Early Prediction of Sepsis From Clinical Data Using Ratio and Power-Based Features

Naimahmed Nesaragi, MTech; Shivnarayan Patidar, PhD

Objectives: Early prediction of sepsis is of utmost importance to provide optimal care at an early stage. This work aims to deploy soft-computing and machine learning techniques for early prediction of sepsis.

Design: An algorithm for early identification of sepsis using ratio and power-based feature transformation of easily obtainable clinical data.

Setting: PhysioNet Challenge 2019 provided ICU data from three separate hospital systems. Publicly shared data from two hospital systems are used for training and validation purposes, whereas sequestered data from all the three systems is used for testing.

Patients: Over 60,000 ICU patients with up to 40 clinical variables are sourced for each hour of their ICU stay. The Sepsis-3 criterion is applied for annotation.

Interventions: None.

Measurements and Main Results: The clinical feature exploration for early prediction of sepsis is achieved using the proposed framework named genetic algorithm optimized ratio and power-based expert algorithm. An optimal feature set containing 46 ratio and power-based features is computed from the given patient covariates using genetic algorithm optimized ratio and power-based expert and grouped with identified 17 raw features and 55 statistical features to form a final feature set of 118 clinical features to predict the onset of sepsis in the proceeding 6 hours. The obtained features are fed to a hybrid Random Under-Sampling-Boosting algorithm, called RUSBoost for alleviating the involved class imbalance. The optimal RUSBoost model has achieved a normalized utility score of 0.318 on full test data.

Conclusions: The proposed study supports the realization of a hospital-specific customized solution in the form of an early-warning system for sepsis. However, an extended analysis is necessary to apply this framework for hospital-independent diagnosis of sepsis in general. Nevertheless, the clinical utility of hospital-specific customized solutions based on the proposed method across a wide range of hospital systems needs to be studied. (*Crit Care Med* 2020; 48:e1343–e1349)

Key Words: early prediction; feature extraction; genetic algorithm; machine learning; sepsis

All authors: Department of Electronics and Communication Engineering, National Institute of Technology Goa, Ponda, India.

DOI: 10.1097/CCM.0000000000004691

Sepsis is an overwhelming bodywide response to an infection that at the beginning might seem symptomatically prosaic but suddenly worsens as the time progresses. It involves a chain of events that includes abnormal inflammation and clotting, leading to major organ failure and even death (1–3).

Against the manual prognosis of sepsis, the development of computer-aided or automated screening tools for sepsis is an active and important area of investigation (4–10). Recently, several studies have considered and implemented versions of common rule-based scoring systems into electronic medical records (EMRs) (11–13). However, in such rule-based scores, the heterogeneous infection source and diversity among the patient's population may not be accurately represented (14). In contrast, machine learning-based prediction tools have the potential to enable early intervention by clinicians by providing advance notice of sepsis risk with more generalizability (15–17).

This work aims to demonstrate that a high-performance machine learning-based model can be derived and applied for early prediction of sepsis using ratio and power-based feature transformation and RusBoost-based classification (18–21).

MATERIALS AND METHODS

The proposed study is developed and evaluated as a part of the PhysioNet Challenge 2019-based official and follow-up submissions.

Study Population and Data Sources

The publicly available PhysioNet Challenge 2019 training data comprise of 1,552,210 observations for 40,336 patient records with 40 clinical variables measured on an hourly basis (22). This dataset was collected with approval from the appropriate institutional review boards (22).

Feature Extraction

A total of 118 features are extracted from the given clinical data and used as input to an optimal RusBoost architecture for early

prediction of sepsis. The proposed framework with various stages of model development and testing approach is shown in **Figure 1** and the flowchart of genetic algorithm optimized ratio and power-based expert (GRAPE) is shown in **Figure 2** (for details on proposed framework and flowchart, see **Appendix A**, Supplemental Digital Content 1, http://links.lww.com/CCM/F856). The main aspects of GRAPE are discussed below.

Feature Exploration Using GRAPE. Capturing static interrelations among the vital signs of clinical time series, rather than considering them independently, can enhance the clinical potential of the extracted feature set for automated analysis. In literature, such interrelations among the covariates in the form of ratio and power have been found to improve the capabilities of anomaly detection systems (23). Given this, we have explored all possible ratio and power-based features by combining the maximum of three covariates, as defined by the following two spaces:

$$R1 = \left\{ \frac{x^k}{y^m - z^n} : x, y, z \in P1; -9 \le k, m, n \le 9 \right\} \quad (1)$$

$$R2 = \left\{ \frac{x^k}{y^m - z^n} : x, y, z \in P2; -3 \le k, m, n \le 3 \right\} \quad (2)$$

where *x*, *y*, and *z* denote any three of the given clinical signs considered to form the said ratios and power-based features. The range of the *k*, *m*, and *n*, that is, power indices, is restricted for feasible exploration from –9 to 9 and –3 to 3 for *R1* and *R2* spaces, respectively, because of involved computational complexity. However, exploration beyond this said limit is open and can be addressed as a part of limitation and future work.

Furthermore, in Equation 1, *P1* is the vector containing 38 of the 40 given patient signs, ignoring the two values of administrative identifier for ICU units. In addition, the specified space is intended for static ratio and power-based exploration that contains ${}^{38}C_3 \times {}^{19}C_3$ number of explorable sample points.

Similarly, in Equation 2, vector *P2* is the same as *P1* excluding length of stay. And the specified space mainly explores derivative interrelations statically that contains ${}^{37}C_3 \times {}^{7}C_3$ number of explorable sample points. The derivative-based operation on covariates would capture information on the dynamics involved among the given clinical signs.

At the onset, genetic algorithm (GA)-based optimization (24) is performed for feature exploration and identification of the clinically influential features that can maximize the early prediction performance of the sepsis in terms of the normalized utility function U_n (for details on GA-based optimization, see Appendix A, Supplemental Digital Content 1, http://links.lww.com/CCM/F856).

Temporal Dynamics-Based Features. To capture the involved temporal dynamics within the covariates, the following 55 features are also identified from the given patient data: statistical and quantile-based features over sliding window of last 5- and 11-hour observations. Here, two different time windows (i.e., 5 and 11 hours) are used to capture the short- and long-term

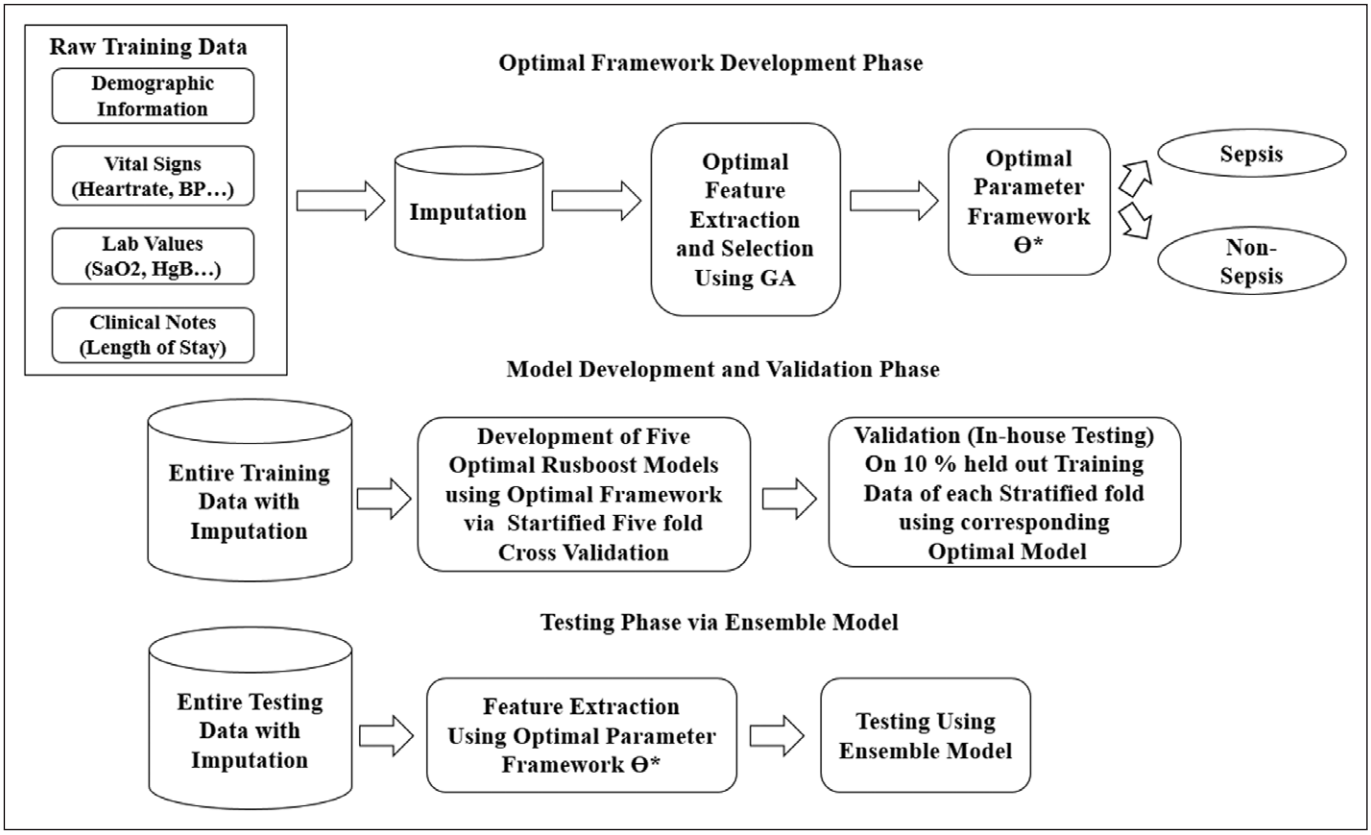

Figure 1. The proposed framework.

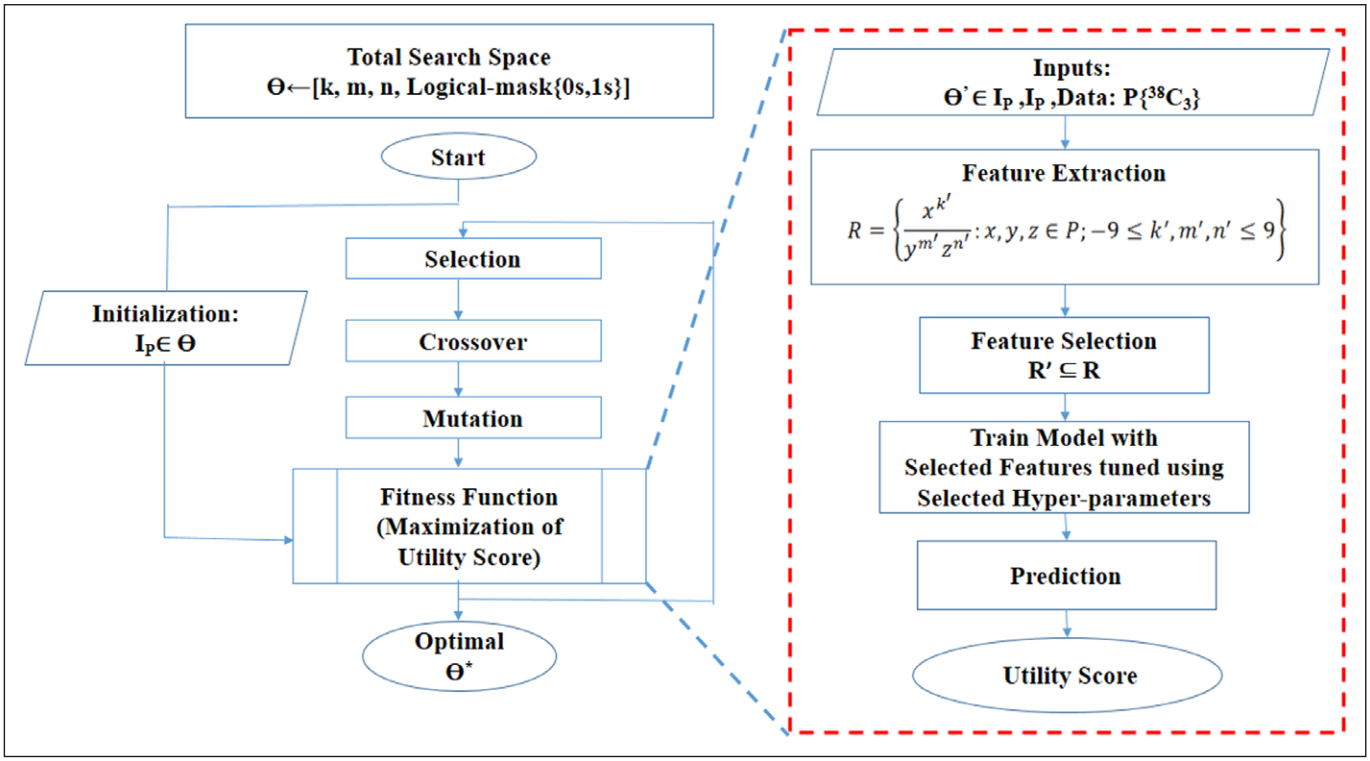

Figure 2. The flowchart of the GRAPE depicting the formulation of the fitness function.

temporal evolutions over each of the first eight covariates only, as they represent vital signs excluding laboratory values and demographics. Furthermore, descriptors like "mean" of the first differences, "energy," and "Shannon entropy" are also calculated from the given frame over nonsliding window-based trend.

Resultant Optimal Feature Set. Finally, the obtained dynamic 55 features are combined with the five out of 17 optimal static features identified from *R1* space, as shown in **Table 1**, along with the selected 17 given clinical signs and 41 derivative features obtained by *R2* space defined in Equation 2 to form a resultant optimal feature set of 118 variables (for a detailed listing of derivative and temporal dynamics-based features, see **Appendix C**, Supplemental Digital Content 1, http://links.lww.com/CCM/F856).

Implementation of GRAPE

Using GA in MATLAB, an optimal sepsis-detection framework called GRAPE is designed and implemented for classifying the records into nonsepsis and sepsis classes. As the given data have a high density (approximately 20%) of missing values, so at the onset, as a preprocessing step, the algorithm begins with the imputation of given clinical data. The missing value is filled by performing linear interpolation of neighboring two nonmissing values. The trailing missing values in a row vector (or feature vector) are filled with the nearest nonmissing value in that row (for details on preprocessing, see **Appendix B**, Supplemental Digital Content 1, http://links.lww.com/CCM/F856).

During model development, 90% of the training set is used for cross-validation, and remaining is used for testing purposes. To find the useful clinical features, GA-based optimization with feature-transformation and feature-selection processes from the given patient signs are applied to characterize sepsis. Optimization is done to maximize the underlying classification performance in terms of utility score. It is to be noted that only indices of the possible combinations of the ratio-based features are passed to the objective function. In addition, actual computation of GA-based selected ratios with powers is performed within the objective function. A supervised ensemble machine learning model called RusBoost is used to handle the class-imbalance problem with default settings for the underlying classification of the sepsis and nonsepsis records (for further details on machine learning model, see Appendix B, and for relevant glossary, see **Appendix D**, Supplemental Digital Content 1, http://links.lww.com/CCM/F856). Nevertheless, a five-fold cross-validation scheme is implemented to get robust performance while doing optimization.

As a result of exploring *R1* space with GA-based optimization, 17 most influential ratio and power-based clinical features are identified (Table 1). In addition, *R2* space has resulted in the identification of 41 most influential derivative-based ratio and power-indexed clinical features (Table C1, Supplemental Digital Content 1, http://links.lww.com/CCM/F856). The settings for optimization include integer populations, integer mutation with stall generation of 50, and population size of 30. The optimal values of *k*, *m*, and *n*, which ultimately form the ratio-based features, are obtained after executing GA.

RESULTS

This study has been evaluated as two challenge submissions as official and follow-up entries on the sequestered data. For the

TABLE 1. List of 17 Most Influential Ratio and Power-Based Clinical Features Identified From R1 Space Using the Proposed Algorithm

S. No	Component of Features	Type
1	End-tidal CO_2/partial thromboplastin time (s)	x/y^2
2	Diastolic BP/gender	x^4/y^8
3[a]	Diastolic BP/gender	x/y^2
4[a]	Heart rate/age	x/y^2
5[a]	Age/gender	x/y^2
6[a]	Heart rate/(systolic BP × age)	x/yz
7	Heart rate (beats/min)	x^5
8	Temperature (°C)	x^4
9	Mean arterial pressure (mm Hg)	$1/y^2$
10	Diastolic BP (mm Hg)	x^2
11	End-tidal CO_2 (mm Hg)	x^4
12	FIO_2 (%)	x^8
13[a]	Alkalinephos (international units/L)	x^6
14	Creatinine (mg/dL)	$1/y^3$
15	Fibrinogen (mg/dL)	x^4
16	Age	x^7
17	ICU length of stay	$1/y$

[a]Feature numbers 3–6 and 13 are selected for the current study.

official phase entry, a single RusBoost model is designed using the entire training data for the prediction of sepsis. Whereas for the follow-up entry, five models are designed based on patientwise stratified five-folds, each containing unique 20% of the entire training set. This ensemble system is used for the early prediction of sepsis on test data.

The hyperparameters of the above RusBoost models that minimize five-fold cross-validation loss are obtained by using automatic hyperparameter optimization utility, as given in MATLAB. The given software utility finds optimal parameters automatically using Bayesian optimization. At the outset, the obtained RusBoost models include 96 classification trees for boosting with a minimum leaf size of 511.

Official Phase Challenge Submission

Table 2 gives a summary of the results of the proposed method obtained during the official phase of the challenge for *R1* space-based features. Basically, obtained 17 features of *R1* space are grouped with the given first 38 out of 40 signs to form the final feature set of 55. Table 2 depicts the results for 90% of the training data in a five-fold cross-validation scheme. Results also include on intersets, individual cohorts, and a full set of the hidden test set. It is noteworthy that the model trained on the combined cohorts A and B outperforms the individual trained models. This approach was officially rated 18th out of 79 teams during the official phase. This work appears in the ranking table with the team name as "Shivpatidar." The final results in the leader board demonstrate a utility score of 0.3090 for the full test data that include a separate test set C along with remaining test subsets of A and B of the 2019 PhysioNet/CinC Challenge dataset.

Follow-Up Challenge Submission

Table 3 gives a summary of the results of the follow-up entry with a resultant optimal feature set of 118 features on the training data in a five-fold cross-validation scheme. Results also include on intersets, individual cohorts, and a full set of the hidden test set. During training with combined cohorts A and B, the five-fold cross-validation produced an average utility score of 0.4013, resulting in a substantial improvement of ~3% compared with the results of *R1* space, as tabulated in Table 2.

Comparison of GRAPE With Baseline

Furthermore, we justify the clinical relevance of our proposed feature set by comparing it with baseline study. As a part of baseline study, we performed two well-tuned simpler methods: the given 40 clinical signs are fed to RusBoost model in a five-fold cross-validation scheme without imputation and with imputation. The results of these two base studies are presented in **Tables 4** and **5**, respectively. As expected, the performance was significantly worse in the case of the study carried without imputation. Even with imputation, results are significantly lower by 2–3% in terms of the utility score compared with the proposed work.

DISCUSSION

This study explores the use of GA-optimized ratio and power-based features for the prediction of sepsis in ICU. The main contributions of this study are: 1) capture ratio and power-based interrelations among the vital signs of clinical time series, rather independently, 2) robust to handle EMRs with dominant missing data, 3) automatic selection of features from the given clinical data, and 4) achieve a significant utility score with a few feature.

As an alternative to gradient-boosting schemes, the multilayer neural network with enough number of neurons and hidden layers can be employed to learn such multiplicative interactions and extract similar features. However, this may be achieved at the penalty of high computational cost. Even in our proposed work, the bounds of total search space for GA are limited because of this constraint and exploring beyond these limited bounds can enhance the results with more features but at the expense of high computational cost.

Furthermore, literature published related to Challenge submissions (25–28) reported that boosting schemes have emerged as better candidates for handling class-imbalance problems as considered in this study when compared with the neural network. Therefore, we feel a baseline study with a multilayer neural network is out of the scope of the current study as we had to resort to RusBoost because of solving the class-imbalance problem.

TABLE 2. Summary of the Training Results of the R1 Space on 90% of the Training Data in a Five-Fold Cross-Validation Scheme

Fold	Area Under the Receiver Operating Characteristic	Accuracy	Utility Score
1	0.8312	0.8951	0.3447
2	0.8282	0.9030	0.3147
3	0.8147	0.8317	0.4138
4	0.8172	0.8320	0.4213
5	0.8104	0.8394	0.3682
Average (SD)	0.8203 (0.0089)	0.8602 (0.0357)	0.37254 (0.0453)
Training on set A and testing on set B			0.2881
Training on set B and testing on set A			0.2984
Utility score of full test			0.3090
Utility score of full test A			0.3900
Utility score of full test B			0.3860
Utility score of full test C			−0.2120

Training results include performance on intersets. Testing results include performance on individual hidden cohorts and on the full set of the hidden test set.

TABLE 3. Summary of the Training Results of the Optimal Resultant Feature Set on 90% of the Training Data in a Five-Fold Cross-Validation Scheme

Fold	Area Under the Receiver Operating Characteristic	Accuracy	Utility Score
1	0.8554	0.9056	0.3906
2	0.8197	0.8562	0.3915
3	0.8354	0.8762	0.3982
4	0.8390	0.8644	0.4429
5	0.8663	0.9163	0.3927
Average (SD)	0.8432 (0.0181)	0.8837 (0.0261)	0.4013 (0.0224)
Training on Set A and testing on Set B			0.3031
Training on Set B and testing on Set A			0.3183
Utility Score of full test			0.3180
Utility Score of full test A			0.4010
Utility Score of full test B			0.3750
Utility Score of full test C			−0.1640

Training results include performance on intersets. Testing results include performance on individual hidden cohorts and on the full set of the hidden test set.

In addition, regarding the justification of clinical relevance of the selected features, these features are created from given 40 clinical signs, and it is to be noted that the clinically significant properties manifest in these clinical signs itself and are a part of simpler measurements adhering up to some extent to Sepsis-3 rules. Therefore, it is ubiquitous that the optimally selected features in the form of powers of the ratios of vital signs that maximize the utility score are also clinically more significant from the feature-engineering perspective.

The proposed framework promises to provide a hospital-specific customized solution to get early warning of sepsis. However, as indicated in Tables 2 and 3, the proposed methods have not rendered requisite performance on hidden test set C, which hints designing and deploying a hospital-independent sepsis-prediction model is a challenging task. Even the clinical utility in the general framework demands the need for a common hospital-independent sepsis-prediction system. In addition, the challenge for addressing the same yet remains open.

TABLE 4. Summary of the Baseline Results Using Given 40 Clinical Signs and RusBoost Classifier Without Imputation in a Five-Fold Cross-Validation Scheme

Fold	Area Under the Receiver Operating Characteristic	Accuracy	Utility Score
1	0.7883	0.9018	0.2719
2	0.7432	0.8365	0.3041
3	0.7679	0.8760	0.3098
4	0.7627	0.8489	0.3474
5	0.7837	0.9052	0.2969
Average (SD)	0.7692 (0.0089)	0.8737 (0.0357)	0.3060 (0.0453)
Training on set A and test on set B			0.2472
Training on set B and test on set A			0.2541

Training results of intersets are also included.

TABLE 5. Summary of the Baseline Results Using Given 40 Clinical Signs and RusBoost Classifier With Imputation in a Five-Fold Cross-Validation Scheme

Fold	Area Under the Receiver Operating Characteristic	Accuracy	Utility Score Baseline (Proposed)
1	0.8161	0.8715	0.3177 (0.3447)
2	0.7830	0.8079	0.3629 (0.3147)
3	0.8094	0.8374	0.3623 (0.4138)
4	0.7922	0.8149	0.3882 (0.4213)
5	0.8264	0.8856	0.3369 (0.3682)
Average	0.8054	0.8435	0.3536 (0.3725)
Training on set A and test on set B			0.2605 (0.2881)
Training on set B and test on set A			0.2729 (0.2984)

Training results of intersets are also included.

CONCLUSIONS

This study explores the strength of ratio and power-based features for automated diagnosis of sepsis. An extended framework can serve as an early-warning system for sepsis. Feature space exploration beyond the considered limits can further improve predictive performance.

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/ccmjournal).

The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: shivnarayan.patidar@nitgoa.ac.in

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:801–810
2. Seymour CW, Liu VX, Iwashyna TJ, et al: Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:762–774
3. Shankar-Hari M, Phillips GS, Levy ML, et al; Sepsis Definitions Task Force: Developing a new definition and assessing new clinical criteria for septic shock: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:775–787
4. 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
5. Calvert J, Desautels T, Chettipally U, et al: High-performance detection and early prediction of septic shock for alcohol-use disorder patients. *Ann Med Surg (Lond)* 2016; 8:50–55
6. Calvert JS, Price DA, Chettipally UK, et al: A computational approach to early sepsis detection. *Comput Biol Med* 2016; 74:69–73
7. 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:e0174708
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:e28
9. Zador Z, Landry A, Cusimano MD, et al: Multimorbidity states associated with higher mortality rates in organ dysfunction and sepsis: A data-driven analysis in critical care. *Crit Care* 2019; 23:247

10. Knaus WA, Marks RD: New phenotypes for sepsis: The promise and problem of applying machine learning and artificial intelligence in clinical research. *JAMA* 2019; 321:1981–1982
11. Berger T, Birnbaum A, Bijur P, et al: A computerized alert screening for severe sepsis in emergency department patients increases lactate testing but does not improve inpatient mortality. *Appl Clin Inform* 2010; 1:394–407
12. Hooper MH, Weavind L, Wheeler AP, et al: Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit*. *Crit Care Med* 2012; 40:2096–2101
13. Semler MW, Weavind L, Hooper MH, et al: An electronic tool for the evaluation and treatment of sepsis in the ICU: A randomized controlled trial. *Crit Care Med* 2015; 43:1595–1602
14. Fohner AE, Greene JD, Lawson BL, et al: Assessing clinical heterogeneity in sepsis through treatment patterns and machine learning. *J Am Med Inform Assoc* 2019; 26:1466–1477
15. Nemati S, Holder A, Razmi F, et al: An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018; 46:547–553
16. Shashikumar SP, Stanley MD, Sadiq I, et al: Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J Electrocardiol* 2017; 50:739–743
17. Shashikumar SP, Li Q, Clifford GD, et al: Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas* 2017; 38:2235–2248
18. Polikar R: Ensemble based systems in decision making. *IEEE Circuits Syst Mag* 2006; 6:21–45
19. Breiman L: Bagging predictors. *Mach Learn* 1996; 24:123–140
20. Schapire RE: The strength of weak learnability. *Machine Learning* 1990; 5:197–227
21. Freund Y, Laboratories TB: Boosting a weak learning algorithm by majority. *Inform Comput* 1995; 121:256–285
22. Reyna MA, Josef CS, Jeter R, et al: Early prediction of sepsis from clinical data: The PhysioNet/computing in cardiology challenge 2019. *Crit Care Med* 2020; 48:210–217
23. Delahanty RJ, Alvarez J, Flynn LM, et al: Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med* 2019; 73:334–344
24. Tang KS, Man KF, Kwong S, et al: Genetic algorithms and their applications. *IEEE Signal Process Mag* 1996; 13:22–37
25. Wang Y, Xiao B, Bi X, et al: Prediction of Sepsis From Clinical Data Using Long Short-Term Memory and eXtreme Gradient Boosting. 2019. Computing in Cardiology Conference (CinC). IEEE, Matrix, Biopolis, Singapore, September 8-11, 2019
26. Zabihi M, Kiranyaz S, Gabbouj M: Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models. 2019. Computing in Cardiology Conference (CinC). IEEE, Matrix, Biopolis, Singapore, September 8-11, 2019
27. Lyra S, Leonhardt S, Hoog Antink C: Early Prediction of Sepsis Using Random Forest Classification for Imbalanced Clinical Data. 2019. Computing in Cardiology Conference (CinC). IEEE, Matrix, Biopolis, Singapore, September 8-11, 2019
28. Sweely B, Park A, Winter L, et al: Time-Padded Random Forest Ensemble to Capture Changes in Physiology Leading to Sepsis Development. 2019. Computing in Cardiology Conference (CinC). IEEE, Matrix, Biopolis, Singapore, September 8-11, 2019