

Statement of Changes and Answers to Comments

We would like to thank the editor and the reviewers for their careful consideration of our paper and for the valuable suggestions on how to improve it. We are also grateful for showing interest and appreciation for our work. We do think that this revised version of the manuscript addresses all the remaining criticisms raised by the editor. Together with the revised version of the manuscript, we also provide a copy of it with all changes highlighted in blue colour to ease the reviewing process. Please find below a Statement of Changes and detailed Answers to the comments.

STATEMENT OF CHANGES

Prompted by the reviewer's suggestions, the revised manuscript includes the following major changes:

- 1- The related works section is improved by adding an introductory paragraph and discussing methods on extracting physiological signals.
- 2- Details of the ML model parameters are described and the architectural diagram of the LSTM model is shown in the supplementary material document.
- 3- Additional figures representing the relative contribution of each feature input in final prediction of SVM, Gaussian Process and XGBoost models are added in the supplementary material document. These figures are the average of feature importance scores represented in Figure 4 along all the patients and they offer a clearer representation of the explainability of the models.
- 4- A new table (Table 5 in the manuscript) is added showing the highest contributing features in the prediction outcome of the trained ML models among each feature set.
- 5- ROC and PR curves for the LSTM models using all the feature sets are added in the supplementary material document.
- 6- The conclusion section is improved by discussing the main highlights of the experiments to support our conclusions and presenting the main challenges of the study.

Response to Reviewer 3 Comments

Point 1: why authors have used ML methods? an explicit explanation is required.

Response 1:

Many thanks for your comment. The main aim of our study is to design an automatic pipeline for the most accurate prediction of the mortality of severe sepsis patients admitted to the hospital using wearable sensors. Although, statistical models are very powerful tools to obtain inferences from the data and to model the relationships between different variables, ML models offer a suitable approach to make the most accurate predictions possible [1].

Moreover, with recent advances in the interpretability algorithms designed for ML models [2], we are able to obtain explanations for their prediction which particularly makes their application more attractive in clinical domain.

[1] Ghahramani, Z., 2015. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553), pp.452-459.

[2] Molnar, C., 2020. Interpretable machine learning. Lulu. com.

Point 2: challenges of this work should be reported in the conclusion section.

Response 2:

Many thanks for your comment. Following the reviewer's suggestion, we have now added the following sentences highlighting our study challenges in the conclusion section.

“The main challenges of our study are the relatively small sample size of the patients and the lack of external validation in the clinical setting “.

Other main challenges are discussed in the Discussion section:

“We reported our results within different cross-validation schemes suitable for each ML model to avoid the over-fitting of the models' due to the limitations of having a larger small sample size in real life LMIC settings. Although, the generalizability of the results can definitely be improved by increasing the sample size of the patient cohort in our study, to the best of our knowledge, this is the first prospective study in an LMIC settings to collect a relatively large dataset of high resolution ECG signals for long time duration (24 hours) from severely ill sepsis patients, and from this data proposes ML based prediction algorithms for a better care management of these patients.

Another study [1] with a similar data collection protocol of sepsis patients in Rwanda performed statistical analysis with data from 43 patients. This attests to the fact that data collection for research purposes from critically ill patients from a critical unit for research purposes is a very cumbersome task particularly in low resource settings.

Another limitation of our study is the imbalanced nature of our dataset (35% for the positive class) which limits the reliance on many performance metrics in supervised learning. However, the task of mortality prediction in clinical practice is often performed with imbalanced datasets and usually with a 5-10% of mortality rate. We tried to overcome this limitation by relying on those performance metrics (e.g., AUCPRC) that are less affected by the low proportion of positive cases in the dataset.

Unlike many other studies, we have not applied extensive pre-processing methods on the ECG signals collected from the ePatch. This choice was firstly to reduce the processing computational costs as much as possible, and secondly to make our prediction pipeline

suitable for real life scenarios, rather than controlled lab conditions. Due to the very low computational cost of the training models used in this study, we believe there is a great potential to directly deploy these models in clinical practice. In fact, in our future studies we aim to implement the proposed decision systems in clinical practice for a more efficient way of monitoring critically ill sepsis patients in the HTD hospital. Moreover, to overcome the limitations caused by sample size, we aim to take advantage of similar datasets with a much larger sample size collected in high-income-settings.”

[1] Garbern, S.C., Mbanjumucyo, G., Umuhoza, C., Sharma, V.K., Mackey, J., Tang, O., Martin, K.D., Twagirumukiza, F.R., Rosman, S.L., McCall, N. and Wegerich, S.W., 2019. Validation of a wearable biosensor device for vital sign monitoring in septic emergency department patients in Rwanda. *Digital Health*, 5, p.2055207619879349.