



# HHS Public Access

Author manuscript

*Crit Care Med.* Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:

*Crit Care Med.* 2018 April ; 46(4): 547–553. doi:10.1097/CCM.0000000000002936.

## An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU

Shamim Nemati, PhD<sup>1,\*</sup>, Andre Holder, MD, MSc<sup>2</sup>, Fereshteh Razmi<sup>1</sup>, Matthew D. Stanley, MD<sup>3</sup>, Gari D. Clifford, PhD<sup>1,4</sup>, and Timothy G. Buchman, PhD, MD<sup>3,5</sup>

<sup>1</sup>Department of Biomedical Informatics; Emory University School of Medicine

<sup>2</sup>Division of Pulmonary, Critical Care, Allergy and Sleep Medicine; Emory University School of Medicine

<sup>3</sup>Department of Surgery, Emory University School of Medicine, Atlanta, GA 30322

<sup>4</sup>Department of Biomedical Engineering, Georgia Institute of Technology

<sup>5</sup>Emory Critical Care Center, Emory Healthcare

### Abstract

**OBJECTIVE**—Sepsis is among the leading causes of morbidity, mortality, and cost overruns in critically ill patients. Early intervention with antibiotics improves survival in septic patients. However, no clinically validated system exists for real-time prediction of sepsis onset. We aimed to develop and validate an Artificial Intelligence Sepsis Expert (AISE) algorithm for early prediction of sepsis.

**DESIGN**—Observational cohort study.

**SETTING**—Academic medical center from January 2013 to December 2015.

**PATIENTS**—Over 31,000 admissions to the intensive care units (ICUs) at two Emory University hospitals (development cohort), in addition to over 52,000 ICU patients from the publicly available

---

\*Corresponding author: Shamim Nemati, PhD, Assistant Professor, Emory Department of Biomedical Informatics, Woodruff Memorial Research Building, 101 Woodruff Circle, Room 4131, Atlanta, GA 30322, Phone: (405) 850-4751, Fax: shamim.nemati@alum.mit.edu.

#### Conflicts of interest

The remaining authors have disclosed that they do not have any remaining conflicts of interest. The opinions or assertions contained herein are the private ones of the author/speaker and are not to be construed as official or reflecting the views of the Department of Defense, the Uniformed Services University of the Health Sciences or any other agency of the U.S. Government.

**Financial support:** none

**Copyright form disclosure:** Drs. Nemati, Stanley, and Clifford received support for article research from the National Institutes of Health (NIH). Dr. Nemati's institution received funding from the NIH, award #K01ES025445. Dr. Holder received funding from CR Bard, Inc.

Dr. Clifford c/f (disclosed: Dr Buchman's research is partially supported by the Surgical Critical Care Initiative (SC2i), funded through the Department of Defense's Health Program – Joint Program Committee 6 / Combat Casualty Care (USUHS HT9404-13-1-0032 and HU0001-15-2-0001).; Dr. Buchman's institution received funding from the Henry M. Jackson Foundation for his role as site director in Surgical Critical Care Institute, [www.sc2i.org](http://www.sc2i.org), funded through the Department of Defense's Health Program – Joint Program Committee 6 / Combat Casualty Care (USUHS HT9404-13-1-0032 and HU0001-15-2-0001).; from Society of Critical Care Medicine for his role as Editor-in-Chief of *Critical Care Medicine*; and from Philips Corporation (unrestricted educational grant to a physician education association in South Korea so he could present the results of his research in eICU). Dr. Buchman received support for article research from the Henry M Jackson Foundation. Ms. Ramzi has disclosed that she does not have any potential conflicts of interest.

MIMIC-III ICU database (validation cohort). Patients who met the Third International Consensus Definitions for Sepsis (sepsis-3) prior to or within 4 hours of their ICU admission were excluded, resulting in roughly 27,000 and 42,000 patients within our development and validation cohorts, respectively.

**INTERVENTIONS**—None

**MEASUREMENTS and MAIN RESULTS**—High-resolution vital signs time series and Electronic Medical Record (EMR) data were extracted. A set of 65 features (variables) were calculated on hourly basis and passed to the AISE algorithm to predict onset of sepsis in the proceeding T hours (where T = 12, 8, 6 or 4). AISE was used to predict onset of sepsis in the proceeding T hours, and to produce a list of the most significant contributing factors. For the 12-hour, 8-hour, 6-hour, and 4-hour ahead prediction of sepsis, AISE achieved area under the receiver operating characteristic (AUROC) in the range of 0.83–0.85. Performance of the AISE on the development and validation cohorts were indistinguishable.

**CONCLUSION**—Using data available in the ICU in real-time, AISE can accurately predict the onset of sepsis in an ICU patient 4 to 12 hours prior to clinical recognition. A prospective study is necessary to determine the clinical utility of the proposed sepsis prediction model.

### Keywords

Sepsis; SOFA; organ failure; prognostication; machine learning; informatics

---

## INTRODUCTION

Sepsis, a dysregulated immune-mediated host response to infection, is prevalent, lethal, and costly. [1–4] Recent literature suggests that early and appropriate antibiotic therapy is the main factor predicting sepsis outcomes. [5] Identifying those at risk for sepsis and initiating appropriate treatment, prior to any clinical manifestations, would have a significant impact on the overall mortality and cost burden of sepsis.

Clinical decision support tools can help identify those at highest risk for future sepsis. Existing work on EMR and laboratory data seem promising, [6–8] but they are limited by being static, or collected at low or inconsistent frequencies. The dynamics of heart rate (HR) and blood pressure (BP) extracted directly from the electrocardiogram (ECG) and arterial waveform can improve mortality prediction over clinical data (demographics or data collected at low frequency) in ICU patients with transient hypotension. [9] The objective of this study is to demonstrate that a high-performance prediction model can be derived from a combination of EMR and high-frequency physiologic data (collected at least once per second). We further test the relationship between the prediction lead time (prediction window) and predictive accuracy of the model, and investigate questions of generalizability and interpretability of the proposed model.

## METHODS

### Study Population and Data Sources

All ICU patients aged 18 years or older were included from two hospitals within the Emory Healthcare system, as well as an external publicly available ICU database. [10] This investigation was conducted according to Emory University IRB approved protocol 33069. Patients were followed throughout their ICU stay until discharge or development of sepsis, according to the Third International Consensus Definitions for Sepsis (sepsis-3). Specifically, all episodes of suspected infection ( $t_{\text{suspicion}}$ ) were identified as the earlier timestamp of antibiotics and blood cultures within a specific time span; if the antibiotic was given first, the culture sampling must have been obtained within 24 hours. If the culture sampling was first, the antibiotic must have been ordered within 72 hours. The onset time of sepsis ( $t_{\text{sepsis}}$ ) was then defined as an episode of suspected infection with two or more points change in the SOFA score ( $t_{\text{SOFA}}$ ) from up to 24 hours before to up to 12 hours after the  $t_{\text{suspicion}}$  ( $t_{\text{SOFA}+24\text{h}} > t_{\text{suspicion}} > t_{\text{SOFA}-12\text{h}}$ ). These definitions were based on a recent assessment of the revised clinical criteria for sepsis. [11] Finally, we defined  $t_{\text{onset}}$  as the minimum of  $t_{\text{sepsis}}$  and  $t_{\text{SOFA}}$ . Though our primary outcome was  $t_{\text{sepsis}}$ , we report the predictive performance of our algorithm also on  $t_{\text{SOFA}}$  and  $t_{\text{onset}}$  for completeness and to facilitate comparison with the existing literature.

Data from the electronic medical record (Cerner, Kansas City, MO) was extracted through a clinical data warehouse (MicroStrategy, Tysons Corner, VA). High resolution heart rate and blood pressure time series at 2 seconds resolution were collected from select ICUs, through the BedMaster system (Excel Medical Electronics, Jupiter FL, USA), which is a third-party software connected to the hospital's General Electric (GE) monitors for the purpose of electronic data extraction and storage of high-resolution waveforms. Patients were excluded if they developed sepsis within the first 4 hours of ICU admission (by analyzing pre-ICU IV antibiotic administration and culture acquisition), or if their length of ICU stay (LOS) was less than 8 hours or more than 20 days.

### Feature Extraction and Machine Learning

A total of 65 features from the electronic medical record and high-resolution bedside monitoring data. These features were used as inputs to a modified Weibull-Cox proportional hazards model, the machine learning algorithm used in this study. See Appendices B and C for further details on feature extraction and machine learning, and Appendix F for a glossary of machine learning-related terms and their meanings.

### Statistical Methods

For all continuous variables we report medians ([25-percentile, 75-percentile]) and utilize a two-sided Wilcoxon Rank-Sum test when comparing two populations. For binary variables we report percentages and utilize a two-sided Chi-square test to assess differences in proportions between two populations. AISE classification results for  $T$  hours ( $T = 12, 8, 6$  or 4 hours) ahead predictions are based on a random split into 80% training 20% testing, and the area under receiver operating characteristic (AUROC) curves statistics for both the

training and the testing sets are reported, as well as specificity (1–false alarm rate) and accuracy at a fixed 85% sensitivity level.

## RESULTS

Our development cohort included a total of 27,527 patients, 2,375 (8.6%) of whom developed sepsis in the ICU with a median lag time of 23.9 hours (see Appendix – Table A1). Those who developed sepsis tended to have a slightly higher percentage of male patients (56.2% vs. 52.4%) and have more comorbidities (CCI 4 vs. 2). Septic patients had longer median lengths of ICU stay (5.9 vs. 1.9 days), higher median SOFA scores (5.0 vs. 1.7), and higher hospital mortality (14.5% vs. 2.9%). Similar patterns were observed within our validation cohort (see Appendix – Table E1).

Both training set and testing set AUROCs for detecting sepsis were 0.79 or higher for every prediction task ( $t_{\text{sepsis}}$ ,  $t_{\text{SOFA}}$ ,  $t_{\text{onset}}$ ) and prediction window ( $n=4, 6, 8,$  and  $12$  hours) (Figure 1). The best performance was achieved for predicting  $t_{\text{SOFA}}$  four hours in advance (AUROC of 0.87), which was slightly higher than predicting  $t_{\text{sepsis}}$  four hours in advance (AUROC of 0.85). In our development cohort, roughly 21% of the time  $t_{\text{SOFA}}$  occurred after  $t_{\text{sepsis}}$ . We therefore defined  $t_{\text{onset}}$  as the earliest timestamp of the two ( $t_{\text{sepsis}}$  and  $t_{\text{SOFA}}$ ), which proved more difficult to predict (four hours prediction AUROC of 0.82). When sensitivity was fixed at 85% (risk score=0.45), specificity in the test cohort was highest for  $T_{\text{SOFA}}$  prediction (72%), followed by  $t_{\text{sepsis}}$  (67%), and the lowest for  $t_{\text{onset}}$  (64%) (see Table 1).

Model performance decreased slightly when prediction occurred over longer time windows regardless of the sepsis time-point of interest (Figure 1 & Table 1). To predict  $t_{\text{sepsis}}$ , model AUROC decreased from 0.85 at a four-hour prediction window, to 0.83 at a twelve-hour window. Specificity demonstrated similar declines when sensitivity was fixed at 85% (67% with four-hour windows vs. 63% at twelve-hour window). These findings were consistent across our development and validation cohorts (see Appendix, Fig. E1 and Table E2).

Hospital mortality increased as the risk score for sepsis ( $t_{\text{sepsis}}$ ) increased from 0 to 1 (Table 2). Those with a risk score of 0–0.2 had a mortality of 0.5%, while those with risk scores of  $>0.8$  had a 32.9% mortality rate. This was true even among those who were false positives, defined as those who did not develop sepsis in the predicted window but had risk scores of 0.45 or greater. In fact, the mortality was higher among false positives assigned the risk score of  $>0.8$ , compared with those given a similar score in the total cohort (56.3% vs. 32.9%). Compared to those who were false negatives, patients who were false positives had higher SOFA scores (4.0 [IQR 2.0,7.0] vs 3.0 [IQR 1.0,5.0];  $p<0.01$ ), higher CCIs (4.0 [IQR 2.0,6.0] vs. 3.0 [IQR 2.0,5.0];  $p<0.01$ ), and higher hospital mortality (15.5% vs. 6.4%;  $p<0.01$ ).

## DISCUSSION

In this study, we demonstrated that a high-performing prediction model (AUROC 0.85) can predict sepsis ( $t_{\text{sepsis}}$ ) 4 hours in advance using EMR data combined with high-resolution time series dynamics of heart rate and blood pressure. This is true no matter the outcome of interest, whether the prediction task involves more objective physiological manifestations of

sepsis (as captured by  $t_{\text{SOFA}}$ ; AUROC 0.87), clinical suspicion of infection (as marked by  $t_{\text{Sepsis}}$ ; AUROC 0.85), or the earlier of the two (namely,  $t_{\text{onset}}$ ; AUROC 0.82). Prediction performance is inversely proportional to size of the prediction window; AUROC, specificity and accuracy for  $t_{\text{sepsis}}$  all decreased slightly as the prediction window lengthened from 4 to 12 hours, but still provided high-performing models (AUROC of at least 0.83). We externally validated all findings in patients from a separate academic center.

As ICU clinicians are inundated with ever-increasing data collected at higher frequencies, machine learning will become more essential to research and clinical practice. Machine learning refers to a body of methods based in computer science that use patterns in data to identify or predict an outcome. Machine Learning provides a powerful set of tools for describing relationships between features and the outcome(s) of interest (e.g. sepsis), particularly when they are nonlinear and complex. It is best used when there are a large number of variables, and overfitting (poor generalizability) can be a problem for traditional statistical methods. We had access to over 65 features in our analysis; we therefore used a modified regularized Weibull-Cox analysis, a type of machine learning approach that results in a more interpretable and generalizable survival model, to predict sepsis in ICU patients. [See Appendix C for more details on this approach]

Machine learning-based clinical decision support (CDS) tools embedded within electronic medical record improve early detection and prompt treatment in those with early sepsis, and can predict septic shock. [13, 16–18] EMR alerts to detect existing sepsis can improve adherence to treatment protocols, decreases time until antibiotic administration and length of hospital stay, and can improve mortality [19–21] They can predict septic shock with 85% accuracy using either EMR data or high-resolution vital sign streams [13, 18] Still, a patient for whom CDS is used for septic shock prediction already has sepsis. Fluid and hemodynamic management would be the only modifiable intervention to provide those at risk for septic shock, but a recent study suggested that this is not associated with lower in-hospital mortality. [5]

This study makes several significant contributions to the existing literature on sepsis prediction. The data used in our model is widely available in current practice. Lukaszewski and colleagues [6] demonstrated that a neural network using only cytokine data predicted sepsis better than a similar algorithm using clinical EMR data. However, cytokines are not routinely measured, making it an impractical tool for contemporary practice. Wang and colleagues [7] used simple EMR features such as white blood cell count, heart rate, and APACHE 2 score and created an estimate of future sepsis severity in ICU patients (scale of 0 to 1). Though their model was very good at classifying severe sepsis (by sepsis-1 definition [22]) and its severity (AUROC 0.94), it averaged repeated measures for each feature during the first 24 hours of ICU stay. This is less useful for real-time use, and one can't identify specific prediction windows in which sepsis would occur since time series inputs are not utilized.

Our algorithm is among the first to predict sepsis by combining data collected at different resolutions (low-resolution EMR data, and high-resolution blood pressures and heart rates). Others have used low-resolution inputs primarily, either as a single input feature by

Author Manuscript

averaging repeated measures, [7] or by retaining time series integrity. [8] Coupling low-resolution with high-resolution data provide complementary information used by our algorithm to predict sepsis risk in our cohort. Two high-resolution features, entropy of heart rate and blood pressure, were important in model development. Multiscale entropy, one of many variability metrics thought to represent neuro-cardiac organ interaction (i.e. adjustments in autonomic tone), improves hospital mortality prediction in those with sepsis. [23] The AISE system can accommodate more input features as the medical community learns more about sepsis. As our biological and physiologic understanding of sepsis improves and new biomarkers are created, it may also allow clinicians to use the algorithm in smarter ways. Since AISE can inform the physician of the most relevant features contributing to the risk score over time (See Appendix C; Visualization and Interpretability), one can use what is known about sepsis and apply it to the clinical context to decide if and when one should act upon the prediction.

Author Manuscript

To our knowledge, this is the first study to demonstrate acceptable performance of a sepsis prediction algorithm over incrementally longer time windows. Desautels and colleagues used a proprietary machine learning algorithm with vital signs, pulse oximetry, GCS and age as features and demonstrated moderate capability to predict sepsis four hours before it occurred (AUROC 0.74). Sepsis was defined as the first time at which there was a 2-point increase in SOFA that preceded evidence of suspicion of infection. Our algorithm was superior to this using a similar sepsis definition ( $t_{\text{SOFA}}$ ) over the same time window (AUROC 0.87), and stayed superior up to a prediction window of 12 hours. The robustness of our model could at least be partly explained by the rich information provided by the different resolutions of our inputs.

Author Manuscript

Defining the onset of sepsis can be very subjective and provider-dependent; we therefore assessed the performance of a range of clinically meaningful outcomes. The  $t_{\text{SOFA}}$  is the most objective of all three markers used and the easiest to predict using EMR and vitals data. In our datasets, roughly 20% of the time  $t_{\text{SOFA}}$  occurred after  $t_{\text{sepsis}}$ . We therefore introduced the  $t_{\text{onset}}$ , as the minimum of  $t_{\text{SOFA}}$  and  $t_{\text{sepsis}}$ . However,  $t_{\text{onset}}$  was the most difficult to predict and had the lowest AUROC. This is not surprising, since prompt prediction of sepsis requires up-to-date clinical measurements, which are more likely to be available if there is already a clinical suspicion of sepsis. High-resolution data can potentially mitigate this problem and provide more timely prediction.

Author Manuscript

Though our algorithm was designed to predict new sepsis, those with positive risk assignments (score of 0.45 or higher) were associated with worse outcomes. Risk of death was over two-fold higher among those who had high-risk scores but didn't develop sepsis (false positives) compared to those who had low risk scores but developed sepsis (false negatives). Many of the input features from the EMR (e.g. lactate) are not specific to sepsis and just indicate poor tissue perfusion. The same is true of high frequency variables like multiscale entropy; loss of organ-organ coupling is a sign of critical illness. It is possible that our algorithm can be extrapolated to all ICU patients to predict clinical decompensation agnostic of cause, but this would require further research.

Much of the contemporary focus in sepsis management is early intervention. AISE shifts the focus towards sepsis interdiction – identifying candidates for treatments before organ failure becomes established and before tissue sampling would be meaningful – thereby mitigating the cost, morbidity and mortality burden of sepsis care. This interdiction system will allow caregivers to identify and treat patients with IV antibiotics, fluids, and other adjunct therapies based on a reliable estimate of their likelihood of developing sepsis in the near future.

AISE approaches have the potential to predict (and interdict) other forms of physiologic decompensation. Shared general experience around alerts and alarms makes it unlikely that simple notification of a bedside staff member will maximize the usefulness of this decision support approach. More likely, notification of a clinician that there is a prediction and the basis for that prediction will prompt an immediate focused review of a patient's physiology. There is no requirement that the clinician be at the bedside; indeed, our initial implementations will leverage a remote monitoring team precisely to avoid further burdens to the bedside staff. There are no restrictions on how – and to whom – the alerts can be transmitted: facilities can choose whom will be alerted based on local factors depending on their available resources and alert volume. We expect that healthcare systems with high numbers of alerts may use a tele-ICU system as an added monitoring “layer”, allowing bedside staff more time to provide quality care and decrease alarm fatigue.

This study comes with some limitations. The data was analyzed retrospectively using electronic medical record data not originally designed for the analyses performed. However, this authenticates our analysis; it confirms its utility in a real-world clinical setting, showing good performance even in the presence of missing data. The retrospective design also means that suspicion of infection had to be inferred from systematic criteria, which may not reflect the true rationale of care in all patients. However, chart review was conducted on a select group of 100 patients with sepsis to validate our results and our method was accurate in that sub-cohort (over 99% accuracy). The model used was trained on EMR data entered by bedside nurses, which may confer some recall and information bias. For instance, blood pressure documentation by humans can be biased towards normal when compared to corresponding blood pressure waveforms, [24] in part due to back-documentation of past data. However, these flawed data-points are appropriate to use in our model since they contain predictive information when combined with other measurements.

Prediction is clearly less important than interdiction: it is not enough to know, we have to take actions that enhance patient-centered outcomes. Accordingly, we will first explore our algorithm's ability to identify the source of sepsis in those with risk scores above 0.45. This iteration of AISE will also be trained and validated on datasets accumulated in other hospitals. Second, AISE accuracy will be prospectively validated for real-time prediction in the clinical setting. We intend to let usual care proceed ignorant of the sepsis score while independent and isolated adjudicators knowing the sepsis score (experienced clinicians) but blinded to the decisions of the bedside team will be asked if and when the patient is sufficiently ill to interdict. Finally, a clinical intervention trial would assign patients to AISE (versus best care lacking AISE monitoring), followed by antibiotic and fluid administration depending on the risk score and the predicted infectious source. Such a trial could follow a

stepped-wedge design, under which the AISE algorithm is sequentially but randomly introduced to multiple units within one or more hospitals over a period of time. Primary outcomes of interest may include the number of vasopressor and ventilator-free days, with mortality and length of hospital stay as secondary outcomes of interest.”

## CONCLUSION

In this two-center retrospective study, we demonstrate that high-performance models can be constructed to predict the onset of sepsis by combining data available from the EMR and high-resolution time series dynamics of blood pressure and heart rate. Predictive performances of these models are inversely proportional to the lead-time of prediction. Patients who are incorrectly labeled as those who will develop sepsis confer significant mortality, making this tool potentially useful in other clinical syndromes and disease processes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

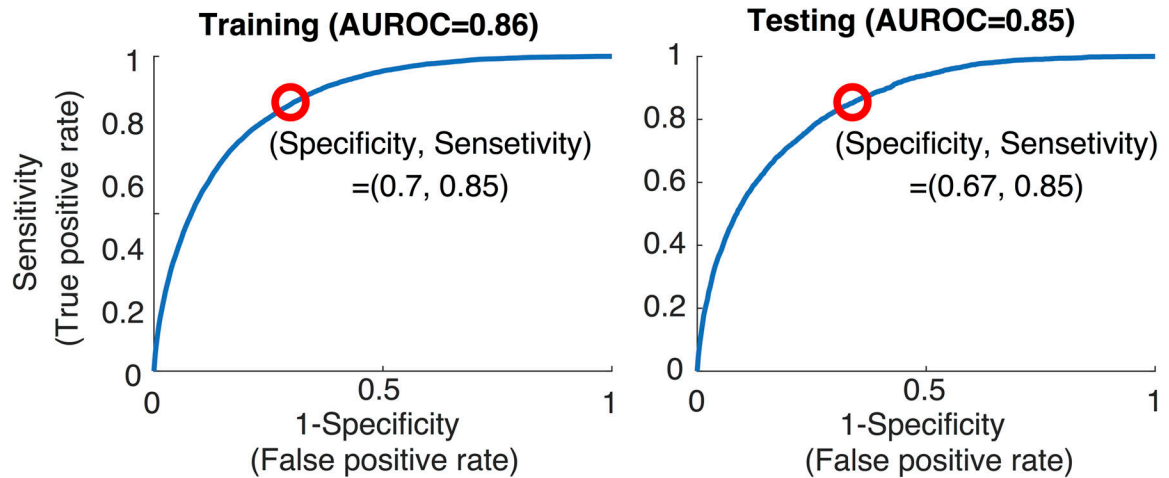
**sources of funding:** Dr. Nemati is funded by the National Institutes of Health, award #K01ES025445. Dr. Holder received support from CR Bard. Dr. Buchman’s research is partially supported by the Surgical Critical Care Initiative (SC2i), funded through the Department of Defense’s Health Program – Joint Program Committee 6 / Combat Casualty Care (USUHS HT9404-13-1-0032 and HU0001-15-2-0001). He is also the Editor-in-Chief of Critical Care Medicine.

## References

1. Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med.* 2001; 29(7):1303–10. [PubMed: 11445675]
2. Arise AAM Committee. The outcome of patients with sepsis and septic shock presenting to emergency departments in Australia and New Zealand. *Crit Care Resusc.* 2007; 9(1):8–18. [PubMed: 17352661]
3. Martin GS, Mannino DM, Eaton S, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med.* 2003; 348(16):1546–54. [PubMed: 12700374]
4. Stoller J, Halpin L, Weis M, et al. Epidemiology of severe sepsis: 2008–2012. *J Crit Care.* 2016; 31(1):58–62. [PubMed: 26601855]
5. 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):2235–2244. [PubMed: 28528569]
6. Lukaszewski RA, Yates AM, Jackson MC, et al. Presymptomatic prediction of sepsis in intensive care unit patients. *Clin Vaccine Immunol.* 2008; 15(7):1089–94. [PubMed: 18480235]
7. Wang SL, Wu F, Wang BH. Prediction of severe sepsis using SVM model. *Adv Exp Med Biol.* 2010; 680:75–81. [PubMed: 20865488]
8. Desautels T, Calvert J, Hoffman J, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform.* 2016; 4(3):e28. [PubMed: 27694098]
9. Mayaud L, Lai PS, Clifford GD, et al. Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Crit Care Med.* 2013; 41(4):954–62. [PubMed: 23385106]

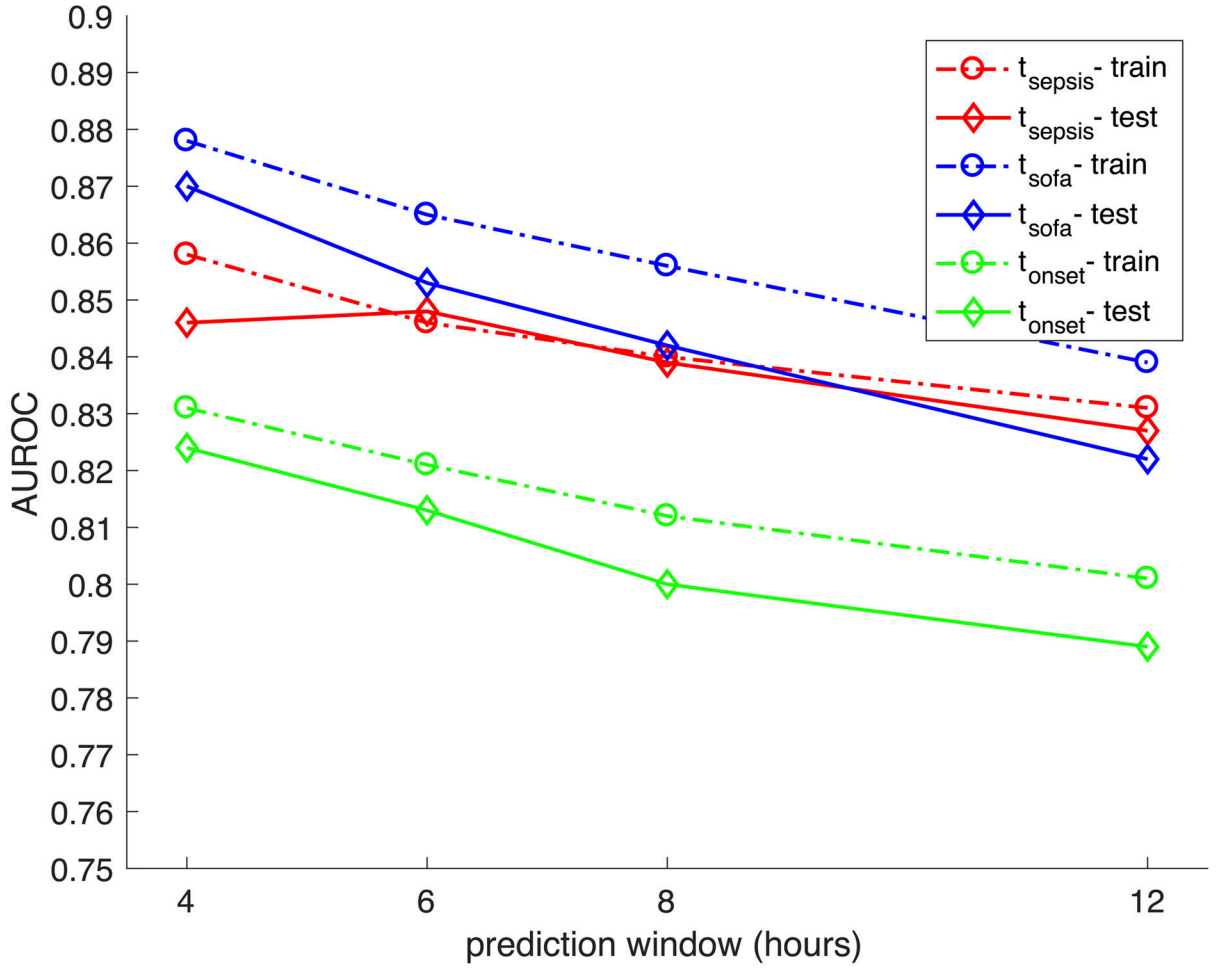


10. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016; 3:160035. [PubMed: 27219127]
11. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016; 315(8):801–10. [PubMed: 26903338]
12. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996; 22(7): 707–10. [PubMed: 8844239]
13. Henry KE, Hager DN, Pronovost PJ, et al. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med*. 2015; 7(299):299ra122.
14. Hinton, G., Srivastava, N., Sewrsky, K. *Neural Networks for Machine Learning*. 2012. Overview of mini-batch gradient descent.
15. Simonyan, K., Vedaldi, A., Zisserman, A. U.o.O. Visual Geometry Group. *Deep inside convolutional networks: Visualising image classification models and saliency maps*. 2014.
16. Horng S, Sontag DA, Halpern Y, et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One*. 2017; 12(4):e0174708. [PubMed: 28384212]
17. Brown SM, Jones J, Kuttler KG, et al. Prospective evaluation of an automated method to identify patients with severe sepsis or septic shock in the emergency department. *BMC Emerg Med*. 2016; 16(1):31. [PubMed: 27549755]
18. Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *J Biomed Inform*. 2017; 66:19–31. [PubMed: 28011233]
19. Tafelski S, Nachtigall I, Deja M, et al. Computer-assisted decision support for changing practice in severe sepsis and septic shock. *J Int Med Res*. 2010; 38(5):1605–16. [PubMed: 21309474]
20. Narayanan N, Gross AK, Pintens M, et al. Effect of an electronic medical record alert for severe sepsis among ED patients. *Am J Emerg Med*. 2016; 34(2):185–8. [PubMed: 26573784]
21. Amland RC, Haley JM, Lyons JJ. A Multidisciplinary Sepsis Program Enabled by a Two-Stage Clinical Decision Support System: Factors That Influence Patient Outcomes. *Am J Med Qual*. 2016; 31(6):501–508. [PubMed: 26491116]
22. Bone RC, Sibbald WJ, Sprung CL. The ACCP-SCCM consensus conference on sepsis and organ failure. *Chest*. 1992; 101(6):1481–3. [PubMed: 1600757]
23. Lehman LW, et al. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE J Biomed Health Inform*. 2015; 19(3):1068–76. [PubMed: 25014976]
24. Hug CW, Clifford GD, Reisner AT. Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension. *Crit Care Med*. 2011; 39(5):1006–14. [PubMed: 21336136]
25. Baehrens D, et al. How to explain individual classification decisions. *JMLR*. 2010; 11:1803–1831.
26. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. *Physical review letters*. 2002 Jul 19. 89(6):068102. [PubMed: 12190613]
27. Nemati S, Edwards BA, Lee J, et al. Respiration and heart rate complexity: effects of age and gender assessed by band-limited transfer entropy. *Respiratory physiology & neurobiology*. 2013 Oct 1; 189(1):27–33. [PubMed: 23811194]

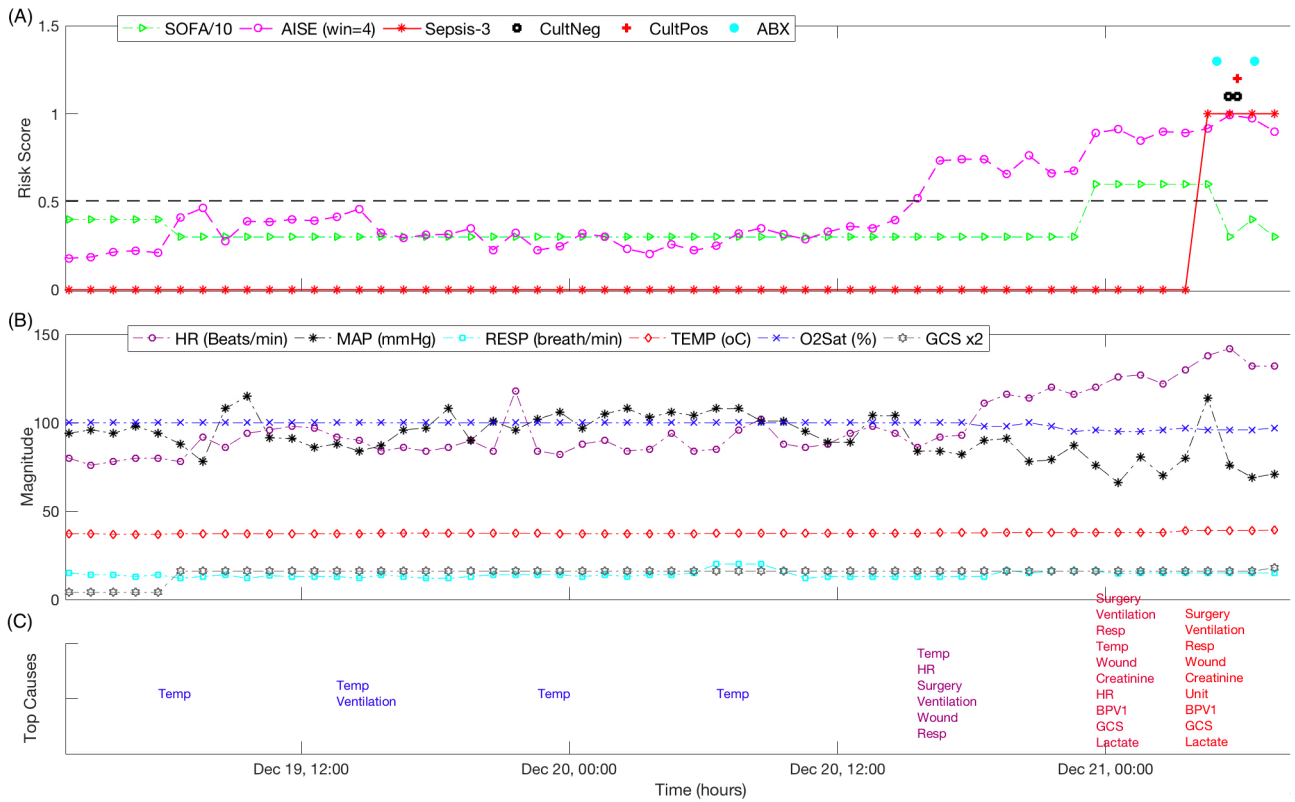


**FIGURE 1.**

Receiver Operating Characteristic (ROC) curves for predicting  $t_{\text{sepsis}}$  4 hours in advance. Catching 85% of the septic patients yielded 30% false alarms (SP=0.70) within the training set (left panel) and 33% false alarms (SP=0.67) within the testing set (right panel). See Table 2 for information on the false alarms.



**FIGURE 2.** Summary of training set (dashed lines) and testing set (solid lines) prediction performance of AISE on the Emory cohort. Area under the ROC curve (AUROC) as a function of prediction window shows a decreasing pattern. Across all windows, the best performance is achieved for predicting t<sub>SOFA</sub>, followed by t<sub>sepsis</sub>, and finally t<sub>onset</sub>. A close agreement between the training set and testing set performance indicates good generalizability.



**FIGURE 3.**

An illustrative example of the prediction performance of AISE. Hourly calculated Sequential Organ Failure Assessment (SOFA) Score, Sepsis-3 definition, and the AISE score are shown for one patient in Panel (A). Superimposed on the figure is the order-time of three blood cultures, and the administration-time of two antibiotics. In Panel (B), commonly recorded hourly vital signs of the patient, including heart rate (HR), Mean Arterial Blood Pressure (MAP), Respiratory Rate (RESP), Temperature (TEMP), Oxygen Saturation (O<sub>2</sub>Sat) and the Glasgow Coma Score (GCS) are shown. Panel (C) shows the most significant features contributing to the AISE score (for clarity of presentation only selected time-points are shown). Notably, around 4pm on December 20<sup>th</sup>, roughly 8 hours prior to any change in the SOFA score, the AISE score starts to increase. The top contributing factors were slight changes in HR, RESP, and TEMP, given that the patient had surgery in the past 12 hours with a contaminated wound, and was on a mechanical ventilator. Close to midnight on December 21<sup>st</sup>, other factors such as multiscale entropy of MAP time series (BPV1), GCS, and Lactate show abnormal changes. Five hours later, the patient met the sepsis-3 definition of sepsis.

**TABLE 1**

Summary of algorithm performance on the Emory cohort

Performance metric	4 hours	6 hours	8 hours	12 hours
<b>t<sub>sepsis</sub> Prediction Testing set (Training set)</b>				
<b>AUROC</b>	0.85 (0.86)	0.85 (0.85)	0.84 (0.84)	0.83 (0.83)
<b>Specificity<sup>*</sup></b>	0.67 (0.70)	0.67 (0.68)	0.65 (0.66)	0.63 (0.65)
<b>Accuracy</b>	0.67 (0.70)	0.67 (0.68)	0.66 (0.67)	0.63 (0.65)
<b>t<sub>SOFA</sub> Prediction Testing set (Training set)</b>				
<b>AUROC</b>	0.87 (0.88)	0.85 (0.87)	0.85 (0.86)	0.82 (0.84)
<b>Specificity<sup>*</sup></b>	0.72 (0.74)	0.67 (0.70)	0.65 (0.68)	0.59 (0.64)
<b>Accuracy</b>	0.72 (0.74)	0.68 (0.71)	0.66 (0.68)	0.60 (0.64)
<b>t<sub>onset</sub> Prediction Testing set (Training set)</b>				
<b>AUROC</b>	0.82 (0.83)	0.81 (0.82)	0.80 (0.81)	0.79 (0.80)
<b>Specificity<sup>*</sup></b>	0.64 (0.64)	0.62 (0.62)	0.58 (0.61)	0.56 (0.58)
<b>Accuracy</b>	0.64 (0.64)	0.62 (0.62)	0.58 (0.61)	0.56 (0.59)

\* Sensitivity was fixed at 0.85 (catching 85% of sepsis cases)

Relationship between probability of sepsis and mortality for the entire cohort, among the false positive cases, and the false negative cases.

**Table 2**

Probability (p) of $t_{sepsis} < 4$ hours	0<p<=0.2	0.2<p<=0.4	0.4<p<=0.6	0.6<p<=0.8	0.8<p<=1.0
<b>Overall Cohort</b>					
Expired (%)	0.5	2.8	8.6	17.3	32.9
Inpatient Hospice (%)	0.9	4.2	10.6	16.2	16.7
<b>False Positives</b>					
Expired (%)	--	--	8.3	20.1	56.3
Inpatient Hospice (%)	--	--	10.7	19.5	23.4
<b>False Negatives</b>					
Expired (%)	5.7	7.3	1.7	--	--
Inpatient Hospice (%)	2.9	8.2	11.7	--	--