robustness in binary prediction problems. Logistic regression (LR) is a widely used machine learning model in medicine for classification tasks, which assumes a linear relationship between the input variables and the outcomes. Support Vector Machine (SVM) is a maximum margin classifier that performs classification by finding a decision boundary, which generates the maximum separation between decision classes [31]. Random Forest [32] and XGBoost [33] are two ensemble techniques, respectively, a bagging and a boosting type of ensemble, characterized by high generalizability and robustness, which are effective at capturing interactions and non-linear relationships between variables, by aggregating sub-models that have no or low correlation with each other [34, 35]. During models' development, a fivefold cross-validation (stratified fivefold CV) routine was defined so that the data were partitioned into five folds of equal size: training occurred on four of the folds, and the remaining fold was used as validation set, to monitor the performance of the algorithm. Folds were created 10 times, each time with a different data split, to remove any bias in selecting training and validation subsets (repeated (stratified) fivefold CV). Models' hyperparameters were defined through a cross-validated grid search, as the combination of parameters that maximized models' performance: for each model, various combinations of parameters were tried and the one with the best cross-validation accuracy was selected. As recommended by Hsu et al. [36], a coarse grid was first used to identify the "best region" of each parameter, followed by a finer grid within this region. This parameter search has been done for each set of features, as the parameters' setting may vary with different set of features.

The median value of the area under the receiver operating characteristic curve computed from the validation folds (validation AUC) was chosen as the summarizing metric. The comparison between the percentages of patients having high or low lung recruitability, based on the different classification models, was performed with the Cochran's Q test. When applicable, significant differences between the percentages were tested using the McNemar test. These tests were implemented using the python library mlxtend. The model with the best validation AUC was retained and its performance was evaluated on the test set, which was kept isolated from the model development process, by calculating AUC, accuracy, sensitivity and specificity.

Results
A total of 221 patients were retrospectively analyzed [9]. According to the median value of radiologically defined and gas exchange-defined recruitability, recruiters and non-recruiters were 110 and 111, respectively. The main clinical characteristics of the population at 5 cmH$_2$O of PEEP divided in recruiters ($n = 110$) and not recruiters ($n = 111$) according to radiologically

Table 1 Baseline characteristics at 5 cmH$_2$O of PEEP in patients divided according to lung potential recruitment (LPR)

Variables	Recruiters (LPR > 15%)	Non-recruiters (LPR ≤ 15%)	*p*-value
Demographic			
Number	110	111	
Age (years)	62 (47, 74)	61 (48, 72)	0.317
Male sex (%)	44	56	0.739
Weight (kg)	75 (61, 85)	75 (61, 89)	0.718
BMI (kg/m2)	25 (22, 28)	25 (22, 29)	0.894
SAPS II	43 ± 17	42 ± 13	0.406
Origin of ARDS			**0.001**
Pulmonary (%)	73	51	
Extrapulmonary (%)	27	49	
ARDS severity			**< 0.001**
Mild (%)	61	27	
Moderate (%)	28	42	
Severe (%)	11	31	
Ventilatory parameters			
Tidal volume (TV, ml)	500 (425, 560)	522 (461, 600)	**0.013**
TV per ideal body weight (mL/kg)	7.9 (6.7, 8.5)	8.0 (7.3, 9.2)	**0.017**
Respiratory rate (breath per minute)	16 (14, 20)	16 (14, 18)	0.541
Minute ventilation (L/min)	8.3 (6.9, 9.9)	8.4 (7.6, 10.0)	0.252
Peak inspiratory pressure (cmH$_2$O)	28 (23, 34)	26 (23, 29)	**0.029**
Plateau pressure (cmH$_2$O)	19 (16, 22)	18 (15, 20)	**0.030**
Physiological dead space	0.66 ± 0.13	0.57 ± 0.12	**< 0.001**
Respiratory mechanics			
Driving pressure (cmH$_2$O)	13 (10, 16)	12 (10, 15)	0.112
Respiratory system elastance (cmH$_2$O/L)	27 (21, 35)	24 (19, 30)	**0.007**
Mechanical power (J/(min kg))	15 (11, 20)	15 (11, 18)	0.890
Respiratory partitioned mechanics			
Lung elastance (cmH$_2$O/L)	21(14, 28)	18 (13, 24)	**0.048**
Chest wall elastance (cmH$_2$O/L)	6 (4, 9)	6 (4, 8)	0.995
Gas exchange			
PaCO$_2$ (mmHg)	47 (41, 53)	42 (38, 49)	**< 0.001**
PaO$_2$ (mmHg)	68 (60, 75)	76 (65, 91)	**< 0.001**
PaO$_2$/FiO$_2$	113 (84, 144)	163 (119, 207)	**< 0.001**
CT parameters			
Total lung volume (mL)	2377 (2073, 2798)	2701 (2056, 3201)	0.022
Total lung weight (g)	1598 (1278, 1968)	1319 (1116, 1472)	**< 0.001**
Total lung gas volume (mL)	736 (504, 979)	1233 (912,1798)	**< 0.001**
Not aerated lung tissue (%)	52 ± 14	35 ± 14	**< 0.001**
Poorly aerated lung tissue (%)	31 (21, 39)	28 (22, 39)	0.652
Well aerated lung tissue (%)	14 (10, 22)	34 (26, 42)	**< 0.001**
Overaerated lung tissue (%)	0.00 (0.00, 0.04)	0.03 (0.00, 0.29)	**< 0.001**

Continuous data are expressed as mean ± SD or median (interquartile range), while categorical data are expressed as %. Student's t test or Mann–Whitney rank-sum tests and Chi-square test were used as appropriate

defined lung recruitability are reported in Table 1. The recruiter group presented a significantly higher percentage of pulmonary ARDS origin compared to the non-recruiter (73% *vs* 51%). The recruiter group was ventilated with a significantly lower tidal volume compared to non-recruiters (500 [425–560] vs 522 [461–600] mL) but with a similar minute ventilation. At 5 cmH$_2$O of PEEP, respiratory system and lung elastances were both significantly higher in recruiters (27 [6, 19–32] vs 24 [6, 17–27] cmH$_2$O/L and 21

Pennati *et al. Annals of Intensive Care*   (2023) 13:60

Page 6 of 11

Table 2 Subsets of the most informative variables selected according to the frequency with which they were chosen after repeating the least absolute shrinkage and selection operator (LASSO) algorithm

Feature group name	Data
Outcome: $\Delta_{45\text{-}5}$non-aerated tissue > 15%	
M5	ARDS origin, tidal volume, plateau pressure, driving pressure
M5 + M15	ARDS origin, tidal volume, plateau pressure, driving pressure, $\Delta_{15\text{-}5}$ mechanical power
M5 + M15 + RPM	ARDS origin, tidal volume, plateau pressure, driving pressure, $\Delta_{15\text{-}5}$ mechanical power, lung elastance
G5	ARDS origin, PaO_2/FiO_2
G5 + G15	ARDS origin, PaO_2/FiO_2, $\Delta_{15\text{-}5}$ PaO_2
CT5	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue
CT5 + G5	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue, PaO_2/FiO_2
CT5 + M5	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue
CT5 + G5 + G15	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue, $\Delta_{15\text{-}5}$ PaO_2
CT5 + M5 + M15 + RPM	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue, $\Delta_{15\text{-}5}$ mechanical power
CT5 + G5 + G15 + M5 + M15 + RPM	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue, $\Delta_{15\text{-}5}$ PaO_2, $\Delta_{15\text{-}5}$ mechanical power
Outcome: $\Delta_{15\text{-}5}$PaO2 > 24 mmHg	
M5	ARDS origin, BMI, mechanical power
M5 + M15	ARDS origin, BMI, mechanical power, $\Delta_{15\text{-}5}$ driving pressure
M5 + M15 + RPM	ARDS origin, BMI, mechanical power, $\Delta_{15\text{-}5}$ driving pressure, chest wall elastance
G5	ARDS origin, BMI, PaO_2
CT5	ARDS origin, total lung weight, poorly aerated lung tissue, well-aerated lung tissue
CT5 + G5	ARDS origin, total lung weight, poorly aerated lung tissue, well-aerated lung tissue, PaO_2
CT5 + M5	ARDS origin, total lung weight, poorly aerated lung tissue, well-aerated lung tissue, mechanical power
CT5 + M5 + M15 + RPM	ARDS origin, total lung weight, poorly aerated lung tissue, well-aerated lung tissue, $\Delta_{15\text{-}5}$ driving pressure, chest wall elastance
CT5 + G5 + M5 + M15 + RPM	ARDS origin, total lung weight, poorly aerated lung tissue, well-aerated lung tissue, $\Delta_{15\text{-}5}$ driving pressure, chest wall elastance, PaO_2
M5	ARDS origin, BMI, mechanical power
CT5 + G5 + G15 + M5 + M15 + RPM	Age, ARDS origin, well-aerated lung tissue, non-aerated lung tissue, $\Delta_{15\text{-}5}$ PaO_2, $\Delta_{15\text{-}5}$ mechanical power

Lung recruitability was defined both as the percent change in not aerated tissue between 5 cmH$_2$O and 45 cmH$_2$O (recruiters: $\Delta_{45\text{-}5}$non-aerated tissue > 15%) and as the change in PaO$_2$ between 5 cmH$_2$O and 15 cmH$_2$O (recruiters: $\Delta_{15\text{-}5}$PaO2 > 24 mmHg). Input parameters included lung mechanics at PEEP 5 cmH$_2$O (M5), lung mechanics at PEEP 15 cmH$_2$O (M15), respiratory partitioned mechanics (RPM), gas exchange measured at PEEP 5 cmH$_2$O (G5), gas exchange measured at PEEP 15 cmH$_2$O (G15), CT imaging acquired at PEEP 5 cmH$_2$O (CT5)

[12–26] vs 18 [11–22] cmH2O/L, respectively). Arterial oxygenation (PaO_2/FiO_2) was significantly lower in recruiters compared to non-recruiters (113 [84–144] vs 163 [119–207]). At 5 cmH$_2$O of PEEP, recruiters had a lower lung gas volume and higher lung weight compared to non-recruiters (736 [504–979] vs 1233 [912–1798] mL and 1598 [1278–1968] vs 1319 [1116–1472] g). Similarly, the percentage of not aerated and well-aerated tissue was, respectively, higher and lower in recruiters compared to non-recruiters (52 ± 14 vs 35 ± 14% and 14 [8, 10–20] vs 34 [6, 11, 24–38] %).

The PEEP test response of the population divided according to radiologically defined lung recruitability is shown in Additional file 1: Table S1 and described in Additional file 1: Section S2.

Development of classification models

From set of predictor variables (M5, M15, RPM, G5, G15, CT5), using LASSO, subsets of the most informative

variables, were produced and used in building ML models, and are summarized in Table 2. See also Additional file 1: Section S3.

Model performances

Radiologically defined lung recruitability: ML algorithm selection

Figure 2 shows the validation AUCs for each pair of set of variables and ML algorithm, when lung recruitability was radiologically defined (recruiters: $\Delta_{45\text{-}5}$non-aerated tissue > 15%). More details on the comparison between the classifiers are reported in the online Additional file 1: Section S4. Based on the results, logistic regression (LR) was chosen as ML classification algorithm (classifier) when lung recruitability was radiologically defined, as faster and more interpretable compared to the other algorithms. Additional metrics for the logistic regression algorithm are reported in Additional file 1: Table S2.
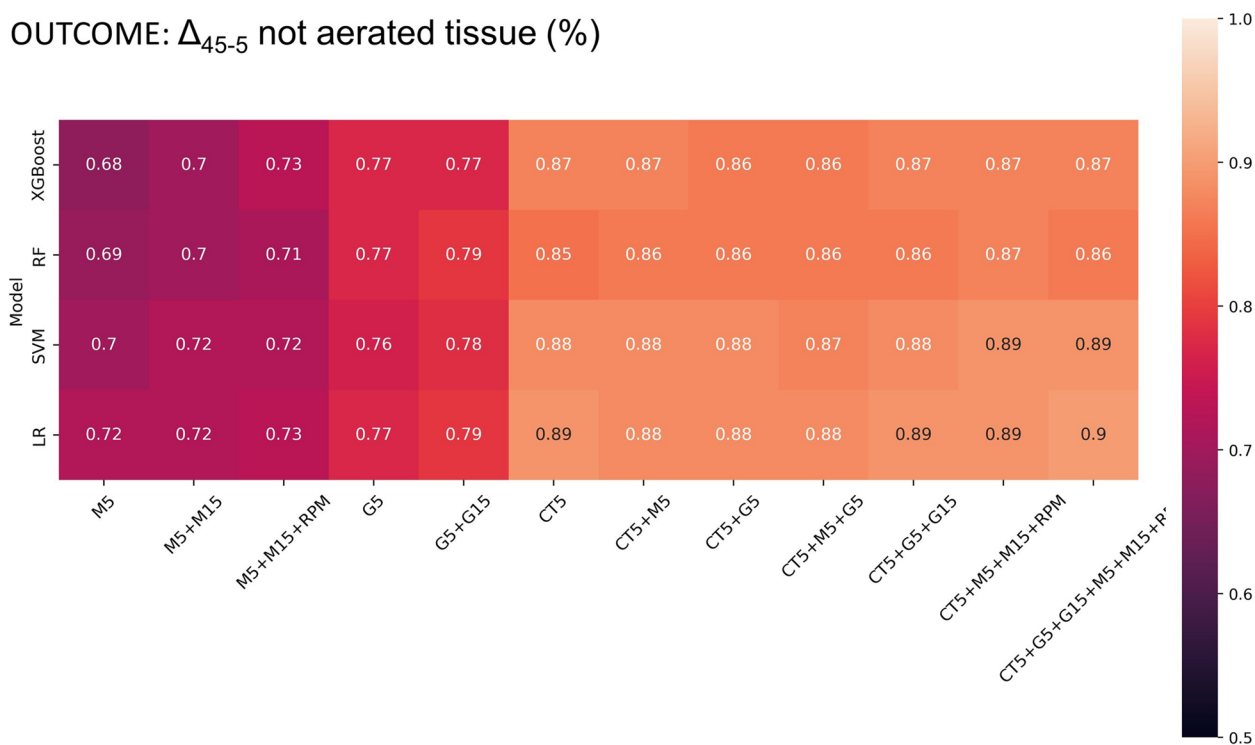
Fig. 2 Validation AUC for each pair of dataset and machine learning algorithm, when lung recruitability was radiologically defined (recruiters: Δ_{45-5}non-aerated tissue > 15%). M5, lung mechanics at PEEP 5 cmH$_2$O, M15, lung mechanics at PEEP 15 cmH$_2$O, RPM, respiratory partitioned mechanics, G5, gas exchange measured at PEEP 5 cmH$_2$O, G15, gas exchange measured at PEEP 15 cmH$_2$O, CT5, CT imaging acquired at PEEP 5 cmH$_2$O. XGBoost, gradient-boosted tree; RF, random forest; LR, logistic regression; SVM, support vector machine

Radiologically defined lung recruitability: logistic regression performance on different set of variables

Figure 2 shows that models based only on lung mechanics and gas exchange variables showed worse AUCs, with the highest values achieved by the models trained on gas exchange variables measured at PEEP 5 cmH$_2$O and 15 cmH$_2$O (0.79 for the LR model). The Cochran test reported no statistically significant difference among mechanical models ($p=0.920$), among gas exchange models ($p=0.317$) and between mechanical and gas exchange models ($p=0.893$). The model based on all features (CT5+G5+G15+M5+M15+RPM) showed the highest validation AUC (0.90), but no significant difference was obtained when only CT data model (CT5; AUC 0.89) was evaluated ($p=0.466$). Models including CT parameters reported statistically significant higher AUCs compared to models based only on lung mechanics (M5, M5+M15, M5+M15+RPM) and gas exchange (G5, G5+G15) ($p<0.001$) (Fig. 2).

Gas exchange-defined lung recruitability: ML algorithm selection

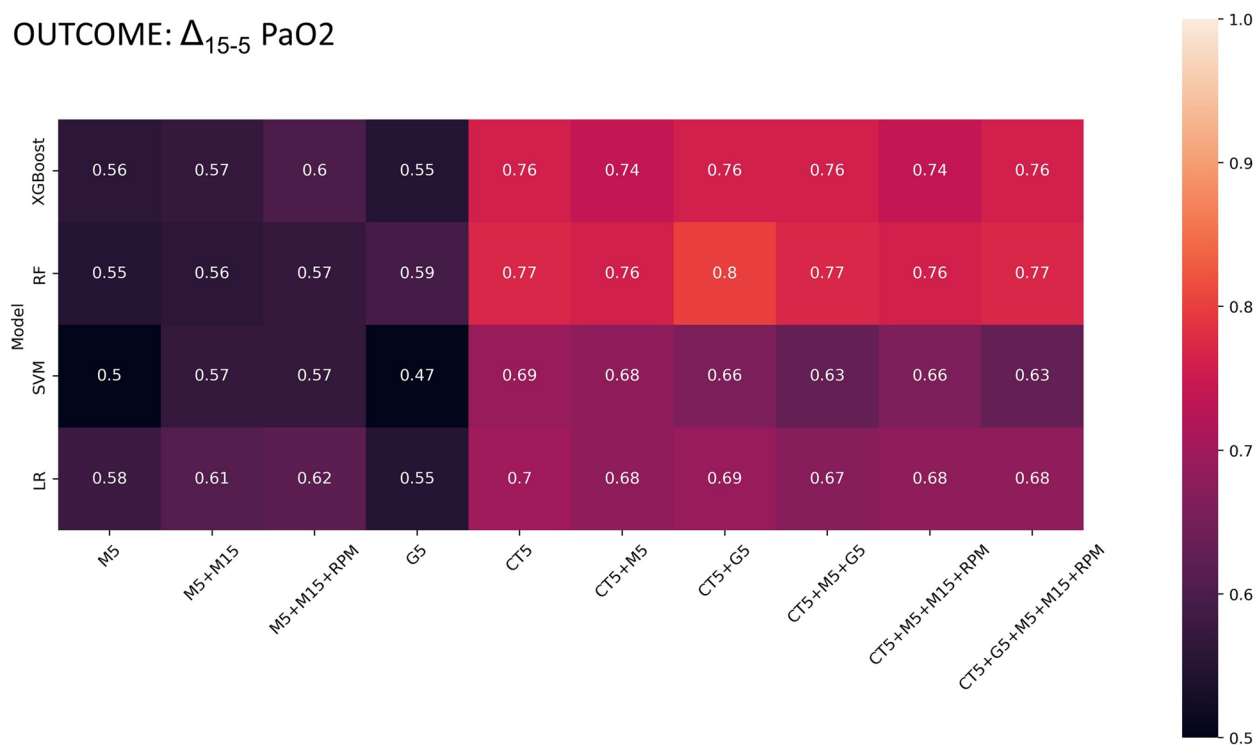When lung recruitability was defined based on gas exchange (recruiters: Δ_{15-5}PaO$_2>24$ mmHg) (Fig. 3), random forest resulted in the statistically highest validation AUCs on gas exchange and CT models and was chosen as classification algorithm. More details on the comparison between the classifiers is reported in the online Additional file 1: Section S4. Additional metrics for the random forest algorithm are reported in Additional file 1: Table S3.

Gas exchange-defined lung recruitability: random forest performance on different set of variables

Figure 3 shows that random forest algorithm based only on lung mechanics (M5, M5+M15, M5+M15+RPM) and gas exchange variables (G5) showed AUCs lower than 0.6. No statistically significant difference among mechanical models was found ($p=0.108$). Significant difference was reported among models including CT data ($p=0.005$), with better performance of CT5 (AUC 0.77) compared to CT5+M5 (AUC 0.76) (McNemar's test $p=0.032$). Models based on CT parameters reported statistically significant higher AUCs compared to models based only on lung mechanics ($p=0.009$) and gas exchange ($p=0.035$).

The models' performance over a range of thresholds of radiologically defined and gas exchange-defined lung

Pennati *et al. Annals of Intensive Care*   (2023) 13:60

Page 8 of 11

OUTCOME: Δ_{15-5} PaO2

Fig. 3 Validation AUC for each pair of dataset and machine learning algorithm, when lung recruitability was gas exchange-defined (recruiters: Δ_{15-5}PaO$_2$ > 24 mmHg). M5, lung mechanics at PEEP 5 cmH$_2$O, M15, lung mechanics at PEEP 15 cmH$_2$O, RPM, respiratory partitioned mechanics, G5, gas exchange measured at PEEP 5 cmH$_2$O, CT5, CT imaging acquired at PEEP 5 cmH$_2$O. XGBoost, gradient-boosted tree; RF, random forest; LR, logistic regression; SVM, support vector machine

Table 3 Models' performance over a range of thresholds defining lung recruitment

	Validation				Test			
	AUC	**acc**	**Sens**	**Spec**	**AUC**	**acc**	**Sens**	**Spec**
Outcome: Δ_{45-5}non-aerated tissue								
> 10% ($n = 142$)	0.85(0.07)	0.77(0.07)	0.78(0.10)	0.76(0.10)	0.90	0.81	0.80	0.82
> 20% ($n = 76$)	0.88(0.05)	0.81(0.06)	0.81(0.12)	0.81(0.08)	0.84	0.78	0.75	0.79
> 30% ($n = 32$)	0.89(0.07)	0.82(0.07)	0.81(0.18)	0.82(0.08)	0.85	0.87	0.37	0.93
Outcome: Δ_{15-5}PaO2								
> 20 mmHg ($n = 124$)	0.78(0.07)	0.71(0.07)	0.71(0.11)	0.73(0.12)	0.77	0.71	0.77	0.60
> 30 mmHg ($n = 88$)	0.75(0.11)	0.74(0.06)	0.73(0.13)	0.75(0.11)	0.76	0.69	0.41	0.87
> 40 mmHg ($n = 63$)	0.75(0.08)	0.70(0.08)	0.57(0.20)	0.75(0.08)	0.79	0.69	0.21	0.95

Lung recruitability was defined both as the percent change in not aerated tissue between 5 cmH$_2$O and 45 cmH$_2$O (recruiters: Δ_{45-5}non-aerated tissue > 15%) and as the change in PaO2 between 5 cmH$_2$O and 15 cmH$_2$O (recruiters: Δ_{15-5}PaO2 > 24 mmHg). Performance is evaluated in terms of mean area under the receiver operating characteristic curve (AUC), accuracy (acc), sensitivity (sens) and specificity (spec). Models were trained using the overall dataset (respectively, CT5 + G5 + G15 + M5 + M15 + RPM and CT5 + G5 + M5 + M15 + RPM) and least absolute shrinkage and selection operator (LASSO) algorithm for feature selection

recruitment is presented in Table 3 using the overall dataset (CT5 + G5 + G15 + M5 + M15 + RPM). For radiologically defined lung recruitment, similar performance was obtained from logistic regression when lung recruitability is defined using a cut-off 10% and 20%, whereas sensitivity worsened when 30% was used as threshold.

Discussion

In the present study, the lung recruitability was defined at first as the change in not aerated tissue between 5 cmH$_2$O and 45 cmH$_2$O to the total lung tissue weight at 5 cmH$_2$O, and secondly as Δ_{15-5} PaO$_2$. Our findings showed that the best approaches to predict lung recruitment were

the one that included lung CT scan taken at 5 cmH$_2$O of PEEP. The addition of respiratory mechanics and gas exchange did not significantly improve accuracy.

Concerning ventilatory management in ARDS, lung protective strategies include the application of lung recruitment and adequate PEEP levels to reopen collapsed lung regions by increasing the transpulmonary pressure [37]. The reopening of perfused collapsed/atelectatic lung regions should improve gas exchange, decrease alveolar shunt and ameliorate gas exchange by promoting CO$_2$ clearance. However, the increase in transpulmonary pressure, by increasing the end-expiratory lung volume, may also promote lung hyperinflation and higher lung stress at the interface between ventilated and not ventilated lung regions [38].

Among the different monitoring techniques (such as changes in respiratory mechanics, pressure–volume curves, lung ultrasound, electrical impedance tomography), quantitative lung CT analysis, although time-consuming and potentially harmful by exposing the patient to ionizing radiation, remains the most accurate method [6, 7, 11]. By using CT technique, lung recruitment potential is computed as the percentual difference of non-aerated lung tissue at two levels of pressure [8]. Previous studies showed that the application of a machine learning algorithm based on chest radiographs or CT at admission showed a good accuracy in detecting the presence of lung pathologies and ARDS, predicting clinical severity, the need of mechanical ventilation and outcome [15, 16, 24, 39–42]. Zampieri et al., according to a machine learning reanalysis of the ART clinical trial, showed that the application of a recruitment maneuver was associated with higher mortality in ARDS patients with pneumonia compared to sepsis [20]. Thus, the assessment of lung recruitability remains a challenge.

In the present study, we evaluated the possible use of machine learning to predict lung recruitment in ARDS patients, starting from clinical data and CT data obtained at 5 cmH$_2$O. The results showed that CT scan at 5 cmH$_2$O is the most accurate tool for evaluating lung recruitment and that adding data on lung mechanics and gas exchange does not increase accuracy. When only data on respiratory mechanics were used, low performances were achieved. Thus, while respiratory mechanics monitoring during the course of ARDS remains of paramount importance for VILI prevention and prognostication, our data may suggest that PEEP-induced changes in respiratory mechanics do not yield sufficient information about the potential for lung recruitment and that caution should be adopted when using PEEP-setting strategy based on compliance maximization [4, 43]. When only gas exchange data were used for training, low-to-moderate accuracy was found, suggesting that PEEP-induced

changes in PaO$_2$ and PaCO$_2$ are complex and may suffer from the interference of interdependent physiologic mechanisms [43]. Logistic regression was the preferred and stable machine-learning method for the differentiation of recruiters and non-recruiters when lung recruitability was radiologically defined, whereas random forest was the preferred method when lung recruitability was gas exchange-defined. In this study, parsimonious algorithms with higher potential for clinical use were investigated. According to the frequency with which each feature was selected, ARDS origin was the feature selected with the highest frequency in all the dataset. Respiratory partitioned mechanics features were rarely retained. When considering lung CT features measured at PEEP 5 cmH$_2$O, age was always selected to predict radiologically defined lung recruitability. We hypothesize that this may be related to the use of fixed thresholds to define lung aeration compartments, whereas age-related changes of lung volume and tissue density may occur [44].

The best lung recruitment potential cutoff to classify recruiters/non-recruiters remains unknown. In the present study we stratified recruiters and non-recruiters according to the median value of the lung recruitment potential of our whole population (15% and 24 mmHg), as previously suggested [8, 9]. To increase the translatability of the study, we investigated also models classifying recruiters using different thresholds. Comparable results were obtained when using thresholds near the median value of the whole population to define CT-based lung recruitability (10% and 20%, probably because the SMOTE algorithm reduces the impact of classes imbalance); when 30% was used, the worst sensitivity was achieved. On the contrary, using different cutoff to classify recruiters/non-recruiters based on gas exchange data (20 mmHg, 30 mmHg and 40 mmHg), led to worsening specificity and sensitivity, suggesting that the SMOTE algorithm is not sufficient to reduce the impact of classes imbalance and that other techniques should be investigated.

Our study has several strengths. First, it is the first study analyzing a large dataset of CT scans data which showed the possibility to use a machine learning algorithm to divide patients in recruiters and non-recruiters. Secondly, we investigated cross-combinations of three feature selection methods and four classification methods which have commonly been used and achieved high performance in previous studies. Thirdly, if applied these models could avoid the need of a second lung CT scan, decreasing the risk of radiation exposure of the patients and reducing the working time to complete the necessary computation.

Pennati *et al. Annals of Intensive Care*   (2023) 13:60

Page 10 of 11

However, in this study it was applied an *hybrid* approach with a manual lung segmentation of CT slides, which is time-consuming, with a quantitative analysis and machine learning algorithm. However, automatic lung CT segmentation is now available, which could significantly reduce the radiologic work [45]. The median imputation method was adopted for missing values, as simple and easy interpretable, but more advanced imputation methods [46, 47] can be explored to account for the existing relationship among features. Also, data were collected from a single center and a multicentre study with a large sample size is needed for further validation.

Conclusions

In conclusion, this study showed the possibility to use a machine learning algorithm based on a single CT imaging at the admission in intensive care unit to classify ARDS patients in responder and not responder to lung recruitment within the first 48 h from the start of mechanical ventilation. The application of this machine learning algorithm with an automatic lung segmentation and quantitative analysis could reduce the computational burden and the ionizing radiation load of the traditional method to assess lung recruitability, helping to improve the tailoring of ventilatory management according to the parenchymal and functional impairment of the ARDS in the acute phase of the disease.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13613-023-01154-5.

Additional file 1: Section S1. Study enrollment process additional details. **Figure S1**. Study protocol flowchart. **Section S2**. Clinical differences in radiologically defined recruiters and non-recruiters in terms of the PEEP test response. **Table S1.** Changes in respiratory mechanics and gas exchange at 5 and 15 cmH_2O of PEEP in patients divided according to lung potential recruitment (LPR). Continuous data are expressed as mean $\pm$ SD or median (interquartile range), while categorical data are expressed as %. Student t test or Mann–Whitney rank-sum tests were used as appropriate. **Section S3.** Additional material and methods. **Figure S2**. Demographic, mechanical, gas exchange and radiological variable considered in during feature selection according to type of variable (M = respiratory system mechanics, G = gas exchange, CT = radiological data, MPR = partitioned respiratory mechanics) and to the measuring condition (5 = 5 cmH_2O of PEEP, 15 = 15 cmH_2O of PEEP). **Section S4.** Model performances. **Figure S3.** Validation AUCs for each pair of dataset and ML algorithm. Lung recruitability is defined both from CT (left, recruiters: Δ_{45-5}non-aerated tissue > 15%) and from gas exchange data (right, recruiters: $\Delta_{15-5}PaO_2$ > 24 mmHg). The feature selection method applied is marked as _no (no feature selection was applied), corr (feature selection based on correlation was applied), _lasso (the least absolute shrinkage and selection operator was applied once). **Table S2.** Additional metrics on validation and test set of the logistic regression classifier, when lung recruitability was radiologically defined (recruiters: Δ_{45-5}non-aerated tissue > 15%). AUC, area under the receiver operating characteristic curve; acc, accuracy; sens, sensitivity; spec, specificity. M5: lung mechanics at PEEP 5 cmH_2O; M15: lung mechanics at PEEP 15 cmH_2O; RPM: respiratory partitioned mechanics; G5: gas exchange measured at PEEP 5 cmH_2O; G15: gas exchange measured at PEEP 15 cmH_2O; CT5: CT imaging acquired at PEEP 5 cmH_2O.

Table S3. Additional metrics on validation and test set, reported for the random forest classifier, when lung recruitability was gas exchange-defined (recruiters: $\Delta_{15-5}PaO_2$ > 24 mmHg).

Availability of data and materials
The dataset is available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study was approved by Comitato Etico Interaziendale Milano Area A (protocol number 2016/ST/143 on the 22ND June 2016, entitled "PEEP test") and informed consent was obtained according to the Italian regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interest.

Author details
[1]Ipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy. [2]Department of Health Sciences, University of Milan, Milan, Italy. [3]Department of Anesthesiology, University Medical Center Göttingen, Göttingen, Germany. [4]Department of Anesthesia and Intensive Care, ASST Santi Paolo e Carlo, San Paolo University Hospital, Via Di Rudini 9, Milan, Italy. [5]Coordinated Research Center on Respiratory Failure, University of Milan, Milan, Italy.

References

1. Gattinoni L, Marini JJ, Pesenti A, Quintel M, Mancebo J, Brochard L. The, "baby lung" became an adult. Intensive Care Med. 2016;42(5):663–73.
2. Fan E, Del Sorbo L, Goligher EC, Hodgson CL, Munshi L, Walkey AJ, et al. An official American Thoracic Society/European Society of intensive care medicine/society of critical care medicine clinical practice guideline: mechanical ventilation in adult patients with acute respiratory distress syndrome. Am J Respir Crit Care Med. 2017;195(9):1253–63.
3. Pensier J, de Jong A, Hajjej Z, Molinari N, Carr J, Belafia F, et al. Effect of lung recruitment maneuver on oxygenation, physiological parameters and mortality in acute respiratory distress syndrome patients: a systematic review and meta-analysis. Intensive Care Med. 2019;45(12):1691–702.
4. Cavalcanti AB, Suzumura ÉA, Laranjeira LN, De Moraes PD, Damiani LP, Guimarães HP, et al. Effect of lung recruitment and titrated Positive End-Expiratory Pressure (PEEP) vs low PEEP on mortality in patients with acute respiratory distress syndrome—a randomized clinical trial. JAMA. 2017;318(14):1335–45.
5. Dianti J, Tisminetzky M, Ferreyro BL, Englesakis M, Del Sorbo L, Sud S, et al. Association of positive end-expiratory pressure and lung recruitment selection strategies with mortality in acute respiratory distress syndrome:

a systematic review and network meta-analysis. Am J Respir Crit Care Med. 2022. https://doi.org/10.1164/rccm.202108-1972OC.

6. Chiumello D, Papa GFS, Artigas A, Bouhemad B, Grgic A, Heunks L, et al. ERS statement on chest imaging in acute respiratory failure. Eur Respir J. 2019;54:1900435.

7. Gattinoni L, Caironi P, Pelosi P, Goodman LR. What has computed tomography taught us about the acute respiratory distress syndrome? Am J Respir Crit Care Med. 2001;164(9):1701–11.

8. Gattinoni L, Caironi P, Cressoni M, Chiumello D, Ranieri VM, Quintel M, et al. Lung recruitment in patients with the acute respiratory distress syndrome. N Engl J Med. 2006;354(17):1775–86.

9. Coppola S, Froio S, Marino A, Brioni M, Cesana BM, Cressoni M, et al. Respiratory mechanics, lung recruitability, and gas exchange in pulmonary and extrapulmonary acute respiratory distress syndrome. Crit Care Med. 2019;47(6):792–9.

10. Pierrakos C, Smit MR, Hagens LA, Heijnen NFL, Hollmann MW, Schultz MJ, et al. Assessment of the effect of recruitment maneuver on lung aeration through imaging analysis in invasively ventilated patients: a systematic review. Front Physiol. 2021;12:666941.

11. Chiumello D, Marino A, Brioni M, Cigada I, Menga F, Colombo A, et al. Lung recruitment assessed by respiratory mechanics and computed tomography in patients with acute respiratory distress syndrome what is the relationship? Am J Respir Crit Care Med. 2016;193(11):1254–63.

12. Cressoni M, Chiumello D, Carlesso E, Chiurazzi C, Amini M, Brioni M, et al. Compressive forces and computed tomography-derived positive end-expiratory pressure in acute respiratory distress syndrome. Anesthesiology. 2014;121(3):572–81.

13. Mlodzinski E, Stone DJ, Celi LA. Machine learning for pulmonary and critical care medicine: a narrative review. Pulm Ther. 2020;6(1):67–77.

14. Wong AKI, Cheung PC, Kamaleswaran R, Martin GS, Holder AL. Machine learning methods to predict acute respiratory failure and acute respiratory distress syndrome. Front Big Data. 2020;3(November):1–18.

15. Maddali MV, Churpek M, Pham T, Rezoagli E, Zhuo H, Zhao W, et al. Validation and utility of ARDS subphenotypes identified by machine-learning models using clinical data: an observational, multicohort, retrospective analysis. Lancet Respir Med. 2022;10(4):367–77.

16. Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. Am J Respir Crit Care Med. 2020;202(7):996–1004.

17. Sayed M, Riaño D, Villar J. Novel criteria to classify ARDS severity using a machine learning approach. Crit Care. 2021;25(1):1–9.

18. Parreco J, Hidalgo A, Parks JJ, Kozol R, Rattan R. Using artificial intelligence to predict prolonged mechanical ventilation and tracheostomy placement. J Surg Res. 2018;228:179–87.

19. Mamandipoor B, Frutos-Vivar F, Peñuelas O, Rezar R, Raymondos K, Muriel A, et al. Machine learning predicts mortality based on analysis of ventilation parameters of critically ill patients: multi-centre validation. BMC Med Inform Decis Mak. 2021;21(1):1–12.

20. Zampieri FG, Costa EL, Iwashyna TJ, Carvalho CRR, Damiani LP, Taniguchi LU, et al. Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the alveolar recruitment for acute respiratory distress syndrome trial. Br J Anaesth. 2019;123(1):88–95.

21. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. Eur Respir J. 2020;56(2):2000775.

22. Wu C, Chen X, Cai Y, Xia J, 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;180(7):934–43.

23. Saba L, Agarwal M, Patrick A, Puvvula A, Gupta SK, Carriero A, et al. Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs. Int J Comput Assist Radiol Surg. 2021;16(3):423–34.

24. Chieregato M, Frangiamore F, Morassi M, Baresi C, Nici S, Bassetti C, et al. A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data. Sci Rep. 2022;12(1):1–15.

25. Wu Q, Wang S, Li L, Wu Q, Qian W, Hu Y, et al. Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19. Theranostics. 2020;10(16):7231.

26. Coppola S, Caccioppola A, Froio S, Formenti P, De Giorgis V, Galanti V, et al. Effect of mechanical power on intensive care mortality in ARDS patients. Crit Care. 2020;24(1):246.

27. Chiumello D, Consonni D, Coppola S, Froio S, Crimella F, Colombo A. The occlusion tests and end-expiratory esophageal pressure: measurements and comparison in controlled and assisted ventilation. Ann Intensive Care. 2016;6(1):13.

28. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, 2011:2825-2830.

29. McKinney W, others. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. 2010:51–6.

30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

31. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.

32. Breiman L. Random forest. Mach Learn. 2001;45(1):5–32.

33. Chen T, Guestrin C. XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2016. p. 785–94.

34. Zhou ZH. Ensemble methods. Foundation and algorithms. Boca Raton: Chapman and Hall/CRC; 2012.

35. Sagi O, Rokach L. Ensemble learning: A survey. WIREs Data Min Knowl Discov. 2018. https://doi.org/10.1002/widm.1249.

36. Hsu CW, Chang CC, Lin CJ. A Practical Guide to Support Vector Classification. Engineering. 2016;321–32.

37. Lachmann B. Open up the lung and keep the lung open. Intensive Care Med. 1992;18(6):319–21.

38. Chiumello D, Carlesso E, Cadringher P, Caironi P, Valenza F, Polli F, et al. Lung stress and strain during mechanical ventilation for acute respiratory distress syndrome. Am J Respir Crit Care Med. 2008;178(4):346–55.

39. Sjoding MW, Taylor D, Motyka J, Lee E, Co I, Claar D, et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. Lancet Digit Health. 2021;3(6):e340–8.

40. Röhrich S, Hofmanninger J, Negrin L, Langs G, Prosch H. Radiomics score predicts acute respiratory distress syndrome based on the initial CT scan after trauma. Eur Radiol. 2021;31(8):5443–53.

41. Reamaroon N, Sjoding MW, Lin K, Iwashyna TJ, Najarian K. Accounting for label uncertainty in machine learning for detection of acute respiratory distress syndrome. IEEE J Biomed Health Inform. 2019;23(1):407–15.

42. Reamaroon N, Sjoding MW, Gryak J, Athey BD, Najarian K, Derksen H. Automated detection of acute respiratory distress syndrome from chest X-rays using directionality Measure and deep learning features. Comput Biol Med. 2021;134:104463.

43. Grieco DL, De Pascale GAM. Lung recruitability and positive end-expiratory pressure setting in ARDS caused by COVID-19. Ann Oncol. 2020;161(4):869–71.

44. Dicente Cid Y, Mamonov A, Beers A, Thomas A, Kovalev V, Kalpathy-Cramer J, et al. Making sense of large data sets without annotations: analyzing age-related correlations from lung CT scans. Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications. 2017; 10138: 52–63

45. Klapsing P, Herrmann P, Quintel M, Moerer O. Automatic quantitative computed tomography segmentation and analysis of aerated lung volumes in acute respiratory distress syndrome—a comparative diagnostic study. J Crit Care. 2017;42:184–91.

46. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. BMC Med Inform Decis Mak. 2016;16(S3):74.

47. Kim JC, Chung K. Multi-modal stacked denoising autoencoder for handling missing data in healthcare big data. IEEE Access. 2020;8:104933–43.

Publisher's Note