2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Exploring Features Contributing to the Early Prediction of Sepsis Using Machine Learning

Esmaeil Shakeri[1], Emad A Mohammed[1], Zahra Shakeri H.A.[2], and Behrouz Far[1]

***Abstract*— The increasing availability of electronic health records and administrative data and the adoption of computer-based technologies in healthcare have significantly focused on medical informatics. Sepsis is a time-critical condition with high mortality, yet it is often not identified in a timely fashion. The early detection and diagnosis of sepsis can increase the likelihood of survival and improve long-term outcomes for patients. In this paper, we use SHapley Additive exPlanations (SHAP) analysis to explore the variables most highly associated with developing sepsis in patients and evaluating different supervised learning models for classification. To develop our predictive models, we used the data collected after the first and the fifth hour of admission and evaluated the contribution of different features to the prediction results for both time intervals. The results of our study show that, while there is a high level of missing data during the early stages of admission, this data can be effectively utilized for the early prediction of sepsis. We also found a high level of inconsistency between the contributing features at different stages of admission, which should be considered when developing machine learning models.**

I. INTRODUCTION

Sepsis is a life-threatening response to infection and a time-critical condition with a high mortality rate worldwide ($\sim$ 20% [1]). Yet, it is often not identified in a timely fashion [2]. It is the most common cause of admission to an intensive care unit (ICU) and the most common cause of death in the ICU [3]. Additionally, the risk of developing sepsis while staying in the ICU for any reason is incredibly high; in one study analyzing 170 post-surgical patients treated in the ICU, 83 (49%) developed sepsis within the 28-day monitoring period [4]. There are also significant long-term side effects associated with sepsis development even if the patient recovers and is discharged from the ICU [5]–[7]. While the markers of mortality risk post-sepsis development are relatively well-known [8], the more significant challenge and more clinically relevant problem are to determine how to detect patients before sepsis development. The importance and difficulty of accurate monitoring for this issue are significant factors in why Grand View Research, Inc. has estimated that the global sepsis diagnostics market is expected to be worth $USD 1.18 billion by 2027 [9].

Without timely intervention, sepsis can rapidly lead to poor outcomes. In the hospital setting, physicians utilize the qSOFA score (aka. quickSOFA) [10] to identify patients with suspected infection who are at a greater risk for a poor outcome outside the ICU. This score uses three criteria, altered mental status, fast respiratory rate, and low blood pressure. Given that qSOFA uses vital signs measurements and that patients are often in the ICU for hours before labs and blood-gas measurements are available, predicting the early diagnosis of sepsis using vital signs could be of particular utility in a clinical setting. In this paper, we looked at demographic and physiologic data from ICU patients who did or did not develop sepsis during their stay to determine if any key variables could be used for monitoring early warning signs of sepsis and hopefully prevent its occurrence. We took two different approaches: (1) classifying sepsis vs. non-sepsis patients at ICU admission during two different time intervals, and (2) exploring different features' contribution to these predictions. The results of our study show that, while there is a high level of missing data during the early stages of admission, this data can be effectively utilized for the early prediction of sepsis. We also found a high level of inconsistency between the contributing features at different stages of admission, which should be considered when developing machine learning models.

[1]Esmaeil Shakeri, Emad A Mohammed, and Behrouz Far are with the Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Canada. esmaeil.shakerihosse@ucalgary.ca, eamohamm@ucalgary.ca, far@ucalgary.ca

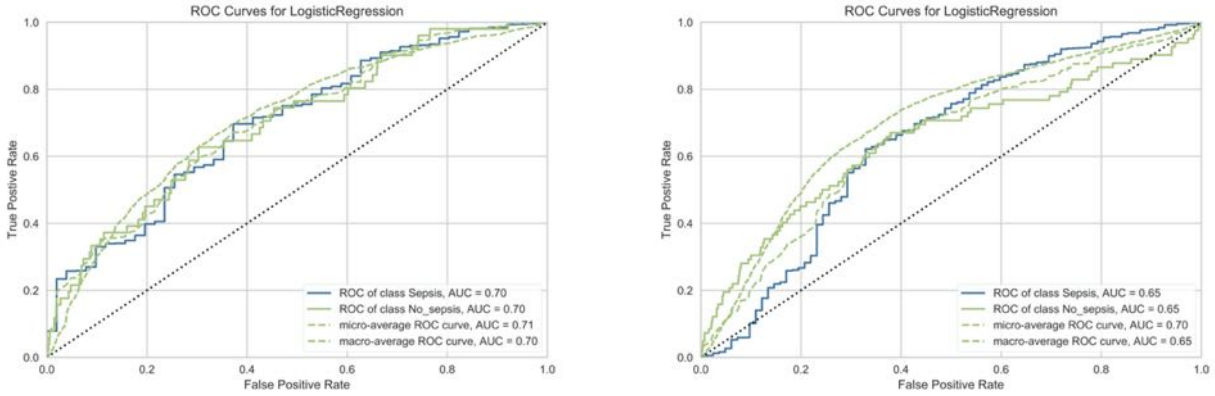
[2]Zahra Shakeri H.S with Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Canada. zshakeri@ucalgary.ca

II. METHODOLOGY

A. Dataset and Data Preparation

The Sepsis dataset is a labelled dataset provided by PhysioNet/Computing in Cardiology Challenge 2019 with 40 columns and 1,552,210 rows combined [11]. There are 43 variables in this dataset, consisting of vital signs and laboratory tests which are time-dependent. Age, gender, patient care unit, the time between hospital admit and ICU admit, and length of stay in ICU were the demographics data. There are 40,336 unique patient identifiers, which we assumed to represent the number of patients in the population. In the exploratory data analysis (EDA), a notable finding was the dataset is highly imbalanced. Records labelled as non-sepsis is approximately 93% of the total records. The volume of null values was also evident. Most of the vital signs had 10% to 15% nulls, except for 'Temp' and 'EtCo2' had 66% and 97%, respectively. In contrast, laboratory results had 94% to 99% nulls. This is expected as not all tests are done at every patient encounter. The population is proportionately distributed in gender and age groups. Patients in the ICU at the onset of their sepsis were mostly within the first 6 hours of ICU stay, which is 54% (1,574 patients).

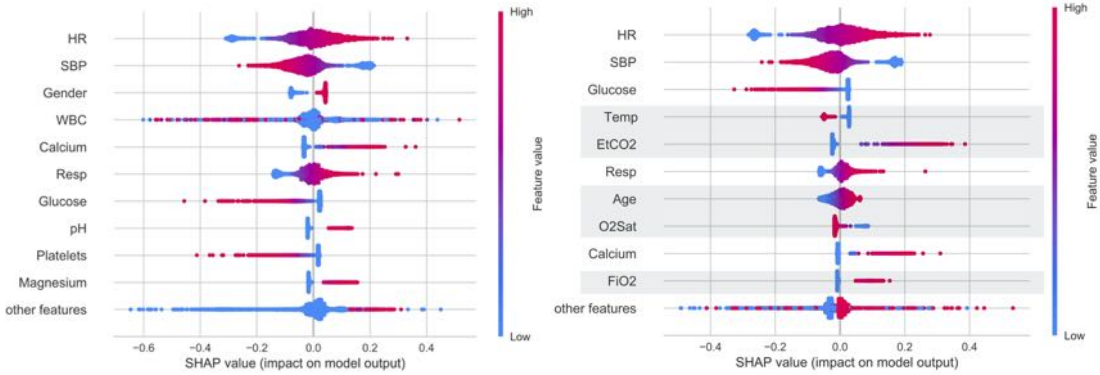
To address the imbalance problem, we used the Synthetic Minority Over-sampling TEchnique-Nominal Contin-



(a) Prediction results after hour 2

(b) Prediction results after hour 6

Fig. 1: Classification AUC score for the early prediction of sepsis, using logistic regression



(a) Global feature importance after Hour 2

(b) Global feature importance after Hour 6

Fig. 2: Global contribution of features to the early prediction of sepsis, using logistic regression.

uous (SMOTE-NC) [12] approach to oversample the minority classes by creating synthetic samples based on their feature-space. Moreover, given the informativeness of the presence/absence of predictor variables in clinical datasets, we incorporated missingness indicators into our models to address the problem of missing data [13].

B. Model Development and Interpretation

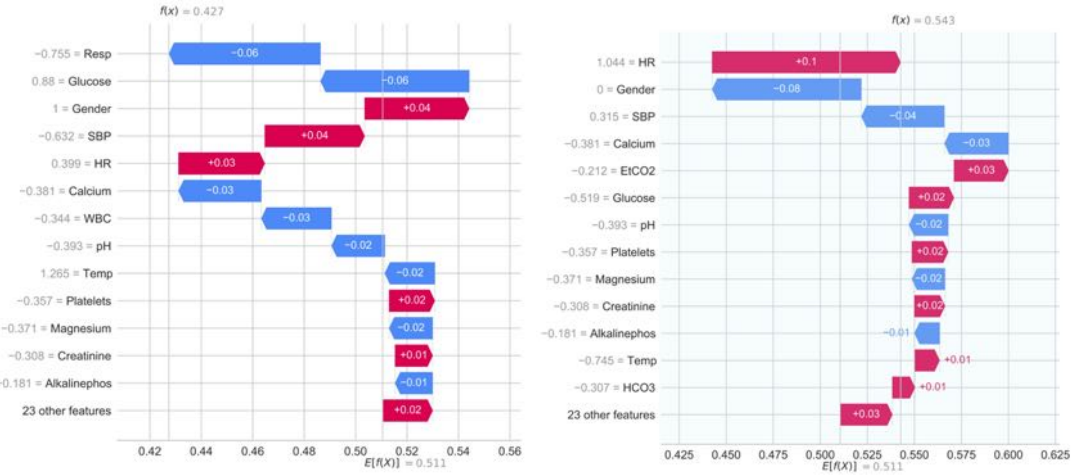
For a timely prediction of sepsis for the cohort of critically ill patients presented in our dataset, we developed a series of machine learning models, including logistic regression, K-nearest neighbour, AdaBoost, XGBoost, EtraTrees, and Random Forest. As these models cover different learning architectures (e.g. tree bases, distance-based, and linear), our learning process will not be biased towards a specific learning architecture.

Moreover, to interpret the prediction results and explore how much each feature contributed to the early prediction of sepsis, we used SHapley Additive exPlanations (SHAP) analysis [14], [15]. Despite the existing feature analysis techniques, which mainly compute the global feature importance, SHAP calculates the local feature importance for every dataset sample and assigns each feature an importance value for a particular prediction. Considering that the importance of a feature may not be consistent across all data points of a dataset, this approach can address the misinterpretations associated with the inconsistency problems in other feature importance techniques.

III. Results

Of the patients who did not arrive in the ICU with sepsis, 1,587 (7.88%) developed sepsis at least one hour into their stay, with 18,546 (92.12%) being released without ever developing sepsis. Among patients who developed sepsis, we found that sepsis occurred more frequently in male patients (e.g. 61% male and 39% female). Older adults ($\geq$ 60) accounted for 61%, with those between 70-80 years demonstrating the most significant risk. The average ICU length of stay for patients who developed sepsis was about 2.3 days (i.e. $\sim$ 56.3 hours). Most patients were diagnosed with sepsis within one day, but some patients stayed in ICU for as long as two weeks.

Among the implemented models in this study, logistic regression yielded the highest performance in early detection of sepsis, with the sensitivity of 66% and AUC of 70% at Hour 2, and sensitivity of 63% and AUC of 65% at



(a) The waterfall feature importance for random patient a with no-sepsis outcome

(b) The waterfall feature importance for random patient b with sepsis outcome

Fig. 3: Waterfall plots presenting the local contribution of features in sepsis prediction. Both of the charts presented in this figure are recorded after Hour 2 predictions.

Hour 6 (see Figures 1 a-b). Interestingly, results from the Hour 2 model contained fewer false negatives, representing a potential benefit of utilizing admission data for early prediction of sepsis.

A. *Influential Features*

Using the additive nature of Shapley values, we integrated all the local feature values for each data point and calculated the global contribution (I) of each feature. Considering $\phi_j^i \in \mathbb{R}$ as the shapely value of feature j for sample i, we can calculate the global importance of this feature as:

$$I_j = \sum_{i=1}^{n} \mid \phi_j^i \mid$$

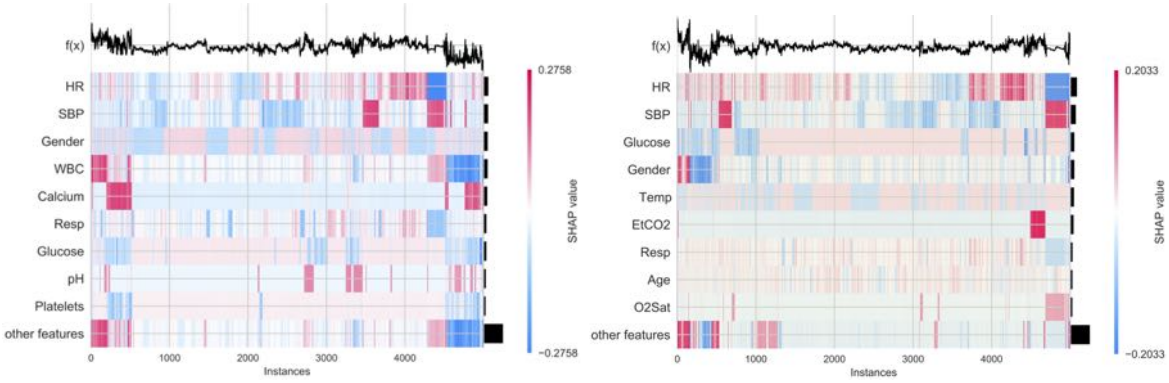
Figures 2a and b show the combination of feature importance (y-axis) and feature effects (coloured points) for the most influential features for the predictions after Hours 2 and 6, ordered based on their importance. Looking at Figure 2a, we can see that heart rate, systolic BP, gender, WBC, and calcium are the top five influential features for Hour 2 prediction, and age is not even among the top 11 features. Highlighted areas in Figure 2b present the features that only contribute to Hour 6 predictions. Interestingly, age plays a crucial role in identifying the sepsis outcome after Hour 6 of ICU admission, and heart rate and systolic BP are still the most influential features. From these results, which are extendable to the other predictive models developed in this study, it can be inferred that regardless of the complexity and the architecture of the predictive models, the features that should be utilized for sepsis prediction of critically ill patients are time-dependent and can differ during different time intervals after ICU admission.

To further investigate the contribution of these features to the prediction results at the patient level, we visualized the local version of the plots presented in Figure 2 for two patients with different outcomes. The waterfall plots presented in Figure 3 present these features based on the predictions using Hour 2 data. The red arrows show the features that contribute to the increase, and the blue arrows present features that contribute to the decrease in the prediction. The width of each arrow shows the height of its impact. These plots powerfully show the association between the prediction results and influential features (i.e. add red values or subtract blue variables to generate the final prediction for each patient).

To explore the pattern of influential features, we further used Shapely values to cluster our dataset based on the explanation similarity of samples, using hierarchical agglomerative clustering (figure 4). Similar patients (aka. data points) are grouped on the x-axis. The $f(x)$ curve, as a line across the top, presents the model predictions for the instances. The colours on the map correspond to the SHAP values. The SHAP first clusters all the data points using hierarchical clustering and then orders the instances on the x-axis. The center of the heatmap presents the base value, which is the mean prediction for all instances. From this figure, we can see that predictions impacted by heart rate also tend to be impacted by SBP, and they are placed in the same cluster. The light colours in the middle of the map present the samples that not swayed far from the base value as the SHAP values are low. Moreover, while the heatmaps presented for Hour 2 and Hour 6 predictions show some levels of similarity between the influential features associated with different samples of the dataset, the discrepancies between the structure and presentation of the maps are very clear and imply the need for different feature sets when developing predictive models for early prediction of sepsis.

IV. Discussion and Conclusions

In this paper, we applied machine learning and SHAP analysis to investigate the efficacy of using machine learning for predicting sepsis among critically ill patients and to explore the features that contribute to these predictions at different time intervals (Hour 2 and Hour 6 of ICU admission). The



(a) Prediction results after hour 2

(b) Prediction results after hour 6

Fig. 4: Heatmap presentation of data points clustered based on the contributing features.

results of our predictive models show that machine learning can be utilized to assist clinicians in distinguishing between patients who would develop sepsis and those who would not by analyzing data collected at admission (AUC: 70%, sensitivity 66%).

We further analyzed the top two features presented in both global feature importance charts (see Figure 2) and found that mean heart rates were very similar between all patients; however, a very rapid heart rate (>110 bpm) was more common in patients who developed sepsis than those without. Interestingly, systolic blood pressure showed similar values between patients with and without sepsis, while it is the second most influential feature in both time intervals under study. Furthermore, we found that a higher percentage of abnormal respiration rate was observed in patients who developed sepsis.

Several limitations should be noted. This study only presents the prediction results based on the data collected after the first and the fifth hour of ICU admission. Including data from later hours of admission might improve the performance of the predictive models presented in this study. However, as giving appropriate treatment to patients with sepsis as soon as possible is critical for improving outcomes for this life-threatening condition, using the data from the hours leading up to sepsis development may not help achieve this goal.

Moreover, other machine learning algorithms, such as Long Short-Term Memory (LSTM), that can model temporal trends of patient data could generate different results in terms of both prediction performance and feature contributions. To mitigate the impact of this limitation on the results of our study, we tried this approach before finalizing the design of the study. As the results of our best-performing model (i.e. logistic regression) were comparable to those of LSTM, we decided to only focus on one best-performing model and design the SHAP analysis section based on only one predictive model.

REFERENCES

[1] WHO. (2020). "Sepsis: Key facts," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/sepsis (visited on 01/30/2021).

[2] U. M. Wallgren, J. Sjölin, H. Järnbert-Pettersson, and L. Kurland, "The predictive value of variables measurable in the ambulance and the development of the predict sepsis screening tools: A prospective cohort study," *Scandinavian journal of trauma, resuscitation and emergency medicine*, vol. 28, no. 1, pp. 1–14, 2020.

[3] A. Perner, A. C. Gordon, D. De Backer, G. Dimopoulos, J. A. Russell, J. Lipman, J.-U. Jensen, J. Myburgh, M. Singer, R. Bellomo, *et al.*, "Sepsis: Frontiers in diagnosis, resuscitation and antibiotic therapy," *Intensive care medicine*, vol. 42, no. 12, pp. 1958–1969, 2016.

[4] D. Pittet, S Rangel-Frausto, N. Li, D Tarara, M Costigan, L Rempe, P Jebson, and R. Wenzel, "Systemic inflammatory response syndrome, sepsis, severe sepsis and septic shock: Incidence, morbidities and outcomes in surgical icu patients," *Intensive care medicine*, vol. 21, no. 4, pp. 302–309, 1995.

[5] T. J. Iwashyna, E. W. Ely, D. M. Smith, and K. M. Langa, "Long-term cognitive impairment and functional disability among survivors of severe sepsis," *Jama*, vol. 304, no. 16, pp. 1787–1794, 2010.

[6] B. D. Winters, M. Eberlein, J. Leung, D. M. Needham, P. J. Pronovost, and J. E. Sevransky, "Long-term mortality and quality of life in sepsis: A systematic review," *Critical care medicine*, vol. 38, no. 5, pp. 1276–1283, 2010.

[7] A. Sun, G. Netzer, D. S. Small, A. Hanish, B. D. Fuchs, D. F. Gaieski, and M. E. Mikkelsen, "Association between index hospitalization and hospital readmission in sepsis survivors," *Critical care medicine*, vol. 44, no. 3, pp. 478–487, 2016.

[8] D. Pittet, B. Thievent, R. P. Wenzel, N. Li, R. Auckenthaler, and P. M. Suter, "Bedside prediction of mortality from bacteremic sepsis. a dynamic analysis of icu patients.," *American journal of respiratory and critical care medicine*, vol. 153, no. 2, pp. 684–693, 1996.

[9] G. V. Research. (2021). "Sepsis diagnostics market size, share trends analysis report by product (blood culture media, software, instruments, assay kits reagents), by technology, by pathogen, by testing type, and segment forecasts, 2021 - 2027," [Online]. Available: https://www.grandviewresearch.com/industry-analysis/sepsis-diagnostics-market (visited on 02/10/2021).

[10] P. E. Marik and A. M. Taeb, "Sirs, qsofa and new sepsis definition," *Journal of thoracic disease*, vol. 9, no. 4, p. 943, 2017.

[11] M. A. Reyna, C. Josef, S. Seyedi, R. Jeter, S. P. Shashikumar, M. B. Westover, A. Sharma, S. Nemati, and G. D. Clifford, "Early prediction of sepsis from clinical data: The physionet/computing in cardiology challenge 2019," in *2019 Computing in Cardiology (CinC)*, IEEE, 2019, Page–1.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] N. Tomašev, X. Glorot, J. W. Rae, M. Zielinski, H. Askham, A. Saraiva, A. Mottram, C. Meyer, S. Ravuri, I. Protsyuk, *et al.*, "A clinically applicable approach to continuous prediction of future acute kidney injury," *Nature*, vol. 572, no. 7767, p. 116, 2019.

[14] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[15] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.