Applying Artificial Intelligence to Identify Physiomarkers Predicting Severe Sepsis in the PICU

Rishikesan Kamaleswaran, PhD[1,2]; Oguz Akbilgic, PhD[1]; Madhura A. Hallman, MD, MPH[2]; Alina N. West, MD, PhD[2]; Robert L. Davis, MD, MPH[1]; Samir H. Shah, MD, MBA, FRCPC[2]

Objectives: We used artificial intelligence to develop a novel algorithm using physiomarkers to predict the onset of severe sepsis in critically ill children.

Design: Observational cohort study.

Setting: PICU.

Patients: Children age between 6 and 18 years old.

Interventions: None.

Measurements and Main Results: Continuous minute-by-minute physiologic data were available for a total of 493 critically ill children admitted to a tertiary care PICU over an 8-month period, 20 of whom developed severe sepsis. Using an alert time stamp generated by an electronic screening algorithm as a reference point, we studied up to 24 prior hours of continuous physiologic data. We identified physiomarkers, including SD of heart rate, systolic and diastolic blood pressure, and symbolic transitions probabilities of those variables that discriminated severe sepsis patients from controls (all other patients admitted to the PICU who did not meet severe sepsis criteria). We used logistic regression, random forests, and deep Convolutional Neural Network methods to derive our models. Analysis was performed using data generated in two windows prior to the firing of the electronic screening algorithm, namely, 2–8 and 8–24 hours. When analyzing the physiomarkers present in the 2–8 hours analysis window, logistic regression performed with specificity of 87.4% and sensitivity of 55.0%, random forest performed with 79.6% specificity and 80.0% sensitivity, and the Convolutional Neural Network performed with 83.0% specificity and 75.0% sensitivity. When analyzing physiomarkers from the 8–24 hours window, logistic regression resulted in 77.1% specificity and 39.3% sensitivity, random forest performed with 82.3% specificity and 61.1% sensitivity, whereas the Convolutional Neural Network method achieved 81% specificity and 76% sensitivity.

Conclusions: Artificial intelligence can be used to predict the onset of severe sepsis using physiomarkers in critically ill children. Further, it may detect severe sepsis as early as 8 hours prior to a real-time electronic severe sepsis screening algorithm. (*Pediatr Crit Care Med* 2018; XX:00–00)

Key Words: artificial intelligence; deep learning; machine learning; physiologic data; predictive analytics; severe sepsis screening

Severe sepsis (SS) is a major public health issue and clinical challenge (1). Over a million Americans are hospitalized each year with SS (2), and SS occurs in approximately 8% of hospitalized children (3). SS is caused by a heightened inflammatory response to infection, quickly progressing to multiple organ failure and death (4). Earlier diagnosis with source control and early goal-directed therapy have been shown to reduce mortality; hence, generalizable approaches to early detection or prediction of SS would be an important advance (5–7). Diagnosis is often delayed in children who develop SS during hospitalization (3) in part because SS is a dynamic continuum with frequent changes in physiologic and laboratory values, making subtle presentations difficult to identify among caregivers with varying levels of clinical experience. Further, SS may be difficult to diagnose due to similar manifestations shared by other general inflammatory pathways (8).

Using variables refined by iterative, virtual processes, our team had previously designed a pediatric SS screening algorithm and tested it retrospectively on clinically validated gold standard cases (9). After implementing further modifications to this algorithm, including the application of appropriate filters, we applied the electronic SS screening algorithm integrated

[1]Center for Biomedical Informatics, Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN.

[2]Division of Pediatric Critical Care Medicine, Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/pccmjournal).

Dr. West received funding from Learn4Life (board member). Dr. Davis received funding from GlaxoSmithKline. The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: rkamales@uthsc.edu

DOI: 10.1097/PCC.0000000000001666

within our electronic health record (EHR) in real time among hospitalized children between 6 and 18 years old (10–12). The algorithm incorporated age-specific thresholds for abnormal heart rate (HR) and respiratory rate, employed a linear temperature correction for each category (**Table 1**) and achieved a positive predictive value (PPV) between 45% and 53.9%. However, this electronic screening algorithm relies in part on intermittent manual data entry potentially contributing to alert fatigue due to data entry errors. Hence, examination of an enhanced screening algorithm to further improve consistency and timeliness of SS diagnosis merits further exploration.

Investigation of patterns of variation or fluctuations in physiomarkers, particularly HR variability (HRV) analysis, may provide clinically useful information in SS (5). Several theories exist regarding the pathophysiology of altered HRV; a state of decreased overall variation, lower HR frequency variation, and decreased entropy have been shown to consistently correlate with the presence and severity of systemic infection (5). Although recent work has demonstrated that physiomarkers such as HRV can be used to identify patients at risk for SS prior to the onset of disease in the neonatal (14) and adult populations (15), similar research in the older pediatric population is lacking. In this study, we evaluated the application of machine learning and deep learning techniques on streaming physiologic data to develop prediction models of SS among children admitted to the PICU at the Le Bonheur Children's Hospital, Memphis, TN. In our analysis, we use primary physiomarkers (defined in this study as HR, blood pressures, and peripheral oxygen saturation [Spo_2]) and secondary physiomarkers (defined in this study as derived HRV, blood pressure variability, and Spo_2 variability) to improve the predictive power of our model. We hypothesized that quantifiable primary and secondary physiomarkers exist that can differentiate between patients with and without SS and that these physiomarkers can be used to accurately predict SS earlier than existing practice.

MATERIALS AND METHODS

This observational cohort study was approved by the Institutional Review Board of the University of Tennessee Health Science Center. We enrolled 493 children admitted to the PICU at Le Bonheur Children's Hospital, Memphis, TN who were screened for SS between January 2017 and September 2017 and had continuous physiologic data of at least 8 hours. We collected continuous physiologic data streams from bedside medical monitoring using the Cerner CareAware iBus (Cerner Corporation, Kansas City, MO) (16). We used minute-by-minute HR, mean blood pressure (MBP), systolic blood pressure (SBP), diastolic blood pressure (DBP), and Spo_2 data streams due to their availability; however, neither continuous respiratory rate nor temperature was available; therefore, these were excluded. Noninvasive blood pressure (NIBP) measurements data included as well as those from an arterial catheter (if available). NIBP was sampled at least once every hour. However, those measurement were repeated more frequently in the event of clinical deterioration or for monitoring during procedures. We identified all true positive alerts generated by the electronic SS screening algorithm system as validated cases, and the remaining PICU patients (all those who did not meet SS criteria which included patients who did not alert for SS and false-positive alerts) were used as controls (**Fig. 1**).

Electronic SS Screening Algorithm Alert Generation and Validation

We prospectively applied a pediatric SS screening algorithm integrated within the EHR, based upon previously described physiologic and laboratory criteria for systemic inflammatory response syndrome (SIRS) and acute organ dysfunction (9). Once a patient's EHR-based physiologic data input and/or laboratory results met criteria for SIRS and acute organ dysfunction to qualify for SS, an electronic alert was automatically generated by the algorithm and sent to the medical response team (MRT) smartphone with alert characteristics. In addition, an "SS screening" alert banner was recorded in the patient's EHR with a time stamp. The MRT team assessed the patient in real time to determine whether the patient met clinical criteria for SS/septic shock and either recommended goal-directed therapy or reconciled with ongoing medical management. Alert-based suggestions by the MRT team were recorded, although implementation of specific goal-directed recommendations was not a requirement of this study.

All alerts, along with an audit of the patient charts, were reviewed by a critical care physician, an author of this article (M.A.H.), who was blinded to the results of the machine learning model. Alerts were defined as true positive if the patient met clinical criteria for SS (as specified in Table 1 along with retrospective chart review). The alerts were defined as false positive, if the alert resulted from either incorrect data entry or physiologic characteristics that, when assessed by the electronic algorithm, mimicked SS including conditions such as acute trauma, status asthmaticus, status epilepticus, or altered mental status secondary to anesthesia or procedural sedation.

Physiomarker Feature Extraction

We calculated five descriptive statistics for each of five physiologic data, yielding 25 descriptive physiomarkers. In addition, we extracted seven probabilistic symbolic pattern recognition (PSPR) (17–19) features for each of the five data streams (**Fig. 2**), yielding an additional 35 PSPR physiomarkers, for a total of 60 physiomarkers. In order to compute the reference set for PSPR, we selected a subset consisting of a random fixed number of controls ($n = 50$). To avoid introducing bias, patients included in the reference data series were excluded in the training and test phases. We separately extracted these features from both the 2–8 hours and 8–24 hours segments (Fig. 2). We performed both nonparametric Mann-Whitney *U* and parametric independent sample *t* tests to assess whether the differences in the distribution of these features between SS and non-SS subjects was statistically significant.

Classification and Feature Selection

All extracted physiomarkers were used as predictors of SS using logistic regression (LR) and the machine learning algorithm random forest (RF) (20). RF is an ensemble of multiple simple decision trees (for details of RF algorithm, see **Appendix 1**, Supplemental Digital Content 1, http://links.lww.com/PCC/A707). In the LR model, we implemented stepwise feature selection

TABLE 1. Criteria to Generate Code Logic Variables Used to Design an Electronic Severe Sepsis Algorithm From Systemic Inflammatory Response Syndrome and Major Organ Dysfunction Syndrome Criteria[ab]

Systemic Inflammatory Response Syndrome (Two or More of the Following)	Major Organ Dysfunction
Required vital sign abnormalities	Either cardiovascular and/or respiratory compromise
Heart rate: pulse < 60 or > 140 within 24 hr (6–12 yr) or pulse < 60 or > 130 within 24 hr (13–18 yr)	Cardiovascular: systolic blood pressure < 70 ([+2[a] age in yr] 6–12 yr) or < 90 yr (13–18 yr) in the last 24 hr
Respiratory rate < 12 and > 20, within 24 hr	Or: vasoactive agent administration
"And/or" any of the following abnormalities	Or: base excess < −5 mEq/L
Temperature (excluding axillary): < 36°C or > 38.5°C	Or: mean arterial pressure < (40 + 1.5 [age])
Axillary temperature: < 35.4°C or > 37.9°C	And: lactic acid > 4 mmol/L
WBC: < 4.5 or > 13.5 yr (6–12 yr); < 4.5 or > 11 yr (13–18 yr)	Or: arterial pH, capillary pH < 7.3, and maximum value within the last 48 hr
Band neutrophils: > 10% within the last 24 hr	Or: venous pH ≤ 7.27 and maximum value within the last 48 hr
	Respiratory: supplemental O_2 requirement > 50% to maintain Sao_2 > 90%
	Or: mechanical ventilation within past 24 hr (unless patient is on chronic ventilatory support)
	And/or:
	Compromise of two other organ systems
	Renal: blood urea nitrogen/C4 creatinine level > 30
	Hepatic: bilirubin ≥ 4.0 mg/dL, in the last 48 hr
	And: alanine aminotransferase > 72 "or" maximum value
	And: aspartate aminotransferase > 92 "or" maximum value in the last 48 hr
	And:
	Hematologic: platelet count < 50,000 "or" maximum value in the last 48 hr
	Or: INR and/or prothrombin time = INR > 1.5 and maximum value within 48 hr
	Or: prothrombin > 18.5 and maximum value
	Neurologic: altered level of central nervous system

INR = international normalized ratio.

[a]Clinical suspicion was not a requirement of this algorithm to avoid subjective lack of caregiver awareness of severe sepsis.

[b]Modified from Goldstein et al (13) and Sepanski et al (9).

methods to identify important physiomarkers of SS. In the RF model, we used genetic algorithm (GA), a heuristic algorithm mimicking natural selection to find near-optimal values of a given fitness function over several generations. GA has been applied to many machine learning methods for the purpose of variable selection (21, 22). We performed GA over the RF model with five-fold cross-validation (for details of cross-validation, see Appendix 1, Supplemental Digital Content 1, http://links.lww.com/PCC/A707) to reduce model complexity and to avoid overfitting.

Deep Learning Using Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a class of hierarchical neural networks consisting of convolutional and subsampling layers that alternate to replicate the human visual system (for details of CNNs, see Appendix 1, Supplemental Digital Content 1, http://links.lww.com/PCC/A707). In our approach to CNN, we generated a multidimensional array of all five physiologic data streams for each patient, as illustrated in Figure 2. We had previously demonstrated its utility in predicting both adult sepsis (23) and cardiac arrhythmias (24). Each physiologic measurement was relatively aligned by a delta expressed as minutes before an alert. A filter variable with a depth of five was then used to perform convolutions over all five physiologic data streams simultaneously. This ensures that the temporal locality of the critically ill patients physiology is considered in the receptive field of the CNN architecture.

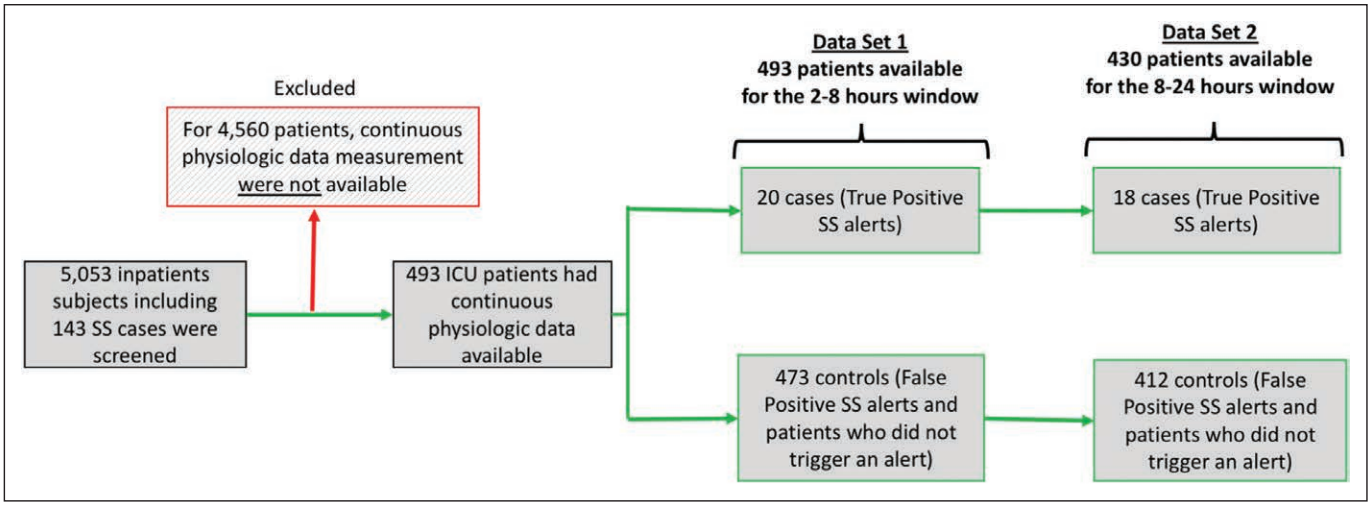

Figure 1. A flow chart of the patients included in the study. SS = severe sepsis.

The dataset was split into 70% training and 30% hidden validation sets. The training set was then further split into five-fold subsets for cross-validation. A model was generated after each cross-validation fold, and the same hidden test set (30% of the original data) was used to evaluate each of the generated models. The model that provided the optimal accuracy, specificity, and sensitivity on the validation dataset was selected as the final model. We aimed to achieve a specificity of at least 75% to ensure accuracy, while minimizing alert fatigue.

RESULTS

Data Characterization

Of the 5,053 patients admitted to the hospital during the application of the electronic SS screening algorithm for inhospital admissions, 4,560 were excluded due to lack of available continuous streaming physiologic data (only available in the PICU). This reduction left a total of 493 patients with continuous data available between 2 and 8 hours. Of these, 430 had continuous data available for 8–24 hours (Fig. 1). Of the 493 patients in the 2–8 hours segment, 20 patients were identified using the electronic SS screening algorithm (9, 10, 12) (Table 1) and confirmed to be true positive for SS based on clinical review. Of the 20 true positive SS patients, two patients did not have data in the 8–24 hours segment, yielding 18 true positive SS patients who had continuous data throughout.

For the machine learning component, we employed template matching features. We randomly selected a subset of 50 non-SS patients for generating a reference database and removed these from the 473-patient control group that was used for analysis, resulting in 423 non-SS patients. In the deep learning component, we did not use a reference subset, thus the number of controls was kept at 473 patients.

Feature Extraction

Of the 60 physiomarkers, 37 had statistically significantly different distributions for SS and non-SS patients. Of these 37 physiomarkers, 35 were PSPR features and the other two were average and minimum HR. **Figure 3** shows the distribution of these two significant HR features across non-SS and SS cases. HR was statistically significantly higher in children who developed SS compared with controls.

Although all 35 PSPR-based features were statistically significant for the SS and non-SS cases, they were highly correlated to each other for each of the five physiologic data streams. Therefore, we ran principle component analysis (PCA) on seven PSPR physiomarkers separately for each physiologic series. For each series, we obtained one principle component representing all seven PSPR features from each physiologic data stream; all had explained variation over 95%. Hence, the total number of physiomarkers was dropped to 30, 25 descriptive and five PCA-PSPR physiomarkers (one per data stream). **Supplemental Figure 1** (Supplemental Digital Content 2, http://links.lww.com/PCC/A708; legend: Statistical analysis of features obtained by applying PCA on PSPR across non-SS and SS cases) shows the distribution of PCA-PSPR features across non-SS and SS patients; PCA-PSPR–based physiomarkers significantly differ for SS cases versus non-SS controls.

LR Model

We ran stepwise selection in two separate LR models to predict the occurrence of SS at a minimum of 1) 2 hours earlier than existing clinical screening methods, using the 2–8 hours segment data and 2) 8 hours earlier than existing clinical screening methods, using the 8–24 hours segment data. For the 2–8 hours segment data, our stepwise LR model selected three variables, namely PCA-PSPR of HR, SD of HR, and maximum DBP (for the details of the model, see **Appendix 2**, Supplemental Digital Content 3, http://links.lww.com/PCC/A709), predicting SS at least 2 hours earlier than the current electronic SS screening algorithm system. This model yielded area under the curve (AUC) of 0.89 with 95% CI of (0.81–0.97) and high classification performance, detailed in **Table 2**. However, when we implemented five-fold cross-validation, the AUC statistic dropped to 0.77, 95% CI (0.63–0.91) with specificity of 87.4% and sensitivity of 55%. For the 8–24 hours segment data, we

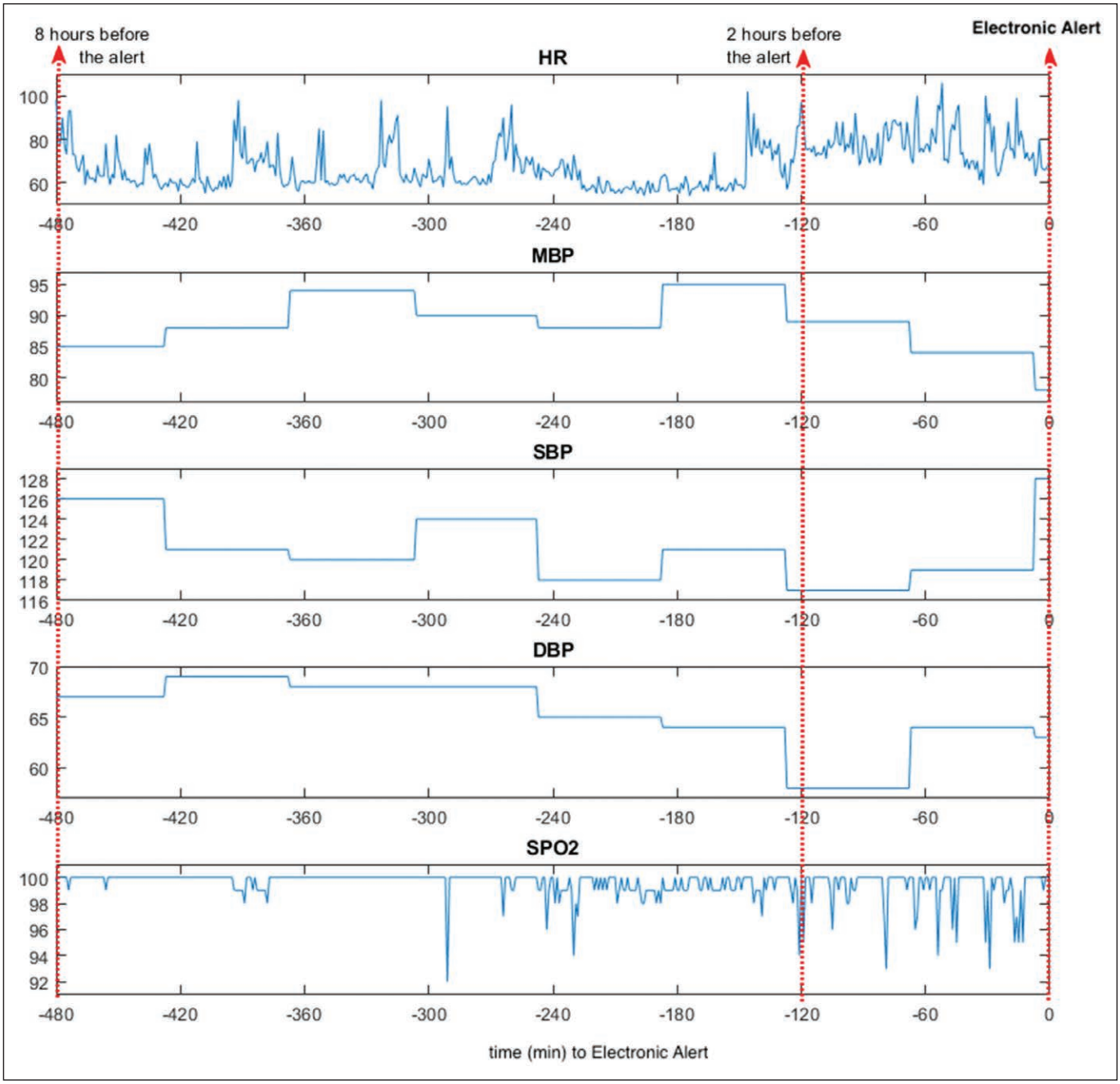

Figure 2. Five streams of physiologic data used to identify early physiomarkers predicting the onset of severe sepsis (SS). This figure illustrates data from a patient who developed SS. The *x*-axis represents relative alignment of physiologic data in minutes, where zero represents nearest to the electronic SS screening algorithm alert and −500 represents the physiologic status 8.5 hr before the alert. DBP = diastolic blood pressure, HR = heart rate, MBP = mean blood pressure, SBP = systolic blood pressure, Spo_2 = oxygen saturation.

investigated the accuracy of another stepwise LR model using the data before the electronic SS screening algorithm alert. This model yielded an AUC of 0.81, 95% CI (0.71–0.91), overall accuracy of 76.3%, specificity of 76.8%, sensitivity of 66.7%, PPV of 12.5%, and negative predictive value (NPV) of 97.9% by selected predictors of SD and PCA-PSPR of DBP, minimum MBP, average HR, and minimum and coefficient of variation of SBP (for the details of the model, see Appendix 2, Supplemental Digital Content 3, http://links.lww.com/PCC/A709). When we implemented five-fold cross-validation, the model performance dramatically decreased with an AUC of 0.56, 95% CI (0.39–0.76), specificity of 77.1%, and sensitivity of 39.3%. We have also calculated positive likelihood ratio (25) with 95% CI as 1.72 (0.40–7.39) and negative likelihood ratio of 0.79 (0.55–1.15) indicating that the LR model for 8–24 hours data is not significant since the CI includes 1. Note that the LR model with five-fold cross-validation for 2–8 hours data was found to be significant with both positive likelihood ratio of 4.36 (1.95–9.76) and negative likelihood ratio of 0.51 (0.32–0.84).

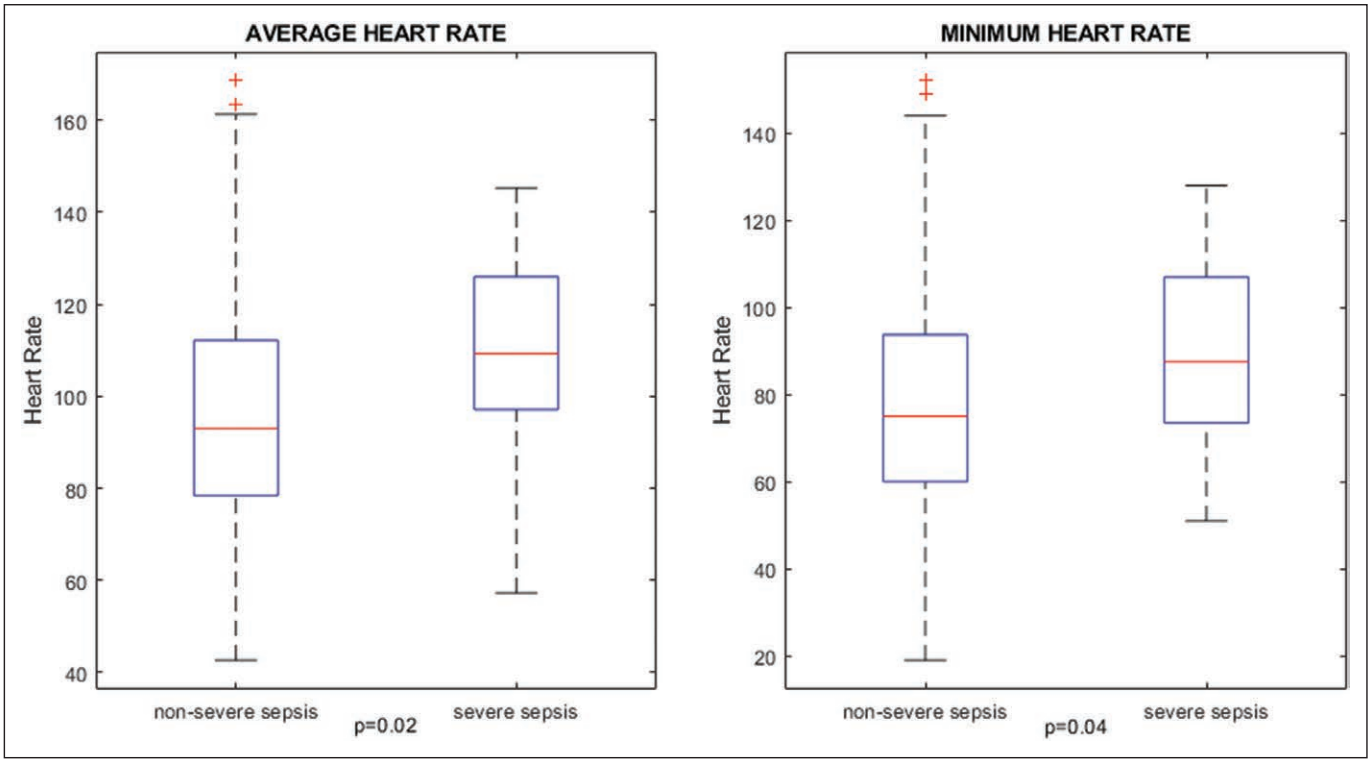

Figure 3. Boxplot of two heart rate features across non-severe sepsis (SS) and SS cases.

TABLE 2. Accuracy Statistics With Data 2–8 Hours Prior to the Severe Sepsis Electronic Screening Algorithm Alert

Machine Learning Methods	Predicted		Total	Accuracy
	Non-SS	SS		
Logistic regression				
Actual				
Non-SS	350	73	423	Specificity = 82.7%
SS	3	17	20	Sensitivity = 85.0%
Total	353 (NPV = 99.1%)	90 (PPV = 18.9%)	443	Overall accuracy = 82.8%
Random forest				
Actual				
Non-SS	337	86	423	Specificity = 79.6%
SS	4	16	20	Sensitivity = 80.0%
Total	341 (NPV = 98.8%)	102 (PPV = 15.7%)	443	Overall accuracy = 79.7

NPV = negative predictive value, PPV = positive predictive value, SS = severe sepsis.

RF Model

We further analyzed our data using RF with 1,100 learners and a maximum split of 220 (refer to Appendix 1, Supplemental Digital Content 1, http://links.lww.com/PCC/A707). We embedded the RF model into GA-based feature selection with population size of 70 with 50 generations. We applied a five-fold cross-validation strategy to ensure that the model was generalizable and to reduce selection bias. Using the same 30 physiomarkers from the 2–8 hours segment, the final five-fold cross-validation RF model consisted of 14 physiomarkers (Table 2). We further evaluated the variable importance of the 14 features used in the RF model (**Fig. 4**). We normalized the importance values by scaling the maximum value to 100 to make comparisons easier. We repeated the RF model on the features extracted from 8 to 24 hours before electronic SS screening algorithm alert. The five-fold cross-validated RF model provided an accuracy of 81.3%, specificity of 82.3%, sensitivity of 61.1%, PPV of 14.7%, and NPV of 97.7. Both RF

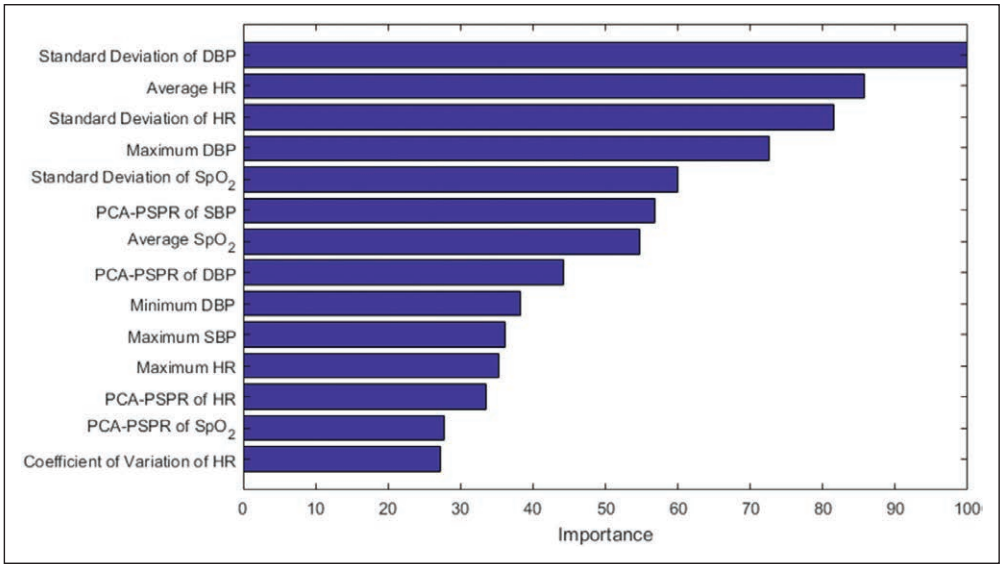

Figure 4. Variable importance analysis for the random forest model illustrating the order of importance of each variable that was selected. DBP = diastolic blood pressure, HR = heart rate, PCA-PSPR = principle component analysis-probabilistic symbolic pattern recognition, SBP = systolic blood pressure, Spo_2 = oxygen saturation.

models, for two to eight and eight to 24 segments, found to be significant with positive and negative likelihood ratios with 95% CI of 1.96 (3.13–4.91) and 3.45 (1.97–6.05), respectively.

CNNs

The CNN architecture was trained on a five-fold cross-validation set abstracted from 70% of the original data (with 381 controls and 17 SS cases in each fold), with 30% of the original data kept aside for a hidden test of the cross-validated models. For the 2–8 hours segment, the mean accuracy in the validation phase was 83%, with a mean sensitivity and specificity of 75% and 83%, respectively. Unlike in the traditional approaches, the CNN performed better during the 8–24 hours segment. For that segment, the mean test accuracy across all five models was 80%, with mean sensitivity and specificity of 89% and 80%. We then evaluated each of the models against a hidden test set generated from 30% of the original data before the cross-validation; the hidden test set (8–24 hr) contained 125 controls and five SS cases. In the hidden test set, the mean accuracy was 81%, mean sensitivity was 76%, and mean specificity was 81% (**Table 3**). CNN model for 2–8 hours segment found to be significant with positive and negative likelihood ratios with 95% CI of 4.41 (2.64–7.36) and 0.24 (0.14–0.42), respectively. However, CNN model for 8–24 hours segment found to be significant based on positive likelihood ratio (4.00 [1.88–8.51]) and not significant based on negative likelihood ratio (0.25 [0.04–1.43]).

DISCUSSION

Our study found distinct quantifiable physiomarkers, such as SD of DBP, SD of HR, and PCA-PSPR of SBP and DBP, that uniquely identify critically ill children with SS well before they develop clinically recognizable characteristics. We further determined that our methods can also predict SS in patients up to 8 hours earlier than a currently implemented electronic SS surveillance algorithm. Our findings indicate a significant opportunity for bedside monitors to be integrated with artificial intelligence to enhance real-time monitoring of SS without reliance on asynchronous data entry within the EHR.

Neonatal studies have demonstrated observable changes in HR characteristics as early as 3–4 days before the onset of SS (26). Using the Kolmogorov-Smirnov test, Cao et al (27) demonstrated that HR becomes more nonstationary with reduced variability and transient decelerations prior to neonatal sepsis. Lake et al (28) introduced the sample entropy method as an algorithm for identifying sepsis up to 24 hours before clinical recognition. Adult sepsis prediction models such as the Targeted Real-Time Early Warning Score (TREWScore) (29)

TABLE 3. Results From Five-Fold Cross-Validation Applied to a Three-Layer Convolutional Neural Network

	Cross-Validation (Five-Fold)			Hidden Test Set (Separate)		
Fold	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	86.5	84.6	86.5	89.2	60.0	90.4
2	79.9	92.3	79.3	79.2	100.0	78.4
3	76.2	84.6	75.9	75.4	100.0	74.4
4	68.3	100.0	66.9	72.3	80.0	72.0
5	90.8	84.6	91.0	89.2	40.0	91.2
Average	80.34	89.22	79.92	81.06	76 (positive predictive value = 47.4%)	81.28 (negative predictive value = 93.84%)

identified sepsis up to 30 hours before clinical signs using EHR data on over 400,000 patients. Kam and Kim (30) used the same dataset to predict sepsis using Multilayer Perceptron and Recurrent Neural Networks up to 3 hours after initial SIRS value greater than two is observed. However, a major limitation of those studies is their reliance on data that comes, on average, only once every hour, and in some cases, once every 24 hours. This results in losing important physiologic information pertaining to the patient within those intervals. Although Ahmad et al (15) used continuous electrocardiogram data to identify the relationship between reduced HRV changes and sepsis in adult transplant patients, our article represents the first pediatric (nonneonate) SS prediction algorithm using continuous physiologic data captured at minute-to-minute intervals.

Four out of five PCA-PSPR based features were selected in the final RF model for the 2–8 hours segment. This can be interpreted as evidence that temporal and morphological changes in physiologic data can indicate the possible progression to SS in critically ill children. We further implemented LR and RF models with five-fold cross-validation on the data from 8 to 24 hours before the electronic SS screening alert. Although the RF model yielded an accuracy of 81.3%, the LR for 8–24 hours data not perform as well in cross-validation in terms of specificity and sensitivity. However, we acknowledge that comparing LR with machine learning algorithms simply based on specificity and sensitivity statistics may not be sufficiently informative of model performance (31, 32). Therefore, we also calculated the positive and negative likelihood ratios for LR models and also for RF and CNN models for consistency. We found that all three models were significant for 2–8 hours segment, and only RF model was significant for 8–24 hours segment as per the likelihood ratios. The biological reasoning for nonsignificant LR and CNN models for 8–24 hours segment may be that the signature of sepsis may be less apparent or more complex in this segment compared with 2–8 hours segment. We also posit that poor performance of LR for 8–24 hours data is potentially due to two factors. First, the number of cases ($n = 18$) was very low considering the larger number of predictors. Second, the underlying statistical relationship between physiologic data beyond 8 hours prior to sepsis onset and the occurrence of SS may be nonlinear, and if this is the case, a generalized linear model such as LR is not suitable for this application.

The deep learning technique demonstrated visible discrimination between SS and non-SS populations. The deep learning technique achieved an accuracy of 79%, with specificity of 78% and sensitivity of 100% in a hidden test set. In the training set, the model accurately predicted 12 of the 13 SS cases 8 hours before the electronic SS screening algorithm alert. We derived the CNN model by biasing the model for maximum sensitivity with an aim of alerting clinicians earlier than currently existing algorithms, thereby enabling earlier administration of goal-directed therapies. One of the key strengths of the CNN approach is that it has been shown to continuously improve with additional training data (33); hence, with the inclusion of additional examples, the specificity may be further improved. We plan to progress this approach in future work, where a significantly larger cohort of SS patients can be recruited. Furthermore, as opposed to traditional machine learning approaches, which rely on manual derivation of features, the deep learning CNN method performs automatic feature extraction using raw data as input. This presents a key advantage, as manual derivation of features can be a complicated and time-consuming process, often influenced by the biases of the feature engineer (34–36).

There are limitations to our results. We developed our models on only 20 children who developed SS in the PICU and who had at least 8 hours of continuous data prior to the clinical validation of SS at the bedside. Given our limited sample size, our model may not capture other potential physiomarkers predictive of SS. Furthermore, given this limited sample size, a small change in the prediction can significantly alter the sensitivity by as far as 5%. In order to overcome this limitation, we plan to continually train our model on a quarterly basis, each time increasing the number of positive and negative cases prior to its prospective, real-time application. In addition, we did not include children under the age of 6, due to the significant complexity that can be introduced by the physiologic heterogeneity that exists within this age group (37). In future work, we plan to study this group and develop new models that encompass children between 1 and 18 years old. We acknowledge that our findings may be affected due to other clinical variables that may have altered the physiomarkers, including medications such as vasopressors or sedation/analgesia. Nevertheless, although the impact of adjuvant therapies would be difficult to exclude, the occurrence of physiomarker variability despite those interventions attest to salient physiologic changes in this patient population and will need further evaluation during prospective studies. Our study used a single clinical reviewer to determine the true clinical state of children in the PICU; therefore, bias may have been introduced due to lack of external validation of the clinical determination. Finally, our data are limited to a single PICU in the mid-south region with predominant African American children accounting for ~50% of admissions; thus, our data might not be widely generalizable to other pediatric populations of similar age groups.

In conclusion, we present results from LR (without cross-validation), machine learning, and deep learning (with cross-validation) on physiomarkers to develop prediction models of SS in critically ill children admitted to the PICU. We use five commonly captured physiologic data streams, including HR, SBP, DBP, MBP, and Spo_2. Our results indicate that quantifiable physiomarkers exist that are predictive of the onset of SS in critically ill children when more granular minute-to-minute continuous physiologic data are used. Furthermore, we demonstrate that our methods are able to use these physiomarkers to identify SS at least 8 hours before an existing hospital deployed electronic SS screening algorithm triggers an alert. The results provide further evidence that early identification of SS is possible in the PICU by applying artificial intelligence using machine learning methods, thereby presenting an opportunity for earlier administration of goal-directed therapy using continuous physiologic data streams.

ACKNOWLEDGMENTS

We would like to acknowledge the efforts of Kimberley Giles, Amy Guynn, Michael Younker, Brian Williams, Don MacMillan for their work in preparing and providing key data elements that were used in this article. We would also like to extend our sincere thanks to the associate editor and anonymous reviewers for their constructive suggestions and comments resulting in a significant improvement to our article.

REFERENCES

1. Proulx F, Fayon M, Farrell CA, et al: Epidemiology of sepsis and multiple organ dysfunction syndrome in children. *Chest* 1996; 109:1033–1037
2. Dombrovskiy VY, Martin AA, Sunderram J, et al: Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: A trend analysis from 1993 to 2003. *Crit Care Med* 2007; 35:1244–1250
3. Weiss SL, Fitzgerald JC, Pappachan J, et al; Sepsis Prevalence, Outcomes, and Therapies (SPROUT) Study Investigators and Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network: Global epidemiology of pediatric severe sepsis: The sepsis prevalence, outcomes, and therapies study. *Am J Respir Crit Care Med* 2015; 191:1147–1157
4. Liu V, Escobar GJ, Greene JD, et al: Hospital deaths in patients with sepsis from 2 independent cohorts. *JAMA* 2014; 312:90–92
5. Ahmad S, Tejuja A, Newman KD, et al: Clinical review: A review and analysis of heart rate variability and the diagnosis and prognosis of infection. *Crit Care* 2009; 13:232
6. Rivers E, Nguyen B, Havstad S, et al; Early Goal-Directed Therapy Collaborative Group: Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345:1368–1377
7. 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; 34:1589–1596
8. Alberti C, Brun-Buisson C, Burchardi H, et al: Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. *Intensive Care Med* 2002; 28:108–121
9. Sepanski RJ, Godambe SA, Mangum CD, et al: Designing a pediatric severe sepsis screening tool. *Front Pediatr* 2014; 2:56
10. West A, Mangubat PM, Giles K, et al: 1343: Analyzing accuracy and early determinants of severe sepsis using real-time EMR integrated algorithm. *Crit Care Med* 2016; 44:411
11. Shah SH, Sepanski RJ, Mangum C, et al: Performance of a Proposed Pediatric Severe Sepsis Screening Tool and Review of Severe Sepsis Outcome. *In*: Pediatric Academic Societies Meeting, Boston, MA, April 28–May 1, 2012
12. West A, Hallman M, Giles K, et al: 1540: Accuracy of detecting clinically relevant severe sepsis in children using a real-time EMR algorithm. *Crit Care Med* 2018; 46:753
13. Goldstein B, Giroir B, Randolph A: International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics*. *Pediatr Crit Care Med* 2005; 6:2–8
14. Griffin MP, Moorman JR: Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. *Pediatrics* 2001; 107:97–104
15. Ahmad S, Ramsay T, Huebsch L, et al: Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. *PLoS One* 2009; 4:e6642
16. Cerner Corporation: CareAware iBus: Increasing Efficiency and Eliminating Error, 2014. Available at: https://www.cerner.com/pages/careaware. Accessed July 10, 2018
17. Akbilgic O, Howe JA: Symbolic pattern recognition for sequential data. *Seq Anal* 2017; 36:1–13
18. Mahajan R, Kamaleswaran R, Akbilgic O: Effects of varying sampling frequency on the analysis of continuous ECG data streams. VLDB Workshop on Data Management and Analytics for Medicine and Healthcare. Cham, Springer, 2017
19. Mahajan R, Viangteeravat T, Akbilgic O: Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics. *Int J Med Inform* 2017; 108:55–63
20. Ho TK: The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 1998; 20:832–844
21. Akbilgic O, Bozdogan H, Balaban ME: A novel hybrid RBF neural networks model as a forecaster. *Stat Comput* 2014; 24:365–375
22. Eiben AE, Smith JE: Introduction to Evolutionary Computing. Berlin, Springer, 2015, pp 37–69
23. Van Wyk F, Khojandi A, Kamaleswaran R, et al: How Much Data Should We Collect? A Case Study in Sepsis Detection Using Deep Learning. *In*: NIH-IEEE 2017 Special Topics Conference on Healthcare Innovations and Point of Care Technologies: Technology in Translation, Bethesda, MD, November 6–8, 2017
24. Kamaleswaran R, Mahajan R, Akbilgic O: A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length. *Physiol Meas* 2018; 39:035006
25. Katz D, Baptista J, Azen SP, et al: Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics* 1978; 34:469
26. Kovatchev BP, Farhy LS, Cao H, et al: Sample asymmetry analysis of heart rate characteristics with application to neonatal sepsis and systemic inflammatory response syndrome. *Pediatr Res* 2003; 54:892–898
27. Cao H, Lake DE, Griffin MP, et al: Increased nonstationarity of neonatal heart rate before the clinical diagnosis of sepsis. *Ann Biomed Eng* 2004; 32:233–244
28. Lake DE, Richman JS, Griffin MP, et al: Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol* 2002; 283:R789–R797
29. Henry KE, Hager DN, Pronovost PJ, et al: A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; 7:299ra122
30. Kam HJ, Kim HY: Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med* 2017; 89:248–255
31. Provost F, Fawcett T, Kohavi R: The Case Against Accuracy Estimation for Comparing Induction Algorithms. *In*: ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning, San Francisco, CA, July 24–27, 1998
32. Davis J, Goadrich M: The Relationship Between Precision-Recall and ROC Curves. *In*: Proceedings of the 23rd International Conference on Machine Learning - ICML '06, Pittsburgh, PA, June 25–29, 2006
33. Simard PY, Steinkraus D, Platt JC: Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. *In*: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings, Edinburgh, United Kingdom, August 6, 2003
34. Bengio Y, Courville AC, Vincent P: Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR* 2012
35. Bengio Y, Courville A, Vincent P: Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2013; 35:1798–1828
36. Langkvist M, Karlsson L, Loutfi A: A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit Lett* 2014; 42:11–24
37. Rijnbeek PR, Witsenburg M, Schrama E, et al: New normal limits for the paediatric electrocardiogram. *Eur Heart J* 2001; 22:702–711