JAMIA Open, 5(4), 2022, 1–11
https://doi.org/10.1093/jamiaopen/ooac080
Research and Applications

Research and Applications

Performance effectiveness of vital parameter combinations for early warning of sepsis—an exhaustive study using machine learning

Ekanath Srihari Rangan [1], Rahul Krishnan Pathinarupothi [2], Kanwaljeet J.S. Anand[3], and Michael P. Snyder [1]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA, [2]Center for Wireless Networks & Applications, School of Engineering, Amrita University, Amritapuri, India, and [3]Division of Critical Care, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA

Corresponding Author: Michael P. Snyder, PhD, Department of Genetics, Stanford University School of Medicine, Stanford Center for Genomics and Personalized Medicine, 240 Pasteur Drive #4452, Stanford, CA 94304, USA; mpsnyder@stanford.edu
Kanwaljeet J.S. Anand and Michael P. Snyder are the senior authors on this manuscript.

Received 29 July 2022; Revised 7 September 2022; Editorial Decision 12 September 2022; Accepted 20 September 2022

ABSTRACT

Objective: To carry out exhaustive data-driven computations for the performance of noninvasive vital signs heart rate (HR), respiratory rate (RR), peripheral oxygen saturation (SpO_2), and temperature (Temp), considered both independently and in all possible combinations, for early detection of sepsis.
Materials and methods: By extracting features interpretable by clinicians, we applied Gradient Boosted Decision Tree machine learning on a dataset of 2630 patients to build 240 models. Validation was performed on a geographically distinct dataset. Relative to onset, predictions were clocked as per 16 pairs of monitoring intervals and prediction times, and the outcomes were ranked.
Results: The combination of HR and Temp was found to be a minimal feature set yielding maximal predictability with area under receiver operating curve 0.94, sensitivity of 0.85, and specificity of 0.90. Whereas HR and RR each directly enhance prediction, the effects of SpO_2 and Temp are significant only when combined with HR or RR. In benchmarking relative to standard methods Systemic Inflammatory Response Syndrome (SIRS), National Early Warning Score (NEWS), and quick-Sequential Organ Failure Assessment (qSOFA), Vital-SEP outperformed all 3 of them.
Conclusion: It can be concluded that using intensive care unit data even 2 vital signs are adequate to predict sepsis upto 6 h in advance with promising accuracy comparable to standard scoring methods and other sepsis predictive tools reported in literature. Vital-SEP can be used for fast-track prediction especially in limited resource hospital settings where laboratory based hematologic or biochemical assays may be unavailable, inaccurate, or entail clinically inordinate delays. A prospective study is essential to determine the clinical impact of the proposed sepsis prediction model and evaluate other outcomes such as mortality and duration of hospital stay.

Key words: vitals parameters, sepsis prediction, machine learning, XGBoost

Lay Summary

Early detection of sepsis is a highly recognized need in healthcare to reduce mortality and improve patient outcomes. In this study, we built and evaluated machine learning models for 240 different combinations of 4 vital signs: heart rate, respiratory rate, peripheral oxygen saturation, and temperature and systematically ranked these vital sign combinations with regard to 2 performance parameters: (1) area under receiver operating curve and 2) prediction times. By using second-order derived features (eg, comparing the hourly differences, cumulative differences, variance, and rate of change in the vital signs measurements), we achieved an area under the curve of 0.94, sensitivity of 82%, specificity of 85% and show that combining just heart rate with temperature predicts sepsis 6 h in advance with accuracy comparable to other sepsis predictive tools reported in literature. This lays the foundation for being able to predict progression to sepsis using noninvasive wearable sensors in limited resource settings.

BACKGROUND AND SIGNIFICANCE

Infectious diseases are showing trends toward significantly amplified severity, with possible causes attributed to antimicrobial resistance, mutated variants, impaired immune responses, advanced age, and other clinical factors. Such infectious diseases put intensive care unit (ICU) patients at increased risk of unnoticed progression to sepsis, ultimately leading to avoidable increases in morbidity and mortality.[1,2]

Sepsis is life-threatening in many bacterial, fungal, and viral infections. Computational methods have helped to glean insights into the pathophysiology of sepsis.[3] Signs of multiorgan system injury occur in approximately 30–50% of those presenting with sepsis.[4] The mortality rate increases by 4–8% with every hour of delay in diagnosis, therefore the early recognition of sepsis is pivotal in preventing mortality.[5,6] Currently, sepsis risk assessment is done using rule-based scoring systems such as the Systemic Inflammatory Response Syndrome (SIRS),[7] quick-Sequential Organ Failure Assessment (qSOFA),[8] and National Early Warning Score (NEWS).[9] Whereas qSOFA requires laboratory variables through a blood draw, SIRS and NEWS require manual beside assessments.

There is a growing body of work in machine learning techniques for sepsis prediction,[10–17] many of which are analyzed by Giacobbe et al[18] and Tayefi et al[19] (see Supplementary Table ST-0). Almost all of them use either laboratory, EMR, or clinical data in addition to vital signs. They require features to be extracted from biochemical parameters, electronic medical records (demographic data and medications), physician and nursing notes, etc. The machine learning techniques used range from Gradient Boosting, Logistic regression, and Long short-term memory to Bayesian networks. In recent years, Lauritsen et al,[20] Yuan et al[21] have added features from image data and comorbid history as well.

There is an increasing need for techniques that rely solely on noninvasive modalities. This will enable a paradigm shift from post-symptomatic manifold invasive investigations to presymptomatic early detection and shorten the timing for sepsis prediction particularly for patients in resource-limited hospital settings.

OBJECTIVES

Our aim was to study vital parameters, heart rate (HR), respiratory rate (RR), peripheral oxygen saturation (SpO_2), and temperature (Temp), individually and in all possible permutations of dual, triple, and quadruple cardinalities, for their effectiveness as early indicators of impending sepsis in adult patients aged 18 years and above and who are admitted to the ICU. The ultimate objective was to come up with the minimal set of vitals that could rely on lightweight sensors for fast-track monitoring, automated periodic execution of sepsis prediction engine, generation of early warning alerts, and expand the time period available for preemptive therapeutic interventions.

MATERIALS AND METHODS

Study design and datasets

An overview of our study design is presented in Figure 1. The study began by acquiring 2 large-scale independent datasets from the widely accepted Physionet database.[22] These datasets have been compiled from the MIMIC-III database by the 2019 Physionet Computing in Cardiology Challenge[23] and made publicly available. The data have been collected over a period of 10 years (2009–2019) and is from ICUs of 2 major hospitals: (1) Beth Israel Deaconess Medical Center (Hospital system A) in which 91.2% of the patients were nonseptic while 8.8% had sepsis in the ICU, and (2) Emory University Hospital (Hospital system B) in which 94.29% were nonseptic while 5.71% had sepsis in the ICU (Supplementary Figure SF-1).

Since our machine learning models considered data starting 12 h before sepsis onset, the main inclusion criteria was the availability of 12 consecutive hourly vital signs measurements. The outliers were removed and patients who had more than 2 consecutive missing data points were excluded from the study. If 2 or less of their consecutive measurements were missing, we employed causal imputation through the last observation carried forward (LOCF) filling technique.[24] LOCF is easy to implement and understand, and given that the missing entries in data belonging to the same patient are relatively less, LOCF will not change the direction of our findings. Furthermore, class imbalance was reduced through random undersampling as recommended in literature,[17,25] so as to match the cases and controls without incurring selection bias.

The data preprocessing (Supplementary Figure SF-1) described above yielded 1130 sepsis and 1500 control patients from Hospital A and 695 sepsis and 9000 control patients from Hospital B. Table 1 (presented in Results section) shows the demographic details of patients in the 2 sepsis and 2 control groups and demonstrates a balanced age and sex distributions. Hospital A data were selected for model building, and Hospital B data for independent unbiased generalizability validation (without any additional tuning). Considering that determining exact onset time of sepsis is very difficult in clinical practice,[26] we have followed the criteria used by most of the existing works[23] and are based on the Sepsis-3 guidelines.[27] This is best estimated to be the earlier of (1) first clinical suspicion of infection (as indicated by ordering of blood cultures and/or administering IV antibiotics) and (2) 2-point increase in the patient's Sequential

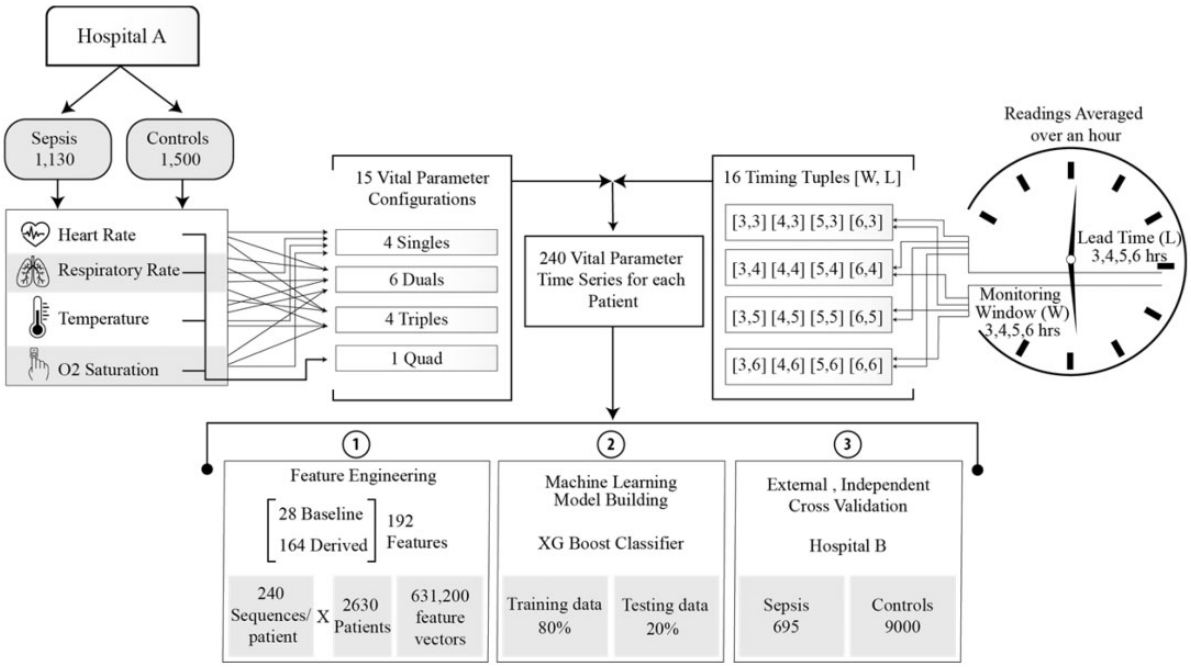

Figure 1. Overview of the Study Design starting with selection of patient datasets (Hospital A data with 1130 sepsis and 1500 control patients), 15 vital parameter configurations, setting of 16 timing tuples, feature engineering, building of machine learning models, and ending with external validation through an independent dataset (Hospital B patients).

Table 1. Demographics of patients in sepsis and control groups showing balanced age and sex distributions

Demographic Factors	Hospital A		Hospital B	
	Sepsis	Controls	Sepsis	Controls
Patients (*N*)	**1130**	**1500**	**695**	**9000**
Age (years)				
Range	18–88	18–89	17–100	15–100
Mean (SD)	63 (16)	63 (16)	61 (17)	61 (17)
Gender distribution				
Female	435 (38%)	586 (39%)	295 (42%)	4209 (47%)
Male	695 (62%)	914 (61%)	400 (58%)	4791 (53%)
ICU Stay (h)				
Mean (SD)	85 (58)	37 (12)	89 (63)	38 (15)

ICU: intensive care unit.

Organ Failure Assessment (SOFA) score within 24 h. Setting the "onset" for nonsepsis controls—referred to as case-control alignment, is a well-debated issue in the literature. Our choice of the data sampling window for nonsepsis controls was guided by the discussion presented in prior literature.[26,28] We applied relative onset matching in which data sampling windows for the controls were chosen relative to the start of ICU admission since they present the hardest case for differentiation between sepsis and controls. Subsequently controls are likely to become more stable with interventions and the differentiation between the control and sepsis group will be more and more distinct making the prediction task easier as a result of which machine learning algorithms might show higher than actual AUCs.

Out of 40 clinical variables available in these datasets, the 4 vital parameters (HR, RR, SpO_2, Temp), recorded as time-stamped nurse-verified physiological measurements, were extracted and downsampled to hourly data points by finding their medians to yield a time series of 26 300 data points per sensor. The 4 vital parameters were then considered both individually (4) and in all possible permutations of dual (6), triple (4), and quad (1) cardinalities, resulting in 15 different vital parameter configurations. Next, relative to the time of onset of sepsis, we considered 2 time intervals of direct relevance to the patient:

- Monitoring Window (W): The minimum number of hours of vital sign data needed to determine sepsis. This window should be as short as possible to make a real-time determination of sepsis and simplify the amount of data needed for the algorithm to perform. We chose to experiment with 4 different monitoring windows, of 3-, 4-, 5-, and 6-h durations.
- Lead Time (L): It is the interval between the end of the monitoring window and time of onset of sepsis. The Lead Time should be adequate to transfer the patient to ICU and to initiate the medical interventions required to reduce mortality, such as, administration of antibiotics and intravenous fluids. Based on this, we chose to experiment with 4 different lead times of 3-, 4-, 5-, and 6-h lengths.

Permutation of the 4 monitoring windows and 4 lead times yields 16 timing tuples (W, L).

Machine learning classifier

Our choice of the machine learning algorithm was guided by 2 desirable attributes. First, it must be nimble to be ultimately installed as an app on portable devices (eg, smartphones) as may be the case in resource-limited hospital settings, which imposes tight constraints on computational complexity and bandwidth. Second, to increase acceptability by physicians, it should be amenable to clinical interpretation. We used an expressive classifier known as Gradient

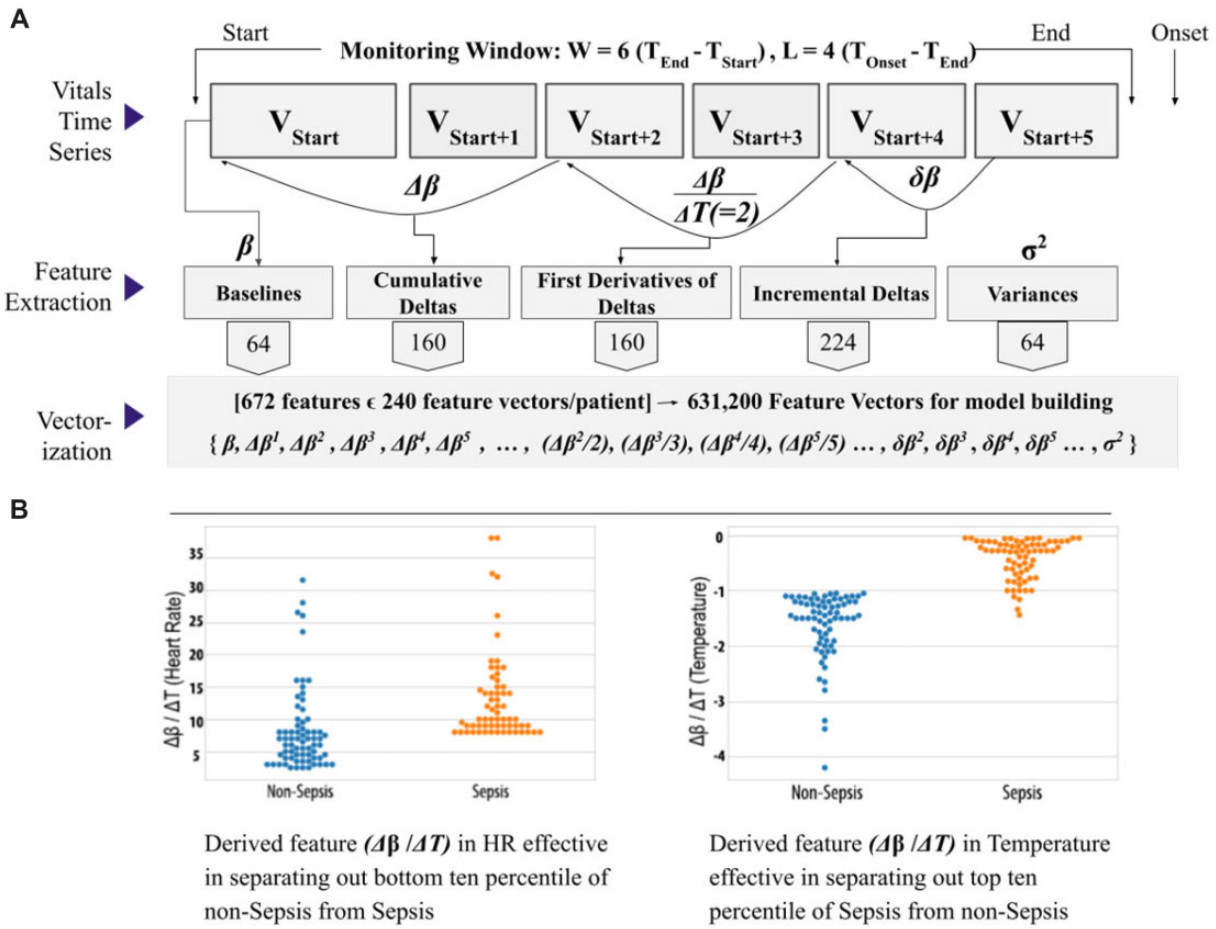

Figure 2. (A) Vital-SEP's innovative Feature Engineering resulting in half a million feature vectors for the dataset; (B) The Differentiating Effect of Derived Feature Engineering.

Boosted Decision Tree (or XGBoost).[17] Multiplexing the data in accordance with the 15 vital parameter configurations and slicing the data to fit the 16 timing tuples required us to build 240 Vital-SEP XGBoost models. The steps involved in building Vital-SEP XGBoost models are Feature Engineering, Training, and Testing.

Feature engineering

Feature engineering gave us a lot of scope for innovation and improved performance. In addition to features that intuitively captured severity trends in the vital parameters, both most recent and cumulative, we experimented with second-order derived and aggregate features (see Figure 2). These features are:

- β: Baseline vital signs values at the start of the monitoring window.
- $\delta\beta$: Incremental deltas in vital signs values between consecutive hours.
- $\Delta\beta$: Cumulative change in vital signs values (representing the physiological deterioration versus improvement) from the baseline.
- $\Delta\beta/\Delta t$: First derivatives from the baseline.
- σ^2: Variance observed in the vital signs values.

An Illustrative example of the above feature extraction is presented in Supplementary Table ST-16.

The above features were computed for all timing tuples and vital parameter configurations, summing up to 672 features (of which 192 are unique considering overlapping intervals), which were then vectorized into 240 feature vectors per patient, resulting in more than half a million feature vectors for the dataset.

Figure 2B illustrates, taking HR and Temp derived features as examples, how Vital-SEP achieves differentiation between sepsis and nonsepsis through innovative feature engineering. Whereas derived feature ($\Delta\beta/\Delta T$) in HR is effective in separating out bottom 10% nonsepsis from sepsis, derived feature ($\Delta\beta/\Delta T$) in Temp is effective in separating out top 10 percentile of sepsis from nonsepsis.

Model building and testing

The curated Hospital A dataset containing 1130 sepsis patients and 1500 nonsepsis controls, after being subjected to feature extraction, the next step was to build/train the 240 Vital-SEP XGBoost models representing the different combinations of vitals, monitoring window (W) and lead times (L).

Vital-SEP Algorithm Pseudocode is presented in Supplementary Figure SF-4. Eighty percent of the training data (Hospital A) was used for building the XGBoost model with 5-fold cross-validation. Prediction was set up as a binary classification task, predicting sepsis versus no sepsis. GridSearchCV was employed to select hyperparameter values to maximize the area under the ROC curve (AUC).

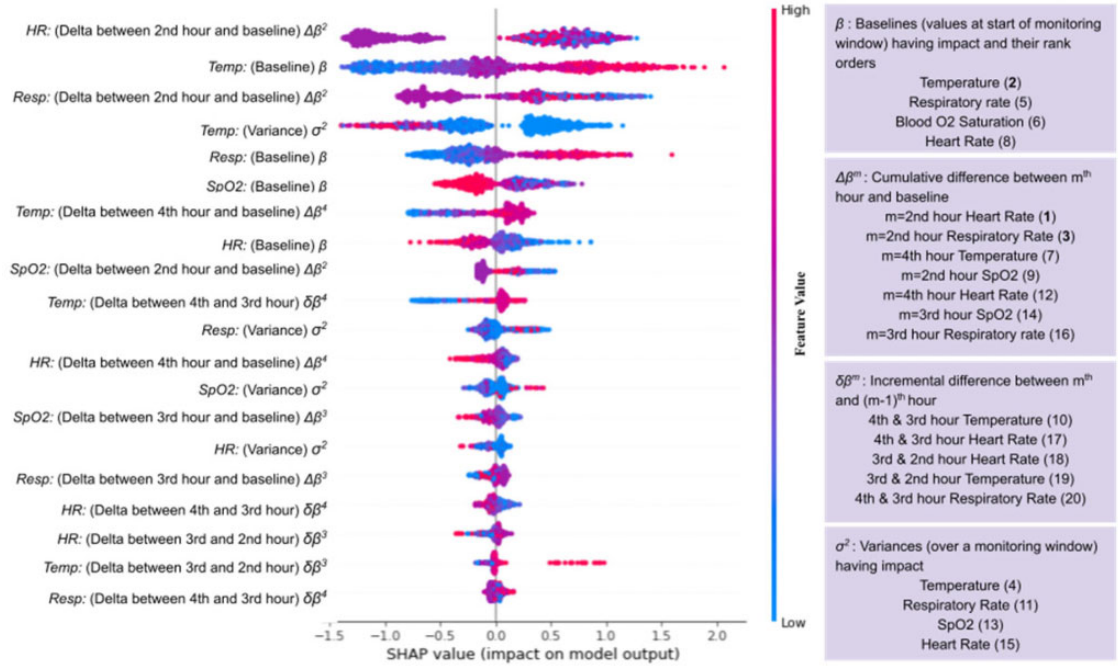

Figure 3. SHapley Additive exPlanations (SHAP) plot comparing the top 20 features that have an impact on performance of Vital-SEP. The features are stacked in descending order of their effect on predictability. Negative or low SHAP values indicate no sepsis, whereas positive or high SHAP values indicate sepsis. The listing on the right shows the features classified and rank ordered (ranks listed in parenthesis).

The remaining 20% data were used for the first-level performance evaluation.

Relative comparison of the impact of features using *SHapley additive exPlanations*

To examine the individual impact of vital parameters and features derived from them, we employed a widely used tool called SHapley Additive exPlanations (SHAP).[29] A Shapley value denotes how much a feature in the context of its interaction with other features contributes to model prediction. The results of this analysis are presented in Figure 3. The top 20 features are rank ordered according to their relative level of contribution to Vital-SEP prediction.

Among all these features, the change in HR between baseline and second hour of the monitoring window (HR: $\Delta\beta^2$) had the highest influence on Vital-SEP. The higher the change in HR values—higher the probability of predicting sepsis over nonsepsis. Next to HR was Temp. Higher temperatures took positive SHAP values indicating sepsis while lower temperatures took negative SHAP indicating no sepsis. Another interesting observation is that sepsis prediction probability was higher with higher baseline temperature and lower temperature variance. Next in order, decreased baseline SpO_2 (SpO_2:β) and increased baseline RR (Resp:β) clearly indicate higher propensity toward sepsis.

Validation method

To assess the reliability of Vital-SEP across healthcare systems, an entirely independent patient dataset from Hospital B was used. Hospital B dataset contained 695 sepsis and 9000 nonsepsis patients. The trained Vital-SEP models were tested on this external dataset without any additional tuning.

RESULTS

The demographic and vital sign characteristics of both datasets are presented in Tables 1 and 2.

Maximizing prediction accuracy through synergistic interplay of vital parameters

For each of the 240 Vital-SEP models, the sepsis prediction performance was evaluated in terms of Precision, Recall, F1 score, and area under receiver operating characteristic curve (AUROC, or further abbreviated as AUC), shown in Supplementary Tables ST-1 to ST-15 and Supplementary Figure SF-2. From each of the Supplementary Tables ST-1 to ST-15, the mean and standard deviation (SD) of the AUC across all timing tuples were computed, as presented in Table 3 and illustrated in Figure 4. In the same tables, the timing tuple which maximizes the AUC was highlighted and all such best-performing timing tuples (W_{AUC}, L_{AUC}) were consolidated in Table 4 and illustrated in Figure 4.

Figure 5 shows, for each vital parameter configuration, the highest AUC achieved among all of the timing tuples for that configuration. Starting with single parameters, HR gives an AUC of 0.82 (95% CI, 0.73–0.83), dips to 0.8 for SpO_2 and 0.79 for Temp, rises back to AUC 0.83 (95% CI, 0.78–0.87) for RR. For dual vitals pairs, [HR + SpO_2] yielded 0.87, [HR + Temp] yielded AUC of 0.90 (95% CI, 0.83–0.93), and then it dips a bit. Continuing with triple vitals sets, [HR + RR + Temp] starts off at AUC of 0.93 (95% CI, 0.89–0.96) and other combinations are pretty close. Using all the 4 vitals (quad configuration), we obtained an AUC of 0.94 (95% CI, 0.90–0.96) with sensitivity and specificity of 0.85 and 0.90, respectively.

Stretching the lead time between prediction and sepsis onset

In hospital settings, it is desirable that the lead time offset between prediction and sepsis onset be adequate enough to account for the

Table 2. Summary of recorded vital parameters in the 4 patient groups: Values are aggregated over all the patients in each group, and represented as the mean and standard deviation (shown in parenthesis)

	Hospital A		Hospital B	
	Sepsis	Controls	Sepsis	Controls
Heart rate: mean (standard deviation) beats/min				
High	114 (20)	103 (16)	115 (19)	101 (19)
Low	69 (14)	70 (11)	68 (14)	69 (13)
Range	20–223	25–167	36–211	20–186
Baseline	88 (14)	85 (12)	88 (14)	82 (14)
Respiratory rate in breaths/min				
High	31 (7)	26 (6)	29 (9)	25 (6)
Low	11 (3)	11 (3)	11 (4)	12 (3)
Range	1–67	1–67	1–100	1–100
Baseline	20 (4)	18 (3)	18 (4)	18 (3)
Temperature in °C				
High	38.1 (0.8)	37.8 (0.7)	38.1 (0.8)	37.4 (0.7)
Low	36.0 (0.8)	36.0 (0.7)	35.8 (0.9)	36.1 (0.6)
Range	26.7–40.6	30.6–41.6	30.9–41.4	30.0–50.0
Baseline	37.1 (0.6)	37.1 (0.5)	37.0 (0.6)	36.8 (0.5)
Blood oxygen saturation in %				
High	99 (0)	99 (0)	99 (0)	99 (0)
Low	89 (7)	92 (5)	87 (9)	91 (6)
Range	22–100	25–100	20–100	20–100
Baseline	97 (1)	97 (1)	97 (1)	97 (1)

The high and the low together with the range show the extent of variation of each vital parameter. The baseline represents the recorded values at the start of the monitoring window.

time required for transfer to ICU and subsequent therapeutic intervention. For each of the 4 lead times, we determined the minimal Vital Parameter configurations that exhibit sufficient prediction accuracy. The best of these results are described below.

Figure 6 shows the increase in lead time with number of vital parameters. A lead time of 3 h is obtained by just monitoring the RR. Adding a second vital sign (temperature) to RR improves the lead time to 4 h and additionally reduces the monitoring window duration from 6 to 4 h. The triple vital signs configuration of HR+Temp+SpO_2 increases the lead time to 5 h. A steeper increase is observed for the triple vital signs configuration of RR+Temp+HR which gives a 2-h jump in lead time taking it to 6 h. When a fourth vital is added, the lead time reaches/stays at its upper limit of 6 h along with a point increment in AUC.

Performance validation of vital-SEP on hospital B dataset

Table 5 and Supplementary Figure SF-3 present a one-to-one comparison of Vital-SEP performance on the 2 datasets A and B. The results were very similar, and for most vital parameter configurations coinciding as well. Just as observed for Hospital A, the quad configuration for Hospital B also yielded the highest AUC of 0.94 (95% CI, 0.92–0.95) with a specificity of 96% and sensitivity of 78%.

Comparison with standard sepsis risk assessment methods

In order to benchmark Vital-SEP, we compared its performance to standard risk assessment methods used for predicting sepsis namely, SIRS, NEWS, and qSOFA methods.[32,33] The AUC, sensitivity, and specificity of sepsis prediction at various lead times are presented in Table 6 and illustrated in Figure 7.

Table 3. Vital Parameter configurations and their AUC (mean and SD across all timing tuples)

Vital parameter configuration	AUC (mean ± SD)
HR	0.79 ± (0.02)
Temp	0.75 ± (0.03)
RR	0.80 ± (0.01)
SpO_2	0.75 ± (0.02)
HR+Temp	0.84 ± (0.04)
RR+Temp	0.85 ± (0.03)
SpO_2+RR	0.86 ± (0.02)
HR+RR	0.87 ± (0.02)
HR+SpO_2	0.84 ± (0.02)
SpO_2+Temp	0.81 ± (0.03)
HR+RR+Temp	0.89 ± (0.03)
SpO_2+Temp+RR	0.86 ± (0.03)
HR+SpO_2+Temp	0.86 ± (0.03)
HR+SpO_2+RR	0.87 ± (0.02)
HR+RR+Temp+SpO_2	0.87 ± (0.03)

Highlighted rows represent the highest mean within each cardinality.
HR: heart rate; RR: respiratory rate; SpO_2: peripheral oxygen saturation; Temp: temperature.

Vital-SEP maintains a much higher AUC hovering around 0.86–0.94 throughout the prediction time range of 3–6 h prior to sepsis onset. For the same prediction times, NEWS has lower AUC of 0.79–0.80, and SIRS and qSOFA are even lower at 0.64–0.70. Notably, the 6-h prediction AUC (0.93) using Vital-SEP is better than even the onset-AUC for all 3 standard methods (NEWS 0.79, SIRS 0.70, and qSOfA 0.66). At 6 h prior to onset, while the sensitivities of SIRS, NEWS, and qSOFA are comparable to Vital-SEP, their specificities are noticeably lower compared to Vital-SEP.

DISCUSSION

In this article, we carried out one of the first exhaustive performance evaluations of various combinations of vital parameters for early sepsis prediction, and externally validated all findings in patients from a separate center. We focused on using the 4 vital signs HR, RR, SpO_2, and Temp which are easily obtainable from noninvasive finger clip photoplethysmography (PPG) sensors. We built Gradient Boosted Decision Tree machine learning models for each of the single, dual, triple, and quad combinations and compared their sensitivity, specificity, and AUCs for different prediction times. While most existing papers take a set of physiological and/or biochemical parameters and evaluate their combined performance in toto, we take a significant step forward by analyzing each and every combination (total 240) of vital parameters and lead times so as to suit different clinical settings and availability of monitoring devices.

We were able to arrive at the best-performing vital parameter set in each class: Among single parameters, RR with AUC 0.83 performed the best. Among dual vitals pairs, [HR + Temp] and [RR + Temp] yielded the highest results with an AUC of 0.90. A surprising role for body temperature seemed to emerge. By itself the AUC of temperature was low (0.79) and the lead time was 4 h, but in combination with HR (HR+Temp), it had an amplifying effect, as is evident from the high AUC (0.9) and lead time of 6 h. A similar observation can be made for SpO_2 as well. Whereas SpO_2 individually had an AUC of 0.8, when combined with RR, increased the

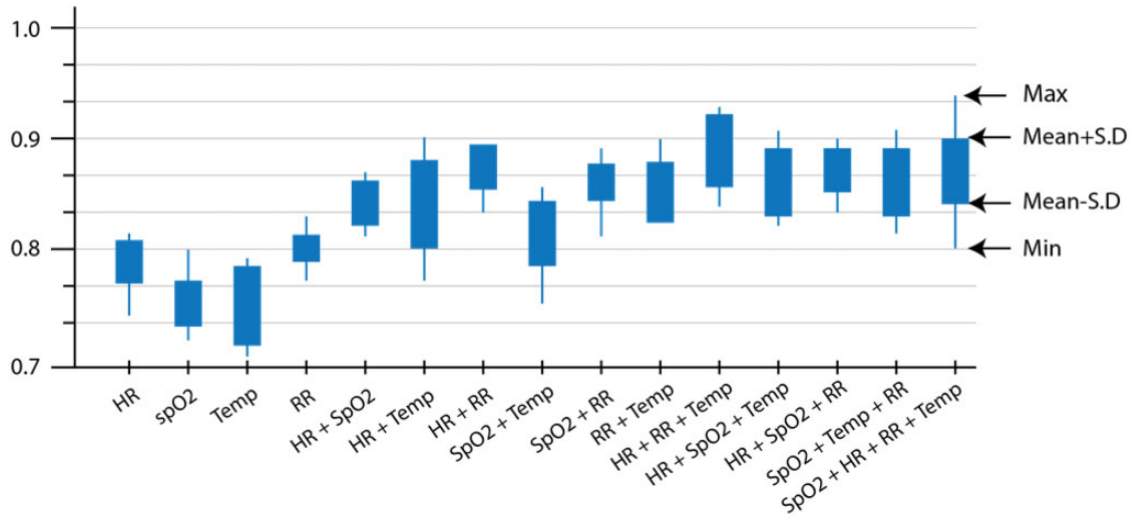

Figure 4. The mean and standard deviation of the AUC across all timing tuples computed for each vital parameter combination.

Table 4. Vital Parameter configurations and their timing tuples (W_{AUC}, L_{AUC}) that maximize AUC

Vital parameter configuration	Timing tuple that maximizes AUC		
Single vital	W_{AUC} (Hrs.)	L_{AUC} (Hrs.)	AUC_{max}
HR	5	5	0.82
Temp	4	4	0.79
RR	6	3	0.83
SpO_2	4	3	0.80
Dual Vitals			
HR+Temp	4	6	0.90
RR+Temp	4	4	0.90
SpO_2+RR	3	5	0.89
HR+RR	4	5	0.89
HR+SpO_2	4	3	0.87
SpO_2+Temp	4	3	0.86
Triple Vitals			
HR+RR+Temp	4	6	0.93
SpO_2+Temp+RR	4	6	0.91
HR+SpO_2+Temp	4	5	0.91
HR+SpO_2+RR	4	6	0.9
Quad Vitals			
HR+RR+Temp+SpO_2	4	6	0.94

Highlighted rows represent the local maxima of AUC_{max} within each cardinality.

HR: heart rate; RR: respiratory rate; SpO_2: peripheral oxygen saturation; Temp: temperature.

AUC to 0.89 and lead time to 5 h. When just the SpO_2 and Temp were combined, there was no satisfactory increase in lead time. There is also a direct relationship between Vital Parameters cardinality and lead times. The additional hourly improvement in lead time is evident as we progress incrementally from using 1 vital sign to using all 4.

In Vital-SEP, among the clinical parameters, the superiority of HR and RR when used either individually or in combination with other vitals, can be attributed to their acute state reactivity, being more sensitive to the pathophysiological changes such as hyperinflammation, dysregulated coagulation, and hypotension that precede sepsis by a few hours.[34] The autonomic nervous systems that regulate HR and RR are tightly coupled with the metabolic demands of all vital organs, being exquisitely responsive to the sepsis induced changes in cellular metabolism and microcirculation. Furthermore, HR and RR variations are also mostly nonspecific with regard to causative factors of sepsis, as well as the organ systems involved.[35] Hence they seem to perform uniformly well for all patient datasets. Among the triple vitals sets, [HR + RR + Temp] performed the best with AUC of 0.93. Using all the 4 vitals (quad configuration), we obtained an AUC of 0.94 with sensitivity and specificity of 0.85 and 0.90, respectively. The above results also reveal a monotonic trend. Starting with HR as a single vital, as additional vitals RR, Temp, and SpO_2 were added, we observed progressive improvements in AUC by almost 15%.

The highest AUROC of 0.94 for 6-h prediction achieved by Vital-SEP compares well with 0.88 achieved by Bedoya et al[16] who use patient demographics, comorbid history, medications, and laboratory data in addition to vital signs. Mao et al,[36] who used systolic blood pressure and diastolic blood pressure (in addition to the 4 vitals used in this paper), achieved an AUROC of 0.92. However, in most hospital ICUs, blood pressure measurements are dependent on manual or automatic inflatable cuffs, both of which require nursing staff and are prone to human errors. This makes it less likely to be monitored hourly particularly in resource constrained settings where invasive mechanisms like arterial lines are also not a possibility. Whereas the other 4 vitals can be measured from finger-clip PPG sensors. In terms of lead time, Vital-SEP's best-performing model achieved 6 h. Though Delahanty et al,[13] achieve a higher prediction time of 24 h, they require the use of features extracted from biochemical and demographic data including medications, as well as, nursing notes.

These results constitute a significant contribution to the existing literature on the effectiveness of vital parameters for sepsis prediction. The innovative feature engineering of Vital-SEP is a novel contribution to the machine learning methods for sepsis detection. This lays the foundation for being able to predict progression to sepsis using noninvasive wearable sensors in limited resource settings. Since the training and validation datasets in our study were from 2 different hospitals and 2 different regions, there is guarded optimism that by extending the training dataset to the Electronic Health Records (EHR) from other hospitals, Vital-SEP can be made suitable for hospitals in diverse regions and populations. After satisfactory prospective studies

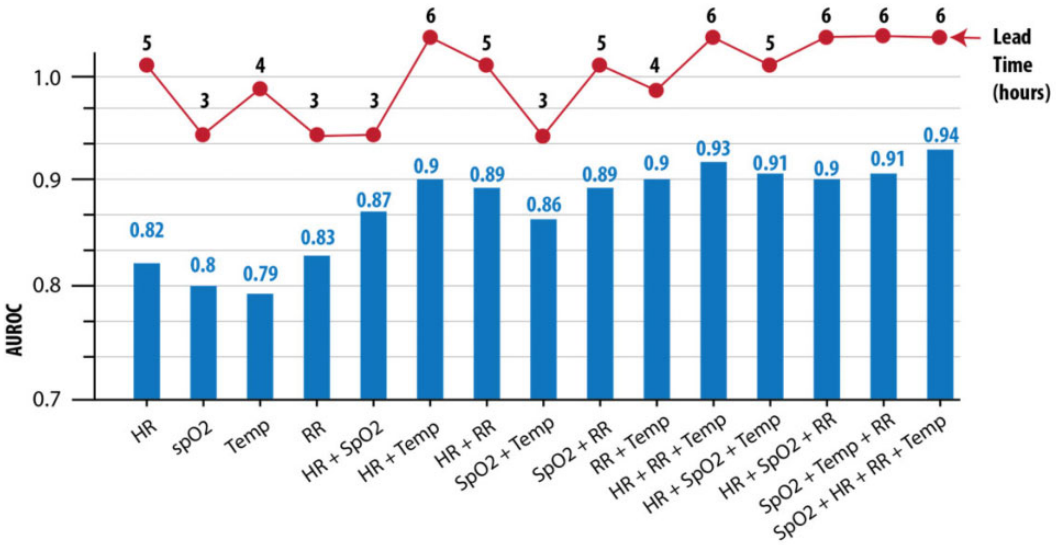

Figure 5. Barplot showing the AUCmax of all Vital Parameter configurations. Also shown in the line tracing are the lead times that result in maximizing AUC for that Vital Parameter configuration.

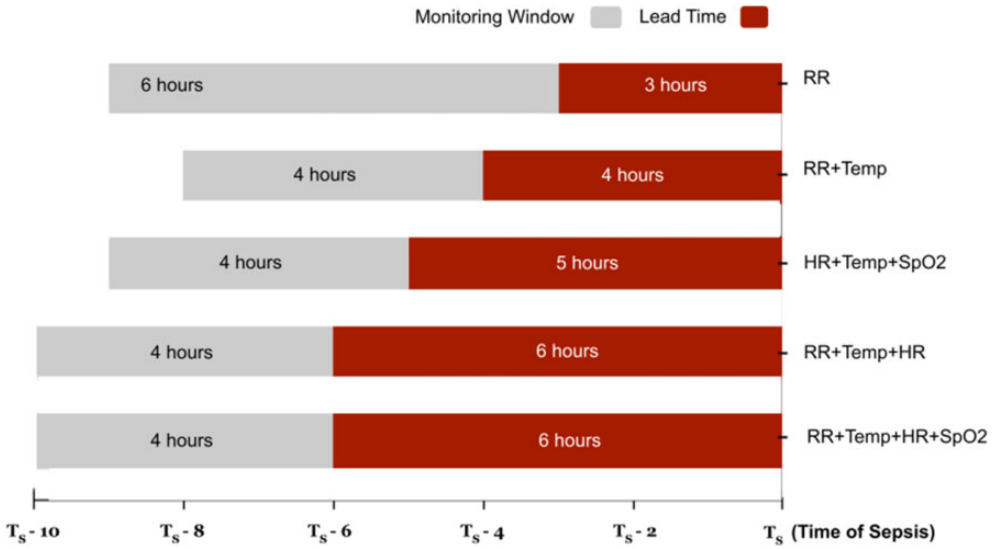

Figure 6. The gain in lead time and the reduction in the monitoring window as we increment the vital parameter cardinality from 1 to 4.

and clinical trials, Vital-SEP can be integrated into patient management in the ICUs of hospitals. ICU patients are monitored with a wearable PPG cum temperature sensor (there are many available in the market but need to be certified as medical grade), which transmit the vitals to the Vital-SEP app installed on the ICU EHR server. A positive prediction of sepsis triggers advance warnings and alarms which are transmitted to the physician's console who can initiate therapeutic intervention.

Limitations

Our study is not without limitations. Our method assumes uninterrupted hourly measurements in the monitoring window. Also, in make-shift limited resource settings, there are possibilities of errors due to nonstandard devices being used for clinical monitoring which may confound the algorithm. Although it would be desirable to incorporate automatic error correction mechanisms, even the commercially available wearable sensors[37] can monitor vital signs with sufficient accuracy to provide an uninterrupted stream of hourly measurements in the monitoring window. There is also scope for further minimizing the false negatives so as to improve the clinical outcomes of all patients at risk for developing sepsis, even as the low false-positive rate of Vital-SEP will conserve hospital resources and manpower.

Since this is a retrospective study, clinical outcomes such as mortality, duration of hospital stay and other quality improvement parameters could not be measured. Our study in this article addresses the limited problem space of predicting sepsis in patients who are already admitted to the ICU. Therefore future studies with data from ward and emergency department patients will be required before implementation in non-ICU settings.

Table 5. Vital parameter configurations and their AUCs for Hospital A and Hospital B patients

Vital Parameter Configurations	AUC (Hospital A) [95% CI]	AUC (Hospital B) [95% CI]	*P* value
Single Vitals			
HR	0.78 [0.73–0.83]	0.88 [0.87–0.90]	.02
Temp	0.78 [0.72–0.83]	0.70 [0.67–0.73]	.32
RR	0.82 [0.78–0.87]	0.84 [0.82–0.86]	.80
SpO_2	0.74 [0.69–0.78]	0.76 [0.74–0.78]	.65
Dual Vitals			
HR+Temp	0.88 [0.84–0.93]	0.84 [0.82–0.86]	.61
RR+Temp	0.87 [0.83–0.92]	0.82 [0.79–0.85]	.38
SpO_2+RR	0.85 [0.81–0.89]	0.88 [0.86–0.89]	.62
HR+RR	0.85 [0.81–0.89]	0.93 [0.91–0.94]	.04
HR + SpO_2	0.87 [0.83–0.91]	0.90 [0.88–0.91]	.53
SpO_2+Temp	0.83 [0.78–0.88]	0.83 [0.81–0.86]	.98
Triple Vitals			
HR+RR+Temp	0.93 [0.89–0.96]	0.92 [0.91–0.94]	.95
SpO_2+Temp+RR	0.91 [0.87–0.95]	0.91 [0.89–0.93]	.90
HR+SpO_2+Temp	0.92 [0.89–0.96]	0.91 [0.90–0.93]	.89
HR+SpO_2+RR	0.89 [0.86–0.93]	0.94 [0.92–0.95]	.14
Quad Vitals			
HR+RR+Temp+SpO_2	0.93 [0.90–0.96]	0.94 [0.92–0.95]	.91

Except for the 2 highlighted rows, all others have no statistically significant difference. *P* values were computed via permutation test.[30,31]
HR: heart rate; RR: respiratory rate; SpO_2: peripheral oxygen saturation; Temp: temperature.

Table 6. Performance comparison of Vital-SEP with standard sepsis risk assessment methods SIRS, NEWS, qSOFA, in terms of Area under Receiver Operating Characteristics (AUC), Sensitivity, and Specificity

Measure	Prediction Time	VitalSEP	SIRS	NEWS	qSOFA
AUC [95% CI]	At onset of sepsis	Intended only for Advance prediction	0.70 [0.66–0.73]	0.79 [0.76–0.82]	0.66 [0.62–0.70]
	3 h prior onset	0.91 [0.90–0.96]	0.69 [0.65–0.74]	0.79 [0.75–0.83]	0.66 [0.61–0.71]
	4 h prior onset	0.86 [0.82–0.90]	0.68 [0.64–0.73]	0.80 [0.76–0.83]	0.66 [0.61–0.71]
	5 h prior onset	0.91 [0.87–0.95]	0.68 [0.64–0.73]	0.79 [0.75–0.83]	0.64 [0.59–0.68]
	6 h prior onset	0.94 [0.90–0.96]	0.68 [0.64–0.73]	0.79 [0.75–0.83]	0.66 [0.61–0.71]
Sensitivity	At onset of sepsis		0.48	0.69	0.83
	3 h prior onset	0.83	0.85	0.86	0.84
	4 h prior onset	0.91	0.85	0.87	0.85
	5 h prior onset	0.80	0.85	0.87	0.82
	6 h prior onset	0.83	0.82	0.87	0.83
Specificity	At onset of sepsis		0.79	0.71	0.41
	3 h prior onset	0.90	0.52	0.40	0.39
	4 h prior onset	0.67	0.52	0.41	0.39
	5 h prior onset	0.87	0.47	0.40	0.39
	6 h prior onset	0.89	0.52	0.41	0.39

NEWS: National Early Warning Score; qSOFA: quick-Sequential Organ Failure Assessment; SIRS: Systemic Inflammatory Response Syndrome.

CONCLUSION

It can be concluded that using Vital-SEP, even 2 vital signs are adequate to predict sepsis upto 6 h prior to clinical diagnosis with promising accuracy comparable to standard scoring methods and other sepsis predictive tools reported in literature. It relies on a minimal set of vital parameters that can be captured from wearable sensors automatically and noninvasively in the background, and these can be installed even in resource-limited hospitals.[38–40] Cases that would otherwise be missed due to delays laboratory testing, or lack of obvious clinical symptoms, or infrequent manual assessment of vital signs (such as in isolation wards), would also be detected at the nascent stage. In order to ascertain the clinical applicability of Vital-SEP and evaluate outcomes and quality improvement, the next step would be a prospective study.

AUTHOR CONTRIBUTIONS

ESR and RKP conceived the study design. ESR did the data preparation, data cleaning, and first-level analysis. ESR and RKP built the machine learning algorithms, conducted the evaluations, and drafted the manuscript. ESR and KSA did the statistical analysis. MPS and KSA guided the study, provided scientific input, enhanced the data and results interpretation, reviewed and revised the manuscript. The final version of the article has been approved by all the authors.

ETHICS APPROVAL

The data used in this study were obtained from the publicly available Physionet's CinC (Computing in Cardiology Challenge 2019) database which is compiled from the Beth Israel Deaconess Medical

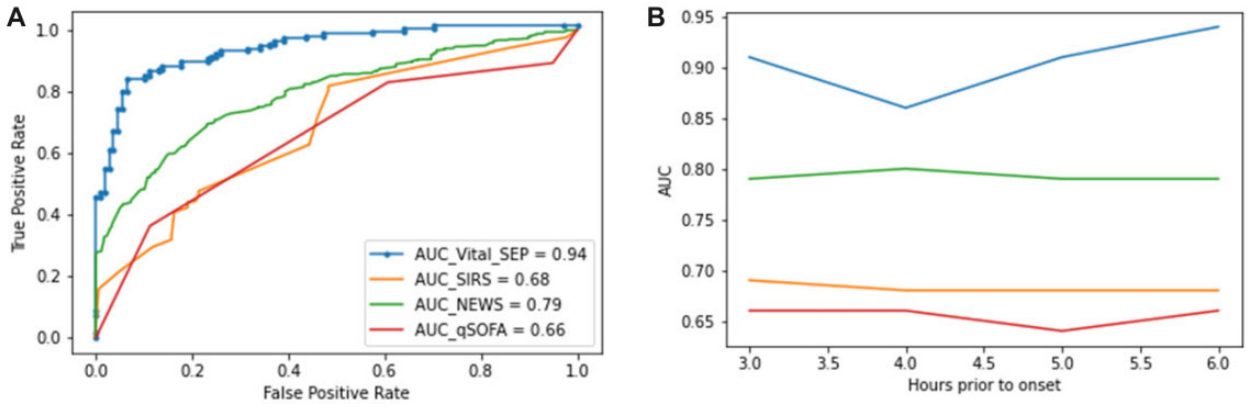

Figure 7. Comparison between Vital-SEP and the standard methods SIRS, NEWS, and qSOFA: (A) receiver operating characteristic curve at 6-h prediction; (B) AUC as a function of prediction times. NEWS: National Early Warning Score; qSOFA: quick-Sequential Organ Failure Assessment; SIRS: Systemic Inflammatory Response Syndrome.

Center (BIDMC), Boston and Emory University Hospital, Atlanta. PhysioNet is a repository of medical research data, managed by the MIT Laboratory for Computational Physiology, under the auspices of NIH. The data in this publicly available database have been deidentified/anonymized in compliance with HIPAA. Data collection for all datasets did not impact patient safety. Therefore, this study constitutes nonhuman subjects research, which does not require Institutional Review Board approval.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *JAMIA Open* online.

ACKNOWLEDGEMENTS

The authors are thankful to Sri Mata Amritanandamayi Devi (Chancellor, Amrita University), for fostering an ambience of collaboration to leverage the scientific strengths of Amrita and Stanford Universities for advancing medical research.

FUNDING

This work was supported by NIH grant 1R01NR02010501.

CONFLICT OF INTEREST STATEMENT

MPS is cofounder and a member of the scientific advisory board of Personalis, Qbio, January, SensOmics, Protos, Mirvie, and Oralome. He is on the scientific advisory board of Danaher, GenapSys, and Jupiter. ESR and RKP have US patents on using wearable devices for remote patient monitoring (US10433726B2 and US10542889B2). No reagents or funding from any of these organizations/sources was used in this study. Hence the authors declare that there are no competing interests.

DATA AVAILABILITY

The code and data used to develop, test, and validate the Vital-SEP engine are available at https://pprahul.github.io/Vital-SEP/.

REFERENCES

1. D'Onofrio V, Heylen D, Pusparum M, *et al.* A prospective observational cohort study to identify inflammatory biomarkers for the diagnosis and prognosis of patients with sepsis. *J Intensive Care* 2022; 10 (1): 13.
2. World Health Organization. Sepsis. https://www.who.int/news-room/fact-sheets/detail/sepsis. Accessed February 10, 2021.
3. An G, Nieman G, Vodovotz Y. Toward computational identification of multiscale "tipping points" in acute inflammation and multiple organ failure. *Ann Biomed Eng* 2012; 40 (11): 2414–24.
4. Rossaint J, Zarbock A. Pathogenesis of multiple organ failure in sepsis. *Crit Rev Immunol* 2015; 35 (4): 277–91.
5. Kumar A, Roberts D, Wood KE, *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34: 1589–96.
6. Seymour CW, Gesten F, Prescott HC, *et al.* Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med* 2017; 376 (23): 2235–44.
7. Marik PE, Taeb AM. Sirs, qsofa and new sepsis definition. *J Thorac Dis* 2017; 9 (4): 943–5.
8. Jones AE, Trzeciak S, Kline JA. The sequential organ failure assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* 2009; 37: 1649.
9. Royal College of Physicians. National Early Warning Score (NEWS) 2. *Standardizing the Assessment of Acute-Illness Severity in the NHS*. London: National Early Warning Score (NEWS); 2012.
10. Fleuren LM, Klausch TLT, Zwager CL, *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020; 46 (3): 383–400.
11. Vellido A, Ribas V, Morales C, Sanmartin AR, Ruiz Rodriguez JC. Machine learning in critical care: state-of-the-art and a sepsis case study. *Biomed Eng Online* 2018; 17(Suppl. 1): 135.
12. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018; 46 (4): 547–53.
13. Delahanty RJ, Alvarez J, Flynn LM, *et al.* Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med* 2019; 73 (4): 334–44.
14. Fagerstrom J, Bang M, Wilhelms D, Chew MS. LiSep LSTM: a machine learning algorithm for early detection of septic shock. *Sci Rep* 2019; 9 (1): 15132.
15. Yee CR, Narain NR, Akmaev VR, Vamulapalli V. A data-driven approach to predicting septic shock in the intensive care unit. *Biomed Inform Insights* 2019; 11: 1178222619885147.

16. Bedoya AD, Futoma J, Clement ME, *et al.* Machine learning for early detection of sepsis: an internal and temporal validation study. *JAMIA Open* 2020; 3 (2): 252–60.
17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16); Association for Computing Machinery, New York, NY; 2016: 785–94.
18. Giacobbe DR, Signori A, Del Puente F, *et al.* Early detection of sepsis with machine learning techniques: a brief clinical perspective. *Front Med (Lausanne)* 2021; 8: 617486.
19. Tayefi M, Ngo P, Chomutare T, *et al.* Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev Comput Stat* 2021; 13: e1549.
20. Lauritsen SM, Kalor ME, Kongsgaard EL, *et al.* Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artif Intell Med* 2020; 104: 101820.
21. Yuan KC, Tsai LW, Lee KH, *et al.* The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform* 2020; 141: 104176.
22. Goldberger AL, Amaral LAN, Glass L, *et al.* Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 2000; 101 (23): E215–20.
23. Reyna MA, Josef CS, Jeter R, *et al.* Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *Crit Care Med* 2020; 48 (2): 210–7.
24. Jakobsen JC, Gluud C, Wetterslev J, *et al.* When and how should multiple imputation be used for handling missing data in randomised clinical trials—a practical guide with flowcharts. *BMC Med Res Methodol* 2017; 17 (1): 162.
25. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Trans Syst Man Cybern C* 2012; 42 (4): 463–84.
26. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. *Front Med (Lausanne)* 2021; 8: 607952.
27. Singer M, Deutschman CS, Seymour CW, *et al.* The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016; 315 (8): 801–10.
28. Futoma J, Hariharan S, Heller K, *et al.* An improved multi-output gaussian process RNN with real-time validation for early sepsis detection. Vol. 68 of Proceedings of Machine Learning Research. Boston: PMLR; 2017: 243–54.
29. Bowen D, Ungar L. *Generalized SHAP: Generating Multiple Types of Explanations in Machine Learning*. New York: CoRR; 2020.
30. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. *Biometrics* 2000; 56 (4): 1134–8.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44 (3): 837–45.
32. Kaukonen K-M, Bailey M, Pilcher D, Cooper DJ, Bellomo R. Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med* 2015; 372: 1629–38.
33. Churpek MM, Zadravecz FJ, Winslow C, Howell MD, Edelson DP. Incidence and prognostic value of the systemic inflammatory response syndrome and organ dysfunctions in ward patients. *Am J Respir Crit Care Med* 2015; 192 (8): 958–64.
34. Theertha M, Sanju S, Priya VV, Jain P, Varma PK, Mony U. Innate lymphoid cells: potent early mediators of the host immune response during sepsis. *Cell Mol Immunol* 2020; 17: 1114–6.
35. Lin GL, McGinley JP, Drysdale SB, Pollard AJ. Epidemiology and immune pathogenesis of viral sepsis. *Front Immunol* 2018; 9: 2147.
36. Mao Q, Jay M, Hoffman JL, *et al.* Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open* 2018; 8 (1): e017833.
37. Alavi A, Bogu GK, Wang M, Rangan ES, et.al. Real-time alerting system for COVID-19 and other stress events using wearable data. *Nat Med* 2022; 28 (1): 175–84.
38. Pathinarupothi RK, Dhara Prathap J, Rangan ES, Gopalakrishnan EA, Vinaykumar R, Soman KP. Single sensor techniques for sleep apnea diagnosis using deep learning. In: 2017 IEEE International Conference on Healthcare Informatics (ICHI). Salt Lake City, UT: IEEE; 2017: 524–9.
39. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Per Med* 2018; 15 (5): 429–48.
40. Pathinarupothi RK, Durga P, Rangan ES. IoT-based smart edge for global health: Remote monitoring with severity detection and alerts transmission. *IEEE Internet Things J* 2019; 6 (2): 2449–62.