

Early Prediction of Sepsis Considering Early Warning Scoring Systems

P Biglarbeigi¹, D McLaughlin¹, K Rjoob², Abdullah¹, N McCallan¹, A Jasinska-Piadlo³, R Bond², D Finlay¹, KY Ng¹, A Kennedy¹, J McLaughlin¹

¹ NIBEC, Ulster University, Newtownabbey, UK

² School of Computing, Ulster University, Newtownabbey, UK

³ Cardiology Registrar, Craigavon Area Hospital

Abstract

Sepsis is a noted cause of mortality in hospitalised patients, particularly patients in the ICU. Early prediction of sepsis facilitates a better targeted therapy which in turn reduces patient mortality rates. This study developed a methodology to allow automatic prediction of sepsis 6 hours prior to its clinical presentation. For this purpose, four vital signs comprising of HR, SBP, Temperature and respiratory rate, along with laboratory results for Platelets, WBC, Glucose and Creatinine are scored using Prehospital Early Sepsis Detection (PRESEP) and Sequential Organ Failure Assessment (SOFA) Early Warning Scoring (EWS) systems or screening tools and Systemic Inflammatory Response Syndrome (SIRS) criteria to allow under-sampling. The weighted scores obtained from the screening tools are also used to categorise patients into 4 groups with different probabilities of facing sepsis in ICU. The hourly data of each group is then trained through a KNN classifier to detect sepsis hours. The ensemble of classifiers are used to predict sepsis in all available dataset. The proposed model developed by UlsterTeam is trained on training setA and evaluated on training setB. The evaluation of the model on the training setB of the publicly available dataset shows the Utility Score, accuracy, AUROC and AUPRC of the model are 0.27, 0.97, 0.71 and 0.07 respectively.

1. Introduction

Sepsis, severe sepsis and septic shock are different terms used to address disorders characterised by a host response to physiological, pathologic and biochemical abnormalities caused by infection [1]. Sepsis is defined as a “life-threatening organ dysfunction caused by a dysregulated host response to infection” [2], and organ dysfunction is defined by an increase in Sequential Organ Failure Assessment (SOFA) score [3].

Sepsis is believed to be one of the leading causes of mortal-

ity for hospitalised patients and has been a very noticeable cost concern for healthcare systems worldwide [4]. However, due to the diverse nature of the possible infection and host reaction, diagnosing sepsis remains to be a difficult task for physicians [1]. Moreover, it has been argued that the survivors of sepsis may suffer from physical and psychological complications in the long-term, which requires more social and healthcare support [5].

Therefore, early prediction of sepsis results in early treatment of patients at-risk and reduces the possibility of severe sepsis and septic shock. It also provides a means to develop more effective preventive methods in treating patients with sepsis [6]. This study aims to provide a numerical methodology to automatically predict sepsis in ICU patients at least 6 hours in advance. The dataset used in this study has been published by Physionet [7] as part of the 2019 Computing in Cardiology challenge .

2. Methodology

The dataset of this study is sourced from 40,336 ICU patients with 40 different time-dependent (i.e. with one-hour time step) variables including Vital Signs, Laboratory (VSL) values and Demographics [7]. This study aims to provide a comprehensive methodology to overcome the challenges concerning the dataset. In this section, each of these challenges and the proposed methodology to overcome them will be addressed.

2.1. Variable selection

This dataset includes about 31.63% of NaN values, corresponding to the hourly values that have not been measured at the time for each patient. Therefore, it is essential to choose the variables that have the least number of NaNs and are clinically important in diagnosing sepsis. In order to choose these variables, sepsis screening tools have been studied. Screening tools are widely used as decision support systems for early detection of sepsis in both medical and surgical patients [8]. In this study, two screen-

ing tools of Prehospital Early Sepsis Detection (PRESEP) and SOFA as well as the Systemic Inflammatory Response Syndrome (SIRS) criteria are considered. PRESEP score is a screening tool that has been developed to identify septic patients in a pre-hospital environment based on variables that are typically assessed [9] with cut-off values that have empirically validated. PRESEP score is defined as the sum of scored weights of Heart Rate (HR), Temperature (Temp), Systolic Blood Pressure (SBP), Respiratory Rate (RR), peripheral oxygen saturation (SaO₂) and Glasgow Coma Scale (GCS). Different studies have classified PRESEP among the best performing screening tools in identifying septic patients considering its sensitivity and specificity [9], [10].

SIRS is an inflammatory response syndrome criteria that was originated in the 90s [11] as a starting point in the definition of sepsis, considering six different variables of Temperature, HR, Leukocyte count (WBC), RR and Blood Glucose. SIRS is not used in clinical applications anymore, however, it is still a relevant identification of infection for the at-risk group of patients [12]. Finally, SOFA Score is considered the official screening tool to be used as the diagnostic criteria in organ dysfunction and sepsis, according to the 4th Sepsis Manual [12]. SOFA scoring considers Creatinine and Platelets as two other important variables in early sepsis prediction.

Furthermore, a variable selection approach has been performed on the available dataset in order to verify the choice of the variables. Variable selection approach considers a set of strategies to find an optimal subset of variables that provide a good performance in model construction [13], [14]. In this study, a hybrid variable selection approach of Joint Mutual Information (JMI) was applied to rank the 40 available variables based on their importance. JMI [15] is the information between the target class and a random variable, x , which is defined by pairing a variable, x_i , with another variable, x_k , where i and k are two out of 40 available variables. JMI ranked Hours between hospital admit and ICU admit (HospAdmTime), Age, HR, SBP, Mean arterial pressure (MAP), Glucose, Diastolic Blood Pressure (DBP), Platelets, RR and Blood urea nitrogen (BUN) as the first 10 variables based on their importance.

Consequently, in this study the four Vital Signs of Temp, HR, SBP and RR are scored according to PRESEP screening tool. WBC and Blood Glucose as well as Creatinine and Platelets are scored according to SIRS and SOFA, respectively. Additionally, the Demographic variables of Age, Gender, HospAdmTime and ICU length-of-stay (ICULOS) are also considered to construct a model for early detection of Sepsis.

2.2. Missing values

The missing values (NaNs) in the 8 selected VSL variables (listed in Table 1), corresponding to missing measurements, have been categorised into two classes. The first class is scattered NaNs defined by missing values with at least one measurement recorded for a specific variable for a patient. For this class, the final measurement of the variable for each patient has been used to fill the missing measurements.

The second class of NaNs is defined as the NaN columns, for which there is no measurement recorded for a specific variable. Overall, about 2.94% of the 8 selected columns of VSL values among all the patients are NaN columns. In order to fill these missing values, the average of each variable over the time of stay in ICU is considered as the defining feature of that variable per patient. Considering the mean of variables, each patient can be defined by 8 features describing their VSL variables. Therefore, each NaN value in the mean feature datasets corresponds to a NaN column of a variable for a patient. Markov Chain Monte Carlo (MCMC) simulation is used to impute these NaN values in the mean feature dataset. MCMC is a combination of techniques to randomly simulate sample values of the dataset with a known state space to create a Markov Chain [16]. The state space is the domain or space from which the randomly simulated data takes its value [17]. The state space in this study is considered as the multivariate normal (Gaussian) distribution of the dataset. Therefore, by considering a multivariate normal distribution for the 8 mean features and using their mean and covariance matrix, 100,000 random samples of data are generated through MCMC. Accordingly, for a patient with a NaN column, all the other 7 non-NaN features are considered. A search engine was then set to look into the MCMC generated dataset to find all the rows with values close to the 7 non-NaN features with a tolerance of 1 unit for each feature. Eventually, the NaN value is imputed by computing the average of the corresponding feature in the obtained rows of the MCMC generated dataset having a relatively close values for the non-NaN features. This feature is then used to fill the whole column of that specific variable in the patient's hourly dataset.

2.3. Under-sampling

After filling the scattered missing values and the NaN columns, it is important to consider the class balance of the dataset. In the current dataset, out of 40,336 only about 7.27% of the patients (2932 patients) face sepsis at some point of their stay in ICU. Therefore, due to the imbalance of the classes, it is necessary to under-sample the dataset in order to obtain an optimal performance in the classification model. To meet this objective, in this study, the

mean features of the 8 variables considered for each patient have been scored using PRESEP, SIRS and SOFA screening tools, as presented in Table 1. Moreover, as most of the

Variable	Screening tool	Cutoff	Weights
Temp	PRESEP	$> 38^{\circ}C$	4
	PRESEP	$< 36^{\circ}C$	1
HR	PRESEP	> 90 BPM	2
SBP	PRESEP	< 90 mmHG	2
RR	PRESEP	> 22 BPM	1
Platelete	SOFA	$< 20 * 10^3 / \mu L$	4
	SOFA	$< 50 * 10^3 / \mu L$	3
	SOFA	$< 100 * 10^3 / \mu L$	2
	SOFA	$< 150 * 10^3 / \mu L$	1
Creatinine	SOFA	> 8.8 mg/dL	4
	SOFA	> 6 mg/dL	3
	SOFA	> 3.42 mg/dL	2
	SOFA	> 2.2 mg/dL	1
Glucose	SIRS	> 138.6 mg/dL	1
WBC	SIRS	$\geq 12 * 10^3 / \mu L$	1
	SIRS	$\leq 4 * 10^3 / \mu L$	1

Table 1. Weighted Scoring System [9], [12]

septic patients face sepsis during the final hours of their stay in ICU, the mean features of the variables have also been compared to the first hour of their ICU admission, in order to represent the increase or decrease of the variable during the ICU hours for each patient. The supplementary scores (SS) for this comparison is presented in Equation 1.

$$SS_i = \begin{cases} 1 & x_{i,avg} - x_{i,1} > 0 \\ -1 & x_{i,avg} - x_{i,1} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where i corresponds to each variable (listed in Column 1 of Table 1), while $x_{i,avg}$ and $x_{i,1}$ are the mean feature of the variable i and the value of the variable at the first hour of admission to ICU, respectively.

Considering this scoring, each patient is defined by a combination of 16 scored values. All the 40,336 patients fall into 15,990 of unique combination of these 16 scores. In order to under-sample the dataset, the groups of patients having the same combination of scores are studied. If none of the patients in a group had sepsis, the oldest patient is added to the undersampled new dataset, as the probability of severe sepsis or septic shock is higher in older population [18]. On the other hand, if only one of the patients of the group faced sepsis, all group members have been considered in the new dataset. By this analysis, the number of patients in the under-sampled new dataset has been reduced to 25,569 patients, and the probability of sepsis in the new dataset has been increased to 11.47%.

2.4. Classification

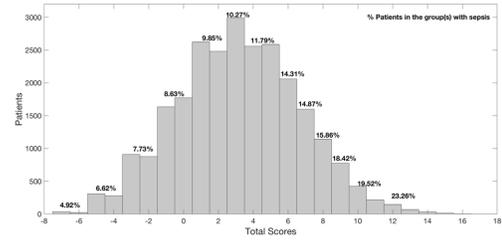
The sum of the 16 weighted scores obtained in Section 2.3 for each patient can be an indicator for patients at risk of

	Score range	Patients	Sepsis ratio
Group 1	overall score < 1	5835	8.11%
Group 2	$1 \leq$ overall score < 4	8096	10%
Group 3	$4 \leq$ overall score < 9	9949	13.28%
Group 4	overall score ≥ 9	1681	19.45%

Table 2. Specifications of the designated groups for training.

sepsis. Figure 1, represents the histogram of the sum of the weighted scores for the selected patients as the under-sampled new dataset. The ratio of patients with sepsis is shown on each range of the weighted scores. It is clear that the higher the sum of the weighted scores, the higher the probability of patients facing sepsis. Therefore, in order to obtain a better performance in a classification model to detect sepsis and also to decrease the training time, the new dataset is categorised into 4 groups. The summary of the specifications corresponding to each group is summarised in Table 2. For each patient in the 4 groups of the

Figure 1. Histogram of the sum of weighted scored for under-sampled new dataset and the percentage of patients with sepsis in different groups



new dataset, the mean value of the variables for six-hour windows with 5-hour overlaps are considered to allow the model to predict sepsis in a 12-hour period. In each 6-hour window, the sepsis label is defined as 1 if at least in one hour of that window the patient faces sepsis. The 6-hour mean values of all the patients in each group are trained with a Fine KNN classifier with 10-fold cross validation.

3. Results and Discussion

As explained in Section 2.4, the undersampled new dataset is categorised into 4 groups, considering the overall scores of the three screening tools and Supplementary Score. Each group is then trained and classified by KNN classifier with 10-fold cross-validation. The True Positive Rate (TPR) and False Negative Rate (FNR) for both classes and the total accuracy of the model for training and test dataset, 80% and 20% of the available dataset, is presented in Table 3, respectively, as the average of all the four groups.

For a better evaluation the same methodology is applied to train a model on training setA (20,336 patients) and to test it on training setB (20,000 patients). These four models

		Predicted Class		Total Accuracy
		0	1	
80% dataset Training	True Class	0	>99 %	99.7 %
		1	4%	
20% dataset Testing	True class	0	99.9 %	99.6 %
		1	4%	

Table 3. Confusion matrix for Training (80% of the dataset) and Testing (20% of the dataset)

are evaluated using the scoring function provided by Physionet for the challenge. Utility function rewards the model for early prediction of sepsis, and penalises for the late predictions or misclassifications. Based on the function, the model shows the Utility score of 0.27, F-measure of 0.21, Accuracy of 0.97, AUPRC and AUROC of 0.07 and 0.71, respectively for training setB, which can be considered as an unseen dataset for the trained model here.

4. Conclusion

The developed algorithm in this study, combines clinical knowledge, numerical and statistical approaches in order to provide a reliable tool in predicting sepsis in ICU patients 6 hours in advance. The proposed methodology is able to overcome the challenges of missing values and the unbalanced dataset, considering the screening tools. It is found that this methodology can be a good method in prediction of 96% of septic patients of the same hospital data. However, it is less accurate considering the dataset of an unknown hospital to the model.

Acknowledgments

This project is supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB).

References

- [1] Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman MD, Barton C, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. *JMIR Medical Informatics* 2016;4(3):e28.
- [2] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama* 2016;315(8):801–810.
- [3] Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart C, Suter P, Thijs L. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine* 1996; 22(7):707–710.
- [4] Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, Angus DC, Reinhart K. Assessment of global incidence and mortality of hospital-treated sepsis. current estimates and limitations. *American Journal of Respiratory and Critical Care Medicine* 2016; 193(3):259–272.
- [5] Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *Jama* 2010;304(16):1787–1794.
- [6] Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL, Jay M, Das R. A computational approach to early sepsis detection. *Computers in Biology and Medicine* 2016;74:69–73.
- [7] Reyna M, Josef C, Jeter R, Shashikumar S, Westover M, Nemati S, Clifford G, Sharma A. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *CriticalCare Medicine* 2019. In Press.
- [8] Gyang E, Shieh L, Forsey L, Maggio P. A nurse-driven screening tool for the early identification of sepsis in an intermediate care unit setting. *J Hospital Medicine* 2015; 10(2):97–103.
- [9] Bayer O, Schwarzkopf D, Stumme C, Stacke A, Hartog CS, Hohenstein C, Kabisch B, Reichel J, Reinhart K, Winning J. An early warning scoring system to identify septic patients in the prehospital setting: the presep score. *J Academic Emergency Medicine* 2015;22(7):868–871.
- [10] Smyth MA, Brace-McDonnell SJ, Perkins GD. Identification of adults with sepsis in the prehospital environment: a systematic review. *BMJ Open* 2016;6(8):e011218.
- [11] Bone R. The systemic inflammatory response syndrome (sirs). In *Clinical Trials for the Treatment of Sepsis*. Springer, 1995; 3–12.
- [12] Daniels R, Nutbeam T. *The sepsis manual 4th edition* 2017-2018, 2018.
- [13] Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Machine Learning Research* 2003;3:1157–1182.
- [14] Nakariyakul S. High-dimensional hybrid feature selection using interaction information-guided search. *J Knowledge Based Systems* 2018;145:59–66.
- [15] Yang HH, Moody J. Data visualization and feature selection: New algorithms for nongaussian data. In *Advances in Neural Information Processing Systems*. 2000; 687–693.
- [16] Ni D, Leonard JD. Markov chain monte carlo multiple imputation using bayesian networks for incomplete intelligent transportation systems data. *Transportation Research Record* 2005;1935(1):57–67.
- [17] Geyer CJ. *Introduction to markov chain monte carlo*. Handbook of Markov Chain Monte Carlo 2011;20116022:45.
- [18] Nasa P, Juneja D, Singh O. Severe sepsis and septic shock in the elderly: an overview. *World Journal of Critical Care Medicine* 2012;1(1):23.

Address for correspondence:

Pardis Biglarbeigi
NIBEC, Ulster University, Newtownabbey, UK, BT37 0QB
p.biglarbeigi@ulster.ac.uk