



# Use of machine learning to identify protective factors for death from COVID-19 in the ICU: a retrospective study

Lander Dos Santos<sup>1</sup>, Lincoln Luis Silva<sup>2</sup>, Fernando Castillo Pelloso<sup>3</sup>, Vinicius Maia<sup>4</sup>, Constanza Pujals<sup>1</sup>, Deise Helena Borghesan<sup>5</sup>, Maria Dalva Carvalho<sup>1</sup>, Raíssa Bocchi Pedroso<sup>1</sup> and Sandra Marisa Pelloso<sup>1</sup>

<sup>1</sup> State University of Maringá, Graduate Program in Health Sciences, Maringá, Paraná, Brazil

<sup>2</sup> Department of Emergency Medicine, Duke University School of Medicine, Durham, NC, United States of America

<sup>3</sup> Department of Medicine, Federal University of Paraná, Curitiba, Paraná, Brazil

<sup>4</sup> Unicesumar, Maringá, Paraná, Brazil

<sup>5</sup> Union of Catholic Colleges of Mato Grosso, Cuiabá, Mato Grosso, Brazil

## ABSTRACT

**Background.** Patients in serious condition due to COVID-19 often require special care in intensive care units (ICUs). This disease has affected over 758 million people and resulted in 6.8 million deaths worldwide. Additionally, the progression of the disease may vary from individual to individual, that is, it is essential to identify the clinical parameters that indicate a good prognosis for the patient. Machine learning (ML) algorithms have been used for analyzing complex medical data and identifying prognostic indicators. However, there is still an urgent need for a model to elucidate the predictors related to patient outcomes. Therefore, this research aimed to verify, through ML, the variables involved in the discharge of patients admitted to the ICU due to COVID-19.

**Methods.** In this study, 126 variables were collected with information on demography, hospital length stay and outcome, chronic diseases and tumors, comorbidities and risk factors, complications and adverse events, health care, and vital indicators of patients admitted to an ICU in southern Brazil. These variables were filtered and then selected by a ML algorithm known as decision trees to identify the optimal set of variables for predicting patient discharge using logistic regression. Finally, a confusion matrix was performed to evaluate the model's performance for the selected variables.

**Results.** Of the 532 patients evaluated, 180 were discharged: female (16.92%), with a central venous catheter (23.68%), with a bladder catheter (26.13%), and with an average of 8.46- and 23.65-days using bladder catheter and submitted to mechanical ventilation, respectively. In addition, the chances of discharge increase by 14% for each additional day in the hospital, by 136% for female patients, 716% when there is no bladder catheter, and 737% when no central venous catheter is used. However, the chances of discharge decrease by 3% for each additional year of age and by 9% for each other day of mechanical ventilation. The performance of the training data presented a balanced accuracy of 0.81, sensitivity of 0.74, specificity of 0.88, and the kappa value was 0.64. The test performance had a balanced accuracy of 0.85, sensitivity 0.75, specificity 0.95, and kappa value of 0.73. The McNemar test found that there were no significant differences in the error rates in the training and test data, suggesting good classification.

Submitted 28 December 2023

Accepted 29 April 2024

Published 12 June 2024

Corresponding author

Lincoln Luis Silva,  
lincoln.silva@duke.edu,  
lincolnluis07@gmail.com

Academic editor

Gökhan Karakülah

Additional Information and  
Declarations can be found on  
page 14

DOI 10.7717/peerj.17428

© Copyright

2024 Dos Santos et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

This work showed that female, the absence of a central venous catheter and bladder catheter, shorter mechanical ventilation, and bladder catheter duration were associated with a greater chance of hospital discharge. These results may help develop measures that lead to a good prognosis for the patient.

**Subjects** Epidemiology, Infectious Diseases, Public Health, Data Mining and Machine Learning, COVID-19

**Keywords** COVID-19, Intensive care unity, Critical care, Machine learning, Epidemiology

## INTRODUCTION

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has been one of the most significant global public health crises in recent history. Since its appearance in Wuhan, China, in December 2019, more than 758 million cases and 6.8 million deaths have been reported (*World Health Organization, WHO*). The most severe patients often must be admitted to intensive care units (ICUs) (*Phua et al., 2020*). Nevertheless, accurately discerning the prognostic variables crucial for predicting patient discharge amidst the heterogeneous manifestations of COVID-19 poses a great challenge (*Esakandari et al., 2020*). In recent times, several studies have employed machine learning (ML) algorithms to forecast patient outcomes (*Islam et al., 2022; Jamshidi et al., 2021; Kar et al., 2021*). Thus, diverse sets of variables yield varying model performances. Utilizing a comprehensive set of variables might offer substantial insights for developing a model with optimal performance. Hence, integrating advanced techniques such as ML into prognostic modeling could provide insights for comprehending this complexity and improving patient care strategies.

ICU patients often have critical conditions associated with high morbidity and mortality rates (*Taylor et al., 2021*), including during the COVID-19 pandemic. According to a multicenter study, the death rate from COVID-19 in patients admitted to the ICU was 41.6% (*Armstrong, Kane & Cook, 2020*). Furthermore, studies carried out in Africa have shown a mortality rate of 31.5% (*African COVID-19 Critical Care Outcomes Study Investigators, 2021*), 24% in Europe (*Wendel Garcia et al., 2020*), and almost 40% in Brazil (*Zeiser et al., 2022*), where the high mortality rates of patients hospitalized in UTI are frequently observed in public hospitals (*Soares Pinheiro et al., 2020*).

The main risk factors of mortality involving critically ill patients admitted to the ICU are extended length of stay, age, sex, heart disease, multiple organ failure, sepsis, severe trauma, and acute respiratory failure (*Auld et al., 2022; Otero et al., 2020; Soares Pinheiro et al., 2020*). Treatment is often challenging as each patient may respond differently; however, in general, treatment may involve the use of advanced life support such as mechanical ventilation (*Fadel et al., 2020*), corticosteroid therapy (*Phua et al., 2020*), renal replacement (*Hajjar et al., 2021*), and use of antiretrovirals (*Beigel et al., 2020*). In addition, continuous monitoring and rigorous clinical evaluation are essential to ensure hemodynamic stability and prevent complications (*Poor et al., 2020*).

Investigating the variables correlated with the death of patients with COVID-19 in the ICU may be fundamental for understanding the disease and improving treatment

strategies (Liu et al., 2020; Pijls et al., 2021). Data of patients with COVID-19 has been explored utilizing ML to forecast their prognosis (Kamel et al., 2023) because it represents a sophisticated and adaptable approach to classification modeling by analyzing large datasets to unveil significant latent relationships or patterns (Zakariaee et al., 2023). Studies indicate that, in predicting clinical outcomes among COVID-19 patients, ML methods demonstrate superior accuracy compared to conventional statistical models (Afrash et al., 2022).

Several approaches exist to identify variables related to a negative outcome, with the decision tree being one of these methods (Elhazmi et al., 2022). The decision tree is a technique based on ML used in medicine capable of identifying and correlating several variables involved with the patient's outcome to assist in elaborating strategies to improve their prognosis (Giotta et al., 2022; Kingsford & Salzberg, 2008). This analysis can help clinicians identify risk factors for mortality in patients with COVID-19 in the ICU, allowing early and effective interventions to be implemented to improve clinical outcomes.

Knowing that, in Brazil, the number of critically ill patients and mortality in the ICU has reached alarming levels and that the identification of variables involved with discharge is crucial for successful treatment and reduction of mortality, the objective of this research was to verify through machine learning the variables involved in the outcome of patients admitted to the intensive care unit due to COVID-19. The results can help develop more effective preventive and therapeutic measures and patient care, providing better treatment and reducing the negative impact on public health.

## MATERIALS & METHODS

### Study design

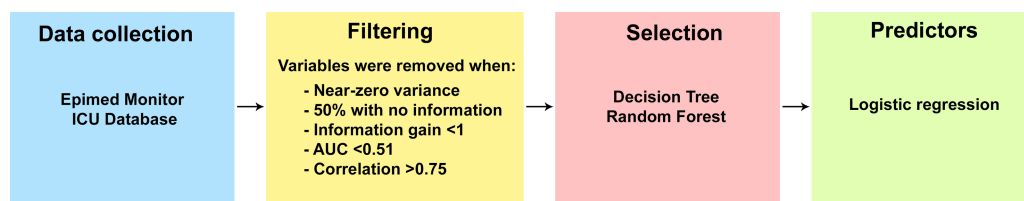
This is an observational study with information from the medical records of patients admitted to the intensive care unit due to COVID-19 in a Municipal Hospital in Paraná between March 2020 and July 2021 to verify the clinical aspects related to patient discharged.

### Settings

This study was divided into four stages: (1) Data collection, (2) filtering of variables to remove those that do not meet the criteria, (3) selection of variables with the best performance to explain the outcome, and forth) identification of variables to estimate the risk for the outcome. The steps developed in this study can be seen in Fig. 1.

### Data collection

A spreadsheet was structured containing four groups of variables for data collection. In the first group are the information related to the sociodemographic characteristics of the patients. As for the second group, information about signs, symptoms, results of imaging, and laboratory tests at the time of hospitalization was gathered. The same variables as the second group were also collected for the third group, but three days after hospitalization. Finally, in the fourth group, there are the same variables as in groups 2 and 3, however, obtained at the time of the patient's outcome, that is, if he was discharged from the hospital or died. Patient information was collected using the Epimed Monitor ICU Database<sup>®</sup>, a platform where all clinical information of patients admitted to the ICU in Brazil is stored



**Figure 1** Steps for identifying variables related to patient outcome.

Full-size DOI: [10.7717/peerj.17428/fig-1](https://doi.org/10.7717/peerj.17428/fig-1)

(Zampieri *et al.*, 2017). Data came from electronic case report forms (eCRF) from which information is gathered through integration between electronic records (medical and/or administrative) of the hospital and manual entry of data made by a manager responsible for entering consecutively the information of each patient in the database (Zampieri *et al.*, 2017).

## Variables

One hundred twenty-six variables were collected from each patient and grouped into seven groups for a better understanding of the variable's nature, shown in [Table 1](#).

## Participants

The study included all patients over 18 years of age admitted to the ICU of a public municipal hospital, a reference in the care of patients diagnosed with COVID-19, in the state of Paraná, southern region of Brazil.

## Statistical analysis

### Filtering

Prior data processing was initially carried out to eliminate variables with little or no information and redundancies in five steps. First, the variance of the variables was calculated, and those that showed almost zero variance were eliminated from the final set. Second, variables with more than 50% missing values were eliminated, that is, variables without information. Third, the information gain expressed by each variable was identified, and those with values lower than one were discarded. Fourth, the sensitivity and specificity of each variable were analyzed to verify its importance for the model using the receiver operating characteristic (ROC) curve, and those below 0.51 were discarded. Finally, a Pearson correlation analysis was used, and variables with solid correlations, *i.e.*, greater than 0.75, were removed to avoid collinearity. Steps 1, 3, and 4 are described below.

To identify variables with zero variance, those variables with the same value in all lines were checked, and those with almost zero variance when meeting the following requirements: (a) regarding the frequency ratio according to the calculation of the frequency of the highest value prevalent divided by the frequency of the second most prevalent value is greater than 19; and (b) when the percentage of unique values given by the number of unique values divided by the total number of samples (times 100) is less than 10%. Therefore, variables classified as almost zero variance were eliminated.

**Table 1** Variables collected from the Epimed Monitor ICU Database for use in this study. Maringá, Paraná, 2023.

Classification	Variables
Personal	Personal: Age, gender, weight, height (cm);
Patient stay and outcome	ICU stay duration, hospital stay duration, unit outcome (treatment/hospitalization outcome);
Chronic diseases and tumors	Insuficiência renal crônica não dialítica, insuficiência renal crônica dialítica, cirrose child A ou B, cirrose child C, insuficiência hepática, tumor sólido locorregional, tumor sólido metastático, sítio do tumor, doença hematológica maligna, tipo de doença hematológica maligna, nome da doença hematológica maligna;
Comorbidities and risk factors	Immunosuppression, severe COPD, AIDS, systemic arterial hypertension, asthma, uncomplicated diabetes (type 1 or 2), complicated diabetes (type 1 or 2), angina, previous myocardial infarction, arrhythmia, hypothyroidism, hyperthyroidism, peripheral arterial disease, chronic arterial fibrillation, rheumatic disease, sequelae of stroke, stroke without sequelae, dementia, smoking, alcoholism, psychiatric disease, morbid obesity, malnutrition, ischemic heart disease, dyslipidemia, history of pneumonia;
Complications and adverse events	Delirium obtundation stupor or coma, seizure or epilepsy, focal neurological deficit, cardiac rhythm disorders, respiratory failure in the first hour, cardiac arrhythmias in the first hour, cardiopulmonary arrest in the first hour, acute kidney injury in the first hour, asystole in the first hour, pulseless electrical activity in the first hour, atrial fibrillation in the first hour, sustained ventricular tachycardia in the first hour, acute respiratory failure in the first hour, arrhythmias, cardiopulmonary arrest, acute kidney injury, pulseless electrical activity, atrial fibrillation, sustained ventricular tachycardia, non-invasive mechanical ventilation, cardiac rhythm disorders, hypovolemic or hemorrhagic shock, septic shock, anaphylactic or undefined shock, BMI;
Healthcare	Failure of non-invasive ventilation, duration of mechanical ventilation, tracheostomy, high-flow mask, duration of hemodialysis, extended hemodialysis in acute kidney injury, hemodialysis, asystole, non-invasive ventilation, vasopressors, hemodialysis in the first hour, mechanical ventilation in the first hour, non-invasive ventilation in the first hour, vasopressors in the first hour, decision on palliative care, chemotherapy, radiotherapy, solid organ transplant, autologous blood transfusion, intestinal transplant, lung transplant, kidney transplant, neurosurgery, central venous catheter, Foley catheter, MAP catheter, intra-aortic balloon, minimally invasive hemodynamic monitoring, red blood cell concentrate transfusion, fresh frozen plasma, thrombolytic agents, NYH classification 2 or 3, use of steroids;

*(continued on next page)*

Table 1 (continued)

Classification	Variables
Vital signs	Lowest SBP 1 h, lowest DBP 1 h, lowest MAP 1 h, highest HR 1 h, highest RR 1 h, highest Temperature 1 h, lowest Glasgow Coma Scale 1 h, highest Leukocyte Count 1 h, lowest Platelet Count 1 h, highest Creatinine 1 h, highest Bilirubin 1 h, highest pH 1 h, lowest pH 1 h, highest PaO <sub>2</sub> 1 h, lowest PaO <sub>2</sub> 1 h, highest PaCO <sub>2</sub> 21 h, lowest PaCO <sub>2</sub> 1 h, highest FiO <sub>2</sub> 1 h, lowest FiO <sub>2</sub> 1 h, highest PaO <sub>2</sub> /FiO <sub>2</sub> Ratio 1 h, lowest PaO <sub>2</sub> /FiO <sub>2</sub> Ratio 1 h, highest Lactate 1 h, Urea, BUN.

Algorithms were applied using the method proposed by [Zawadzki & Kosinski \(2019\)](#), which finds important ranks of discrete attributes based on their entropy with a continuous class attribute. Therefore, variables with little or no information gain, that is, values below one, were ignored.

A function called `filterVarImp` from the `caret` package in R was used, which applies an algorithm to calculate the two probabilities of dichotomous variables to generate an ROC curve and calculate the area under the curve (AUC) ([Kuhn & Johnson, 2013](#)). The AUC ranges from 0 to 1 and serves as a measure of variable importance, as an AUC of 0.5 would indicate that a variable cannot discriminate between the two possibilities. An AUC of 1, on the other hand, suggests a variable capable of completely separating the possibilities ([Silveira et al., 2020](#)). Therefore, those variables with AUC values lower than 0.51 were eliminated.

### Variable selection

For the preparation of decision tree and Random Forest, we segmented the filtered dataset into 80% training and 20% testing sets, to ensure model generalization. This method facilitates an internal validation process: the model is trained on one subset of the data and subsequently tested on another independent subset. This division of the dataset enables an evaluation of the model's capacity to extrapolate predictions to novel instances.

A decision tree is a data analysis technique that uses a set of rules to divide data into smaller, more homogeneous groups. It is a supervised machine learning model used to predict the value of a target variable based on various characteristics. The tree is constructed by dividing the data set into smaller subsets based on the characteristic that best separates the values of the target variable. This division is done through an algorithm where each branch represents a choice or condition that can lead to a specific result ([Venkatasubramaniam et al., 2017](#)). Then, the subsets are split again until the final leaves of the tree contain enough information to predict the target variable. The goal is to identify patterns and relationships between variables in a data set to make predictions and decisions. When applied to complex problems, the decision tree can help simplify data interpretation and find solutions more efficiently. As a result, the decision tree has been widely used in areas such as medicine ([Hajje et al., 2022](#); [Rokach & Maimon, 2005](#)).

Subsequently, Random Forest (RF) was employed to identify potential variables associated with patient discharge. The RF classification model underwent training using

10-fold cross-validation on the training dataset. We also used the XGBoost algorithm for predicting patient outcomes and subsequently compared its performance against RF (*Denisko & Hoffman, 2018; Montomoli et al., 2021; Shanbehzadeh, Nopour & Kazemi-Arpanahi, 2022*). However, the XGBoost algorithm demonstrated inferior performance compared to decision trees, prompting us to maintain the latter as our chosen method for classification. Ultimately, the model's performance was verified using the confusion matrix to evaluate the parameters of sensitivity, specificity, accuracy, kappa agreement test, and McNemar's Test. These variables were then inserted into the logistic regression to determine the chances of the patient progressing to discharge.

All results from the variable selection performed in each of the steps described above are available for further analysis in the supplementary document (<https://doi.org/10.6084/m9.figshare.c.6975699>).

### Logistics regression

Logistic regression (LR) was employed due to its interpretability, simplicity, and widespread usage in medical research settings, making it feasible to develop prognostic models for COVID-19 patients (*Gutiérrez-Pérez et al., 2022; Zapata et al., 2023*). This analysis is an efficient technique that can be used for the classification of two classes. Variables potentially associated with discharge indicated by RF were jointly evaluated using the LR to estimate the odds ratio. Although variables that were highly correlated with each other during data processing were discarded to avoid collinearity problems, variance inflation factor (VIF) was also used in order to detect multicollinearity in the data, and variables with VIF values above ten were removed (*Menard, 2002*). Finally, the quality of the logistic regression fit was verified with the assistance of the half-normal plot and simulated envelope from the “hnp” package in the R software (*Moral, Hinde & Demétrio, 2017*), and the absence of extreme outliers was confirmed through residual analysis. Statistical analyses were conducted using R software version 4.2 (*R Core Team, 2022*) and were considered statistically significant when  $p < 0.05$ .

### Ethic

This study was approved by the Human Research Ethics Committee of Centro Universitário Ingá according to process no 4.276.900.

### Data availability

Data can be accessed at <https://doi.org/10.6084/m9.figshare.c.6975699>.

## RESULTS

### Population characteristic

In total, 532 patients were included in the study between March 2020 and July 2021. According to [Table 2](#), most patients were male (57.14%), without severe COPD (93.05%), and with arterial hypertension (70.11%). The majority, 66.92%, of patients did not have complicated diabetes (type 1 or 2), 92.29% did not have complicated diabetes (type 1 or 2), 79.51% were not morbidly obese, and 89.85% did not have hypothyroidism. Regarding

patient care, most did not need non-invasive ventilation in the first hour (68.8%) or vasopressors (62.03%). However, 59.77% required vasopressors, 86.65% a central venous catheter, 90.41% a bladder catheter, and 73.87% an arterial catheter. Only 9.21% needed a transfusion. Regarding relative frequencies, the people who died presented one or more conditions: hypertension, need for vasopressors, central venous catheter, bladder catheter, and/or arterial catheter.

Regarding the continuous variables, in [Table 3](#), it can be seen that the mean and median values for age, highest creatinine at 1 hour, BUN, and duration of mechanical ventilation were lower in the group of patients who were discharged compared to those who died. Conversely, the length of hospital stay had lower mean and median values in the group of deceased patients, possibly because the clinical progression to death is faster than the recovery process in COVID-19 patients admitted to the ICU. Of the 126 variables collected in this study, only 22 remained to classify patients progressing to discharge in order of importance: duration of mechanical ventilation, hospital length stay, age, BUN, highest creatinine 1 h, central venous catheter, bladder catheter, true vasopressors, acute respiratory failure in the first hour, non-invasive ventilation failure, sex, vasopressors in the first hour, uncomplicated diabetes (type 1 or 2), tracheostomy, hypothyroidism, systemic arterial hypertension, non-invasive ventilation in the first hour, morbid obesity, arterial catheter, transfusion, complicated diabetes, and severe COPD. Therefore, these variables were used in logistic regression to estimate the Odds Ratio for discharge. The results of the logistic regression are presented in [Table 4](#).

According to [Table 4](#), the chances of patients being discharged were 136% higher (OR 2.36; 95% CI [1.28–4.44];  $p = 0.007$ ) in females, 14% higher (OR 1.14; 95% CI [1.10–1.19];  $p < 0.001$ ) as hospital length stay increases, 737% higher when there was no central venous catheter (OR 8.37; 95% CI [2.41–30.44];  $p < 0.001$ ), 716% higher when there was no bladder catheter (OR 8.16; 95% CI [2.30–32.14];  $p = 0.002$ ). However, they decrease by 3% for each additional year of age (OR 1.97; 95% CI [0.95–0.99];  $p = 0.011$ ), and 9% smaller (OR 0.91; 95% CI [0.86–0.96];  $p = 0.001$ ) for each additional day of mechanical ventilation.

[Table 5](#) presents the confusion matrix as a performance report of the variables used in RF and logistic regression to verify the variables associated with the patient's discharge. Regarding the training data of RF, the balanced accuracy (proportion of correctly classified cases) could predict 80.5% of the classification, with a sensitivity (true positive rate) of 90.1% and specificity (true negative rate) of 70.9%. The positive predictive value (PPV) expresses the proportion of true positive cases among all positive predictions, which was 84.9%, and the negative predictive value does the same for true negatives (NPV), which was 79.8%. The kappa (a measure of agreement between observed and predicted classifications) was 62.6%, and the McNemar's test (used to compare error proportions between models) was not significant ( $p > 0.07$ ). For the test data, balanced accuracy was 98.2%, sensitivity of 100%, specificity of 96.5%, PPV of 98.4%, NPV of 100%, kappa of 97.4% and McNemar's test was not significant.

Considering the training data, the logistic regression presents a balanced accuracy of 81.7%; that is, the model corrected 81% of the predictions, with a sensitivity of 74.6%



**Table 2** Absolute and relative frequency of clinical variables of patients admitted to the intensive care unit with COVID-19, Maringá, Paraná, 2023.

Variables	N (%)		
	Total	Discharge	Death
<b>Sex</b>			
Female	228 (42.86%)	90 (16.92%)	138 (25.94%)
Male	304 (57.14%)	90 (16.92%)	214 (40.23%)
<b>COPD severe</b>			
No	495 (93.05%)	171 (32.14%)	324 (60.9%)
Yes	37 (6.95%)	9 (1.69%)	28 (5.26%)
<b>Systemic arterial hypertension</b>			
No	159 (29.89%)	64 (12.03%)	95 (17.86%)
Yes	373 (70.11%)	116 (21.8%)	257 (48.31%)
<b>Uncomplicated diabetes</b>			
No	356 (66.92%)	116 (21.8%)	240 (45.11%)
Yes	176 (33.08%)	64 (12.03%)	112 (21.05%)
<b>Complicated diabetes</b>			
No	491 (92.29%)	169 (31.77%)	322 (60.53%)
Yes	41 (7.71%)	11 (2.07%)	30 (5.64%)
<b>Morbid obesity</b>			
No	423 (79.51%)	136 (25.56%)	287 (53.95%)
Yes	109 (20.49%)	44 (8.27%)	65 (12.22%)
<b>Hypothyroidism</b>			
No	478 (89.85%)	164 (30.83%)	314 (59.02%)
Yes	54 (10.15%)	16 (3.01%)	38 (7.14%)
<b>Non-invasive ventilation in the first hour</b>			
No	366 (68.8%)	121 (22.74%)	245 (46.05%)
Yes	166 (31.2%)	59 (11.09%)	107 (20.11%)
<b>Vasopressors in the first hour</b>			
No	330 (62.03%)	130 (24.44%)	200 (37.59%)
Yes	202 (37.97%)	50 (9.4%)	152 (28.57%)
<b>Acute respiratory failure in the first hour</b>			
No	329 (61.84%)	119 (22.37%)	210 (39.47%)
Yes	203 (38.16%)	61 (11.47%)	142 (26.69%)
<b>Vasopressors</b>			
No	214 (40.23%)	95 (17.86%)	119 (22.37%)
Yes	318 (59.77%)	85 (15.98%)	233 (43.8%)
<b>Failure of non-invasive ventilation</b>			
No	428 (80.45%)	148 (27.82%)	280 (52.63%)
Yes	104 (19.55%)	32 (6.02%)	72 (13.53%)
<b>Tracheostomy</b>			
No	440 (82.71%)	136 (25.56%)	304 (57.14%)
Yes	92 (17.29%)	44 (8.27%)	48 (9.02%)

(continued on next page)

Table 2 (continued)

Variables	N (%)		
	Total	Discharge	Death
<b>Central venous catheter</b>			
No	71 (13.35%)	54 (10.15%)	17 (3.2%)
Yes	461 (86.65%)	126 (23.68%)	335 (62.97%)
<b>Indwelling bladder catheter</b>			
No	51 (9.59%)	41 (7.71%)	10 (1.88%)
Yes	481 (90.41%)	139 (26.13%)	342 (64.29%)
<b>PAM catheter</b>			
No	139 (26.13%)	72 (13.53%)	67 (12.59%)
Yes	393 (73.87%)	108 (20.3%)	285 (53.57%)
<b>Transfusion</b>			
No	483 (90.79%)	167 (31.39%)	316 (59.4%)
Yes	49 (9.21%)	13 (2.44%)	36 (6.77%)

Table 3 Descriptive of continuous variables.

Variables	Discharge	Death
<b>Age (years)</b>		
Mean	57.32	63.72
Median	60	65
Standard deviation	16.71	14.01
<b>Length of hospital stay (days)</b>		
Mean	23.65	14.08
Median	17.5	12
Standard deviation	40.24	20.99
<b>Highest creatinine 1 h (mg/dL)</b>		
Mean	1.03	1.36
Median	0.7	0.8
Standard deviation	1.43	2.07
<b>BUN (mg/dL)</b>		
Mean	28.59	37.92
Median	22.65	30.84
Standard deviation	21.13	26.77
<b>Duration of mechanical ventilation (days)</b>		
Mean	8.46	9.7
Median	6	8
Standard deviation	10.2	8.12

regarding the ability to classify truly positive cases, and the specificity of 88.9% showing the ability to organize negative instances correctly and the kappa value of 64.5%. The test data's balanced accuracy was 85.5%, sensitivity 75.8%, and specificity 95.3%. The kappa value was 73.9%, which indicates a substantial agreement between the predictions and the valid classifications. At the same time, the McNemar test showed no significant differences between the proportions of errors made by the models.

**Table 4** Multivariate logistic regression to detect variables related to discharge. Maringá, PR, 2023.

Variables	Odds ratio	Standard error	95% IC	p-value
(Intercept)	0.25	0.44	0.01–7.39	0.428
Age	0.97**	0.01	0.95–0.99	0.011
Sex [F]	2.36***	0.75	1.28–4.44	0.007
Length of hospital stay	1.14***	0.02	1.10–1.19	<0.001
COPD severe [No]	2.94	2.42	0.63–15.58	0.190
Systemic Arterial Hypertension [No]	1.19	0.42	0.60–2.36	0.613
Uncomplicated diabetes [No]	0.91	0.30	0.48–1.72	0.766
Complicated diabetes [No]	0.70	0.42	0.22–2.42	0.552
Morbid obesity [No]	0.77	0.28	0.37–1.60	0.469
Hypothyroidism [No]	2.14	1.06	0.84–5.96	0.124
Non-invasive ventilation in the first hour [No]	1.38	0.54	0.64–3.01	0.417
Vasopressors in the first hour [No]	0.95	0.48	0.35–2.55	0.924
Acute respiratory failure in the first hour [No]	0.49	0.20	0.22–1.08	0.075
Vasopressors [No]	1.03	0.51	0.38–2.74	0.956
Highest creatinine 1 h	1.02	0.07	0.85–1.16	0.799
BUN	1.00	0.01	0.98–1.01	0.677
Non-invasive Ventilation Failure [No]	0.58	0.25	0.25–1.36	0.212
Duration of mechanical ventilation	0.91***	0.03	0.86–0.96	0.001
Tracheostomy [No]	0.47	0.21	0.19–1.14	0.098
Central Venous Catheter [No]	8.37***	5.37	2.41–30.44	0.001
Indwelling bladder catheter [No]	8.16***	5.40	2.30–32.14	0.002
PAM catheter [No]	1.95	0.88	0.79–4.71	0.140
Transfusion [No]	1.60	0.81	0.61–4.46	0.350

**Notes.**\*\* $p < 0.05$ .\*\*\* $p < 0.01$ .**DISCUSSION**

Machine learning can identify patterns in the variables involved in the outcome of patients admitted to the intensive care unit due to COVID-19 and propose interventions and decision-making to improve epidemiological data and build care plans.

Investigating and knowing the variables involved in the death and discharge of patients in serious situations are relevant metrics for improving treatment, diagnosis, and quality of care, reducing length of stay and mortality. This is the first study to collect 126 variables to determine associated factors that contributed to the death or discharge of these patients. Among the main factors, the variables, hospital length stay, central venous catheter, and bladder catheter, were related to the discharge of patients admitted to the ICU with COVID-19. In contrast, age and duration of mechanical ventilation were related to greater chances of going to death. The model's performance that detected these variables was satisfactory since the training and test data results are similar.

*Chimbunde et al. (2023)* utilized RF to predict determinants of COVID-19 mortality in South Africa. Surprisingly, their analysis revealed that being female was associated with

**Table 5** Confusion matrix showing the performance of training and test data with the variables used in logistic regression.

Outcome	Random forest				Logistic regression			
	Train		Test		Train		Test	
	Discharge	Death	Discharge	Death	Discharge	Death	Discharge	Death
Discharge	95	24	28	0	100	27	22	3
Death	39	220	1	64	34	217	7	61
<b>Performance</b>								
Accuracy	0.833		0.989		0.838		0.892	
95% Confidence interval	0.791–0.869		0.941–0.999		0.797–0.874		0.811–0.947	
Kappa	0.626		0.974		0.645		0.739	
McNemar's test <i>p</i> -value	0.07		1.000		0.442		0.342	
Precision	0.849		0.984		0.787		0.880	
Recall	0.901		1.000		0.746		0.758	
F1 score	0.874		0.992		0.716		0.814	
Sensitivity	0.901		1.000		0.746		0.758	
Specificity	0.709		0.965		0.889		0.953	
Positive predictive value	0.849		0.984		0.787		0.880	
Negative predictive value	0.798		1.000		0.864		0.897	
Accuracy balance	0.805		0.982		0.817		0.855	

increased mortality, contrary to findings from our study and others, which suggested that being female was protective (*Elhazmi et al., 2022; Kar et al., 2021*). It is noteworthy that their study incorporated different variables than ours. This highlights that the choice of variables can significantly impact the variables identified as associated with the outcome. Moreover, their model exhibited different performance metrics compared to ours, achieving a recall of 76% and a precision of 87%, whereas our model achieved a higher recall of 90.1% and precision of 84.9%, indicating that our model correctly identified more true positives (*Chimbunde et al., 2023; Hicks et al., 2022*).

Prolonging hospital length stay was related to patient improvement, and this can be corroborated by *Rees et al. (2020)*, who carried out a systematic review of the length of hospital stay and found that patients who were discharged had a more extended hospital stay than those who died after admission (*Rees et al., 2020*). However, this variable is complex since other studies have detected the opposite. This difference may be related to different criteria used to admit patients to ICUs and the particularities adopted regarding patient care in each location where the studies were conducted (*da Costa Sousa et al., 2022; Gupta et al., 2020*).

Non-use of a central venous catheter was another variable correlated with patient discharge. Around 5% to 10% of individuals infected with COVID-19 have clinical conditions that require hospitalization in the ICU and the use of mechanical ventilation (*Pericàs et al., 2020; Wu & McGoogan, 2020*). In severe cases, the occurrence rate of acute kidney injury and septic shock is 15% and 20%, respectively (*Ng et al., 2020*). Consequently, administering vasoactive agents or hemodialysis is often necessary, and central venous access is expected (*Chun et al., 2020*). The central venous catheter is a medical device used

in patients who require long-term treatments or large-volume infusions, enabling rapid infusions of volumes and medications that cannot be administered through peripheral vascular access (*Dias et al., 2022*). In this sense, the importance of this instrument in stabilizing the patient and increasing the survival rate is observed. However, not many specific studies have evaluated the use of this instrument in the ICU for patients with COVID-19; on the contrary, some studies have shown a risk of contracting infections and other thrombolytic problems arising from this procedure (*Lugon et al., 2022*).

Another interesting finding in this study was the non-use of the bladder catheter as a predictive variable for patient discharge. Although it is also an instrument capable of evaluating specific renal parameters of patients, it can help doctors make decisions for the patient's clinical improvement. However, it is frequently associated with urinary tract infections in ICUs (*Díaz Pollán et al., 2022; Ong et al., 2021*). This underscores the importance of minimizing invasive procedures and implementing strategies to reduce the risk of catheter-associated complications in critically ill patients with COVID-19. Addressing these factors may potentially improve patient outcomes and shorten hospital stays.

Regarding the risks of death, studies have investigated the relationship between age and duration of mechanical ventilation with the death of patients with COVID-19 admitted to the ICU (*Grasselli et al., 2020; Richardson et al., 2020; Wang et al., 2020*). In general, there is a tendency for older patients to have a higher risk of dying, and patients who require mechanical ventilation for more prolonged periods also have a higher risk of death (*Auld et al., 2020; Domecq et al., 2021*). The same was corroborated in this study, in which the average age of death was people over 60 years old.

It was possible through technology to verify the presence of these variables involved in the discharge and deaths of patients admitted to the intensive care unit. This study suggests that paying more attention to these variables is necessary to prevent the condition's progression and increase complications. The use of machine learning in the ICU is evolving but is still limited to diagnostic and prognostic values. *Henry et al. (2015)* developed a machine learning methodology using the MIMIC Clinical Database (Multiparametric Intelligent Monitoring in Intensive Care) –II. They created a model that considers the censoring effects of clinical interventions on the clinical outcomes of patients. This model, called targeted real-time early warning score (TREWScore), can identify patients at high risk of developing septic shock; it was created in 2015 and is already being applied in some hospitals worldwide (*Henry et al., 2015; Niemantsverdriet et al., 2021*). Therefore, it is likely that the use of machine learning tools will become more and more frequent in the future. In this way, this research confirms the findings of previous studies where machine learning demonstrated reliability in determining death and discharge variables.

This study has some limitations. The first is that it is an observational study in which the limits are due to the possibility that some essential variables have not been taken into account or the identification of risk variables. However, more than 126 variables were used here and were carefully selected to create a robust analysis model. The second limitation is that this is a study from a hospital in southern Brazil, which makes it challenging to make generalizations at a global level. However, as technical procedures are standardized

across the national territory, and considering that Brazil has continental dimensions and the diversification of the Brazilian people, it is possible to infer that results can be obtained in other locations within the national territory. Finally, we evaluated the generalizability using the same dataset divided into training and testing sets. This practice is discouraged due to potential demographic biases, as well as biases introduced during data collection, processing, and organization. However, validating the model on additional datasets beyond the original development dataset may not always be feasible due to ethical, technical, or financial constraints associated with sharing clinical data (Yang, Soltan & Clifton, 2022).

## CONCLUSIONS

The use of machine learning techniques was possible to identify variables associated with the discharge of patients admitted to the ICU due to COVID-19. The results indicate that factors such as gender, length of hospital stay, presence of central venous catheter, and bladder catheter were significantly related to the likelihood of hospital discharge. Additionally, the study demonstrated that age and duration of mechanical ventilation had a negative impact on the chances of patient discharge. Both the Random Forest and logistic regression models showed satisfactory performance in predicting hospital discharge, with consistent results between training and test data. These findings suggest that the use of ML can be a valuable tool in identifying prognostic factors and supporting clinical decision-making in critically ill COVID-19 patients in the ICU.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Lander Dos Santos conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Lincoln Luis Silva conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

- Fernando Castilho Pelloso conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Vinicius Maia performed the experiments, prepared figures and/or tables, and approved the final draft.
- Constanza Pujals performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Deise Helena Borghesan analyzed the data, prepared figures and/or tables, and approved the final draft.
- Maria Dalva Carvalho conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Raíssa Bocchi Pedroso conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Sandra Marisa Pelloso conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

### Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Centro Universitário Ingá granted Ethical approval to carry out the study (Ethical Application Ref: 4.276.900)

### Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Centro Universitário Ingá granted Ethical approval to carry out the study (Ethical Application Ref: 4.276.900)

### Data Availability

The following information was supplied regarding data availability:

The data is available at figshare: Silva, Lincoln (2023). Use of machine learning to identify protective factors for death from COVID-19 in the ICU. figshare. Collection. Available at <https://doi.org/10.6084/m9.figshare.c.6975699.v1>.

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.17428#supplemental-information>.

## REFERENCES

- Afrash MR, Kazemi-Arpanahi H, Shanbehzadeh M, Nopour R, Mirbagheri E.** 2022. Predicting hospital readmission risk in patients with COVID-19: a machine learning approach. *Informatics in Medicine Unlocked* 30:100908 DOI 10.1016/j.imu.2022.100908.

- African COVID-19 Critical Care Outcomes Study (ACCCOS) Investigators.** 2021. Patient care and clinical outcomes for patients with COVID-19 infection admitted to African high-care or intensive care units (ACCCOS): a multicentre, prospective, observational cohort study. *Lancet* **397**:1885–1894 DOI [10.1016/s0140-6736\(21\)00441-4](https://doi.org/10.1016/s0140-6736(21)00441-4).
- Armstrong RA, Kane AD, Cook TM.** 2020. Outcomes from intensive care in patients with COVID-19: a systematic review and meta-analysis of observational studies. *Anaesthesia* **75**:1340–1349 DOI [10.1111/anae.15201](https://doi.org/10.1111/anae.15201).
- Auld SC, Caridi-Scheible M, Blum JM, Robichaux C, Kraft C, Jacob JT, Jabaley CS, Carpenter D, Kaplow R, Hernandez-Romieu AC, Adelman MW, Martin GS, Coopersmith CM, Murphy DJ.** 2020. ICU and ventilator mortality among critically ill adults with coronavirus disease 2019. *Critical Care Medicine* **48**:e799–e804 DOI [10.1097/ccm.0000000000004457](https://doi.org/10.1097/ccm.0000000000004457).
- Auld SC, Harrington KRV, Adelman MW, Robichaux CJ, Overton EC, Caridi-Scheible M, Coopersmith CM, Murphy DJ.** 2022. Trends in ICU mortality from coronavirus disease 2019: a tale of three surges. *Critical Care Medicine* **50**:245–255 DOI [10.1097/ccm.0000000000005185](https://doi.org/10.1097/ccm.0000000000005185).
- Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, Hohmann E, Chu HY, Luetkemeyer A, Kline S, Lopez de Castilla D, Finberg RW, Dierberg K, Tapson V, Hsieh L, Patterson TF, Paredes R, Sweeney DA, Short WR, Touloumi G, Lye DC, Ohmagari N, Oh MD, Ruiz-Palacios GM, Benfield T, Fätkenheuer G, Kortepeter MG, Atmar RL, Creech CB, Lundgren J, Babiker AG, Pett S, Neaton JD, Burgess TH, Bonnett T, Green M, Makowski M, Osinusi A, Nayak S, Lane HC.** 2020. Remdesivir for the treatment of Covid-19—final report. *New England Journal of Medicine* **383**:1813–1826 DOI [10.1056/NEJMoa2007764](https://doi.org/10.1056/NEJMoa2007764).
- Chimbunde E, Sigwadhi LN, Tamuzi JL, Okango EL, Daramola O, Ngah VD, Nyasulu PS.** 2023. Machine learning algorithms for predicting determinants of COVID-19 mortality in South Africa. *Frontiers in Artificial Intelligence* **6**:1171256 DOI [10.3389/frai.2023.1171256](https://doi.org/10.3389/frai.2023.1171256).
- Chun TT, Judelson DR, Rigberg D, Lawrence PF, Cuff R, Shalhub S, Wohlauser M, Abularrage CJ, Anastasios P, Arya S, Aulivola B, Baldwin M, Baril D, Bechara CF, Beckerman WE, Behrendt CA, Benedetto F, Bennett LF, Charlton-Ouw KM, Chawla A, Chia MC, Cho S, Choong A, Chou EL, Christiana A, Coscas R, De Caridi G, Ellozy S, Etkin Y, Faries P, Fung AT, Gonzalez A, Griffin CL, Guidry L, Gunawansa N, Gwertzman G, Han DK, Hicks CW, Hinojosa CA, Hsiang Y, Ilonzo N, Jayakumar L, Joh JH, Johnson AP, Kabbani LS, Keller MR, Khashram M, Koleilat I, Krueger B, Kumar A, Lee CJ, Lee A, Levy MM, Lewis CT, Lind B, Lopez-Pena G, Mohebbi J, Molnar RG, Morrissey NJ, Motaganahalli RL, Mouawad NJ, Newton DH, Ng JJ, O'Banion LA, Phair J, Rancic Z, Rao A, Ray HM, Rivera AG, Rodriguez L, Sales CM, Salzman G, Sarfati M, Savlania A, Schanzer A, Sharafuddin MJ, Sheahan M, Siada S, Siracuse JJ, Smith BK, Smith M, Soh I, Sorber R, Sundaram V, Sundick S, Tomita TM, Trinidad B, Tsai S, Vouyouka AG, Westin GG, Williams Jr MS, Wren SM, Yang JK, Yi J, Zhou W, Zia S, Woo K.**



2020. Managing central venous access during a health care crisis. *Journal of Vascular Surgery* 72:1184–1195 DOI 10.1016/j.jvs.2020.06.112.
- da Costa Sousa V, da Silva MC, De Mello MP, Guimarães JAM, Perini JA. 2022. Factors associated with mortality, length of hospital stay and diagnosis of COVID-19: data from a field hospital. *Journal of Infection and Public Health* 15:800–805 DOI 10.1016/j.jiph.2022.06.010.
- Denisko D, Hoffman MM. 2018. Classification and interaction in random forests. *Proceedings of the National Academy of Sciences of the United States of America* 115:1690–1692 DOI 10.1073/pnas.1800256115.
- Dias TO, Assad LG, Paula VG, Almeida LF, Moraes EB, Nassar PRB. 2022. Good practices in central venous catheter maintenance in time of covid-19: an observational study. *Revista Brasileira de Enfermagem* 75:e20210397 DOI 10.1590/0034-7167-2021-0397.
- Díaz Pollán B, Guedez López GV, García Clemente PM, Jiménez González M, García Bujalance S, Gómez-Gil Mirá MR. 2022. Urinary tract infections in hospitalized COVID-19 patients, what's up, doc? *Journal of Clinical Medicine* 11(7):1815 DOI 10.3390/jcm11071815.
- Domecq JP, Lal A, Sheldrick CR, Kumar VK, Boman K, Bolesta S, Bansal V, Harhay MO, Garcia MA, Kaufman M, Danesh V, Cheruku S, Banner-Goodspeed VM, Anderson 3rd HL, Milligan PS, Denson JL, St Hill CA, Dodd KW, Martin GS, Gajic O, Walkey AJ, Kashyap R. 2021. Outcomes of patients with coronavirus disease 2019 receiving organ support therapies: the international viral infection and respiratory illness universal study registry. *Critical Care Medicine* 49:437–448 DOI 10.1097/ccm.0000000000004879.
- Elhazmi A, Al-Omari A, Sallam H, Mufti HN, Rabie AA, Alshahrani M, Mady A, Alghamdi A, Altalaq A, Azzam MH, Sindi A, Kharaba A, Al-Aseri ZA, Almekhlafi GA, Tashkandi W, Alajmi SA, Faqihi F, Alharthy A, Al-Tawfiq JA, Melibari RG, Al-Hazzani W, Arabi YM. 2022. Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU. *Journal of Infection and Public Health* 15:826–834 DOI 10.1016/j.jiph.2022.06.008.
- Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, Farahmandian N, Miresmaeili SM, Bahreini E. 2020. A comprehensive review of COVID-19 characteristics. *Biological Procedures Online* 22:19 DOI 10.1186/s12575-020-00128-2.
- Fadel R, Morrison AR, Vahia A, Smith ZR, Chaudhry Z, Bhargava P, Miller J, Kenney RM, Alangaden G, Ramesh MS. 2020. Early short-course corticosteroids in hospitalized patients with COVID-19. *Clinical Infectious Diseases* 71:2114–2120 DOI 10.1093/cid/ciaa601.
- Giotto M, Trerotoli P, Palmieri VO, Passerini F, Portincasa P, Dargenio I, Mokhtari J, Montagna MT, De Vito D. 2022. Application of a decision tree model to predict the outcome of non-intensive inpatients hospitalized for COVID-19. *International Journal of Environmental Research and Public Health* 19(20):13016 DOI 10.3390/ijerph192013016.

- Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, Cereda D, Coluccello A, Foti G, Fumagalli R, Iotti G, Latronico N, Lorini L, Merler S, Natalini G, Piatti A, Ranieri MV, Scandroglio AM, Storti E, Cecconi M, Pesenti A. 2020. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA* 323(16):1574–1581 DOI 10.1001/jama.2020.5394.
- Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, Brenner SK, Leonberg-Yoo A, Schenck EJ, Radbel J, Reiser J, Bansal A, Srivastava A, Zhou Y, Sutherland A, Green A, Shehata AM, Goyal N, Vijayan A, Velez JCQ, Shaefi S, Parikh CR, Arunthamakun J, Athavale AM, Friedman AN, Short SAP, Kibbelaar ZA, Abu Omar S, Admon AJ, Donnelly JP, Gershengorn HB, Hernán MA, Semler MW, Leaf DE. 2020. Factors associated with death in critically ill patients with coronavirus disease 2019 in the US. *JAMA Internal Medicine* 180:1436–1447 DOI 10.1001/jamainternmed.2020.3596.
- Gutiérrez-Pérez IA, Buendía-Roldán I, Pérez-Rubio G, Chávez-Galán L, Hernández-Zenteno RJ, Aguilar-Duran H, Fricke-Galindo I, Zaragoza-García O, Falfán-Valencia R, Guzmán-Guzmán IP. 2022. Outcome predictors in COVID-19: an analysis of emergent systemic inflammation indices in Mexican population. *Frontiers in Medicine* 9:1000147 DOI 10.3389/fmed.2022.1000147.
- Hajjar LA, Costa I, Rizk SI, Biselli B, Gomes BR, Bittar CS, De Oliveira GQ, De Almeida JP, De Oliveira Bello MV, Garzillo C, Leme AC, Elena M, Val F, De Almeida Lopes M, Lacerda MVG, Ramires JAF, Kalil Filho R, Teboul JL, Landoni G. 2021. Intensive care management of patients with COVID-19: a practical approach. *Annals of Intensive Care* 11:36 DOI 10.1016/s2213-2600(20)30161-2.
- Hajje F, Alohali MA, Badr M, Rahman MA. 2022. A comparison of decision tree algorithms in the assessment of biomedical data. *BioMed Research International* 2022:9449497 DOI 10.1155/2022/9449497.
- Henry KE, Hager DN, Pronovost PJ, Saria S. 2015. A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine* 7:299ra122 DOI 10.1126/scitranslmed.aab3719.
- Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. 2022. On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports* 12:5979 DOI 10.1038/s41598-022-09954-8.
- Islam KR, Kumar J, Tan TL, Reaz MBI, Rahman T, Khandakar A, Abbas T, Hossain MSA, Zughair SM, Chowdhury MEH. 2022. Prognostic model of ICU admission risk in patients with COVID-19 infection using machine learning. *Diagnostics* 12(9):2144 DOI 10.3390/diagnostics12092144.
- Jamshidi E, Asgary A, Tavakoli N, Zali A, Setareh S, Esmaily H, Jamaldini SH, Daaee A, Babajani A, Sendani Kashi MA, Jamshidi M, Jamal Rahi S, Mansouri N. 2021. Using machine learning to predict mortality for COVID-19 patients on day 0 in the ICU. *Frontiers in Digital Health* 3:681608 DOI 10.3389/fdgth.2021.681608.
- Kamel FO, Magadmi R, Qutub S, Badawi M, Madani TA, Alhothali A, Abozinadah EA, Bakhshwin DM, Jamal MH, Burzangi AS, Bazuhair M, Alqutub H, Alqutub

- A, Felemban SM, Al-Sayes F, Adam S. 2023. Machine learning-based prediction of COVID-19 prognosis using clinical and hematologic data. *Cureus* 15:e50212 DOI 10.7759/cureus.50212.
- Kar S, Chawla R, Haranath SP, Ramasubban S, Ramakrishnan N, Vaishya R, Sibal A, Reddy S. 2021. Multivariable mortality risk prediction using machine learning for COVID-19 patients at admission (AICOVID). *Scientific Reports* 11:12801 DOI 10.1038/s41598-021-92146-7.
- Kingsford C, Salzberg SL. 2008. What are decision trees? *Nature Biotechnology* 26:1011–1013 DOI 10.1038/nbt0908-1011.
- Kuhn M, Johnson K. 2013. *Applied predictive modeling*. New York: Springer DOI 10.1007/978-1-4614-6849-3.
- Liu Y, Mao B, Liang S, Yang JW, Lu HW, Chai YH, Wang L, Zhang L, Li QH, Zhao L, He Y, Gu XL, Ji XB, Li L, Jie ZJ, Li Q, Li XY, Lu HZ, Zhang WH, Song YL, Qu JM, Xu JF, Shanghai Clinical Treatment Experts Group for C. 2020. Association between age and clinical characteristics and outcomes of COVID-19. *European Respiratory Journal England* 55(5):2001112 DOI 10.1183/13993003.01112-2020.
- Lugon JR, Neves P, Pio-Abreu A, do Nascimento MM, Sesso R. 2022. Evaluation of central venous catheter and other risk factors for mortality in chronic hemodialysis patients with COVID-19 in Brazil. *International Urology and Nephrology* 54:193–199 DOI 10.1007/s11255-021-02920-9.
- Menard S. 2002. *Applied logistic regression analysis*. Thousand Oaks: Sage.
- Montomoli J, Romeo L, Moccia S, Bernardini M, Migliorelli L, Berardini D, Donati A, Carsetti A, Bocci MG, Wendel Garcia PD, Fumeaux T, Guerci P, Schüpbach RA, Ince C, Frontoni E, Hilty MP. 2021. Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients. *Journal of Intensive Medicine* 1:110–116 DOI 10.1016/j.jointm.2021.09.002.
- Moral RA, Hinde J, Demétrio CGB. 2017. Half-normal plots and overdispersed models in R: the hnp package. *Journal of Statistical Software* 81:1–23 DOI 10.18637/jss.v081.i10.
- Ng JJ, Luo Y, Phua K, Choong A. 2020. Acute kidney injury in hospitalized patients with coronavirus disease 2019 (COVID-19): a meta-analysis. *Journal of Infection* 81:647–679 DOI 10.1016/j.jinf.2020.05.009.
- Niemantsverdriet MSA, Varkila MRJ, Vromen-Wijsman JLP, Hoefler IE, Bellomo D, Van Vliet MH, Van Solinge WW, Cremer OL, Haitjema S. 2021. Transportability and implementation challenges of early warning scores for septic shock in the ICU: a perspective on the TREWScore. *Frontiers in Medicine* 8:793815 DOI 10.3389/fmed.2021.793815.
- Ong CCH, Farhanah S, Linn KZ, Tang YW, Poon CY, Lim AY, Tan HR, Binte Hamed NH, Huan X, Puah SH, Ho BCH, Soon MML, Ang BSP, Vasoo S, Chan M, Leo YS, Ng OT, Marimuthu K. 2021. Nosocomial infections among COVID-19 patients: an analysis of intensive care unit surveillance data. *Antimicrobial Resistance & Infection Control* 10:119 DOI 10.1186/s13756-021-00988-7.

- Otero ML, Menezes RC, Ferreira IBB, Issa FL, Agareno G, Carmo TA, Arriaga MB, Fukutani KF, Pamplona Neto L, Agareno S, Filgueiras Filho NM, Akrami KM, Andrade BB. 2020. Factors associated with mortality in critically ill patients diagnosed with hospital acquired infections. *Infection and Drug Resistance* 13:2811–2817 DOI 10.2147/idr.s264276.
- Pericàs JM, Hernandez-Meneses M, Sheahan TP, Quintana E, Ambrosioni J, Sandoval E, Falces C, Marcos MA, Tuset M, Vilella A, Moreno A, Miro JM. 2020. COVID-19: from epidemiology to treatment. *European Heart Journal* 41:2092–2112 DOI 10.1093/eurheartj/ehaa462.
- Phua J, Weng L, Ling L, Egi M, Lim CM, Divatia JV, Shrestha BR, Arabi YM, Ng J, Gomersall CD, Nishimura M, Koh Y, Du B. 2020. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *The Lancet Respiratory Medicine* 8:506–517 DOI 10.1016/s2213-2600(20)30161-2.
- Pijls BG, Jolani S, Atherley A, Derckx RT, Dijkstra JIR, Franssen GHJ, Hendriks S, Richters A, Venemans-Jellema A, Zalpuri S, Zeegers MP. 2021. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open* 11:e044640 DOI 10.1136/bmjopen-2020-044640.
- Poor HD, Ventetuolo CE, Tolbert T, Chun G, Serrao G, Zeidman A, Dangayach NS, Olin J, Kohli-Seth R, Powell CA. 2020. COVID-19 critical illness pathophysiology driven by diffuse pulmonary thrombi and pulmonary endothelial dysfunction responsive to thrombolysis. *Clinical and Translational Medicine* 10:e44 DOI 10.1002/ctm2.44.
- R Core Team. 2022. R: A language and environment for statistical computing. Version 4.2.0. Vienna: R Foundation for Statistical Computing. Available at <https://www.r-project.org>.
- Rees EM, Nightingale ES, Jafari Y, Waterlow NR, Clifford S, BP CA, Group CW, Jombart T, Procter SR, Knight GM. 2020. COVID-19 length of hospital stay: a systematic review and data synthesis. *BMC Medicine* 18:270 DOI 10.1186/s12916-020-01726-3.
- Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, Cohen SL, Cookingham J, Coppa K, Diefenbach MA, Dominello AJ, Duer-Hefe J, Falzon L, Gitlin J, Hajizadeh N, Harvin TG, Hirschwerk DA, Kim EJ, Kozel ZM, Marrast LM, Mogavero JN, Osorio GA, Qiu M, Zanos TP. 2020. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA* 20:2052–2059 DOI 10.1001/jama.2020.6775.
- Rokach L, Maimon O. 2005. Decision trees. In: Maimon O, Rokach L, eds. *Data Mining and Knowledge Discovery Handbook*. Boston: Springer, 165–192.
- Shanbehzadeh M, Nopour R, Kazemi-Arpanahi H. 2022. Using decision tree algorithms for estimating ICU admission of COVID-19 patients. *Informatics in Medicine Unlocked* 30:100919 DOI 10.1016/j.imu.2022.100919.
- Silveira MVF, Petri CA, Broggio IS, Chagas GO, Macul MS, Leite C, Ferrari EMM, Amim CGV, Freitas ALR, Motta AZV. 2020. Drivers of fire anomalies in

- the Brazilian Amazon: lessons learned from the 2019 fire crisis. *Land* **9**:516  
DOI 10.3390/land9120516.
- Soares Pinheiro FGM, Santana Santos E, Barreto IDC, Weiss C, Vaez AC, Oliveira JC, Melo MS, Silva FA. 2020.** Mortality predictors and associated factors in patients in the intensive care unit: a cross-sectional study. *Critical Care Research and Practice* **2020**:1483827 DOI 10.1155/2020/1483827.
- Taylor EH, Marson EJ, Elhadi M, Macleod KDM, Yu YC, Davids R, Boden R, Overmeyer RC, Ramakrishnan R, Thomson DA, Coetzee J, Biccard BM. 2021.** Factors associated with mortality in patients with COVID-19 admitted to intensive care: a systematic review and meta-analysis. *Anaesthesia* **76**:1224–1232  
DOI 10.1111/anae.15532.
- Venkatasubramaniam A, Wolfson J, Mitchell N, Barnes T, JaKa M, French S. 2017.** Decision trees in epidemiological research. *Emerging Themes in Epidemiology* **14**:11  
DOI 10.1186/s12982-017-0064-4.
- Wang Y, Lu X, Li Y, Chen H, Chen T, Su N, Huang F, Zhou J, Zhang B, Yan F, Wang J. 2020.** Clinical course and outcomes of 344 intensive care patients with COVID-19. *American Journal of Respiratory and Critical Care Medicine* **201**:1430–1434  
DOI 10.1164/rccm.202003-0736LE.
- Wendel Garcia PD, Fumeaux T, Guerci P, Heuberger DM, Montomoli J, Roche-Campo F, Schuepbach RA, Hilty MP. 2020.** Prognostic factors associated with mortality risk and disease progression in 639 critically ill patients with COVID-19 in Europe: initial report of the international RISC-19-ICU prospective observational cohort. *EClinicalMedicine* **25**:100449 DOI 10.1016/j.eclinm.2020.100449.
- World Health Organization (WHO). 2023.** WHO COVID-19 Dashboard. Available at <https://covid19.who.int> (accessed on 19 January 2023).
- Wu Z, McGoogan JM. 2020.** Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72,314 cases from the chinese center for disease control and prevention. *Jama* **323**:1239–1242 DOI 10.1001/jama.2020.2648.
- Yang J, Soltan AAS, Clifton DA. 2022.** Machine learning generalizability across health-care settings: insights from multi-site COVID-19 screening. *NPJ Digital Medicine* **5**:69 DOI 10.1038/s41746-022-00614-9.
- Zakariaee SS, Naderi N, Ebrahimi M, Kazemi-Arpanahi H. 2023.** Comparing machine learning algorithms to predict COVID-19 mortality using a dataset including chest computed tomography severity score data. *Scientific Reports* **13**:11343  
DOI 10.1038/s41598-023-38133-6.
- Zampieri FG, Soares M, Borges LP, Salluh JIF, Ranzani OT. 2017.** The Epimed Monitor ICU Database<sup>®</sup>: a cloud-based national registry for adult intensive care unit patients in Brazil. *Revista Brasileira de Terapia Intensiva* **29**:418–426  
DOI 10.5935/0103-507x.20170062.
- Zapata RD, Huang S, Morris E, Wang C, Harle C, Magoc T, Mardini M, Loftus T, Modave F. 2023.** Machine learning-based prediction models for home discharge

in patients with COVID-19: development and evaluation using electronic health records. *PLOS ONE* **18**:e0292888 DOI [10.1371/journal.pone.0292888](https://doi.org/10.1371/journal.pone.0292888).

**Zawadzki Z, Kosinski M. 2019.** FSelectorRcpp: 'Rcpp' implementation of 'FSelector' entropy-based feature selection algorithms with a sparse matrix support. Available at <https://cran.r-project.org/web/packages/FSelectorRcpp/>.

**Zeiser FA, Donida B, da Costa CA, Ramos GO, Scherer JN, Barcellos NT, Alegretti AP, Ikeda MLR, Müller A, Bohn HC, Santos I, Boni L, Antunes RS, Righi RDR, Rigo SJ. 2022.** First and second COVID-19 waves in Brazil: a cross-sectional study of patients' characteristics related to hospitalization and in-hospital mortality. *The Lancet Regional Health* **6**:100107 DOI [10.1016/j.lana.2021.100107](https://doi.org/10.1016/j.lana.2021.100107).