

Machine Learning Based Algorithms to Impute PaO₂ from SpO₂ Values and Development of an Online Calculator

Shuangxia Ren

University of Pittsburgh

Jill A. Zupetic

University of Pittsburgh

Mohammadreza Tabary

University of Pittsburgh

Rebecca DeSensi

University of Pittsburgh

Mehdi Nouraie

University of Pittsburgh

Xinghua Lu

University of Pittsburgh

Richard D. Boyce

University of Pittsburgh

Janet S. Lee (✉ leejs3@upmc.edu)

University of Pittsburgh

Research Article

Keywords: Machine learning, acute respiratory distress syndrome, respiratory failure, imputation

Posted Date: November 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1053360/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

We created an online calculator using machine learning algorithms to impute the partial pressure of oxygen (PaO₂)/fraction of delivered oxygen (FiO₂) ratio using the non-invasive peripheral saturation of oxygen (SpO₂) and compared the accuracy of the machine learning models we developed to previously published equations. We generated three machine learning algorithms (neural network, regression, and kernel-based methods) using 7 clinical variable features (N=9,900 ICU events) and subsequently 3 features (N=20,198 ICU events) as input into the models. Data from mechanically ventilated ICU patients were obtained from the publicly available Medical Information Mart for Intensive Care (MIMIC III) database and used for analysis. Compared to seven features, three features (SpO₂, FiO₂ and PEEP) were sufficient to impute PaO₂ from the SpO₂. Any of the tested machine learning models enabled imputation of PaO₂ from the SpO₂ with lower error and showed greater accuracy in predicting PaO₂/FiO₂ < 150 compared to the previously published log-linear and non-linear equations. Imputation using data from an independent validation cohort of ICU patients (N = 133) from 2 hospitals within the University of Pittsburgh Medical Center (UPMC) showed greater accuracy with the neural network and kernel-based machine learning models compared to the previously published non-linear equation.

Introduction

The ratio of the partial pressure of oxygen (PaO₂) to the fraction of oxygen (FiO₂) delivered, or the PaO₂/FiO₂, is the reference standard measurement for the assessment of low blood oxygen levels, or hypoxemia, in mechanically ventilated patients with respiratory failure. The PaO₂/FiO₂ ratio (PF ratio) has predictive value for mortality in patients with acute respiratory distress syndrome (ARDS)¹ and is also part of a severity index scoring system called the Sequential Organ Failure Assessment (SOFA) score that is used to predict mortality in patients with critical illness²⁻⁴. Additionally, the PF ratio is relevant to clinical decision-making including the decision to initiate prone positioning in ARDS patients with PF ratios ≤ 150⁵. Currently, measurement of the PF ratio requires invasive arterial blood gas (ABG) sampling and does not provide a continuous measure of the patient's oxygenation. Increasingly, non-invasive monitoring with pulse oximetry is utilized instead of ABGs^{6,7}, particularly in low-resource settings where ABG monitoring may not be readily available. In contrast to invasive blood gas sampling, the SpO₂ (peripheral saturation of oxygen)/FiO₂ ratio can be calculated without blood collection, arterial puncture, or blood gas analyzers and may serve as a surrogate for the PaO₂/FiO₂ ratio. Notably several studies have evaluated the SF ratio in children where non-invasive measurements are increasingly favored⁸⁻¹⁰.

A few studies have examined non-linear imputation of PaO₂/FiO₂ from SpO₂/FiO₂ measurements recorded at the same time^{11,12}. These studies have reported that the accuracy of non-linear imputation is superior to log-linear or linear imputation, especially for moderate to severe hypoxemic respiratory failure with ARDS where the PF ratio is <200^{11,13}. However, in patients with respiratory failure requiring mechanical ventilation, the optimal equation for imputation of PaO₂/FiO₂ from the SpO₂/FiO₂ remains unclear. An algorithm to accurately impute the PaO₂ from the SpO₂ in mechanically ventilated patients would be beneficial for predictive modeling and clinical research to facilitate recruitment of patients for clinical trials if an ABG is not available. Ideally, this approach would include only variables that contribute to the relationship between SpO₂ and PaO₂ but would not require the same invasive ABG measurement as the PaO₂. From the clinical perspective, SF ratio can be utilized as a surrogate for PF ratio to diagnose ARDS or ALI with less invasive nature and comparable reliability¹⁴.

The objective of this study is to develop a calculator utilizing machine learning algorithms to impute PaO₂ using non-invasive SpO₂ measurements from mechanically ventilated patients in the Medical Information Mart for Intensive Care (MIMIC) III database¹⁵ and to compare the accuracy of the machine learning models to the previously published non-linear and log-linear equations^{11,13}. In this study, three common machine learning approaches (neural network¹⁶, regression¹⁷, and kernel-based methods^{18,19}) were tested for regression and classification tasks using data available in MIMIC III²⁰ with 7 clinical variable features and a subsequent 3-feature model. We created models to perform a regression task to impute PaO₂

from SpO₂ values and a classification task to predict patients with moderate to severe hypoxemic respiratory failure based on a cut-off of a predicted PF ratio ≤ 150 ¹¹. Our overall hypothesis is that a machine learning algorithm would perform better in predicting the PaO₂ from SpO₂ across the entire span of SpO₂ values when compared to the previously published equations.

Methods

The MIMIC-III database v1.4 (<https://mimic.physionet.org>) is an openly available dataset developed by the Massachusetts Institute of Technology Lab for Computational Physiology¹⁵. It contains de-identified health data associated with approximately 40,000 intensive care unit admissions for patients admitted to critical care units in the Beth Israel Deaconess Medical Center between 2001 and 2012. MIMIC-III is a relational database that contains information on demographics, vital signs, mechanical ventilation status, laboratory tests, medications, and mortality. We also utilized a validation cohort obtained from an existing database of de-identified clinical information from intensive care unit patients with *Pseudomonas aeruginosa* respiratory isolates from 2 hospitals within the University of Pittsburgh Medical Center (UPMC). This dataset similarly contains information of demographics, mechanical ventilation status, ventilator parameters and laboratory tests. Our study utilizing the MIMIC-III database was determined as exempt by the University of Pittsburgh Institutional Review Board (STUDY19100068). The University of Pittsburgh Institutional Review Board approved the *Pseudomonas aeruginosa* ICU respiratory isolates database as a waiver of informed consent (STUDY21030010) and also approved the use of this database as an independent validation cohort (STUDY21090073).

Data processing

For the MIMIC-III database, we identified unique ICU encounters (icustay_id) with mechanical ventilation status. We next identified the lab event PaO₂ and chart event SpO₂ occurring at the same time of the mechanical ventilation status. In order to minimize error between matched PaO₂ and SpO₂, we constrained the time gap between the lab event PaO₂ and the chart event SpO₂ to be no more than 30 minutes. To minimize repeated sampling from the same subjects, we restricted the search of PaO₂ measurements to the first 24 hours of mechanical ventilation and obtained the first PaO₂ recorded within this timeframe. For chart events including tidal volume (TV), positive end-expiratory pressure (PEEP), FiO₂, temperature, and mean arterial pressure (MAP), we constrained the time gap to within 2 hours of the selected SpO₂ measurement. If a patient was treated with vasoactive infusions, it was recorded as a categorical variable. Data extraction and processing methods are available at <https://github.com/renshuangxia/PaO2PredictorDjango>²¹. The online calculator is available at <https://dikb.org/pa02-predictor>.

For the 3-feature model in the UPMC validation cohort, the database was queried for unique ICU patients requiring mechanical ventilation. The validation set cases include 133 discrete individuals with ABGs obtained within 30 minutes of an SpO₂ reading similar to the constraints defined in the MIMICS III derivation cohort.

Machine learning methods for regression task

For the regression task we implemented 3 different models – a neural network model, a linear regression model, and support vector regression (SVR), a type of kernel-based modeling. For each model, we applied a 10-fold cross-validation²².

For the neural network model, we tested different network structures and various numbers of features to arrive at two models used for comparison with the linear and support vector regression models. One model used seven input features and three hidden layers (16, 8, 5 neurons for layers 1 to 3). The other model used only three input features and two hidden layers (6, 3 neurons for layers 1 and 2). Both final models used a tangent activation function for all layers except the output layer which used a linear function in both models. Also, both models were trained for 200 epochs with Adam optimizer using gradient descent. The learning rate was 0.001 and the batch sizes were 50 for both models.

For the linear regression model, the output variable can be computed by a linear combination of the input variables. We trained the linear regression equation by the Ordinary Least Squares approach. We used the `linear_model.LinearRegression` method from scikit-learn 0.22 (<https://scikit-learn.org/stable/>) with default hyperparameters for predicting PaO₂ values.

For the SVR model, we tested multiple kernels including linear kernel, polynomial kernel, and radial basis function kernel (RBF). Based on the performance in the training data, the RBF kernel was selected.

Machine learning methods for classification task

We utilized PaO₂/FiO₂ ≤ 150, an accepted threshold previously utilized to capture patients with moderate to severe disease meeting the criteria for ARDS^{11,13}. We utilized this cut-off to test machine learning methods to predict this diagnostic threshold PaO₂/FiO₂ ≤ 150 for the different imputation techniques. We implemented 3 classification models including neural network, logistic regression, and a kernel-based model, SVM.

For each machine learning model, we applied a 10-fold cross-validation and calculated the sensitivity, specificity, likelihood ratios, diagnostic Odds Ratio (OR), Area Under Receiver Operating Characteristic curve (AUROC), F1 score and Bayesian Information Criterion (BIC) to compare across models. The two neural network models for classification were similar to the neural networks used in regression, except the output layer used the sigmoid function. As with the regression models, various topologies were tested to arrive at the final two multi-layer perceptron (MLP) classifiers, one with an input size of 7 features and the other with an input size of 3 features. The hidden layer size is (12, 8, 6, 4, 4) for the model with 7 input features. For the other model which utilizes only 3 input features, we used two hidden layers of size 6 and 3. All hidden layers used the tangent activation function. We trained both models for 200 iterations with Adam optimizer, setting 7 feature classifier momentum value as 0.8 and 3 feature classifier momentum value as 0.6. The learning rate was 0.001 and the batch size was 200 for both models.

In addition, we implemented a basic logistic regression model for classification purposes as well as the SVM model which classifies examples with an optimal hyperplane. For the logistic regression, it uses logistic function to model a binary dependent variable. We utilized the `linear_model.LogisticRegression` method provided in the scikit-learn library without regularization, and other arguments were set as default. For the SVM model, we compared the results by applying different kernels and the RBF kernel outperformed other kernels. Methods were similar to those used in the regression task.

Comparison of machine-learning based algorithm to published non-linear and log-linear equations

We compared the performance of our machine learning algorithms to the previously published equations. For the non-linear equation from Brown *et al*¹¹ the PaO₂ was imputed from the SpO₂, where $PO_2 = PaO_2$, $S = SpO_2$ and $F = FiO_2$ which is illustrated in the equation 1. For situations where the recorded SpO₂ was 100% (or, 1.0), the SpO₂ was substituted with 0.996 given that the equation would not permit the calculation of S=1.0.

Equation 1. Non-linear equation to impute PaO₂ from the SpO₂.

$$PO_2 = \left\{ \frac{11,700}{(1/S - 1)} + \left[50^3 + \left(\frac{11,700}{1/S - 1} \right)^2 \right]^{1/2} \right\}^{1/3}$$

$$+ \left\{ \frac{11,700}{(1/S - 1)} - \left[50^3 + \left(\frac{11,700}{1/S - 1} \right)^2 \right]^{1/2} \right\}^{1/3}$$

For the log-linear equation from Pandharipande, et al^{11,13}, the PaO₂/FiO₂ was imputed from SpO₂/FiO₂ utilizing the equation 2:

Equation 2. Log-linear equation to impute PaO₂ from the SpO₂.

$$PO_2 = F \cdot 10^{(0.48 + 0.78 \cdot \log_{10}(\frac{S}{F}))}$$

Results

A parsimonious three features model is sufficient to impute PaO₂/FiO₂ ratio using a large dataset

An overview of the machine learning tasks is outlined in Figure 1. We initially chose 7 relevant features from the chart events (SpO₂, FiO₂, TV, MAP, temperature, PEEP and vasopressor administration) representing recorded bedside measurements that were independent from an invasive arterial blood gas measurement. When applying the 7 features to impute the PaO₂, the final data set contained 9,900 unique ICU encounters from 9,302 mechanically ventilated patients (Supplementary Table e1). The relationship between SpO₂/FiO₂ (S/F) and the PaO₂/FiO₂ (P/F) was examined in dataset 1 containing 9,900 unique ICU events from the MIMIC-III database and was best described by a log-linear relationship between the transformed logarithmic value of the SF and PF ratios as previously described by Pandharipande, et al¹³ (Supplementary Figure e1). The relationship between S/F and P/F ratios showed high variance across the distribution of mechanically ventilated subjects (R² = 0.21).

For the regression task, we derived the RMSE (root-mean-square deviation) and BIC for each of the different 7 feature machine learning models (neural network, linear regression, support vector regression) to assess the performance of the imputation techniques. The RMSE and BIC of the three machine learning methods are shown in Supplementary Table e2. All the machine learning models outperformed the previously published non-linear and log-linear equations as shown by lower RMSE score; the same was observed for subset 1 (SpO₂ < 97%). For the classification task, the three machine learning methods achieved similar classification performance according to F1 scores, as shown in Supplementary Table e3; the same pattern was observed for subset 1 (SpO₂ < 97%).

To improve practicality of the method at the bedside, we attempted to use the smallest number of features possible to predict the PaO₂ or PaO₂/FiO₂ ratio from the regression and classification tasks, respectively. Compared to the other measured variables, PEEP had the strongest correlation with PaO₂/FiO₂ (r = -0.31) outside of the SF ratio (SpO₂/FiO₂) (Table 1). Using this information, we created a 3-feature model using SpO₂, FiO₂ and PEEP. As compared to seven features, three features were sufficient to impute PaO₂/FiO₂ ratio with a similar degree of accuracy. The 3-feature model was therefore utilized in the remainder of the analysis for the machine learning algorithms. The final 3-feature data set (dataset 2) contained 20,198 ICU encounters from 17,818 unique patients (Table 2). Forty percent of subjects were of female sex and the mean age was 64 years. The degree of hypoxemic respiratory failure, as measured by the PaO₂/FiO₂ ratio¹, showed a

distribution in which 26% had mild respiratory failure ($\text{PaO}_2/\text{FiO}_2 = 201\text{-}300$), 22% had moderate respiratory failure ($\text{PaO}_2/\text{FiO}_2 = 101\text{-}200$), and 8% had severe respiratory failure ($\text{PaO}_2/\text{FiO}_2 \leq 100$).

Machine learning models show improved performance when compared to the prior published equations for regression

We quantitatively derived the RMSE for all of the machine learning and previously published models and the BIC for each of the three machine learning models to assess the performance of the different imputation techniques (Table 3). The RMSE of the neural network, linear regression and SVR machine learning models were 84.7, 88.8 and 85.9, respectively, compared to 117.7 and 91.8 for the log-linear and non-linear equations. The lower RMSE values indicate that the 3 machine learning models outperformed the previously published equations. Of the machine learning models, the neural network method showed the lowest RMSE as well as the lowest BIC in both the whole dataset (dataset 2) and for $\text{SpO}_2 < 97\%$ (subset 2). A Bland-Altman Plot suggests that the neural network model is comparable to the published equations (Supplementary Figure e2). There was decreasing accuracy at higher $\text{PaO}_2/\text{FiO}_2$ ratios for all the methods examined.

Machine learning models show improved performance for the classification task

We compared the performance of the machine learning models with the log-linear and non-linear equations using F1 scores. Similar to the findings for the regression task, all three machine learning models performed better in the whole dataset than log-linear and non-linear equations (Table 4). When the dataset was limited to $\text{SpO}_2 < 97\%$ (subset 2), the machine-learning methods performed slightly better than log-linear and better than non-linear equations, respectively (Table 4). The F1 scores for all three machine learning methods were similar when using the whole dataset (dataset 2) and for subset 2 where $\text{SpO}_2 < 97\%$. As shown in Figure 2, when comparing the 3 machine learning models to one another, the neural network performed slightly better in the whole dataset (area under the precision recall curve = 0.94 for the neural network compared to 0.93 and 0.91 for the logistic regression and support vector machine model, respectively). The 3 models had similar performance in subset 2.

Machine learning algorithms show a better accuracy in the validation cohort

We developed an online calculator using the 3 machine learning algorithms requiring three inputs (SpO_2 , FiO_2 , and PEEP): <https://dikb.org/pa02-predictor>. The calculator was then utilized in an independent validation cohort of 133 mechanically ventilated ICU patients to impute the PaO_2 in a regression task. The imputed PaO_2 was compared to the actual PaO_2 obtained by ABG. The accuracy of the machine learning algorithms was compared to the non-linear equation and was reported as the RMSE and adjusted R-squared (Table 5). The neural network and SMV models had lower RMSE than the previously published non-linear equation, demonstrating improved performance in the imputation of PaO_2 . Adjusted R-squared was also higher in the neural network and SMV models. To clarify the models proposed in this study, the following example is worth mentioning: with the assumption of $\text{SpO}_2 = 100\%$, $\text{FiO}_2 = 0.6$, and $\text{PEEP} = 5 \text{ cmH}_2\text{O}$ (observed $\text{PaO}_2/\text{FiO}_2 = 190$), the predicted PaO_2 is estimated as 203.0, 186.2, 188.4 using neural network, SVM, and regression models, respectively, while the estimate of conventional non-linear model is 167 (Table 6).

Discussion

We used the publicly available MIMIC-III database as a derivation cohort to develop and evaluate machine-learning algorithms to impute PaO_2 utilizing non-invasive SpO_2 in patients who are mechanically ventilated. We tested three machine learning models (neural network, linear regression and SVR) first using seven available clinical variables SpO_2 , FiO_2 , PEEP, TV, MAP, temperature, and vasopressor administration to impute the PaO_2 . We subsequently used a parsimonious model with three clinical variables (SpO_2 , FiO_2 and PEEP) to non-invasively impute PaO_2 in both a derivation and validation cohort. The imputation of PaO_2 from the regression tasks enabled us to derive the $\text{PaO}_2/\text{FiO}_2$, a clinically meaningful ratio with predictive value^{1,23}. Additionally, we performed a classification task to predict $\text{PaO}_2/\text{FiO}_2 \leq 150$, a cut off that has been used

to capture those patients with moderate to severe respiratory failure in ARDS cohorts^{11,13} and to guide patient management⁵. To increase the clinical applicability of our work, we also developed an open-access online calculator to impute the PaO₂ using the 3-feature model requiring only non-invasive bedside parameters in mechanically ventilated patients. Our calculator showed improved accuracy in the imputation of the PaO₂ when compared to the previously published non-linear equation in both our initial cohort and the validation cohort.

To develop the machine learning algorithms, we initially evaluated clinical variables such as PEEP, TV, MAP, temperature, and vasopressor administration that are easily obtained at the bedside. We considered other clinical variables such as skin pigmentation, pulse oximeter location, oximeter manufacturer, vasopressor infusion, and laboratory variables such as serum bicarbonate, serum chloride, serum creatinine, serum sodium but these variables provided negligible improvement in the accuracy of imputation in a prior prospective study¹¹ and were therefore not included. However, it is worth mentioning that recent data have shown that SpO₂ may underestimate the degree of hypoxemia in patients with darker skin²⁴. TV, MAP, Temperature and vasopressor use demonstrated a stochastic distribution and did not significantly alter the accuracy of the machine-learning based algorithms and were therefore removed to create the 3 features model (SpO₂, FiO₂, PEEP). This 3-feature model provides a framework for generalizability using large datasets of mechanically ventilated patients.

Our study shows that a machine learning based method for both the regression and classification task, when applied to the MIMIC-III critical care database, improved the accuracy compared to the previously published non-linear and log-linear imputation methods. As is evidenced by comparing the F1 and discrimination measures in Table 4, the performance improvement was more modest for the classification task in subset 2 where SpO₂ < 97%. A possible explanation is that there were fewer ICU events (smaller N) per group in the subset.

Prior studies have examined the relationship between SF and PF ratios for patients with ARDS to determine whether the non-invasive SF ratio can be substituted for the invasively obtained PF ratio^{11,13,25}. Panharipande, et al studied matched measurements of SpO₂ and PaO₂ in a more heterogeneous population (i.e., patients undergoing general anesthesia and patients with ARDS) to determine the association between SF and PF ratios in order to calculate the respiratory parameter of the SOFA score¹³. In their study, matched SpO₂ and PaO₂ values were obtained from two groups of patients: Group 1 comprised of the derivation set and was obtained from patients undergoing general anesthesia from a single center and Group 2 comprised a validation set utilizing data from patients enrolled in a multi-center randomized clinical trial examining low versus high tidal volume for acute respiratory management of ARDS (ARMA)²⁶. All SpO₂ values > 97% were also excluded from analysis in order to maximize matched data to those values likely to be within the linear range of the oxyhemoglobin dissociation curve. Data from 4,728 matched SpO₂ and PaO₂ measurements showed that the relationship was best described by a log-linear equation with slight variation based upon the level of PEEP. In the setting of a more heterogeneous population, a poorer correlation was noted between SF and PF ratios. The regression equation of $\text{Log}(\text{PF}) = 0.48 + 0.78 \times \text{Log}(\text{SF})$ yielded an R-square of 0.31¹³.

Additionally, a retrospective analysis of arterial blood gas measurements from three ARDS Network studies compared the performance of non-linear, log-linear and linear imputation methods to derive PaO₂ from the SpO₂¹². In all patients (N=1,184), the nonlinear imputation was equivalent to log-linear imputation. However, in those patients with SpO₂ < 97% (N=707), the nonlinear imputation showed lower error than either linear or log-linear equations. A prospective study was subsequently conducted in patients enrolled in the Prevention and Early Treatment of Acute Lung Injury network¹¹ to assess the performance of the non-linear equation to impute PaO₂ from the SpO₂ and compare it to the prior log-linear and linear equations^{11,13,25}. This study included 1034 arterial blood gases from 703 patients, of which 650 arterial blood gases had matched SpO₂ < 97%. The non-linear equation showed lower error and better identified moderate to severe ARDS patients (defined in the study as PaO₂/FiO₂ ≤ 150) when compared to log-linear or linear imputation methods.

In our study, we similarly found a high degree of variance across SpO₂ values and corresponding measured PaO₂ values which was noted when we formally examined the relationship between SF and PF. This may be attributed to the retrospective nature of the data collection and the numerous variables that may confound the reliability of a recorded SpO₂ measured non-invasively to reflect the arterial SaO₂^{8,10,12}. Despite this limitation, the machine learning algorithms performed better on both regression and classification tasks when compared to the log-linear and non-linear published equations.

We used a validation cohort to show improved accuracy for the neural network and kernel-based machine learning algorithms when compared to the previously published non-linear equation. Another strength of our study is the development of an online calculator that can be used to impute the PaO₂ from three noninvasive parameters (SpO₂, FiO₂ and PEEP) and may serve as a tool for future studies in large electronic health record datasets. Additionally, our machine learning models allow for the evaluation of all mechanically ventilated patients with available data rather than narrowing the analysis to a specific population such as those with ARDS. Given the inclusion of all mechanically ventilated patients, a significant number of SpO₂ values were > 97% (N=8,510 for 7 features and N=16,918 for 3 features). While this reduced the accuracy of the imputed PF ratio, particularly above a certain threshold, the machine learning models were applied to the data without a pre-defined restriction placed upon the range of SpO₂ values and showed better performance than both the log-linear and non-linear equations on both the regression and classification tasks.

In summary, any of the tested machine learning models applied to MIMIC-III dataset enabled imputation of PaO₂ from the SpO₂ with lower error and provided greater accuracy in predicting PaO₂/FiO₂ ≤ 150 across the entire range of SpO₂ examined when compared to that of published equations in two independent cohorts. Additionally, our study provides a clinically relevant online calculator for the imputation of the PaO₂ from the 3-feature machine learning models. The calculator requires the input of SpO₂, FiO₂, and PEEP all of which are non-invasive and readily available at the bedside of mechanically ventilated patients.

Abbreviations

PaO₂: Partial pressure of oxygen

FiO₂: Fraction of inspired oxygen

SpO₂: Peripheral saturation of oxygen

PF ratio: PaO₂/FiO₂

SF ratio: SpO₂/FiO₂

ARDS: Acute respiratory distress syndrome

SOFA: Sequential Organ Failure Assessment

ABG: Arterial blood gas

TV: Tidal volume

PEEP: Positive end-expiratory pressure

MAP: Mean arterial pressure

SVR: Support vector regression

RBF: Radical basis function kernel

AUROC:Area under receiver operating characteristic curve

BIC:Bayesian information criterion

RMSE: Root-mean-square deviation

Declarations

Acknowledgements:

We thank Dr. William G. Bain for providing thoughtful edits.

Funding:

This work was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Numbers F32 HL152504 (J.Z.); P01 HL114453, R01 HL136143, R01 HL142084, K24 HL143285 (J.S.L.), and R01 LM012011 (X.L. and S.R.). The University of Pittsburgh holds a Physician-Scientist Institutional Award from the Burroughs Wellcome Fund (J.Z.); content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or any other sponsoring agency.

Author Contributions:

*Contribution:*S.R. performed the data extraction and processing, analysis, and interpreted the data. J.Z.and M.T. performed data analysis, interpreted the data and wrote the manuscript. R.B., W.B. and X.L. interpreted the data and revised the work for important intellectual content. R.D. performed the data extraction for the validation cohort. M.N. provided critical statistical expertise, designed, analyzed, interpreted the data, and wrote the manuscript. J.S.L. conceived, designed, analyzed, interpreted the data, and wrote the manuscript. S.R. and M.T. are the guarantors of the paper.

Competing interest statement:

J.S. Lee discloses a paid consultantship with Janssen Pharmaceuticals, Inc. unrelated to this study.The authors have no other relevant conflicts of interest to disclose.

References

- 1 Force, A. D. T. *et al.* Acute respiratory distress syndrome: the Berlin Definition. *JAMA***307**, 2526-2533, doi:10.1001/jama.2012.5669 (2012).
- 2 Vincent, J. L. *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med***22**, 707-710, doi:10.1007/bf01709751 (1996).
- 3 Ferreira, F. L., Bota, D. P., Bross, A., Melot, C. & Vincent, J. L. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA***286**, 1754-1758, doi:10.1001/jama.286.14.1754 (2001).
- 4 Vincent, J. L. *et al.* Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med***26**, 1793-1800, doi:10.1097/00003246-199811000-00016 (1998).
- 5 Guerin, C. *et al.* Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med***368**, 2159-2168, doi:10.1056/NEJMoa1214103 (2013).

- 6 Garland, A. & Connors, A. F., Jr. Indwelling arterial catheters in the intensive care unit: necessary and beneficial, or a harmful crutch? *Am J Respir Crit Care Med***182**, 133-134, doi:10.1164/rccm.201003-0410ED (2010).
- 7 Garland, A. Arterial lines in the ICU: a call for rigorous controlled trials. *Chest***146**, 1155-1158, doi:10.1378/chest.14-1212 (2014).
- 8 Khemani, R. G., Patel, N. R., Bart, R. D., 3rd & Newth, C. J. L. Comparison of the pulse oximetric saturation/fraction of inspired oxygen ratio and the PaO₂/fraction of inspired oxygen ratio in children. *Chest***135**, 662-668, doi:10.1378/chest.08-2239 (2009).
- 9 Khemani, R. G. *et al.* Comparison of SpO₂ to PaO₂ based markers of lung disease severity for children with acute lung injury. *Crit Care Med***40**, 1309-1316, doi:10.1097/CCM.0b013e31823bc61b (2012).
- 10 Lobete, C. *et al.* Correlation of oxygen saturation as measured by pulse oximetry/fraction of inspired oxygen ratio with Pao₂/fraction of inspired oxygen ratio in a heterogeneous sample of critically ill children. *J Crit Care***28**, 538 e531-537, doi:10.1016/j.jcrc.2012.12.006 (2013).
- 11 Brown, S. M. *et al.* Nonlinear Imputation of PaO₂/FIO₂ From SpO₂/FIO₂ Among Mechanically Ventilated Patients in the ICU: A Prospective, Observational Study. *Crit Care Med***45**, 1317-1324, doi:10.1097/CCM.0000000000002514 (2017).
- 12 Brown, S. M. *et al.* Nonlinear Imputation of Pao₂/Fio₂ From Spo₂/Fio₂ Among Patients With Acute Respiratory Distress Syndrome. *Chest***150**, 307-313, doi:10.1016/j.chest.2016.01.003 (2016).
- 13 Pandharipande, P. P. *et al.* Derivation and validation of Spo₂/Fio₂ ratio to impute for Pao₂/Fio₂ ratio in the respiratory component of the Sequential Organ Failure Assessment score. *Crit Care Med***37**, 1317-1321, doi:10.1097/CCM.0b013e31819cefa9 (2009).
- 14 Bilan, N., Dastranji, A. & Ghalehgalab Behbahani, A. Comparison of the spo₂/fio₂ ratio and the pao₂/fio₂ ratio in patients with acute lung injury or acute respiratory distress syndrome. *J Cardiovasc Thorac Res***7**, 28-31, doi:10.15171/jcvtr.2014.06 (2015).
- 15 Johnson, A. E. *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data***3**, 160035, doi:10.1038/sdata.2016.35 (2016).
- 16 Cheng, B. & Titterington, D. M. Neural Network: A Review from a Statistical Perspective. *Statist. Sci***9**, 2-30 (1994).
- 17 Friedman, J. H. & Popescu, B. E. Gradient Directed Regularization for Linear Regression and Classification. *Technical Report, Department of Statistics and Stanford Linear Accelerator Center, Stanford University* (2004).
- 18 Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J. & Vapnik, V. Support vector regression machines. *NIPS'96: Proceedings of the 9th International Conference on Neural Information Processing Systems*, 155-161 (1996).
- 19 Suykens, J. & Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters***9**, 293-300, doi:https://doi.org/10.1023/A:1018628609742 (1999).
- 20 Peng, C. Y., Lee, K. L. & Ingersoll, G. M. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research***Vol 96, 2002**, 3-14, doi:https://doi.org/10.1080/00220670209598786 (2010).
- 21 Code for the Imputation of PaO₂/FIO₂ from SpO₂ values from the MIMIC-III Critical Care Database Using Machine-Learning Based Algorithms (Github.com, 2020).

22 Krough, A. & Vedelsby, J. Neural network ensembles, cross validation and active learning. *NIPS'94: Proceedings of the 7th International Conference on Neural Information Processing Systems*, 231-238 (1994).

23 Bellani, G. *et al.* Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA***315**, 788-800, doi:10.1001/jama.2016.0291 (2016).

24 Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E. & Valley, T. S. Racial Bias in Pulse Oximetry Measurement. *New England Journal of Medicine***383**, 2477-2478, doi:10.1056/NEJMc2029240 (2020).

25 Rice, T. W. *et al.* Comparison of the SpO₂/FiO₂ ratio and the PaO₂/FiO₂ ratio in patients with acute lung injury or ARDS. *Chest***132**, 410-417, doi:10.1378/chest.07-0617 (2007).

26 Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med***342**, 1301-1308 (2000).

Tables

Table 1. Correlation coefficients between PF ratios and variables. Correlation coefficients between measured PF ratios and the 6 other measured variables (SpO₂/FiO₂ = SF ratio, PEEP, MAP, Temperature, Vasopressor Administration and TV) were performed. The variable with the strongest correlation coefficient (r) was chosen for the 3-feature model.

	SF ratio	PEEP	MAP	Temperature	Vasopressor Administration	TV
PF ratio	0.44	-0.31	0.06	-0.06	-0.04	0.02

Abbreviations: PF ratio: PaO₂/FiO₂, SF ratio: SpO₂/FiO₂, TV: Tidal volume, PEEP: Positive end-expiratory pressure, MAP: Mean arterial pressure

Table 2. Subject characteristics based on 3 features. The 3-feature models captured 20,198 ICU events from 17,818 unique patients. Variables included in the 3-feature machine learning models are SpO₂, FiO₂, and PEEP. *For subjects older than 89 years, the age was assigned as 90 years of age.

Total ICU events, N	20,198
Female sex, n (%)	8,084 (40.0)
Age in years, mean (\pm SD)*	64.0 (\pm 16.2)
PaO ₂ /FiO ₂ , mean (\pm SD)	310.4 (\pm 184.4)
Available mean PaO ₂ /FIO ₂ , N	20,198
PaO ₂ /FiO ₂ >300, n	8996
PaO ₂ /FiO ₂ = 201-300, n	5226
PaO ₂ /FiO ₂ = 101-200, n	4448
PaO ₂ /FiO ₂ < 100, n	1528
Available SpO ₂ measurements per unique patient, N	17,818
1 measurement, n	16,065
2 measurements, n	1,367
3 measurements, n	262
4 measurements, n	77
5 measurements, n	29
6 measurements, n	14
7 measurements, n	4

Abbreviations: ICU: intensive care unit,

Table 3. RMSE and BIC of the 3-feature machine learning models regression tasks compared to published methods. The RMSE and BIC for the 3-feature machine learning models were calculated for the entire dataset (20,198 ICU events) and a subset of the dataset with SpO₂< 97% (3,280 ICU events) and compared to the published log-linear and non-linear models.

	Entire Dataset 2 (20,198 events)		Subset 2(SpO ₂ < 97%) (3,280 events)	
	RMSE	BIC	RMSE	BIC
Neural Network	84.7	17952.7	67.5	2778.9
LinearRegression	88.8	18144.3	68.0	2783.5
Support Vector Regression	85.9	18013.6	70.3	2805.0
Log-linear	117.7	NA	72.2	NA
Non-linear	91.8	NA	81.2	NA

Abbreviations: BIC: Bayesian information criterion, RMSE: Root-mean-square deviation

RMSE: An estimate of the differences between values predicted by a model and the values observed. The lower RMSE is, the lower the difference that exists between the predicted and observed values.

BIC: A criterion used in Bayesian statistics to choose between models. The model with the lowest BIC is supposed to be the best.

Table 4. Prediction performance of machine learning classification models based on 3 features. Prediction performance statistics were calculated for the machine learning models based on 3 features and compared to the Log-linear and Non-linear methods for the entire dataset (20,198 ICU events; entire dataset 2) and for a subset of the events where SpO₂<97% (3,280 events; subset 2). Variables included in the 3-feature machine learning models are SpO₂, FiO₂, and PEEP.

	Entire Dataset 2 (20,198 events)					Subset 2(SpO ₂ < 97%) (3,280 events)				
	Neural Network	Logistic Regression	SVM	Log-linear	Non-linear	Neural Network	Logistic regression	SVM	Log-linear	Non-linear
Total, No.	20198	20198	20198	20198	20198	3280	3280	3280	3280	3280
Sensitivity	0.96	0.97	0.98	0.84	0.93	0.80	0.87	0.83	0.85	0.58
Specificity	0.39	0.26	0.33	0.56	0.49	0.76	0.59	0.69	0.59	0.89
Positive LR	1.59	1.32	1.46	1.90	1.83	3.37	2.13	2.75	2.09	5.16
Negative LR	0.09	0.10	0.07	0.29	0.15	0.27	0.23	0.25	0.25	0.47
Diagnostic OR	17.12	13.16	19.68	6.49	12.61	12.53	9.46	10.96	8.44	10.94
AUROC	0.83	0.81	0.74	NA	NA	0.85	0.83	0.84	NA	NA
F1	0.92	0.92	0.92	0.87	0.91	0.81	0.80	0.81	0.79	0.70
BIC	-4612.6	-4440.7	-4446.0	NA	NA	-591.8	-567.0	-580.0	NA	NA

Abbreviations:PaO₂:SVR: Support vector regression, AUROC:Area under receiver operating characteristic curve, BIC:Bayesian information criterion, LR, likelihood ratio, OR: odds ratio

Table 5. RMSE of the 3-feature machine learning models regression task compared to the published non-linear equation. The PaO₂ was imputed using an online calculator of the 3 machine learning models using SpO₂, PEEP, and FiO₂ from a validation cohort of 133 mechanically ventilated ICU patients. Subsequently, the RMSE and adjusted R²for the 3-feature machine learning models were calculated and compared to the published non-linear equation. A lower RMSE and higher adjusted R²indicate higher accuracy.

N = 133	Neural network	SVR	Regression	Non-linear
RMSE (adjusted R²)	65.0 (0.35)	64.9(0.35)	74.1(0.16)	67.1(0.31)

Abbreviations: SVR: Support vector regression, RMSE: Root-mean-square deviation

Table 6. Examples of comparing four models applied to four cases from different categories of PaO₂ (<150, 150-200, 200-300, >300).

PaO ₂	SpO ₂	FiO ₂	PEEP	Neural Network-imputed	Regression-imputed	SVR-imputed	Nonlinear- imputed
113	96	40	5	115.3	136.7	101.6	82
190	100	60	5	203.0	186.2	188.4	167
217	100	90	5	226.8	220.1	194.2	167
304	100	100	5	259.3	231.4	260.5	167

Abbreviations:PaO₂: Partial pressure of oxygen, FiO₂: Fraction of inspired oxygen, SpO₂: Peripheral saturation of oxygen, PEEP: Positive end-expiratory pressure, SVR: Support vector regression

Figures

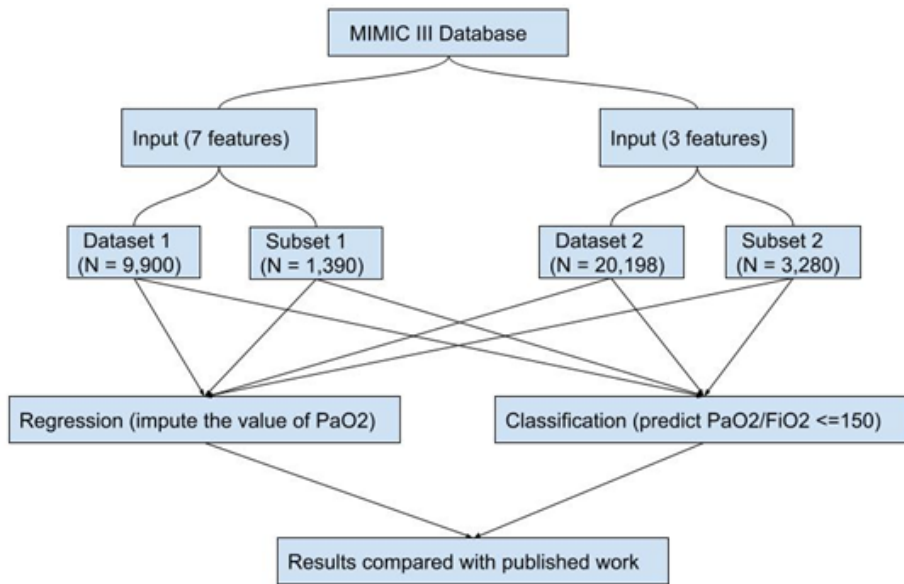


Figure 1

Overview of the experimental study design.

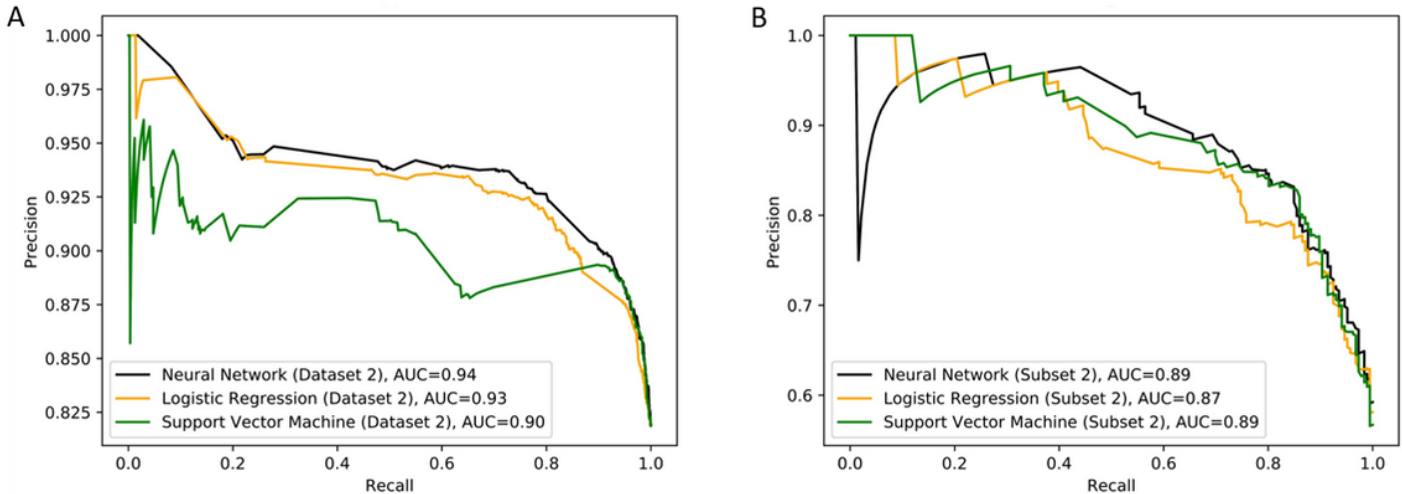


Figure 2

Precision-recall curves of machine learning models in Dataset 2 and Subset 2 using 3 features. The precision recall curves, where improved performance is demonstrated if the curve is closer to the upper right-hand corner or has the highest area under the curve (AUC), are shown for the 3 machine learning models for A) the entire Dataset 2 (N = 20,198) ICU events) and B) Subset 2 where SpO₂<97% (N = 3,280 ICU events).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalMaterials.docx](#)