Dear Reviewer,

Thank you very much for your valuable and insighful comments.

The state-of-the-art of machine learning studies for mortality prediction have been incorporated into both the introduction and the discussion of the article – with more up-to-date references added. The excerpts referring to these additions follow:

> In introduction:
>
> The use of the machine learning has been consolidating as an alternative for the development of predictive models of mortality in the critical care setting. For instance, 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 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-day 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 predictive model by machine learning 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 Sars-CoV-2 pandemic, the application of these predictive models using machine learning technology have been employed on various grounds such as for risk of critical covid-19[14], need for ICU transfer[15] and the prognosis of intensive care covid-19 patients[16]. The latter one associated eight main component factors, namely: 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.

In discussion:

Furthermore, the machine learning approach to predict mortality in ICU patients has been documented. For example, Veith *et al.*[23] developed a LazyKStar model to predict mortality and ICU patients at 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 hours of ICU patient admission proposed by Thorsen-Meyer *et al.*[24] achieved an AUC of 0.82. The extreme gradient boosted trees classifier proposed 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 outcome for specific diseases or morbid conditions, like sepsis[13,28] or death from pulmonary tuberculosis[11], which lead to an assessment of parameters specific for the disease studied, somewhat restricting the research.

In **MATERIALS AND METHODS**, a topic was added referring to data acquisition ("**Data acquisition**") and a topic referring to data preprocessing and exploratory data analysis ("**Data preprocessing and exploratory data analysis**"), and the results of this analysis are reported in table 1. A topic describing the selected algorithm ("**Machine learning algorithm selection**"), Random Forest – which is a decision tree ensemble –, and a topic describing the process of training and evaluation of the proposed model ("**Model training and evaluation**"), were also added.

The configuration of the hyperparameters is presented in the second paragraph of the **RESULTS**, presented below:

> The search for the best hyperparameters in our random forest model training was done using random 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 splitted in 6 parts, ein each iteration a different part was used for validation. Ultimately, during training we performed 600 fits, obtaining the following hyperparameters: (i) number of estimators = 213; (ii) maximum depth = 23; (iii) maximum leaf nodes = 24; (iv) minimum samples split = 5; (v) class weights = 3.9; (vi) bootstrap = True.

The discussion paragraph regarding the limitations of the paper was also expanded, resulting in the following final version (consisting of two paragraphs):

Despite the good results found, this study faces as its main limitation the incompletude of the original dataset for many instances regarding important clinical and laboratorial variables, which lead to the use of a relatively small quantity of data 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 in 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 important 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.

Re-reviewer

Comments: The paper can be accepted in current form.

Reply: Thanks for your comments.

Once again, thank you very much for your contribution.

Best regards.