

Predicting the Onset of Sepsis Using Vital Signs Data: A Machine Learning Approach

Clinical Nursing Research
1–10
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10547738231183207
journals.sagepub.com/home/cnr



An Tran, MS¹, Robert Topp, RN, PhD²,
Ebrahim Tarshizi, PhD, MBA³, and Anthony Shao¹

Abstract

Sepsis is a major cause of mortality among hospitalized patients. Existing sepsis prediction methods face limitations due to their reliance on laboratory results and Electronic Medical Records (EMRs). This work aimed to develop a sepsis prediction model utilizing continuous vital signs monitoring, offering an innovative approach to sepsis prediction. Data from 48,886 Intensive Care Unit (ICU) patient stays were extracted from the Medical Information Mart for Intensive Care -IV dataset. A machine learning model was developed to predict sepsis onset based solely on vital signs. The model's efficacy was compared with the existing scoring systems of SIRS, qSOFA, and a Logistic Regression model. The machine learning model demonstrated superior performance at 6 hrs prior to sepsis onset, achieving 88.1% sensitivity and 81.3% specificity, surpassing existing scoring systems. This novel approach offers clinicians a timely assessment of patients' likelihood of developing sepsis.

Keywords

artificial intelligence, intensive care unit, infections, diseases, disease prevention, acute care setting

Introduction

Sepsis is a severe and life-threatening systemic infection that can rapidly progress to septic shock and death. Sepsis is a significant contributor to hospital admissions and is associated with considerable morbidity and mortality. In the United States alone, sepsis affects an estimated 1.7 million patients and results in 350,000 deaths annually (Centers for Disease Control and Prevention, 2016). When compared to patients hospitalized for non-sepsis conditions, individuals with sepsis have a 75% longer average length of stay, leading to substantially higher healthcare costs (Paoli et al., 2018). Furthermore, sepsis represents the most expensive condition treated in hospitals, accounting for 6.2% of aggregate costs for all hospitalizations, despite comprising only 3.6% of hospital stays (Torio & Moore, 2016).

Early identification of signs of sepsis development by clinicians in the hospital is crucial to the timely initiation of treatment to reduce morbidity and mortality associated with this condition. Treatment of sepsis involves fluid and antibiotic administration as soon as signs of sepsis are identified. Previous investigators have reported that during the 6-hour period following the identification of sepsis, each hour of

delay in treatment is associated with an average decrease in survival rate of 7.6% (Kumar et al., 2006). The Systemic Inflammatory Response Syndrome (SIRS) criteria have been considered the gold standard for diagnosing sepsis, requiring any two of the following criteria: body temperature over 38°C or under 36°C, heart rate greater than 90 beats/min, respiratory rate greater than 20 breaths/min or partial pressure of CO₂ less than 32 mmHg, and leukocyte count greater than 12,000 or less than 4,000/microliters or over 10% immature forms or bands (Levy et al., 2003). However, in recent years, the definition of sepsis has been updated to better identify patients at risk for poor outcomes. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) defines sepsis as a life-threatening organ dysfunction caused by a dysregulated host response to

¹Darroch Medical Solutions, Inc., San Diego, CA, USA

²University of Toledo, OH, USA

³University of San Diego, CA, USA

Corresponding Author:

An Tran, Darroch Medical Solutions, Inc., 4445 Eastgate Mall, Ste 200, San Diego, CA 92121, USA.

Email: antran@darrochmedical.com

Table 1. The Sequential Organ Failure Assessment (SOFA) Score.

Variables	SOFA score				
	0	1	2	3	4
Respiratory (mmHg)	PaO ₂ /FiO ₂ > 400 SpO ₂ /FiO ₂ > 302	PaO ₂ /FiO ₂ < 400 SpO ₂ /FiO ₂ < 302	PaO ₂ /FiO ₂ < 300 SpO ₂ /FiO ₂ < 221	PaO ₂ /FiO ₂ < 200 SpO ₂ /FiO ₂ < 142	PaO ₂ /FiO ₂ < 100 SpO ₂ /FiO ₂ < 67
Cardiovascular (doses in mcg/kg/min)	MAP ≥ 70 mmHg	MAP ≥ 70 mmHg	Dopamine ≤ 5 or ANY dobutamine	Dopamine > 5 or Norepinephrine ≤ 0.1 or Phenylephrine ≤ 0.8	Dopamine > 15 or Norepinephrine > 0.1 or Phenylephrine > 0.8
Liver (bilirubin, mg/dL)	< 1.2	1.2–1.9	2.0–5.9	6.0–11.9	> 12
Renal (creatinine, mg/dL)	< 1.2	1.2–1.9	2.0–3.4	3.5–4.9	> 5.0
Coagulation (platelets × 10 ³ /mm ³)	≥ 150	< 150	< 100	< 50	< 20
Neurologic (Glasgow Coma Score)	15	13–14	10–12	6–9	< 6

MAP=mean arterial pressure; SOFA=Sequential Organ Failure Assessment.

infection. Organ dysfunction is quantified using the Sequential Organ Failure Assessment (SOFA) score, and a patient is considered to have sepsis if there is a two-point or more increase in their SOFA score due to infection (Singer et al., 2016). This updated definition emphasizes the importance of early detection and intervention for sepsis management, leading to the replacement of the SIRS criteria with the Sepsis-3 criteria for several key reasons.

One of the most significant reasons is the improved specificity of the Sepsis-3 criteria. Although SIRS criteria are sensitive to detecting inflammation or infection in general, they are not specific enough to differentiate sepsis from other inflammatory conditions. This low specificity can lead to overdiagnosis and overtreatment of patients who do not have sepsis but meet the SIRS criteria due to other non-septic inflammatory processes. Conversely, the Sepsis-3 criteria offer a more specific definition by identifying sepsis as life-threatening organ dysfunction caused by a dysregulated host response to infection. Additionally, the incorporation of the SOFA score, which evaluates six different organ systems, including respiratory, cardiovascular, liver, coagulation, renal, and neurological (see Table 1), provides a comprehensive assessment of a patient's condition and allows for more targeted and effective treatment plans. The transition from the SIRS criteria to the Sepsis-3 criteria reflects advancements in understanding the pathophysiology of sepsis and the need for a more precise diagnostic tool (Singer et al., 2016).

In addition to the SOFA score, the qSOFA (quick Sequential Organ Failure Assessment) score is designed for the rapid identification of patients with suspected infection who are at higher risk of developing sepsis outside of intensive care units (ICUs). The qSOFA score includes three criteria: altered mental status, systolic blood pressure below 100 mmHg, and respiratory rate above 22 breaths/min. A qSOFA score of two or higher indicates an increased risk of poor outcomes and suggests that the patient may require further evaluation and more intensive care. While the SOFA score assesses the severity of organ dysfunction in patients

with sepsis or other critical illnesses, the qSOFA score serves as a quick screening tool for non-ICU settings to identify patients with a higher risk of sepsis-related complications (Singer et al., 2016).

Prior research utilizing manual computations for sepsis scoring systems in sepsis detection has encountered inherent challenges, such as errors in transcription and calculation, delays in identification and treatment due to intermittent calculations, and an inability to examine trends in patient data or connections between measurements (Gyang et al., 2015). While automated screening tools and technology integrated into electronic medical records (EMRs) have been suggested as efficient means for identifying sepsis risks, the dependence on non-immediate lab results and delays in recording vital signs into the EMR can hinder their timeliness (Despins, 2017). To date, no effective methodology exists for predicting the onset of sepsis among hospitalized patients using continuous vital sign monitoring for timely assessments.

Objectives

The primary aim of this innovative study was to develop a sepsis prediction model using data exclusively available from continuous vital signs monitoring, applying robust machine learning methods to create a novel and original approach to sepsis prediction. A secondary aim was to compare the performance of this model with established scoring systems, including qSOFA and SIRS criteria, as well as a baseline logistic regression (LR) model in predicting the development of sepsis among hospitalized patients.

Methodology

Dataset

In this study, data were extracted from the Medical Information Mart for Intensive Care (MIMIC)-IV database, an open-access resource developed by the MIT Lab for

Computational Physiology (Johnson et al., 2023). The MIMIC-IV database contains demographics, vital signs, laboratory tests, and medications for ICU patients at Beth Israel Deaconess Medical Center between 2008 and 2019. Patient privacy was maintained by removing all Protected Health Information in compliance with Health Insurance Portability and Accountability Act (HIPAA) standards; therefore, IRB oversight of this study was not required.

The study included adult patients aged 18 years of age and older. Vital signs, including heart rate, temperature, respiratory rate, SpO₂, systolic and diastolic blood pressure, were extracted from the MIMIC-IV database for the included hospitalized patients. Leukocyte (white blood cell) counts and altered mental status were also extracted from the dataset for SIRS and qSOFA score comparisons. In cases where multiple readings for a vital sign were taken within an hour, the patient's vital sign measurements were averaged on an hourly basis for each patient. If a patient's vital sign measurement was missing for a specific hour, the absent data points were imputed using the patient's most recent available data. In this study, the last observation carried forward, and the next observation carried backward methods were employed for imputing missing values. These approaches are commonly used in healthcare research to address missing data where some follow-up observations may be missing, as they take into account the temporal nature of the data and introduce less bias compared to simple mean or median imputation (Hamer & Simpson, 2009).

Patients in this dataset were monitored from the time of ICU admission until discharge or death. The suspicion of infection was established based on the presence of both a culture draw and antibiotic administration throughout the patient's entire stay (Seymour et al., 2016). To determine the time of suspicion of infection, the initial time was identified as the earliest occurrence of either a culture draw followed by antibiotic administration within 72 hrs or an antibiotic administration followed by a culture draw within 24 hrs. This approach helped identify when clinicians suspected infection in the patient. A suspicion window was then defined, extending 48 hrs before and 24 hrs after the initial time of suspicion of infection.

To determine if a patient was septic, additional assessments using sepsis criteria, such as SOFA or qSOFA scores, were required. In this study, the onset of sepsis was determined using the Sepsis-3 criteria, which define sepsis as a life-threatening organ dysfunction due to a dysregulated host response to infection, as indicated by an acute change in the total SOFA score of equal to or greater than two points resulting from infection. The SOFA score at the beginning of the suspicion window was compared to its hourly value during the window. If the hourly value was equal to or greater than two points higher than the initial value, the first such hour was deemed the onset of sepsis, and the patient was classified as septic (Seymour et al., 2016). Patients who did not experience such an event throughout their entire ICU stay

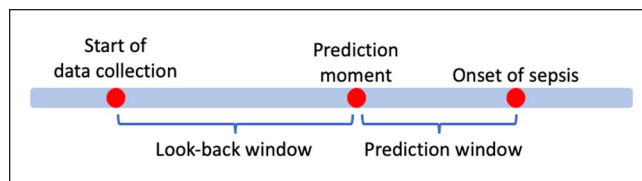


Figure 1. Time intervals for analysis.

were considered non-septic. In instances where the data needed to compute one of the SOFA sub-scores were unavailable in the imputed data, a value of 0 (i.e., “normal”) was assigned to that sub-score.

To leverage the dataset's temporal nature, various data collection window durations were employed to discern patterns that could distinguish sepsis from non-sepsis patients and predict sepsis onset hours before clinical recognition. After determining sepsis onset, vital measurements recorded after the onset were removed from the dataset. Because the primary objective of the study is to predict the onset of sepsis, any data collected after the onset is not relevant for this purpose. By focusing on the data leading up to the sepsis onset, the analysis can better identify patterns and trends that may indicate an impending sepsis event.

The period before onset is divided into two time intervals, including the look-back window, which refers to the data collection interval, and the prediction window, denoting the interval between the prediction moment and sepsis onset (see Figure 1). Sepsis patients with an onset time within 6 hrs after the start of their record were excluded to ensure the prediction windows of up to 6 hrs before onset. Look-back windows of 3, 6, and 12 hrs were assessed to predict sepsis onset from 3 to 6 hrs prior, focusing on providing clinicians with valuable lead time to prepare for and initiate appropriate interventions. Prediction windows of 0, 1, or 2 hrs before onset may not offer sufficient preparation time for effective clinical response. The final dataset consisted of hourly vital sign records for sepsis and non-sepsis patients, featuring a sepsis prevalence of 16.7% with 8,149 sepsis cases and 40,737 control cases.

Analysis Plan

Besides the initial six vital signs extracted from the dataset, two additional indicators, mean arterial pressure (MAP) and shock index (SI), were derived from the original vital signs, as they have been previously recognized as potential predictors for sepsis onset (Berger et al., 2013; Corrêa et al., 2013).

The rolling window technique was adopted to systematically extract the statistical characteristics within each window for every patient. This technique allowed for continuous analysis of vital signs data, capturing the dynamic nature of the patients' conditions. For sepsis patients, only the last rolling window was considered, as it encompassed the onset of sepsis, which is the critical point for prediction. Including

earlier windows does not contribute to the immediate prediction of sepsis onset, as the focus of the study was on the period immediately preceding the event. For non-sepsis patients, look-back windows before the end of their record were used to maintain consistency in the analysis approach with sepsis patients and to capture any relevant patterns or trends in their vital signs.

Original vital signs data, along with MAP and SI, were scrutinized within each look-back window to extract the statistical characteristics, including mean, median, standard deviation, minimum, and maximum. In addition to the standard statistical features, the study introduced proprietary features that emphasized trend changes and instability in patients' vital signs, shedding light on the subtle differences between sepsis and non-sepsis look-back windows.

The proprietary features specifically focus on capturing the frequency and magnitude of trend changes in vital signs, such as the number of times a vital sign transitions from increasing to decreasing or vice versa, and the extent of these changes. The frequency of trend changes feature is calculated by identifying consecutive periods of increasing and decreasing values, counting the number of transitions between these periods, and the magnitude of trend changes measure the difference in values at each transition point. By capturing both the frequency and magnitude of trend changes, these features are crafted to quantify the level of instability in patients' vital signs and improve the model's ability to discern between the two conditions.

Model Training

The dataset used for analysis was meticulously divided into two distinct subsets, including the training and test subsets. This division process preserved an 80/20 ratio, ensuring that the original sepsis prevalence of 16.7% was maintained in both subsets. The main purpose of this partitioning was to distinguish the data used for model training from the data set aside for assessing the model's performance on previously unseen data.

Training a machine learning model entails teaching the computer to recognize patterns and generate predictions based on data. The model seeks to determine the optimal set of internal parameters, or weights, which allow the model to accurately predict the classes (in this case, sepsis, and non-sepsis) based on the features or variables inputted into the model. Throughout the training process, the model continually refines its weights by comparing its predictions to the actual classes within the training data and adjusting accordingly to minimize the discrepancy between its predictions and the true classes. The objective of this process is to reduce misclassifications and find the optimal weights that lead to the lowest possible error.

The training subset, which comprises 80% of the total patient stays, was employed to train the machine learning

models. The models learned to identify patterns and trends in vital signs data, enabling them to predict sepsis onset. To enhance performance and accuracy, a technique called 5-fold cross-validation was implemented, dividing the training subset into five equal-sized, non-overlapping subsets or "folds." The model is then iteratively trained and validated using these folds.

After the model has been fully trained and optimized using the 5-fold cross-validation technique, it is evaluated on the test set, which consists of unseen data not used during training. This test subset, constituting the remaining 20% of the dataset, allows for an unbiased assessment of the model's prediction sensitivity and specificity, providing a better understanding of its real-world effectiveness. The test set also ensures that the model generalizes well to new data, offering a reliable estimate of its real-world performance.

In this study, two machine learning models, the Explainable Boosting Machine (EBM) and LR, were employed to predict sepsis onset based on vital signs data. Once the models were trained and optimized, they were evaluated on a test subset comprising unseen data not used during training to allow for an unbiased assessment of their prediction sensitivity and specificity. The performance of these models on the test subset were compared to the two well-established scoring systems, the qSOFA and the SIRS criteria with the aim of assessing the effectiveness and accuracy of the EBM and LR models in sepsis onset prediction.

Explainable Boosting Machine. The primary machine learning model employed for sepsis onset prediction is the EBM, a decision tree-based model trained on patterns extracted from patients' vital sign characteristics (Caruana et al., 2015). EBM was selected for its powerful interpretability, making it highly suitable for high-stake decision fields like healthcare settings. The main idea behind this model is to create a Generalized Additive Model with interactions (GA²M) by performing a round-robin training process on one input feature at a time. The general form of the EBM model is given as follows:

$$g(E[y]) = \beta_0 + \sum f_j(x_j) + \sum f_{ij}(x_i, x_j) \quad (1)$$

In this equation, $E[y]$ represents the expected value of the target variable y , which can be either non-sepsis (0) or sepsis (1). The link function, denoted by g , is responsible for mapping the model's output, derived from the weighted sum of input features and their interactions, to the expected value of the target variable. In this case, the link function transforms the model's output into a probability value between 0 and 1, corresponding to the likelihood of sepsis. The term β_0 represents the model's intercept, $f_j(x_j)$ represents the individual contribution of a single feature x_j to the model, while $f_{ij}(x_i, x_j)$ represents the contribution of the interaction between two features x_i and x_j to the model (Caruana et al., 2015).

The EBM model offers strong predictive performance while being a white-box approach, making it easier to interpret the model's predictions in clinical scenarios. The model's additive nature allows for a clear evaluation of each feature or interaction's contribution while facilitating local explanations, which can be determined by assessing the impact of individual features on the outcome. Moreover, the EBM model excels in capturing feature interactions, or the combined effects of two features that become relevant to the model. In real-world situations, feature interactions often influence the outcome, as a change in one feature may impact the value of other features. Many algorithms overlook these interactions, relying instead on local linear approximations. By taking feature interactions into account, the EBM model balances accuracy and interpretability.

The EBM model generates a probability value, ranging from 0 to 1, that represents the likelihood of a look-back window belonging to a particular sepsis or non-sepsis class based on observed trends in patients' vital signs in the window. To make predictions more accurate, a decision threshold is used to convert these probabilities into discrete class assignments, maximizing both sensitivity and specificity. The optimal decision threshold is determined using the training dataset, providing a basis for evaluating the model's performance at various thresholds and ensuring the best balance between sensitivity and specificity before assessment on the test sets. When the generated probability is compared against the decision threshold by the model, patterns within a specific look-back window are classified as either sepsis positive or non-sepsis. If the probability is higher than the threshold, the patterns are classified as sepsis positive, indicating a high likelihood of sepsis onset. Conversely, if the probability is lower than the threshold, the patterns are classified as non-sepsis, suggesting a lower likelihood of sepsis development within the given time frame.

Logistic Regression. In addition to the EBM primary machine learning model, a LR model is utilized as a baseline for comparative analysis. LR has been widely used for classification tasks and is known for its inherent interpretability as it transforms input features into a linear regression problem (Fan et al., 2008). The general form of the LR model is given as follows:

$$p(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (2)$$

In this equation, $p(y)$ represents the probability of the output variable y being equal to sepsis (1), given the input variables x_1 to x_n and the associated weights β_1 to β_n . The LR model allows for a global interpretation of feature importance by analyzing the values of the β feature weights. However, due to its linear nature, the LR model is not able to provide direct assessment of local explanations, as the feature importances are adjusted as being the same for the entire

features distribution. The LR model is trained on the same patterns extracted from the vital sign characteristics as the EBM model to ensure a fair and direct comparison between the two models in terms of their effectiveness and potential advantages in sepsis prediction.

Results

The sensitivity and specificity values of the models were analyzed, corresponding to various sepsis prediction scenarios with prediction windows from 3 to 6 hrs before sepsis onset and look-back windows of 3, 6, and 12 hrs. The EBM model achieved its highest sensitivity of 88.1% and specificity of 81.3% for a prediction window of 6 hrs and a look-back window of 12 hrs. Its lowest sensitivity was 85.15% with a specificity of 78.56% for a prediction window of 3 hrs and a look-back window of 3 hrs (see Table 2).

In comparison, the LR model exhibited the highest sensitivity of 84.17% with a specificity of 75.58% for a prediction window of 5 hrs and a look-back window of 6 hrs. The lowest sensitivity observed for this model was 82.33% with a specificity of 77.43% for a prediction window of 3 hrs and a look-back window of 12 hrs.

When compared to the machine learning models, the qSOFA and SIRS scoring systems exhibited substantially lower sensitivity values. For qSOFA, the highest sensitivity recorded was 10.55%, with a specificity of 92.46% for a prediction window of 3 hrs. This sensitivity remained relatively constant across all prediction windows, with only minor fluctuations. Similarly, the SIRS scoring system showed the highest sensitivity of 38.83% with a specificity of 73.15% for a prediction window of 6 hrs. As with qSOFA, the sensitivity values for SIRS were relatively consistent across different prediction windows.

Discussion

The study aimed to evaluate the performance of the EBM model in comparison to a LR model and two well-established scoring systems, qSOFA and SIRS, for the prediction of sepsis onset. The results obtained from the test subset demonstrated that the EBM model consistently outperformed the LR model, qSOFA, and SIRS in terms of sensitivity and specificity across various combinations of prediction and look-back windows.

One of the key advantages of the EBM model is its inherent interpretability, allowing clinicians to comprehend the model's predictions and derive insights from feature contributions. Figure 2 depicts two examples from the test subset, representing both sepsis and non-sepsis classes, and highlights the differences between these cases in terms of individual feature contributions. For the sepsis observation, all feature contributions are positive on the x-axis and arranged in descending order of contribution, suggesting that these features collectively increase the probability of sepsis and

Table 2. Performance Results on Test Subset.

Algorithm	Prediction window	Look-back window	Sensitivity	Specificity
EBM	3	3	85.15	78.56
		6	85.46	81.58
		12	86.75	79.87
	4	3	86.38	79.96
		6	86.93	80.55
		12	87.24	81.25
	5	3	86.13	81.39
		6	87.18	82.33
		12	87.79	82.52
	6	3	86.44	83.01
		6	87.55	80.71
		12	88.1	81.3
LR	3	3	83.01	74.45
		6	82.94	75.27
		12	82.33	77.43
	4	3	83.37	75.55
		6	83.99	74.78
		12	83.62	76.71
	5	3	83.25	76.14
		6	84.17	75.58
		12	83.19	78.38
	6	3	84.05	77.04
		6	83.62	76.73
		12	83.8	80.06
qSOFA	3	–	10.55	92.46
	4	–	9.02	92.46
	5	–	8.96	92.46
	6	–	9.51	92.46
SIRS	3	–	37.61	73.15
	4	–	37.98	73.15
	5	–	38.4	73.15
	6	–	38.83	73.15

steer the model's prediction toward the sepsis class. In contrast, the non-sepsis observation displays some features with negative contributions on the x-axis, implying a decrease in the probability of sepsis and directing the model's prediction toward the non-sepsis class. This variation in feature contributions underscores the distinct patterns and trends the EBM model captures to differentiate between sepsis and non-sepsis cases.

The log of odds values represented on the x-axis indicate the relative contribution of each feature toward the model's prediction, with a higher positive value denoting a stronger influence on the probability of sepsis and a more negative value denoting a stronger influence on the probability of non-sepsis. The sum of the log of odds for all feature contributions is passed through the link function, transforming the model's output into a probability between 0 and 1. The intercept term in the EBM model represents the baseline log of odds when all feature contributions are zero, serving as a starting point for the model's predictions.

Through these local explanations, the EBM model offers a transparent and interpretable representation of the relationships between features and their respective contributions to the model's predictions. By providing this level of interpretability, the EBM model allows clinicians to gain a better understanding of the factors influencing the model's predictions and make more informed decisions regarding patient care. Furthermore, it was observed that the EBM model achieved the highest sensitivity and specificity for a prediction window of 6 hrs and a look-back window of 12 hrs, with values of 88.1% and 81.3%, respectively. These results indicate that the EBM model is capable of effectively identifying sepsis onset within a clinically relevant timeframe, while offering a transparent representation of the relationships between features and their respective contributions to the model's predictions. In comparison, the LR model, which was trained on the same patterns extracted from vital sign characteristics as the EBM model, exhibited lower sensitivity and specificity values. The highest sensitivity and specificity recorded for the LR model were 84.17% and 75.58%, respectively, for a prediction window of 5 hrs and a look-back window of 6 hrs.

The results also demonstrated that the performance of the EBM model improved with increasing look-back windows. This finding suggests that the incorporation of more extensive historical data on vital sign characteristics enhances the model's ability to identify patterns and trends indicative of sepsis onset. A larger look-back window allows the model to capture longer-term fluctuations in vital signs, which may contribute to improved prediction accuracy.

An interesting observation from the study's results is that the larger prediction window of 6 hrs provided better performance than smaller prediction windows, which were closer to the onset of sepsis. This finding may initially seem counterintuitive, as one might expect a model to perform better when predicting events closer to their actual occurrence. However, there are several possible explanations for this phenomenon.

First, it is worth noting that as the prediction window increases, the model is provided with a longer time frame in which to identify patterns and trends in the vital signs data that may be indicative of sepsis. This additional time may enable the model to more accurately capture the progression of physiological changes leading up to sepsis onset, ultimately enhancing its predictive performance.

Second, a larger prediction window could potentially reduce the impact of short-term fluctuations in the vital signs data. When predicting events further into the future, the model may be less influenced by transient changes in the data that are unrelated to sepsis, allowing it to better focus on the underlying patterns that are indicative of the condition.

Third, in real-world clinical settings, clinicians often require some lead time to initiate appropriate interventions for sepsis patients. A model that can accurately predict sepsis onset within a larger prediction window may be more valuable in practice, as it allows clinicians to proactively initiate

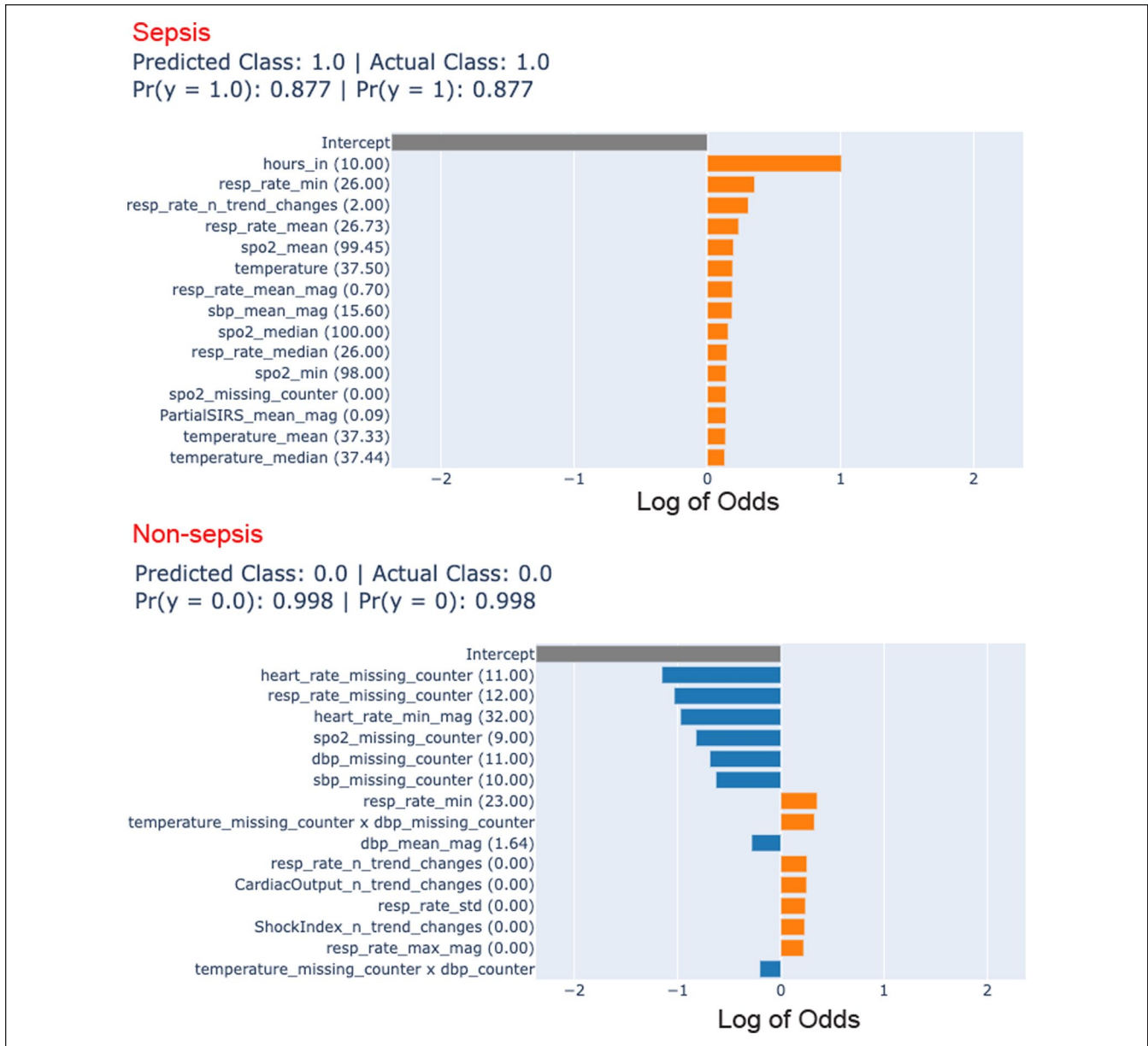


Figure 2. Interpretability of sample sepsis and non-sepsis observations from test subset.

treatment plans and allocate resources in anticipation of a patient's deteriorating condition.

Lastly, the improved performance observed with a larger prediction window may also be related to the specific characteristics of the EBM model. As a decision tree-based model, the EBM is designed to effectively capture complex, non-linear relationships between input features, which may be more evident over a longer time frame. The enhanced performance at a 6-hrs prediction window may be an indication of the model's ability to capitalize on these relationships to make more accurate predictions.

Despite these potential explanations, it is important to recognize that the choice of an appropriate prediction window will depend on various factors, including the specific

clinical context, the nature of the data being analyzed, and the desired balance between sensitivity and specificity. Further research is needed to better understand the optimal prediction window for sepsis prediction models and to explore the potential benefits and limitations of different window configurations in diverse clinical settings.

In contrast to the machine learning models, the qSOFA and SIRS scoring systems demonstrated substantially lower sensitivity values. For qSOFA, the highest sensitivity recorded was 10.55%, with a specificity of 92.46% for a prediction window of 3 hrs. This sensitivity remained relatively constant across all prediction windows, with only minor fluctuations. Similarly, the SIRS scoring system showed the highest sensitivity of 38.83% with a specificity of 73.15%

for a prediction window of 6 hrs. As with qSOFA, the sensitivity values for SIRS were relatively consistent across different prediction windows.

It is important to note that the qSOFA and SIRS scoring systems do not rely on look-back windows to calculate their scores. Instead, these scores are calculated based on the patient's physiological parameters at a specific hour prior to sepsis onset. This approach differs from the machine learning models, which leverage information from a series of data points within a look-back window to identify patterns and trends indicative of sepsis. The qSOFA and SIRS systems focus on the patient's physiological state at a single point in time, which may limit their ability to capture the full range of changes leading up to sepsis onset and, consequently, result in lower sensitivity values compared to the machine learning models.

The superior performance of the EBM model in comparison to the LR model, qSOFA, and SIRS underscores the potential advantages of employing a more sophisticated machine learning approach for sepsis prediction. The qSOFA and SIRS are linear approaches, with points assigned to each organ system or criteria based on the severity of dysfunction. While the qSOFA and SIRS scores are valuable tools for quantifying dysfunction, they do not capture the full complexity of the relationships between variables that contribute to sepsis. The EBM model, on the other hand, allows for a more nuanced understanding of these relationships, which leads to improved early detection and intervention strategies. The higher sensitivity values achieved by the EBM model suggest that it is capable of accurately detecting the majority of true sepsis cases, thus minimizing the risk of false negatives and enabling clinicians to initiate timely interventions. Furthermore, the higher specificity values imply that the EBM model is less likely to generate false-positive results, which could lead to unnecessary treatments, increased healthcare costs, and potential patient harm due to unwarranted interventions. Striving for excellence in both sensitivity and specificity is crucial for the success of sepsis prediction models and their implementation in clinical settings.

Limitations

Despite the promising findings of this study, several limitations should be acknowledged to provide a comprehensive understanding of the results and their implications. First, the study utilized a single dataset from a specific population, which may limit the generalizability of the findings to other populations and clinical settings. The performance of the EBM model, as well as the other models, may vary across different populations with diverse demographic characteristics, underlying health conditions, and risk factors. Future research should aim to validate the EBM model using multiple datasets from various populations and clinical environments to ensure its applicability and reliability across diverse settings.

Second, while the study compared the performance of the EBM model with LR, qSOFA, and SIRS, it did not investigate other advanced machine learning models or alternative scoring systems. The inclusion of additional machine learning models, such as deep learning and ensemble methods, could provide further insights into the comparative performance of these approaches for sepsis prediction. Similarly, evaluating alternative clinical scoring systems might reveal other potential candidates for early sepsis detection and intervention.

Additionally, the study did not consider the impact of early treatment, such as fluid administration, on patients' vital signs and outcomes. Incorporating treatment-related factors into the feature set could potentially improve the model's performance and provide a more comprehensive understanding of the factors contributing to sepsis onset and progression.

Finally, the study primarily focused on sensitivity and specificity as performance measures, which, although essential, may not provide a comprehensive understanding of the model's effectiveness in clinical settings. Other performance metrics, including positive predictive value (PPV) and negative predictive value (NPV) could offer additional insights into the model's performance and its suitability for implementation in clinical practice. For instance, PPV can help determine the likelihood of a positive sepsis look-back window being correct, while NPV can assess the probability of a non-sepsis look-back window being accurate. These metrics can assist stakeholders in better comprehending the model's real-world implications and making informed decisions about its use in clinical settings.

In light of these limitations, it is essential to interpret the findings of this study with caution. While the results demonstrate the potential of the EBM model for sepsis prediction, further research is required to address the limitations and validate the model across different populations, clinical settings, and in comparison to other machine learning models and clinical scoring systems.

Conclusion

This study highlights the potential of the EBM model, an innovative and novel approach, for improving sepsis prediction and facilitating early intervention efforts by exclusively utilizing continuous vital signs monitoring data. The EBM model demonstrates superior performance in terms of sensitivity and specificity compared to the LR model, qSOFA, and SIRS, suggesting its effectiveness in enhancing patient outcomes associated with sepsis without relying on laboratory results or EMRs. The study's findings also emphasize the importance of selecting appropriate prediction and look-back windows to optimize the performance of sepsis prediction models in clinical settings. The integration of machine learning-based approaches, such as the EBM model, into clinical decision-making processes has the potential to improve sepsis prediction and early intervention strategies.

By utilizing the EBM model's white-box approach and powerful interpretability, clinicians can make informed decisions in high-stakes healthcare settings. As clinicians continue to strive for excellence in patient care, the adoption of innovative and accurate prediction models that do not depend on lab results and EMRs can play a vital role in enhancing sepsis management and improving patient outcomes.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the views of their affiliated organizations.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Consideration

This study protocol involved a review and secondary analysis of a publicly available dataset consisting of de-identified data.

ORCID iDs

An Tran  <https://orcid.org/0009-0001-8915-780X>

Robert Topp  <https://orcid.org/0000-0001-6667-5789>

Data Availability

The data used in this study were obtained from a publicly available dataset. Access to and use of this dataset are subject to the data provider's terms and conditions. Please visit <https://mimic.mit.edu/> for information on using the Medical Information Mart for Intensive Care (MIMIC)-IV database.

Bibliography

- Berger, T., Green, J., Horeczko, T., Hagar, Y., Garg, N., Suarez, A., Panacek, E., & Shapiro, N. (2013). Shock index and early recognition of sepsis in the emergency department: Pilot study. *The Western Journal of Emergency Medicine*, *14*(2), 168–174.
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, August 10–13, 2015, Sydney, Australia. <https://doi.org/10.1145/2783258.2788613>.
- Centers for Disease Control and Prevention. (2016). *Sepsis: Data & reports*. <https://www.cdc.gov/sepsis/what-is-sepsis.html>
- Corrêa, T. D., Vuda, M., Takala, J., Djafarzadeh, S., Silva, E., & Jakob, S. M. (2013). Increasing mean arterial blood pressure in sepsis: Effects on fluid balance, vasopressor load and renal function. *Critical Care (London, England)*, *17*(1), R21.

- Despins, L. A. (2017). Automated detection of sepsis using electronic medical record data: A systematic review. *Journal for Healthcare Quality : Official Publication of the National Association for Healthcare Quality*, *39*(6), 322–333.
- Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, *9*, 1871–1874.
- Gyang, E., Shieh, L., Forsey, L., & Maggio, P. (2015). A nurse-driven screening tool for the early identification of sepsis in an intermediate care unit setting. *Journal of Hospital Medicine*, *10*(2), 97–103.
- Hamer, R. M., & Simpson, P. M. (2009). Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials. *The American Journal of Psychiatry*, *166*(6), 639–641.
- Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2023). MIMIC-IV (version 2.2). *PhysioNet*. <https://doi.org/10.13026/6mml1-ek67>
- Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., Taiberg, L., Gurka, D., Kumar, A., & Cheang, M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*, *34*(6), 1589–1596.
- Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., Cohen, J., Opal, S. M., Vincent, J. L., Ramsay, G., & SCCM/ESICM/ACCP/ATS/SIS. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Medicine*, *31*(4), 1250–1256.
- Paoli, C. J., Reynolds, M. A., Sinha, M., Gitlin, M., & Crouser, E. (2018). Epidemiology and costs of sepsis in the United States—An analysis based on timing of diagnosis and severity level. *Critical Care Medicine*, *46*(12), 1889–1897.
- Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., Rubenfeld, G., Kahn, J. M., Shankar-Hari, M., Singer, M., Deutschman, C. S., Escobar, G. J., & Angus, D. C. (2016). Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, *315*(8), 762–774.
- Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J. D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., & van de, P. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, *315*(8), 801–810.
- Torio, C. M., & Moore, B. J. (2016). *National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013: Statistical Brief #204*. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Agency for Healthcare Research and Quality (US).

Author Biographies

An Tran is a data scientist for Darroch Medical Solutions, Inc. An obtained his first Master's degree in Computer Engineering where he specialized in biomedical signals processing and a second Master's degree in Data Science. He contributed to this study by conducting data processing and analysis, developing and validating the machine learning models, and contributing to the writing of the manuscript.

Robert Topp is a professor of Nursing at the University of Toledo. He contributed to this study by developing the protocol, conducting data analysis and contributing to the writing of the manuscript.

Ebrahim Tarshizi is a Professor of Practice and Academic Director of the Applied Data Science (MS-ADS) & Applied Artificial Intelligence (MS-AAI) programs at the University of San Diego's

Shiley-Marcos School of Engineering. Ebrahim contributed to this study by advising the technicality of the machine learning development part and contributing to the editing and revision of the manuscript.

Anthony Shao is the CEO of Darroch Medical Solutions, Inc. He contributed to the writing of the manuscript.