

69252_Auto_Edited-check.docx

1

Name of Journal: *World Journal of Critical Care Medicine*

Manuscript NO: 69252

Manuscript Type: ORIGINAL ARTICLE

Retrospective Study

Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: A machine learning approach

Caires Silveira E *et al.* Prediction of mortality in ICU patients

Abstract

BACKGROUND

Intensive care unit (ICU) patients demand continuous monitoring of several clinical and laboratory parameters that directly influence their medical progress and the staff's decision-making. Those data are vital in the assistance of these patients, being already used by several scoring systems. In this context, machine learning approaches have been used for medical predictions based on clinical data, which includes patient outcomes.

AIM

To develop a binary classifier for the outcome of death in ICU patients based on clinical and laboratory parameters. Methods: A set formed by 1087 instances and 50 variables from ICU patients admitted to the emergency department was obtained in the “WiDS (Women in Data Science) Datathon 2020: ICU Mortality Prediction” dataset.

METHODS

For categorical variables, frequencies and risk ratios were calculated. For numerical variables, were computed means and standard deviations and performed Mann-Whitney *U* test. We then divided the data into a training (80%) and test (20%) set, and the first one was used to train a predictive model based on the Random Forest algorithm. The test set was used to evaluate the predictive effectiveness of the model.

RESULTS

A statistically significant association was identified between submission to intubation, as well predominant systemic cardiovascular and hospital death. A number of the numerical variables analyzed (for instance Glasgow Coma Score punctuations, mean arterial pressure, temperature, pH, and lactate, creatinine, albumin and bilirubin values) were also significantly associated with death outcome. The proposed binary Random Forest classifier obtained on the test set ($n = 218$) accuracy of 80.28%, sensitivity of

81.82%, specificity of 79.43%, positive predictive value of 73.26%, negative predictive value of 84.85%, F1 score of 0.74, and area under the curve score of 0.85. The predictive variables of the greatest importance were the maximum and minimum lactate values, adding up to a predictive importance of 15.54%.

CONCLUSION

We demonstrated the efficacy of a Random Forest machine learning algorithm for handling clinical and laboratory data from patients under intensive monitoring. Therefore, we endorse the emerging notion that machine learning has great potential to provide us support to critically question existing methodologies, allowing improvements that reduce mortality.

Key Words: Hospital mortality; Machine learning; Patient outcome assessment; Routinely collected health data; Intensive care units; Critical care outcomes

Caires Silveira E, Mattos Pretti S, Santos BA, Santos Corrêa CF, Madureira Silva L, Freire de Melo F. Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: A machine learning approach. *World J Crit Care Med* 2022; In press

Core Tip: Considering the critical nature of patients admitted to intensive care units (ICUs), this study seeks to analyze clinical and laboratory data using a machine learning model based on a Random Forest algorithm. Consequently, we developed a binary classifier that forecasts death outcome, achieving a relevant area under the curve value of 0.85 and also identifying the variables that contributed the most to the prediction. With this, we aim to contribute to the improvement and methodological advancement in the development of clinically relevant machine learning tools, seeking to make medical practice decisions more accurate and reduce mortality in ICU patients.

INTRODUCTION

The intensive care unit (ICU) is the section of the hospital responsible for monitoring acute patients, and it relies on specialized multidisciplinary staff and high-technology equipment to ensure the best support for these patients, who are usually unstable and at high risk of death. These patients demand continuous monitoring of the most diverse clinical and laboratory parameters that directly influence their medical progress and the staff's decision-making. Lactate levels obtained from arterial blood samples, for example, may indicate the levels of tissue hypoxia and severity^[1]. The elevation in serum lactate levels (hyperlactatemia) is associated with increased mortality^[2,3]. Another important parameter in critically ill patients is the prothrombin time expressed in international normalized ratio (INR), which reveals abnormalities in the coagulation status^[4]. This parameter is also associated with an increased mortality when at altered levels. Besides these, many other laboratory and clinical data like temperature, oxygen and carbon dioxide pressure, systolic and diastolic pressure, motor, ocular, and verbal responses, among others, require team supervision since they are all related in some way to the severity of these ill patients^[5].

These data are so vital in the assistance of these patients that they are already used by several scoring systems, including the Acute Physiology and Chronic Health Evaluation (APACHE) and the Simplified Acute Physiology Score (SAPS), which are designed to assess and predict the patient's prognosis and allow for appropriate interventions^[6]. The APACHE score, for example, which has been widely used since its creation in the 1980s and has been undergoing updates ever since, relies on the use of parameters evaluated in three major groups: Demographic characteristics, comorbidities, and physiological measures. From these data, numerical weights are assigned to each one and then summed to assign a severity classification and predict outcomes^[7].

Machine learning may be understood as a scientific discipline by which a computer system is enabled to cross-reference numerous data in order to build statistical prediction models through pattern recognition^[8]. To reach this pattern perception capability, it is essential during the use of the supervised machine learning approach to

separate the data subsets for training and for testing. The training data are presented to the algorithm in order to create the model, and then the test data is also presented after the creation of the model in order to simulate this model prediction and evaluate its performance. The machine learning approach is already used for medical predictions based on clinical data, which includes patient outcome. Heo *et al*^[9] used it to predict the long-term outcome of patients who suffered an ischemic stroke. In another study, Lynch *et al*^[10] sought a survival prediction of lung cancer patients using machine learning by providing a series of patient data such as age, tumor size, type of intervention, and more.

The use of machine learning has been consolidated as an alternative for the development of predictive models of mortality in the critical care setting. An example is the retrospective study by Liu *et al*^[11], who developed a logistic model of the death risk grade in patients with pulmonary tuberculosis using data from patients admitted to ICUs in three hospitals. In this multivariate analysis study, where the sensitivity was 83.3% and specificity was 73.1%, the Apache II score, C-reactive protein levels, albumin levels and pressure of oxygen in arterial blood (PaO₂) were considered the main factors influencing the outcome. However, a registered limitation was the small dataset utilized.

The limiting matter caused by the database used in machine learning predictive models was also observed in the study by Hou *et al*^[12], who developed a model regarding 30-d mortality in patients who fit the Third International Consensus Definitions for Sepsis (Sepsis-3). This paper used a public database **Medical Information Mart for Intensive Care III (MIMIC III)** from **a single-center critical care database**. Another study that also relates the development of a predictive machine learning model in the context of patients with sepsis is the one proposed by Nemati *et al*^[13] that in addition to using the aforementioned MIMIC III also relied on ICU admission data from two hospital centers. In this study, as well as in the two previously mentioned, it is observed the potentiality of the use of this tool in the early identification of severity of

cases and the possibility of making fundamental decisions to the positive outcome for patients.

In addition, more recently, in light of the advent of the severe acute respiratory syndrome coronavirus 2 pandemic, the application of these predictive models using machine learning technology have been employed on various grounds such as for risk of critical coronavirus disease 2019 (COVID-19)^[14], need for ICU transfer and the prognosis of intensive care COVID-19 patients^[15,16]. The latter one associated to eight main component factors, namely: ³ Lymphocyte percentage, prothrombin time, lactate dehydrogenase, total bilirubin, eosinophil percentage, creatinine and neutrophil percentage. And although it also emphasized the difficulties of small databases, they pointed out the significance of this approach in critical patients with a panel of such complicated parameters.

Understanding a clinical setting as complex and full of variables as the ICU, identifying existing patterns, and enabling outcome prediction is a valuable tool for the improvement of health assistance to these patients. Therefore, the aim of the current paper is to develop a predictive model for the outcome of death in ICU patients based on clinical and laboratory parameters using a binary classifier, with predicted outcome consisting of in-hospital death and discharge.

MATERIALS AND METHODS

Data acquisition

We used anonymized retrospective data from ICU patients admitted to the emergency department to build a predictive model geared towards predicting death outcomes in these patients. For this purpose, a dataset to be used in the study was created from the larger “WiDS (Women in Data Science) Datathon 2020: ICU Mortality Prediction” dataset^[17], which presents clinical and laboratory data pertaining to the first 24 h of ICU patient admission. The criteria for inclusion of instances (*i.e.*, patients) in the study dataset were: (1) ICU admission and emergency department admission; and (2) Completeness (*i.e.*, absence of missing data) with respect to the variables of interest.

Since all the data were obtained from a public and anonymized dataset^[16], it was not necessary to submit this study to the ethics committee, being in accordance with all the established precepts by the Committee on Publication Ethics.

Data preprocessing and exploratory data analysis

Aligned with the goal of building an interpretable predictive model from clinical and laboratory data, variables related to the clinical status of patients (such as vital signs, clinical score scores, blood count and biochemical test results) were prioritized in the definition of variables of interest - with exclusion of variables of this type only when redundant or when they represented the application of formulas instead of measured or scored values - to the detriment of anthropometric and demographic variables, with age being the only representative of this group of variables included. Additionally, were also not included among the variables of interest those referring to logistical aspects of hospitalization (such as source and type of admission and readmission status).

This way, a set formed by 1087 instances and 50 variables was obtained, of which 49 were assumed as predictive variables and 1 as predicted variable (outcome variable). The predictive numerical variables were: (1) Age; (2) Disease score; (3) Eye opening score on the Glasgow coma scale (GCS); (4) Heart rate; (5) Hematocrit; (6) Mean arterial pressure; (7) Maximum albumin; (8) Maximum bilirubin; (9) Maximum blood urea nitrogen; (10) Maximum calcium; (11) Maximum creatinine; (12) Maximum diastolic blood pressure; (13) Maximum glucose; (14) Maximum HCO_3 ; (15) Maximum hemoglobin; (16) Maximum INR; (17) Maximum lactate; (18) Maximum platelets; (19) Maximum potassium; (20) Maximum sodium; (21) Minimum systolic blood pressure; (22) Maximum saturation of peripheral oxygen (SpO_2); (23) Maximum white blood cells (WBC); (24) Minimum albumin; (25) Minimum bilirubin; (26) Maximum blood urea nitrogen; (27) Minimum calcium; (28) Minimum creatinine; (29) Minimum diastolic blood pressure; (30) Minimum glucose; (31) Minimum HCO_3 ; (32) Minimum hemoglobin; (33) Minimum INR; (34) Minimum lactate; (35) Minimum platelets; (36) Minimum potassium; (37) Minimum sodium; (38) Minimum systolic blood pressure;

(39) Minimum SpO₂; (40) Minimum WBC; (41) Motor response on the GCS; (42) Partial PaO₂; (43) Partial pressure of carbonic gas in arterial blood (PaCO₂); (44) pH; (45) Respiratory rate; (46) Temperature; and (47) Verbal response on the GCS. The predictive categorical variables were: (1) Need for intubation or not; and (2) Predominant systemic involvement. The outcome variable was the evolution or not with hospital death.

The disease score corresponded to the number of diseases present among the following conditions: (1) Acquired immunodeficiency syndrome; (2) Cirrhosis; (3) Diabetes; (4) Hepatic failure; (5) Immunosuppression; (6) Leukemia; (7) Lymphoma; and (8) Solid tumor. The categories of predominant systemic involvement considered were: (1) Cardiovascular involvement; (2) Gastrointestinal involvement; (3) Genitourinary involvement; (4) Hematological involvement; (5) Metabolic involvement; (6) Musculoskeletal/skin involvement; (7) Neurological involvement; (8) Respiratory involvement; (9) Sepsis; and (10) Trauma.

Initially, a descriptive and comparative analysis of the data was performed. The data were categorized according to the outcome variable. After that, the occurrence frequencies of each category for categorical predictive variables and the means and standard deviations for all numerical predictive variables in both groups were computed. Finally, the differences for each variable between the groups were analyzed using the χ^2 test for risk ratios (for categorical variables) and the Mann-Whitney U test (for numerical variables). Since a decision tree ensemble algorithm was chosen to constitute our predictive model, it was not necessary to normalize or standardize the data, since tree partitioning algorithms are insensitive to scaling.

Machine learning algorithm selection

To perform our predictive analysis, we chose to build a Random Forest algorithm, a model consisting of an ensemble of randomized decision trees. As an extension of bootstrap aggregation (bagging) of decision trees, in Random Forest algorithms each individual model in the ensemble is employed to generate a prediction for a new sample, and these individual models predictions are averaged to give the forest's

prediction, resulting in better performance than any single tree. By combining individual models, the ensemble model tends to be more flexible and efficient. Accordingly, random forests have been incredibly successful in a variety of classification and regression problems with clinical applications. Furthermore, the algorithm does not require any feature scaling since decision trees predictions are partitioning-based instead of distance-based.

Model training and evaluation

We then proceeded to the development of the predictive model for the outcome variable. The data were divided into a training set (80%) and a test set (20%). The training set was used to train a predictive model based on the Random Forest algorithm^[18], implemented here through the Scikit-learn open source library^[19]. The test set was used to evaluate the predictive effectiveness of the model. The metrics used for such evaluation were accuracy, sensitivity, specificity, area under the curve (AUC) score, positive predictive value, and negative predictive value. The adopted methodology is schematically summarized in Figure 1. Besides the predictive performance, the feature importance attributed by the model to each variable was also considered, which not only adds explainability to the model, but also potentially provides insights regarding the evaluation of critically ill patients and the factors associated with higher mortality in this clinical setting. All steps of statistical analysis and development of the predictive model were performed in Python (version 3.6.9), using SciPy and Scikit-learn libraries.

RESULTS

Data from 1087 ICU patients were analyzed and used in the construction of the predictive model, of which 388 evolved with hospital death, while the remaining 699 did not. With regard to the predictive variables categories - need or not of intubation and predominantly affected body system -, among the 388 patients who evolved with hospital death: 275 were intubated and 63 were not; 106 had sepsis as predominant

systemic involvement, 18 respiratory involvement, 4 metabolic involvement, 154 cardiovascular involvement, 11 trauma, 16 neurological involvement, 25 gastrointestinal involvement, 2 genitourinary involvement, 1 musculoskeletal/skin involvement and 1 hematological involvement. Among the 699 patients who did not progress to hospital death: 534 were intubated and 215 were not; 206 had sepsis as predominant systemic involvement, 107 respiratory involvement, 79 metabolic involvement, 167 cardiovascular involvement, 38 trauma, 49 neurological involvement, 74 gastrointestinal involvement, 17 genitourinary involvement, 9 musculoskeletal/skin involvement and 3 hematological involvement. A statistically significant association was identified between submission to intubation and hospital death (risk ratio = 1.5, $\chi^2 = 11.87$, $P < 0.001$), as well as between the predominant systemic cardiovascular involvement and hospital death compared to the musculoskeletal system/skin, related to lower rate of hospital death (risk ratio = 4.80, $\chi^2 = 4.20$, $P = 0.04$). With regard to numerical predictive variables, their mean \pm SD, and the respective comparison between both outcome groups (performed using the Mann-Whitney *U* test) are shown in Table 1.

The search for the best hyperparameters in our Random Forest model training was done using randomized search. In this way, 100 random combinations of hyperparameters were tested. Each combination was iterated 6 times, as a 6-fold validation scheme was adopted. In this scheme, the training set ($n = 869$) was split in 6 parts, and in each iteration a different part was used for validation. Ultimately, during training we performed 600 fits, obtaining the following hyperparameters: (1) Number of estimators = 213; (2) Maximum depth = 23; (3) Maximum leaf nodes = 24; (4) Minimum samples split = 5; (5) Class weights = 3.9; and (6) Bootstrap = true.

The model obtained accuracy of 80.28%, sensitivity of 81.82%, specificity of 79.43%, positive predictive value of 73.26%, negative predictive value of 84.85%, F1 score of 0.74, and AUC score of 0.85 on the test set ($n = 218$). The confusion matrix for the model is shown in Figure 2, and its receiver operating characteristic (ROC) curve is shown in Figure 3. The predictive variables with the greatest importance were the maximum and

minimum lactate values, adding up to a predictive importance of 15.54%, followed by temperature (6.47%), motor punctuation in GCS (5.25%), maximum blood urea nitrogen (4.35%) and minimum WBC (3.31%). The percentage importances of the other variables in the prediction are listed in Table 2.

DISCUSSION

The presented predictive model, a Random Forest binary classifier, was able to predict in the test set the occurrence or not of hospital death with an accuracy of 80.28%, sensitivity of 81.82%, and specificity of 79.43%. It is well established in the literature that this type of classifier is generally well suited for high-dimensional problems with highly correlated features (a frequent situation when it comes to medical data)^[20]. Our results are consistent with that, as they demonstrate the potential for using random forests to handle clinical and laboratory data from patients under intensive monitoring.

The ICU mortality is high, and the patients require interventions that are cost-effective in order to avoid mortality without inputting unnecessary costs or demand to the medical team. Mortality prediction models work with the objective to assess the severity of the patients so that based on its findings, the treatment needed can be directed. The analysis presented in this study works in the same way; if we identify those patients that have major mortality rates, faster and better care can be provided in order to prevent the worse outcome^[21]. For this purpose a variety of assessment scores already exist, like APACHE, SAPS or Mortality Probability Model (MPM). The ROC value of our model (0.85) was comparable with some of these highly used models, like 0.836 for APACHE II, or 0.826 for SAPS II^[22], which showcase the good results obtained.

Furthermore, the machine learning approach to predict mortality in ICU patients has been documented. For example, Veith and Steele^[23] developed a LazyKStar model to predict mortality in ICU patients at the time of hospital admission, obtaining a 10-fold validation AUC value of 0.75. A recurrent neural network inputted with 44 clinical and laboratory features from the first 24 h of ICU patient admission proposed by Thorsen-Meyer *et al*^[24] achieved an AUC of 0.82. The extreme gradient boosted trees classifier

developed by Chia *et al*^[25] reached an AUC of 0.83 using 42 predictive variables. The formats and results of these last two studies are comparable to ours, since we reached an AUC of 0.85 using a random forest fed by 50 features.

Due to the COVID-19 pandemic, there was a great growth of publications focused on machine learning models for predicting ICU mortality in a disease-specific manner, such as those by Pan *et al*^[16], Lichtner *et al*^[26], and Subudhi *et al*^[27]. Meanwhile, many of the previous studies in this field also focus on predicting ICU outcomes for specific diseases or morbid conditions, like sepsis or death from pulmonary tuberculosis^[11,13,28], which lead to an assessment of parameters specific for the disease studied, somewhat restricting the research. Many of the renowned models and scales for ICU mortality prediction demand a series of measurements to make their use possible, but not always all the data required are available. In this sense, it is important to understand what are the main variables involved related to the outcome of interest (and its prediction), so that they can be closely monitored. In our study, lactate level proved to be the most influential one, which is in accordance with its physiological role that indicates poor oxygenation, anaerobic metabolism, acidosis and muscle fatigue, involved in a systemic response of the organism, corroborating the finds by Bou Chebl *et al*^[29], Villar *et al*^[30] and Vincent *et al*^[2]. Despite its predictive importance found in our study (15.54%), lactate is not a variable of most scores used, not being part of APACHE, SAPS or MPM.

Temperature was the second variable that influenced the most the outcome prediction, and is part of APACHE and SAPS, its variation (hyper or hypothermia) is related with a loss of control of body homeostasis, the mean valor for death outcome was 35.2 ± 1.9 , while we have an increase of nearly 1 point in the mean value for the survival outcome, these could represent that an increase of the temperature or even fever could be a positive body response, indicating an immune system attempt to fight the pathology^[31,32].

The third variable of major impact is the motor GCS punctuation, which is part of GCS, a widely known scale for neurologic damage used in hospital admissions as well as assessment models^[33], this motor element has a specific field only in APACHE IV.

Lower punctuations in GCS are related with greater neurologic damage, with 3 and 1 as its bottom punctuation for the global and motor scale respectively, the mean of 2.9 ± 2.2 for the death outcome in contrast with the value of 4.3 ± 2.0 for the survival mean demonstrate a considerable difference between those patients, since the greatest value possible for the motor component is 6. The stratification of the data based on its predictive value is a great contribution since the variables above discussed account for approximately 27% of the result, while the other 45 for the remaining 73%, indicating that continuous monitoring of them may be of great value. Considering their importance, a detailed survey with either a dataset with per hour measurement of parameters or the data separated by ICU type could lead to more specific approaches for the medical staff.

Despite the good results found, this study faces as its main limitation the incompleteness of the original dataset for many instances regarding important clinical and laboratory variables, which lead to the use of a relatively small quantity of instances to train the predictive model. Since machine learning algorithms are essentially data-driven, a larger amount of data could lead to greater accuracy and a wider generalizability of the model, thus being useful for additional testing and refinement. Another potential limitation is related to the clinically broad nature of the variables analyzed, since the purpose was to study the possible parameters available in the ICU, which contrasts with research focused on the outcomes for a specific disease and, therefore, fed with more specific variables in regard to the considered pathophysiological process.

Although the use of a wide range of clinical and laboratory parameters was critical for our purpose of assessing the predictive significance of the variables - in the context of building a model that is not only explainable but also clinically interpretable - this factor may restrict the possibilities of potential datasets to be used to ascertain the reproducibility of the findings, since some parameters may be unavailable. However, since these are variables commonly evaluated in critically ill patients in the ICU, for whom the prognostic evaluation of mortality is more important (in view of their higher

mortality rates), we believe that this should not be a limiting factor to the clinical applicability of the proposed model.

CONCLUSION

In the study, it was possible to develop a reliable model for predicting mortality in the ICU, in which the influence of lactate level stands out as the main variable involved in the outcome prediction, followed by temperature and motor GCS. What can be perceived through the research is that machine learning comes to contribute and to make medical practice more efficient, as it allows faster analysis that otherwise would be complex and time-consuming. More than that, it also allows us to critically question existing parameters and methodologies through the results it provides, in order to allow improvements that reduce the mortality of patients and are time and cost-effective. This study also highlights the importance of complete and organized registers of ICU patient data, in order to enable the development of predictive models towards prevention and prediction of in-hospital bad outcomes.

ARTICLE HIGHLIGHTS

Research background

The monitoring of clinical and laboratory parameters of patients in the intensive care unit (ICU) is an extremely important part of the routine of intensive care staff. Additionally, several scores already utilize these parameters to guide the assistance of these patients. In the meantime, the advance of technological resources such as the machine learning approach allows the development of predictive models capable of being applied to medical practice.

Research motivation

Mortality in ICU is something that worries and drives the search for alternatives that can help the team in directing treatment to avoid this negative outcome. Therefore, a

predictive model that uses the patient's parameters can precisely influence this treatment guidance, improving the cost-effectiveness quickly and safely.

Research objectives

The objective of our study is the development of a binary classifier predictive model between the outcomes of death and non-death in ICU patients. This paper demonstrates the potency of emerging technological realities within the medical field and how it is possible to harness them to improve healthcare practices.

Research methods

Initially, we obtained a set of 1087 instances and 50 variables related to patients admitted to an ICU by using a public database. We calculated frequency and risk rate for categorical variables and means, standard deviations and the Mann-Whitney *U* test for numerical variables. Afterwards, we divided the data for the application in training of the predictive model based on the Random Forest algorithm and then to test the effectiveness of the model.

Research results

Among the 50 variables associated with death outcome, the maximum and minimum lactate values were the most important predictors (15.54%) followed by temperature(6.47%) and motor Glasgow coma scale punctuation(5.25%). The Random Forest binary classifier predictive model (death and no death) showed accuracy of 80.28%, sensitivity of 81.82%, specificity of 79.43%, positive predictive value of 73.26%, negative predictive value of 84.85%, F1 score of 0.74, and area under the curve score of 0.85.

Research conclusions

This study demonstrated the development of a predictive model with high accuracy, sensitivity and specificity for ICU patients by applying a machine learning approach - the Random Forest algorithm - to clinical and laboratory data.

Research perspectives

The proper registration of patient parameters as well as the availability of more and larger databases and even further development of digital tools can enhance machine learning approaches, enabling the refinement of predictive models as well as patient care.

ORIGINALITY REPORT

2%

SIMILARITY INDEX

PRIMARY SOURCES

1	f6publishing.blob.core.windows.net Internet	17 words — < 1%
2	jhsmr.org Internet	16 words — < 1%
3	www.ncbi.nlm.nih.gov Internet	14 words — < 1%
4	www.frontiersin.org Internet	13 words — < 1%
5	Elena Caires Silveira, Caio Fellipe Santos Corrêa, Leonardo Madureira Silva, Bruna Almeida Santos et al. "Recognition of esophagitis in endoscopic images using transfer learning", World Journal of Gastrointestinal Endoscopy, 2022 Crossref	12 words — < 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 12 WORDS

EXCLUDE MATCHES < 12 WORDS