Computer Methods and Programs in Biomedicine 260 (2025) 108574

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Machine learning-based forecast of Helmet-CPAP therapy failure in Acute Respiratory Distress Syndrome patients

Riccardo Campi [a],*, Antonio De Santis [a], Paolo Colombo [b], Paolo Scarpazza [b], Marco Masseroli [a]

[a] Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Piazza Leonardo Da Vinci 32, Milano, MI, 20133, Italy
[b] Azienda Socio Sanitaria Territoriale della Brianza, Via Santi Cosma e Damiano 10, Vimercate, MB, 20871, Italy

ARTICLE INFO

Keywords:
Acute Respiratory Distress Syndrome
Helmet-Continuous Positive Airway Pressure
Predicting H-CPAP failure in ARDS patients
Machine Learning
COVID-19

ABSTRACT

Background and Objective: Helmet-Continuous Positive Airway Pressure (H-CPAP) is a non-invasive respiratory support that is used for the treatment of Acute Respiratory Distress Syndrome (ARDS), a severe medical condition diagnosed when symptoms like profound hypoxemia, pulmonary opacities on radiography, or unexplained respiratory failure are present. It can be classified as mild, moderate or severe. H-CPAP therapy is recommended as the initial treatment approach for mild ARDS. Even though the efficacy of H-CPAP in managing patients with moderate-to-severe hypoxemia remains unclear, its use has increased for these cases in response to the emergence of the COVID-19 Pandemic. Using the electronic medical records (EMR) from the Pulmonology Department of Vimercate Hospital, in this study we develop and evaluate a Machine Learning (ML) system able to predict the failure of H-CPAP therapy on ARDS patients.

Methods: The Vimercate Hospital EMR provides demographic information, blood tests, and vital parameters of all hospitalizations of patients who are treated with H-CPAP and diagnosed with ARDS. This data is used to create a dataset of 622 records and 38 features, with 70%–30% split between training and test sets. Different ML models such as *SVM*, *XGBoost*, *Neural Network*, *Random Forest*, and *Logistic Regression* are iteratively trained in a cross-validation fashion. We also apply a feature selection algorithm to improve predictions quality and reduce the number of features.

Results and Conclusions: The *SVM* and *Neural Network* models proved to be the most effective, achieving final accuracies of 95.19% and 94.65%, respectively. In terms of F1-score, the models scored 88.61% and 87.18%, respectively. Additionally, the *SVM* and *XGBoost* models performed well with a reduced number of features (23 and 13, respectively). The *PaO2/FiO2 Ratio*, *C-Reactive Protein*, and *O2 Saturation* resulted as the most important features, followed by *Heartbeats*, *White Blood Cells*, and *D-Dimer*, in accordance with the clinical scientific literature.

1. Introduction

Acute Respiratory Distress Syndrome (ARDS) is a type of noncardiogenic (i.e., not caused by heart failure or fluid overload) pulmonary edema that causes breathing difficulties, fast breathing, and low oxygen levels. It can quickly lead to respiratory failure and occurs when an injury to the lungs or another part of the body triggers the release of inflammatory mediators. These mediators cause inflammation in the lung air sacs and blood vessels, leading to damage and fluid buildup. This results in decreased lung function and impaired gas exchange [1]. ARDS is a serious condition that accounts for 10% of Intensive Care Unit (ICU) admissions and 25% of mechanical ventilation procedures in the United States. It has a high mortality rate, with in-hospital death rates in the range of 46%–60% for severe cases [2,3]. According to a retrospective U.S. study [4], the estimated incidence of ARDS in 2014 was 193.4 cases per 100,000 people. The Berlin Criteria [1] are used for the diagnosis of ARDS in adults. Diagnosis is based on the timing of symptom onset, bilateral opacities on chest imaging, the likely source of pulmonary edema, and oxygenation as measured by the PaO2/FiO2 *Ratio*. ARDS can be classified as mild, moderate, or severe:

- Mild: 200 < PaO2/FiO2 *Ratio* ≤ 300 mmHg;
- Moderate: 100 < PaO2/FiO2 *Ratio* ≤ 200 mmHg;

* Corresponding author.
E-mail addresses: riccardo.campi@polimi.it (R. Campi), antonio.desantis@polimi.it (A. De Santis), paolo.colombo@asst-brianza.it (P. Colombo), paolo.scarpazza@asst-brianza.it (P. Scarpazza), marco.masseroli@polimi.it (M. Masseroli).

https://doi.org/10.1016/j.cmpb.2024.108574
Received 5 April 2024; Received in revised form 16 December 2024; Accepted 23 December 2024
Available online 30 December 2024

• Severe: PaO2/FiO2 *Ratio* ≤ 100 mmHg.

According to Grieco et al. [5], Non-Invasive Respiratory Support (NIRS) strategies, such as Helmet-Continuous Positive Airway Pressure (H-CPAP) therapy, are recommended as the first-line treatment for ARDS. Compared to standard oxygen therapy, these NIRS therapies have been shown to prevent endotracheal intubation in patients with mild hypoxemia [6–8]. Despite the uncertain effectiveness of NIRS strategies in treating patients with moderate-to-severe hypoxemia, these therapies have been increasingly employed for moderate and severe ARDS cases [9–11].

SARS-CoV-2 has caused a global pandemic of immense proportions, resulting in varying mortality rates across countries for reasons that remain unknown. The Italian national database records more than 4 million SARS-CoV-2-positive cases diagnosed between January 2020 and July 2021, including over 415 thousand patients hospitalized for coronavirus disease-19 (COVID-19) and more than 127 thousand deaths [12]. At Vimercate Hospital, Piluso et al. [13] demonstrated how H-CPAP therapy "*during the three COVID-19 waves*" had an average success rate of 69.33% and an average mortality rate of about 20.66%. Symptoms of the standard ARDS and the one caused by COVID-19 are slightly different [14]. However, our study did not differentiate between the two distinct forms of ARDS.

The advent of Machine Learning (ML), with its ability to construct mathematical models that classify items based on data analysis, has significantly advanced medical diagnostics [15]. This progress has undoubtedly been facilitated by the digital revolution of recent years, which has enabled the recording, archiving, and easy accessibility of various types of clinical data in a digital format. While to our knowledge there are no direct relevant studies on ML applied to H-CPAP and ARDS, Eguchi et al. [16] studied Obstructive Sleep Apnea (OSA) focusing on factors related to poor Mask-CPAP adherence using ML. They employed therapy logs from 219 patients, achieving an 86.4% F1-score. Another study by Mamandipoor et al. [17] predicted the success of Mechanical Ventilation (MV) therapy in ICU patients. Using the VENTILA dataset, they achieved predictive models, primarily using Recurrent Neural Networks (RNN), with an AUC of 72%. Le et al. [18] worked on predicting ARDS onset using XGBoost models. Their study, using the MIMIC-III database, achieved AUROC values of 90.5% for detecting ARDS at onset and of 82.7% for predicting ARDS up to 48 h before onset. Lastly, Singhal et al. [19] predicted early onset ARDS in COVID-19 patients. They used data from ICU patients, achieving AUROC of 89%, sensitivity of 77%, and specificity of 85%. Their eARDS system predicted ARDS development at least 12 h before the Berlin clinical criteria. They incorporated statistical variables along with numerical values of features, identifying important predictors such as O2 saturation and blood pressure.

Taking advantage of the digital data in the Electronic Medical Records (EMR) of the Pulmonology Department of the Vimercate Hospital, our study aims to demonstrate that a new data-driven approach to the treatment with H-CPAP therapy of ARDS patients is possible. This leads to the creation of a system capable of predicting the failure of H-CPAP therapy before it occurs, with positive consequences for the patient and the care provider entity. At the same time, it aims to understand what clinical and physiological conditions of the patient most significantly contribute to the failure of H-CPAP therapy. Preliminary results indicate that ML systems developed in this study can accurately detect therapy failures with over 93% accuracy. This leads to improved patient care and more efficient financial management for the hospital.

2. Materials and methods

This section explains the methodology employed in the development of our ML models. After clarifying the choice of the target variable for prediction, we provide a summary of the dataset used for cross-validation and testing of the models, along with an outline of the pre-processing steps adopted. Next, we describe the techniques for feature selection, the ML models considered, and the metrics applied.

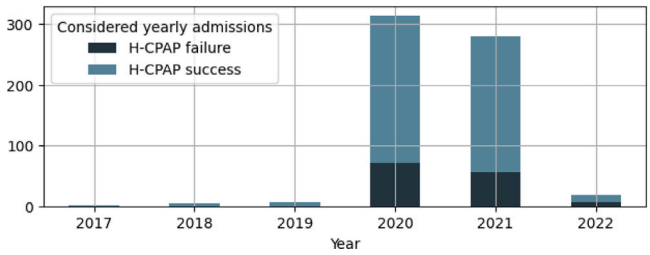


Fig. 1. Stratified distribution of the considered yearly admissions at the Vimercate Hospital of patients diagnosed with ARDS and treated with H-CPAP. It is reasonable to think that the increases in the years 2020 and 2021 were due to the SARS-CoV-2 disease.

2.1. Addressed medical question and target variable

Our study focuses on the outcomes of H-CPAP therapy in patients suffering from ARDS, regardless of whether it is caused by COVID-19 or other factors. The reasoning behind this decision is that COVID-19 complications often lead to ARDS-like conditions [14]. The choice of the therapy for which to predict the failure is driven by the ever-increasing interest of the clinical community in H-CPAP therapy, which had a crucial role during the COVID-19 Pandemic in 2020 and 2021 [9,20].

In light of this consideration, the medical question can be formulated as: "*given a patient affected by ARDS (regardless of whether it is caused by COVID-19 or not) and treated with H-CPAP, what is the therapy failure probability?*". The target variable *H-CPAP failure* is therefore defined as the likelihood of a patient seeing their therapy fail. More specifically, we set *H-CPAP failure* to *True* if the patient passes away during the course of the therapy or within 15 days following discharge, or if the patient requires intubation subsequent to H-CPAP [6].

2.2. Considered data

This study considers patients admitted at the Vimercate Hospital who were diagnosed with ARDS and treated with H-CPAP from January 2011 until about April 2022. The total number of such cases is 1000, and the data is collected from the hospital EMR system. Only admissions occurred after July 2017 include the Arterial Blood Gas (ABG) analyses, as they were digitized in that year. Out of these 1000 cases, about 23% experienced *H-CPAP failure*, while in about 77% of cases the therapy was successful.

Since ARDS is an acute pathology [1], multiple admissions of the same patient can occur and can be considered stochastically independent. In our dataset, only very few patients experienced double admissions. However, to avoid any possible correlated personal features, we retained only the last admission for each patient with multiple admissions (four out of eight).

In order to reduce missing values, we excluded from the evaluated dataset those cases with missing PaO2/FiO2 *Ratio* analyses (most of these were prior to 2017). This resulted in a final dataset of 622 cases, with 21.54% showing positive *H-CPAP failure* and 78.46% indicating negative *H-CPAP failure*. Fig. 1 shows the yearly admission stratified distribution of these cases.

The data is extracted from various hospital databases, including the *Centralized Hospital System* and the *Laboratory System*. Clinically relevant features are extracted from these structured databases for each patient admission. They include *Sex* and *Age* from the *Hospital Discharge Table*, PaO2/FiO2 *Ratio*, *Saturation in ABG*, *Creatinine*, and other blood tests from the *Laboratory Results Table*, the number of days in therapy (*CPAP days*) from the *Therapies Table*, and *Temperature*, *Saturation in Pulse Oximetry*, *Respiratory Rate*, and other parameters from the *Vital Signs Table*. Since there are many multiple values for each of these features, we retain for each patient only those closest in time to the outcome of the therapy.

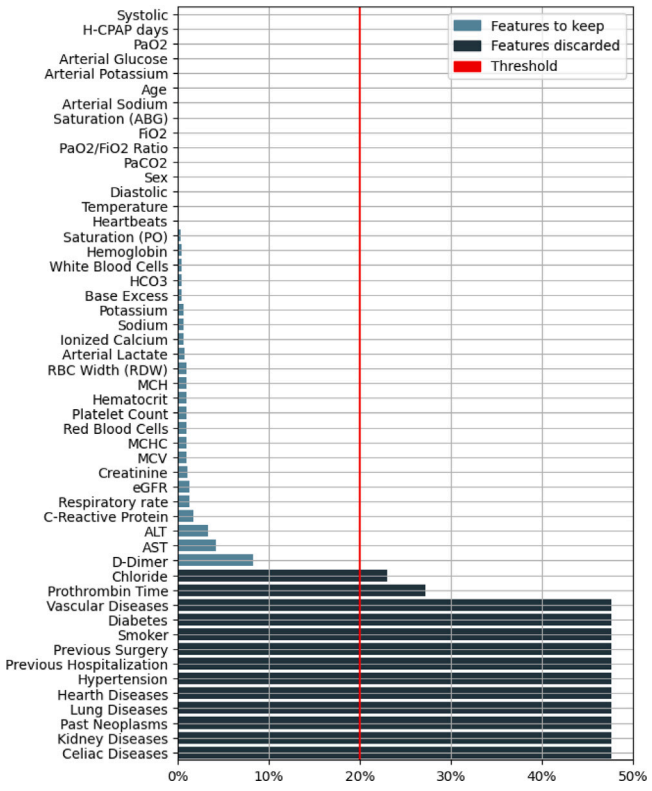


Fig. 2. Percentage on the training set of missing values for each feature. Features with percentage greater than 20% are discarded.

Dataset clinical homogeneity is ensured by the implementation of uniform guidelines on the use of H-CPAP throughout the study period. Since the publication of a foundational study on H-CPAP by Cosentini et al. [7] in 2010, clinical management has remained consistent.

2.3. Data pre-processing

We then split the obtained data in training and test sets. According to Dobbin and Simon [21] and Rácz et al. [22], a good percentage for the test set in the medical domain is at least 20% or higher, depending on the size of the dataset. In our case, we allocate 70% of our dataset to the training set and the remaining 30% to the test set, stratifying the split on the target variable.

Studies utilizing EMR typically need pre-processing steps to transform raw data into a refined dataset [23,24]. For this purpose, we use the deletion method for features with a percentage of missing values on the training set greater than 20% by discarding *Chloride*, *Prothrombin Time*, *Vascular Diseases*, *Diabetes*, and other features (Fig. 2). The imputation of the 38 remaining ones is performed with a multiple iterative imputer; we use the *IterativeImputer* implementation by *scikit-learn*, setting the parameters according to the official guidelines[1]: *estimator = BayesianRidge*, *initial strategy = mean*, and *max iter = 1000*. It is trained on the training set only and applied to both the training and test sets.

The percentile method is used for evaluating outliers. It starts by determining the 99th and 1st percentiles for each feature, which are referred to as the high and low limits, respectively. Any values outside these limits are clipped to these boundaries.

[1] https://scikit-learn.org/1.5/auto_examples/impute/.

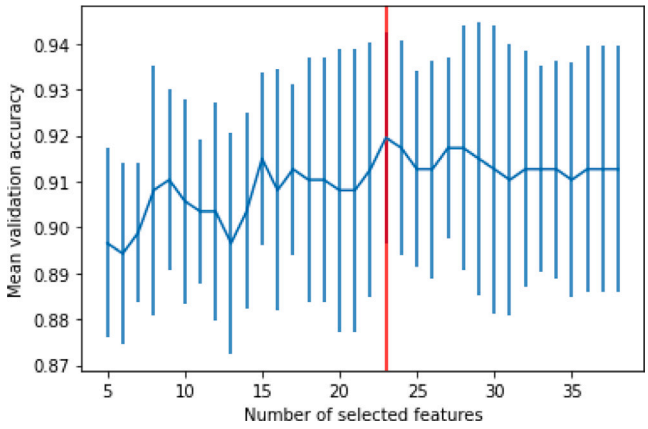


Fig. 3. Example execution of the *RFECV* method, illustrating that a further decrease below the identified threshold (red line) of the number of features used during model training corresponds to a decrease in the accuracy of the model results.

2.4. Feature selection

To reduce the likelihood of overfitting and increase the generalization capabilities of the trained models, we perform a selection of the features previously extracted and pre-processed. Performing feature reduction has also the advantage of making the ML models more usable by lowering the needed information. We use *Recursive Feature Elimination with Cross Validation* (*RFECV*) method, initially proposed by Guyon et al. [25], to reduce the number of features needed by the models. All ML models that have a built-in ranking function for feature importance can utilize the *RFECV* algorithm: it can be considered as a wrapper of the ML model being trained. It starts by training a model with the complete set of *n* features. Then, it calculates the importance of each feature and the mean and standard deviation of the cross-validation accuracies. The least important feature is removed, leaving *n-1* features. This process is repeated until only one feature remains, ending with *n* trained models in a cross-validation fashion. At each step, the model cross-validation performance metrics are recorded. Finally, the method automatically determines the optimal number of features by choosing the model that maximizes a performance metric considering both mean and standard deviation. Fig. 3 shows an example of model performance graph generated during the execution of the *RFECV* method on a ML model. To prevent biases, we run this algorithm on the training set only.

2.5. Machine learning algorithms and metrics

The prediction task at hand regards the development of a binary classifier, using a relatively small dataset and dealing with imbalanced data. Thus, we focus on the following supervised learning algorithms:

- *Logistic Regression* (*LR*), a straightforward binary classifier; we use a "*SAGA*" solver and "*L1*" penalty (i.e., incorporating a *LASSO* regression). It performs well when the dataset is linearly separable;
- *Random Forest* (*RF*) and *eXtreme Gradient Boosting* (*XGB*), both decision tree ensemble techniques that aim to improve the performance of a single decision tree;
- *Linear Support Vector Machines* (*SVM*), a non probabilistic classifier method that works best when there is a clear margin of separation between the classes. Due to the use of *RFECV*, our evaluation is limited to the *SVM* with *linear* kernel. The use of other kernels would have hindered the model's built-in feature importance system, thereby making the use of *RFECV* unfeasible.
- *Fully Connected Neural Network* (*FC-NN*) with one hidden layer and *Perceptron Classifier* (*PC*), both based on *Neural Networks* (*NN*) algorithms.

Table 1
Distribution of the target variable in training and test sets.

H-CPAP failure	Training set		Test set	
False	341	78.39%	147	78.61%
True	94	21.61%	40	21.39%
All	435	69.94%	187	30.06%

Table 2
Cross-validation performances of the ML algorithms using all 38 features.

Model	Accuracy	Precision	Recall	F1-score
RF	90.34%	83.33%	69.15%	75.58%
	±1.87%	±8.46%	±6.60%	±5.06%
XGB	90.80%	81.40%	74.47%	77.78%
	±3.08%	±4.78%	±15.11%	±9.75%
SVM	91.95%	83.91%	77.66%	80.66%
	±2.37%	±6.63%	±9.72%	±6.74%
LR	90.57%	79.78%	75.53%	77.60%
	±3.28%	±9.23%	±12.72%	±8.26%
FC-NN	92.18%	84.09%	78.72%	81.32%
	±1.34%	±3.16%	±6.53%	±3.67%
PC	89.43%	72.64%	81.91%	77.00%
	±2.85%	±9.97%	±12.65%	±5.63%

Table 3
Cross-validation performances of the ML algorithms trained with a number of features (Nf) reduced through the *RFECV* method.

Model	Nf	Accuracy	Precision	Recall	F1-score
RF	7	90.57%	81.18%	73.40%	77.09%
XGB	13	91.49%	82.02%	77.66%	79.78%
SVM	23	92.64%	86.90%	77.66%	82.02%
LR	9	89.89%	79.07%	72.34%	75.56%

Table 4
Testing performances of the ML algorithms using the number of features (Nf) selected in cross-validation.

Model	Nf	Accuracy	Precision	Recall	F1-score
RF	38	93.05%	88.57%	77.50%	82.67%
	7	91.98%	82.05%	80.00%	81.01%
XGB	38	93.05%	84.62%	85.50%	83.54%
	13	94.12%	89.19%	82.50%	85.71%
SVM	38	95.19%	89.74%	87.50%	88.61%
	23	93.58%	85.00%	85.00%	85.00%
LR	38	91.98%	83.78%	77.50%	80.52%
	9	89.84%	78.38%	72.50%	75.32%
FC-NN	38	94.65%	89.47%	85.00%	87.18%
PC	38	91.98%	93.10%	67.50%	78.26%

The various algorithms are evaluated using standard classification performance metrics, including overall accuracy, precision, recall (sensitivity), F1-score, and specificity. Confusion matrices are employed to examine classifications, and performances are also evaluated using Receiver Operating Characteristic (ROC) curves.

We employ cross-validation on the training set for hyper-parameter tuning, which aids in determining the optimal hyper-parameters of the ML algorithms, such as the number of trees in our ensembles.

3. Results

3.1. Data sets

Our processed dataset split results in 435 training set samples and 187 test set samples, all with the 38 features kept as detailed in Fig. 2. Being the split stratified based on the *H-CPAP failure* target variable, both resulting sets have approximately 78% negative and 22% positive *H-CPAP failure* samples (Table 1).

The training set is composed of 74.25% male and 25.75% female patients, with an average age of 65.12 years and a consecutive therapy duration of 8.64 days. The average PaO2/FiO2 *Ratio* is 234.59 mmHg, *Saturation in ABG* is 97.79% and *Pulse Oximetry* is 96.73%. The body *Temperature* averages at 36.3°, the *D-Dimer* at 1276.46 mg/L FEU, the *Creatinine* at 0.95 mg/dL, and the *C-Reactive Protein* at 44.72 pg/mL. Test set sample values are comparable with those in training set. More information on all features are provided in Tables A.5 and A.6.

3.2. Evaluation of approaches and algorithms

We train all the considered ML algorithms with the training set data using a 5-fold cross-validation approach and optimizing the F1-score. Table 2 reports the cross-validation results of the algorithms, providing an evaluation of their performance metrics using all 38 features.

Medium complexity ensemble models, such as *RF* and *XGB*, achieve good cross-validation results in terms of accuracy, but lack in recall, which affects the F1-score. This is also noticeable when looking at the high recall standard deviation. The same is also true for *LR* with *LASSO*. *PC* does not perform well in terms of accuracy, probably because it is too simple to learn from these features. *SVM* model, particularly suited for binary classification, performs very well. At the cost of being more complex and less interpretable, this is the most performant non-NN model for the case of 38 features. Lastly, the *FC-NN* model is the only one to exceed 92% accuracy, maintaining high both precision and recall, and a low standard deviation. Its architecture, with just one hidden layer, makes it very suitable for the dataset.

The results for a reduced number of features, obtained by training the models (except the *NN*-based ones, as they do not have built-in feature importances) with the *RFECV* method, are shown in Table 3. The outcomes achieved by these models align with those of models trained with all 38 features.

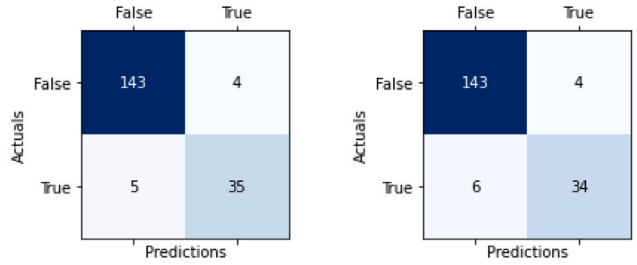
The cross-validation methodology selected the following hyper-parameters for our models using all 38 features. The *RF* had *max depth* = 4, *min samples leaf* = 2 (among 2, 4, 8, 10), *min samples split* = 4 (2, 4, 8, 10), and *n estimators* = 1000 (100, 250, 500, 1000). The *XGB* model had instead *learning rate* = 0.1 (0.1, 0.01, 0.001), *max depth* = 3 (3, 5, 7), *subsample* = 0.7 (0.7, 1.0), and *n estimators* = 100 (50, 100, 250, 500). The *SVM* model had *gamma* = 0.01 (0.1, 1.0, 10, 100, 1000) and *C* = 0.1 (1.0, 0.1, 0.01, 0.001, 0.0001). The *LR* model with *LASSO* had *C* = 10 (100, 10, 1.0, 0.1, 0.01), with a fixed *penalty* of *L1* and *solver* = *SAGA*. Lastly, for the *FC-NN* model the hyper-parameters were *solver* = *adam*, *activation* = *tanh*, *max iter* = 200 (50, 100, 150, 200), *hidden layer sizes* = (50,) ((25,), (50,), (100,)), and *alpha* = 0.01 (0.01, 0.005, 0.001).

3.3. Best performing models

Comparing Tables 2 and 3, it is evident that the *SVM* model performs better than the others in terms of accuracy. It still maintains good results also in terms of overall F1-score. However, since this type of model is more effective in high-dimensional spaces, it cannot significantly reduce the number of features used. The *XGB* model performs well, even when the number of features is reduced to almost one third, particularly in terms of accuracy, recall, and F1-score. The *RF* and *LR* with *LASSO* models tend to overly reduce the number of features used, which leads to a lower performance.

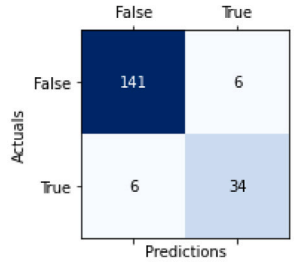
Table 4 reveals that all the trained models are quite good in classifying when using the test set.

In line with the cross-validation outcomes described in Table 2, the *SVM* and the *FC-NN* models exceed the others when considering all the 38 input features. In particular, in terms of accuracy the first achieves more than 95%, while the second scores more than 94%. Considering the amount of features required, *XGB* and *SVM* seem the most powerful models, which aligns with the cross-validation results in Table 3.

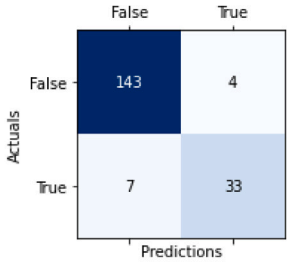


(a) *SVM*, 38 features.

(b) *FC-NN*, 38 features.

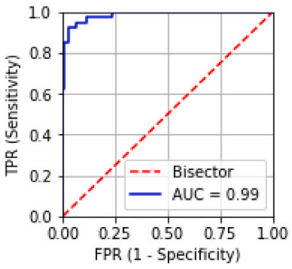


(c) *SVM*, 23 features.

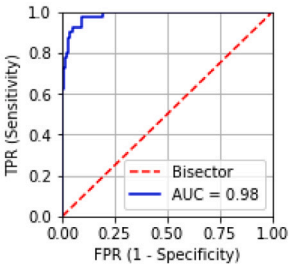


(d) *XGB*, 13 features.

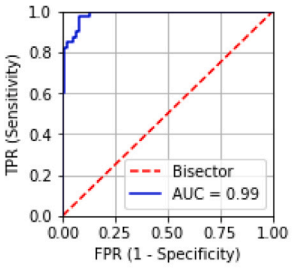
Fig. 4. Test set confusion matrices of the *SVMs* (left), *FC-NN* (top right) and *XGB* (bottom right) models. On the top, the matrices of the models trained with all 38 features, while on the bottom the ones of the models trained with 23 or 13 features, respectively.



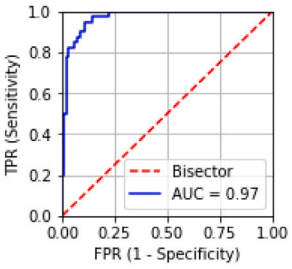
(a) *SVM*, 38 features.



(b) *FC-NN*, 38 features.



(c) *SVM*, 23 features.



(d) *XGB*, 13 features.

Fig. 5. Test set ROC curves of the *SVMs* (left) and *FC-NN* and *XGB* (right) models, including the Area Under the Curve (AUC).

Considering the aforementioned models and their recall performance, all fluctuate around 85%, making them suitable for the clinical use case under study. The same models also performs well in precision, which contributes to relatively high F1-scores, which fluctuate around 87%. Instead, *RF*, *LR* with *LASSO*, and *PC* learn worsley, as evidenced by both cross-validation and testing results.

Fig. 4 shows the test set confusion matrices of the *SVMs* with all 38 and with fewer features, and the ones of *FC-NN* and *XGB* models with all 38 and with fewer features, respectively. The specificity is quite high for the *SVM* and the *FC-NN* models with all 38 features, and for the *XGB* model with 13 features, scoring 97.28% for all three. Conversely, the specificity of the *SVM* model with 23 features achieves 95.92%. Fig. 5 shows the ROC curves.

3.4. Relevant factors

Fig. 6 presents the correlations in the training set between each feature and the target variable *H-CPAP failure*, as determined by the *Pearson's Correlation Coefficient*. This provides an initial understanding of the potential impact that each feature may have on the models.

The feature importances extracted from the *SVM* model that includes all 38 features are depicted in Fig. 7. Overall, the feature selection *RFECV* method is good in pinpointing the features that provide the most relevant informational value to the model, effectively discarding most of those of lesser importance. The *SHapley Additive exPlanations* (*SHAP*) [26] values of the *SVM* model (Fig. 8) provide additional verification for the feature importances.

Among the most important features, there are the *Saturation in PO* and *Saturation in ABG*, which Wick et al. [27] demonstrated correlated with ARDS severity, as well as the PaO2/FiO2 *Ratio*, which represents the relationship between arterial oxygen pressure (PaO2) and the percentage of inhaled oxygen (FiO2). *C-Reactive Protein* is an inflammation marker, which here proves to be significant in predicting the H-CPAP therapy failure. The levels of this protein have a positive correlation with the target variable, meaning that the higher the concentration of *C-Reactive Protein* in the blood, the higher the likelihood of H-CPAP

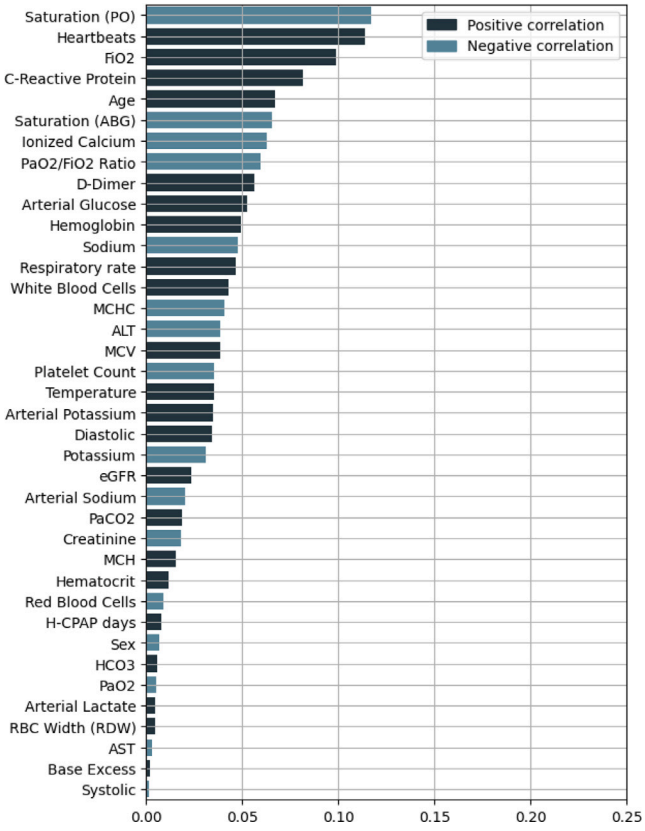


Fig. 6. *Pearson's Correlation Coefficient* values between the *H-CPAP failure* variable and each feature in the training set.

therapy failing. *White Blood Cells* (*WBCs*) count is another inflammation marker, whose increase has been associated with the presence of ARDS [28]. On the other hand, a sudden drop in its level is associated with a severe impairment of the immune system called leukopenia, with a very high risk of contracting infections and poor health [29].

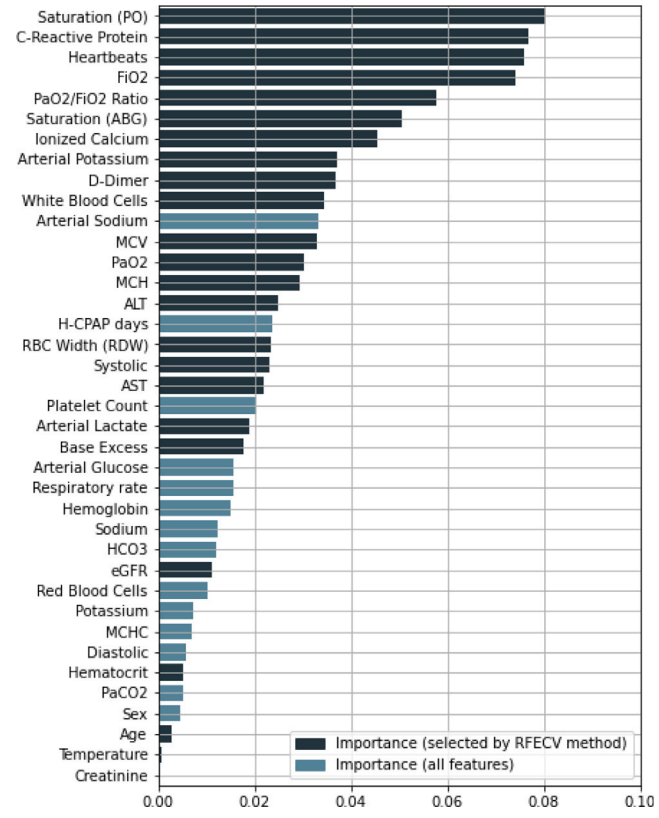


Fig. 7. Feature importance results for the *SVM* model trained using all 38 features.

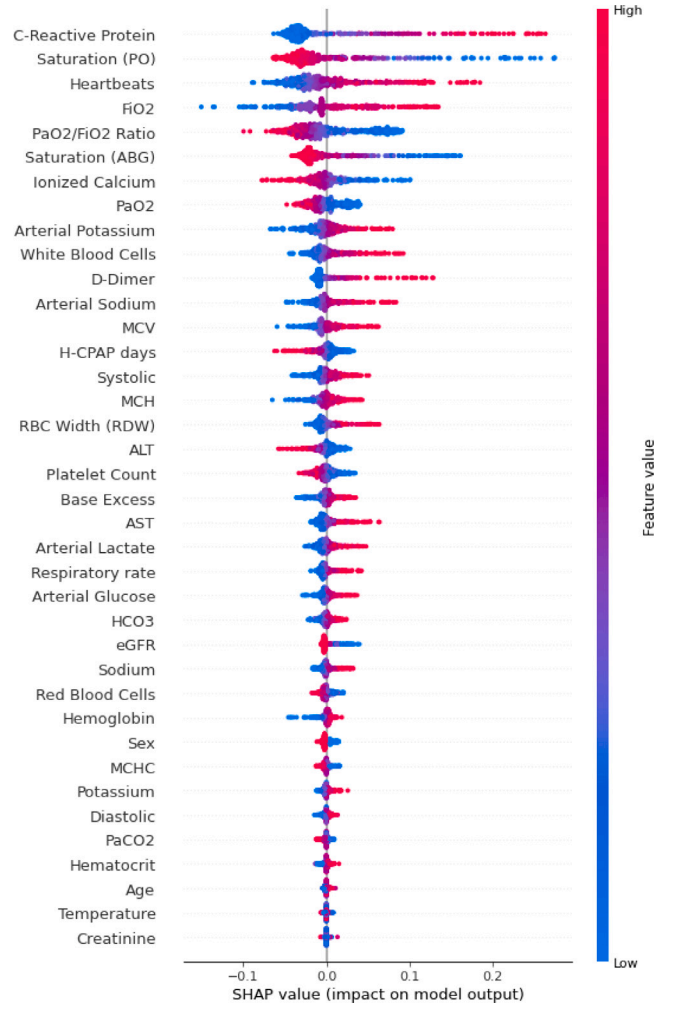


Fig. 8. *SHAP* explainability algorithm results for the 38 features used by the *SVM* model. The higher the *SHAP* value of a feature is, the more important the feature is in determining the model output.

D-Dimer is used to detect the presence of thrombosis or embolism, as well as to identify potential causes of other conditions such as pulmonary thromboembolism, which are all signs of occurred vessel clotting. Long et al. [30] demonstrated that *D-Dimer* can be employed as a marker of COVID-19 progression, since the SARS-CoV-2 disease is linked with the development of Disseminated Intravascular Coagulation. Considering that approximately 79.4% of the admissions in our dataset regarded COVID-19-affected patients, this explains the significance of the coagulation analysis in our predictions. *eGFR* and *Creatinine* act as markers of kidney function and filtration capacity, thereby providing insights into a patient's overall renal health. Darmon et al. [31] have established a connection between the risk of acute kidney diseases and the presence of ARDS. Regarding *Potassium* and *Sodium* levels, both arterial and otherwise, as well as *Ionized Calcium*, Jomova et al. [32] demonstrated that either a deficiency or an excess of these elements may result in a decline in health conditions, and therefore can be indicators of an increased failure probability. *Respiratory Rate* and *Heartbeats* are both important for the prediction of *H-CPAP failure*. This is explained by the fact that altered breathing is positively correlated with a severe ARDS, while constant tachycardia is a sign of poor health, particularly in cardiovascular and cardiorespiratory terms.

3.5. Implementation and availability

To make our study reproducible, our dataset, along with the Python code used for pre-processing the data and training the models, as well as the trained models, are provided at https://github.com/riccamper/ARDS_H-CPAP_ML.

4. Discussion

The developed method enables for the creation of a ML-based system capable of predicting the possible failure of H-CPAP therapy before it occurs. The effectiveness of this system is determined not only by its performance, but also by the number of features it requires to make predictions. For these reasons, the best model is the one based on *SVM*. With 38 features, it has an accuracy of 95.19%, an F1-score of 88.61%, and it correctly identifies 87.50% of actual failures (recall). Its low-features version identifies correctly 93.58% of the times, with a F1-score of 85.00% and 23 features. The *FC-NN* model also proved to be effective, achieving 94.65% of accuracy with all the features, as well as the *XGBoost*, with 94.12% of accuracy using only 13 features.

In our study, several factors resulted influencing the outcome of the H-CPAP therapy for ARDS patients, all supported by relevant literature and medical knowledge. Key indicators such as recent PaO2/FiO2 *Ratio*, O2 *Saturation* (*ABG* and *PO*), and *C-Reactive Protein* evaluations, along with other easily measurable parameters like *Heartbeats* and *Respiratory Rate*, are identified as major contributors to the outcomes.

Furthermore, the study shows that while a primary clinical parameter for monitoring ARDS patients such as *PaO2/FiO2 Ratio* is important, it is not sufficient for accurately predicting the failure of H-CPAP. In fact, this ratio does not provide comprehensive information about the patient's health status, particularly regarding inflammatory processes and other vital aspects such as coagulation, liver and kidney health.

The use of ML techniques to predict H-CPAP therapy failure is a data-driven approach that, when combined with medical expertise, enhances the accuracy in predicting the patient's clinical trajectory. This allows for strategic planning and the delivery of more personalized

patient care. The feature reduction with the *RFECV* method on the training set, besides limiting curse of dimensionality and overfitting (thus providing more robust predictive models), can allow clinicians to use ML algorithms more effectively; in fact, the handling of a lower number of features is more suitable for the proper clinical management of a patient.

Regarding the trade-off between sensitivity (recall) and specificity, also considering precision, we point out what follows. In our clinical use case, identifying all high-risk H-CPAP therapy patients with high sensitivity is useful, as it informs intensivists about which patients are likely to require intubation. However, it is important to avoid overwhelming intensivists with low-risk cases that are not classified as such, meaning both specificity and precision should also be high. Intubation will still be performed in the event of unexpected failure. By training the models optimizing the F1-score, we balanced precision and sensitivity. This causes the specificity to be high. Thus, with both specificity and sensitivity results being high, the study findings are well-aligned with the clinical guidelines.

We acknowledge the limitations of our study, particularly its reliance on data from a single public assistance organization. The specificity of this information makes it challenging to find comparable datasets from other organizations and, to our knowledge, at the time of writing in the literature there are no similar studies for comparison. Future research could apply the same methodology to other assistance organizations data to better understand its impact on service quality. Moreover, validation on an external dataset or data from multiple organizations could enhance the study generalizability and models robustness.

5. Conclusions

Predicting the potential failure of a therapy is a key aspect of designing a patient's health treatment plan. In particular, anticipating negative results from H-CPAP therapy for ARDS patients (whether linked to COVID-19 or not) can be complex for medical professionals. Our ML-based system can assist doctors in conventional diagnostic techniques. This support can help prevent the deterioration of patients' health and avoid needless expenses for the healthcare institution.

The system we propose allows predictions using 38 features with an accuracy of 95.19%, a balanced precision and recall ratio, and an F1-score of 88.61%. The same system achieves 93.58% accuracy and an F1-score of 85.00% with 23 features. This is achieved by training an *SVM* model via cross-validation, optimizing the F1-score, and employing *RFECV* to minimize the features. The training data comes from EMRs of Pulmonology Department admissions at Vimercate Hospital.

The most important features for our ML-based system are the PaO2/FiO2 *Ratio*, *C-Reactive Protein*, and O2 *Saturation*. These are followed in importance by *Heartbeats*, *White Blood Cells*, and *D-Dimer*, in line with clinical literature, as well as *Calcium*, *Potassium*, and *Sodium*.

The promising results achieved could potentially be improved by incorporating additional features, such as the dimension of the PEEP Valve, which is commonly connected to the expiratory port of the CPAP Helmet and maintains positive end-expiratory pressure [10].

Institutional review board statement

Human participants were involved in this research; the study was conducted in accordance with the Declaration of Helsinki. Our study was approved by the local institution, Vimercate Hospital, ASST-Brianza, according to the legal requirements concerning observational studies (Resolutions 0000573 27/07/2021 and 0000133 22/02/2023).

Informed consent statement

Due to the nature of the present observational study and data anonymization, the patients' consent to participate was not required, as declared by the ASST Brianza Ethics Committee.

Table A.5
Overview of the training set, 38 features and 435 records.

Feature	Mean (SD) / Percentages	Unit
Sex	74.25% / 25.75%	M/F
Age	65.12 (±12.00)	years
CPAP days	8.64 (±5.89)	days
Potassium	4.16 (±0.42)	mmol/L
Sodium	139.25 (±3.74)	mmol/L
Creatinine	0.95 (±0.46)	mg/dL
eGFR	56.43 (±9.48)	mL/min
MCHC	32.79 (±1.27)	g/dL
RBC	4.33 (±0.58)	cells/mcL
Hematocrit	39.04 (±4.55)	%
Platelet Count	282.28 (±106.70)	cells/mcL
MCH	29.77 (±2.03)	pg
WBCs	10.47 (±4.38)	cells/mcL
Hemoglobin	12.32 (±2.82)	g/dL
MCV	90.87 (±5.63)	fL
C-Reactive Protein	44.72 (±60.59)	pg/mL
RDW	13.50 (±1.51)	%
ALT	52.64 (±38.87)	U/L
AST	32.89 (±19.77)	U/L
D-Dimer	1276.46 (±1949.91)	mg/L FEU
PaO2	145.69 (±73.43)	mmHg
Arterial Glucose	152.73 (±70.09)	mg/dL
Arterial Potassium	3.98 (±0.43)	mmol/L
PaCO2	39.05 (±5.70)	mmHg
Arterial Sodium	134.76 (±3.91)	mmol/L
Saturation (ABG)	97.79 (±3.08)	%
Base Excess	4.44 (±3.66)	mmol/L
HCO3	28.44 (±4.10)	mmol/L
Ionized Calcium	1.14 (±0.06)	mg/dL
Arterial Lactate	1.61 (±0.65)	mmol/L
FiO2	67.43 (±15.71)	mmHg
PaO2/FiO2 Ratio	234.59 (±116.46)	mmHg
Resp. Rate	15.84 (±2.67)	acts/min
Heartbeats	76.41 (±15.01)	BPM
Saturation (PO)	96.73 (±4.05)	%
Temperature	36.26 (±0.44)	C°
Diastolic	73.09 (±11.72)	mmHg
Systolic	127.49 (±17.69)	mmHg
H-CPAP failure	78.39% / 21.61%	Neg/Pos

CRediT authorship contribution statement

Riccardo Campi: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Antonio De Santis:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Paolo Colombo:** Writing – review & editing, Resources, Data curation, Conceptualization. **Paolo Scarpazza:** Writing – review & editing, Supervision, Resources, Conceptualization. **Marco Masseroli:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We like to express our gratitude to the Information Systems team at Vimercate Hospital and to the Pulmonology Department team.

Appendix. Additional material

In the appendix, we provide an overview of the training and test sets used in this study. Refer to Tables A.5 and A.6 for more details.

Table A.6
Overview of the test set, 38 features and 187 records.

Feature	Mean (SD) / Percentages	Unit
Sex	75.40% / 24.60%	M/F
Age	65.13 (±11.33)	years
CPAP days	8.90 (±5.27)	days
Potassium	4.21 (±0.47)	mmol/L
Sodium	139.50 (±3.55)	mmol/L
Creatinine	0.98 (±0.53)	mg/dL
eGFR	56.02 (±10.43)	mL/min
MCHC	32.67 (±1.21)	g/dL
RBC	4.29 (±0.63)	cells/mcL
Hematocrit	38.75 (±4.69)	%
Platelet Count	287.70 (±110.90)	cells/mcL
MCH	29.66 (±1.91)	pg
WBCs	10.46 (±4.40)	cells/mcL
Hemoglobin	12.38 (±2.38)	g/dL
MCV	90.90 (±5.54)	fL
C-Reactive Protein	45.92 (±60.52)	pg/mL
RDW	13.51 (±1.51)	%
ALT	50.47 (±36.05)	U/L
AST	32.77 (±20.91)	U/L
D-Dimer	1333.30 (±1990.44)	mg/L FEU
PaO2	136.92 (±70.79)	mmHg
Arterial Glucose	154.91 (±65.15)	mg/dL
Arterial Potassium	4.02 (±0.48)	mmol/L
PaCO2	38.67 (±5.56)	mmHg
Arterial Sodium	134.76 (±3.86)	mmol/L
Saturation (ABG)	97.54 (±3.14)	%
Base Excess	4.16 (±3.38)	mmol/L
HCO3	28.09 (±3.80)	mmol/L
Ionized Calcium	1.14 (±0.06)	mg/dL
Arterial Lactate	1.68 (±0.71)	mmol/L
FiO2	67.43 (±16.73)	mmHg
PaO2/FiO2 Ratio	220.63 (±109.96)	mmHg
Resp. Rate	15.91 (±3.28)	acts/min
Heartbeats	73.82 (±14.42)	BPM
Saturation (PO)	96.70 (±4.21)	%
Temperature	36.26 (±0.44)	C°
Diastolic	71.80 (±11.02)	mmHg
Systolic	125.46 (±17.90)	mmHg
H-CPAP failure	78.61% / 21.39%	Neg/Pos

References

[1] A. Saguil, M.V. Fargo, Acute respiratory distress syndrome: Diagnosis and management, Am. Fam. Physician 101 (12) (2020) 730–738.

[2] G. Bellani, J.G. Laffey, T. Pham, E. Fan, L. Brochard, A. Esteban, L. Gattinoni, F. van Haren, A. Larsson, D.F. McAuley, M. Ranieri, G. Rubenfeld, B.T. Thompson, H. Wrigge, A.S. Slutsky, A. Pesenti, LUNG SAFE Investigators, ESICM Trials Group, Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries, JAMA 315 (8) (2016) 788–800, http://dx.doi.org/10.1001/jama.2016.0291.

[3] K.K.S. Parhar, K. Zjadewicz, A. Soo, A. Sutton, M. Zjadewicz, L. Doig, C. Lam, A. Ferland, D.J. Niven, K.M. Fiest, H.T. Stelfox, C.J. Doig, Epidemiology, mechanical power, and 3-year outcomes in acute respiratory distress syndrome patients using standardized screening. An observational cohort study, Ann. Am. Thorac. Soc. 16 (10) (2019) 1263–1272, http://dx.doi.org/10.1513/AnnalsATS.201812-910OC.

[4] E. Eworuke, J.M. Major, L.I. Gilbert McClain, National incidence rates for acute respiratory distress syndrome (ARDS) and ARDS cause-specific factors in the United States (2006–2014), J. Crit. Care 47 (2018) 192–197, http://dx.doi.org/10.1016/j.jcrc.2018.07.002.

[5] D.L. Grieco, S.M. Maggiore, O. Roca, E. Spinelli, B.K. Patel, A.W. Thille, C.S.V. Barbas, M.G. de Acilu, S.L. Cutuli, F. Bongiovanni, M. Amato, J.-P. Frat, T. Mauri, J.P. Kress, J. Mancebo, M. Antonelli, Non-invasive ventilatory support and high-flow nasal oxygen as first-line treatment of acute hypoxemic respiratory failure and ARDS, Intensive Care Med. 47 (8) (2021) 851–866, http://dx.doi.org/10.1007/s00134-021-06459-2.

[6] L. Ball, C. Robba, J. Herrmann, S.E. Gerard, Y. Xin, M. Pigati, A. Berardino, F. Iannuzzi, D. Battaglini, I. Brunetti, G. Minetti, S. Seitun, A. Vena, D.R. Giacobbe, M. Bassetti, P.R.M. Rocco, M. Cereda, L. Castellan, N. Patroniti, P. Pelosi, GECOVID Group, Early versus late intubation in COVID-19 patients failing helmet CPAP: A quantitative computed tomography study, Respir. Physiol. Neurobiol. 301 (103889) (2022) 103889, http://dx.doi.org/10.1016/j.resp.2022.103889.

[7] R. Cosentini, A.M. Brambilla, S. Aliberti, A. Bignamini, S. Nava, A. Maffei, R. Martinotti, P. Tarsia, V. Monzani, P. Pelosi, Helmet continuous positive airway pressure vs oxygen therapy to improve oxygenation in community-acquired pneumonia: A randomized, controlled trial, Chest 138 (1) (2010) 114–120, http://dx.doi.org/10.1378/chest.09-2290, URL https://www.sciencedirect.com/science/article/pii/S0012369210603558.

[8] A.M. Brambilla, S. Aliberti, E. Prina, F. Nicoli, M.D. Forno, S. Nava, G. Ferrari, F. Corradi, P. Pelosi, A. Bignamini, P. Tarsia, R. Cosentini, Helmet CPAP vs. oxygen therapy in severe hypoxemic respiratory failure due to pneumonia, Intensive Care Med. 40 (7) (2014) 942–949, http://dx.doi.org/10.1007/s00134-014-3325-5.

[9] C. Brusasco, F. Corradi, A. Di Domenico, F. Raggi, G. Timossi, G. Santori, V. Brusasco, Galliera CPAP-COVID-19 study group, collaborators of the Galliera CPAP-COVID-19 study group, Continuous positive airway pressure in COVID-19 patients with moderate-to-severe respiratory failure, Eur. Respir. J. 57 (2) (2021) 2002524, http://dx.doi.org/10.1183/13993003.02524-2020.

[10] A. Coppadoro, A. Benini, R. Fruscio, L. Verga, P. Mazzola, G. Bellelli, M. Carbone, G. Mulinacci, A. Soria, B. Noè, E. Beck, R. Di Sciacca, D. Ippolito, G. Citerio, M.G. Valsecchi, A. Biondi, A. Pesci, P. Bonfanti, D. Gaudesi, G. Bellani, G. Foti, Helmet CPAP to treat hypoxic pneumonia outside the ICU: an observational study during the COVID-19 outbreak, Crit. Care 25 (1) (2021) 80, http://dx.doi.org/10.1186/s13054-021-03502-y.

[11] A. Coppadoro, E. Zago, F. Pavan, G. Foti, G. Bellani, The use of head helmets to deliver noninvasive ventilatory support: a comprehensive review of technical aspects and clinical findings, Crit. Care 25 (1) (2021) 327, http://dx.doi.org/10.1186/s13054-021-03746-8.

[12] F. Minnai, G. De Bellis, T.A. Dragani, F. Colombo, COVID-19 mortality in Italy varies by patient age, sex and pandemic wave, Sci. Rep. 12 (1) (2022) 4604, http://dx.doi.org/10.1038/s41598-022-08573-7.

[13] M. Piluso, C. Ferrari, S. Pagani, P. Usai, S. Raschi, L. Parachini, E. Oggionni, C. Melacini, F. D'Arcangelo, R. Cattaneo, C. Bonacina, M. Bernareggi, S. Bencini, M. Nadalin, M. Borelli, R. Bellini, M.C. Salandini, P. Scarpazza, COVID-19 acute respiratory distress syndrome: Treatment with helmet CPAP in respiratory intermediate care unit by pulmonologists in the three Italian pandemic waves, Adv. Respir. Med. 91 (5) (2023) 383–396, http://dx.doi.org/10.3390/arm91050030.

[14] W. Bain, H. Yang, F.A. Shah, T. Suber, C. Drohan, N. Al-Yousif, R.S. DeSensi, N. Bensen, C. Schaefer, B.R. Rosborough, A. Somasundaram, C.J. Workman, C. Lampenfeld, A.R. Cillo, C. Cardello, F. Shan, T.C. Bruno, D.A.A. Vignali, P. Ray, A. Ray, Y. Zhang, J.S. Lee, B. Methé, B.J. McVerry, A. Morris, G.D. Kitsios, COVID-19 versus non-COVID-19 acute respiratory distress syndrome: Comparison of demographics, physiologic parameters, inflammatory biomarkers, and clinical outcomes, Ann. Am. Thorac. Soc. 18 (7) (2021) 1202–1210, http://dx.doi.org/10.1513/AnnalsATS.202008-1026OC.

[15] T. Panch, P. Szolovits, R. Atun, Artificial intelligence, machine learning and health systems, J. Glob. Health 8 (2) (2018) 020303, http://dx.doi.org/10.7189/jogh.08.020303.

[16] K. Eguchi, T. Yabuuchi, M. Nambu, H. Takeyama, S. Azuma, K. Chin, T. Kuroda, Investigation on factors related to poor CPAP adherence using machine learning: a pilot study, Sci. Rep. 12 (1) (2022) 19563, http://dx.doi.org/10.1038/s41598-022-21932-8.

[17] B. Mamandipoor, F. Frutos-Vivar, O. Peñuelas, R. Rezar, K. Raymondos, A. Muriel, B. Du, A.W. Thille, F. Ríos, M. González, L. del Sorbo, M. del Carmen Marín, B.V. Pinheiro, M.A. Soares, N. Nin, S.M. Maggiore, A. Bersten, M. Kelm, R.R. Bruno, P. Amin, N. Cakar, G.Y. Suh, F. Abroug, M. Jibaja, D. Matamis, A.A. Zeggwagh, Y. Sutherasan, A. Anzueto, B. Wernly, A. Esteban, C. Jung, V. Osmani, Machine learning predicts mortality based on analysis of ventilation parameters of critically ill patients: multi-centre validation, BMC Med. Inform. Decis. Mak. 21 (1) (2021) 152, http://dx.doi.org/10.1186/s12911-021-01506-w.

[18] S. Le, E. Pellegrini, A. Green-Saxena, C. Summers, J. Hoffman, J. Calvert, R. Das, Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS), J. Crit. Care 60 (2020) 96–102, http://dx.doi.org/10.1016/j.jcrc.2020.07.019.

[19] L. Singhal, Y. Garg, P. Yang, A. Tabaie, A.I. Wong, A. Mohammed, L. Chinthala, D. Kadaria, A. Sodhi, A.L. Holder, A. Esper, J.M. Blum, R.L. Davis, G.D. Clifford, G.S. Martin, R. Kamaleswaran, eARDS: A multi-center validation of an interpretable machine learning algorithm of early onset acute respiratory distress syndrome (ARDS) among critically ill adults with COVID-19, PLoS One 16 (9) (2021) e0257056, http://dx.doi.org/10.1371/journal.pone.0257056.

[20] A. Pagano, G. Porta, G. Bosso, E. Allegorico, C. Serra, F. Dello Vicario, V. Minerva, T. Russo, C. Altruda, P. Arbo, V. Mercurio, F.G. Numis, Non-invasive CPAP in mild and moderate ARDS secondary to SARS-CoV-2, Respir. Physiol. Neurobiol. 280 (103489) (2020) 103489, http://dx.doi.org/10.1016/j.resp.2020.103489.

[21] K.K. Dobbin, R.M. Simon, Optimally splitting cases for training and testing high dimensional classifiers, BMC Med. Genom. 4 (1) (2011) 31, http://dx.doi.org/10.1186/1755-8794-4-31.

[22] A. Rácz, D. Bajusz, K. Héberger, Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification, Molecules 26 (4) (2021) 1111, http://dx.doi.org/10.3390/molecules26041111.

[23] B. Malley, D. Ramazzotti, J.T.-Y. Wu, Data pre-processing, in: Secondary Analysis of Electronic Health Records, Springer International Publishing, Cham, 2016, pp. 115–141, http://dx.doi.org/10.1007/978-3-319-43742-2_12.

[24] C.M. Salgado, C. Azevedo, H. Proença, S.M. Vieira, Missing data, in: Secondary Analysis of Electronic Health Records, Springer International Publishing, Cham, 2016, pp. 143–162, http://dx.doi.org/10.1007/978-3-319-43742-2_13.

[25] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (1/3) (2002) 389–422, http://dx.doi.org/10.1023/A:1012487302797.

[26] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, 2017, http://dx.doi.org/10.48550/arXiv.1705.07874, arXiv.

[27] K.D. Wick, M.A. Matthay, L.B. Ware, Pulse oximetry for the diagnosis and management of acute respiratory distress syndrome, Lancet Respir. Med. 10 (11) (2022) 1086–1098, http://dx.doi.org/10.1016/S2213-2600(22)00058-3.

[28] J. Morales-Ortíz, M.T. Rondina, S.M. Brown, C. Grissom, A.V. Washington, High levels of soluble triggering receptor expressed on myeloid cells-like transcript (TLT)-1 are associated with acute respiratory distress syndrome, Clin. Appl. Thromb. Hemost. 24 (7) (2018) 1122–1127, http://dx.doi.org/10.1177/1076029618774149.

[29] V.W. Ing, The etiology and management of leukopenia, Can. Fam. Physician 30 (1984) 1835–1839.

[30] H. Long, L. Nie, X. Xiang, H. Li, X. Zhang, X. Fu, H. Ren, W. Liu, Q. Wang, Q. Wu, D-dimer and prothrombin time are the significant indicators of severe COVID-19 and poor prognosis, Biomed Res. Int. 2020 (2020) 6159720, http://dx.doi.org/10.1155/2020/6159720.

[31] M. Darmon, C. Clec'h, C. Adrie, L. Argaud, B. Allaouchiche, E. Azoulay, L. Bouadma, M. Garrouste-Orgeas, H. Haouache, C. Schwebel, D. Goldgran-Toledano, H. Khallel, A.-S. Dumenil, S. Jamali, B. Souweine, F. Zeni, Y. Cohen, J.-F. Timsit, Acute respiratory distress syndrome and risk of AKI among critically ill patients, Clin. J. Am. Soc. Nephrol. 9 (8) (2014) 1347–1353, http://dx.doi.org/10.2215/CJN.08300813.

[32] K. Jomova, M. Makova, S.Y. Alomar, S.H. Alwasel, E. Nepovimova, K. Kuca, C.J. Rhodes, M. Valko, Essential metals in health and disease, Chem. Biol. Interact. 367 (2022) 110173.