Early Prediction of Sepsis: Using State-of-the-art Machine Learning Techniques on Vital Sign Inputs

Manmay Nakhashi[1*], Anoop Toffy[1*], Achuth P V[1], Lingaselvan Palanichamy[1], Vikas C M[1]

[1]Tricog Health India Pvt Ltd, Bangalore, India

Abstract

Electronic Health Records (EHRs) give a lot of information regarding a patient's progress in health, who is admitted to an Intensive Care Unit (ICU). Sepsis is a critical condition suffered by a patient who, if not treated in a timely manner can cause casualties. Machine learning algorithms have evolved to utilize EHRs to help doctors detect the onset of sepsis. In this work, we present a random forest-based ensemble machine learning technique to work on patient data, also called vital sign input, from ICU. The data we used is published as a part of the Physionet Challenge 2019 [11]. The proposed technique performs well on data that contain a major chunk as missing values due to the sparsity of measurement taken in an ICU. We used a combined classifier and an early predictor approach to accomplish the task. The classifier does the job of classification when the early prediction is not possible due to a lack of data. While early predictor predicts the onset of sepsis based on the patient's information it received from previous recordings of vital sign inputs. A utility metric score is used to evaluate the early predictor. The score increases with early predictions and decreases with late predictions as well as false alarms. Our team named 'Tricog' finished 58th in the challenge with a utility score of 0.149 in the official phase on the full test set data.

Key Words: Critical care; Electronic health records; Organ failure; Sepsis; Machine learning; Random forest

1. Introduction

Sepsis is a medical condition where the host organs start failing in a life-threatening manner because of the way the body responds to infection [1]. It causes significant public health expenditure, as noted in 2011 [2], where it accounted for over 5 percent of total US hospital expenditure, which is over $20 billion. It is said sepsis is a major concern for the population in the coming decade, that if not treated in a timely manner can affect the mortality rate and critical illness worldwide. A patient who survives sepsis can have serious health disorders and cognitive disabilities leading to healthcare and social implications [3]. The early treatment of sepsis hence plays an important role in minimizing the side effects caused by the sepsis condition to a patient. Various definitions are developed over the years for identifying sepsis. From Bone *et. al.* [5] in 1991, the International Sepsis Definitions Conference developed initial definitions that focused sepsis as the host's systemic inflammatory response syndrome (SIRS) to infection. Later scoring methods were introduced for quantifying organ dysfunction leading to the introduction of Sequential Organ Failure Assessment (SOFA) [6] scores. A higher SOFA score leads to an increased mortality rate for the patient. In 1991, an international consensus panel described and codified sepsis instances with complicated acute organ failures as 'severe sepsis' or 'septic shock' [7]. A SOFA score ≥ 2 reflects a case of sepsis for the general hospital population suspected with sepsis symptoms. Although another metric qSOFA [1] was introduced in order to identify sepsis with high sensitivity, it never replaced the SOFA score.

Many works exist in the literature on the early detection of sepsis by accurately predicting the onset of sepsis using vital sign inputs taken from patients since the time of admission. Studies conducted on EHRs using machine learning algorithms to accurately predict the onset of sepsis show that early detection of sepsis is possible using vital sign inputs [8, 9, 10]. Futoma *et. al.* [8] uses time-series data from patients, imputes missing values, and feeds it to a multi-output Gaussian processes model. The model helps in maintaining uncertainty about the patient's physiological state. The study shows it improved the early clinical detection of sepsis. But it

* All authors marked, contributed equally for the work

 ISSN: 2325-887X DOI: 10.22489/CinC.2019.259

provided room for improvement with more in-depth analysis of EHRs. In Horng *et. al*. [9], apart from vital signs and demographic data, they use free-text data from the emergency department to identify the infection. The work improved the receiver operating characteristic (ROC) curve for the vitals model. Lauritsen *et. al.* [4] proposes a deep learning-based early detection method which learns features by itself from clinical time-series data. It overcomes the shortcomings of machine learning using a deep learning approach on a diverse multicenter dataset. Their work uses a combination of a convolutional neural network and a long short-term memory network.

This paper is organized into the following sections. A short description of the content of the work is given in the abstract. Section 1 gives an introduction to sepsis by its definitions. It also gives different measures to identify the presence of sepsis. Section 2 describes the dataset used for getting the results. Section 3 gives a detailed overview of the methodology used in this paper to early predict sepsis. In section 4, we cover the results and accompanying discussion. Final section concludes the paper identifying the possible future directions.

2. Dataset and Preprocessing

The dataset [11] used for building our model is extracted three geographically distinct U.S. hospital systems with three different electronic medical record systems. Out of three so collected data, two sets comprising of 40336 patient data is used for training the machine learning models. The third set consisting of 24,819 patients from three hospital systems were sequestered as hidden test sets. The dataset used is very sparse and contains a lot of missing values. It covers features related to the physiology and demographics of the patient admitted to ICU. A detailed list of features and their units of measurement is given in Table 1 of [11]. The dataset used for training is labeled using sepsis-3 clinical criteria [1, 12, 13].

Apart from vital signs and laboratory values, the dataset provides demographic values such as age, gender, hospital admission time and ICU length of stay. Since the measurements are done on a need-to basis, most of the values of the vitals and others were filled with not applicable (NA). We used mean values from across the dataset to fill the NA since their presence can skew the predictions.

3. Methodology

The data was first scanned for outliers since a lot of garbage values were present in a few records. Box plots were used to clean up the outliers and remove the ambiguous records from the data set that can affect the prediction. The cleaned dataset is then imputed using mean values across the columns from the overall dataset. Since only 5% of the dataset consists of sepsis cases, the data is very sparse.

The method developed uses a two classifier-based approach in which one classifier serves to do the job of classification when the number of the recording of the admitted patient is less than a window size (w). The other classifier is invoked when the w^{th} measurement is done for the patient in ICU after the n^{th} hour. This classifier, so-called predictor, does the job of early prediction from the given window size, w. The functionality of the system is given in figure 1.

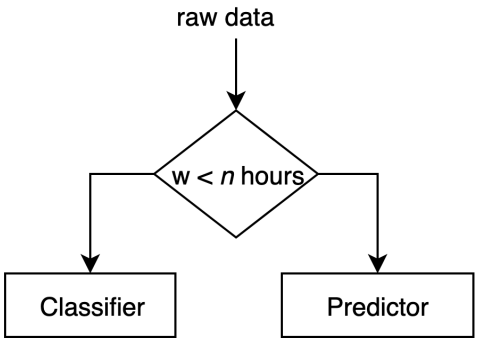

Figure 1: Approach to early detection of sepsis

3.1. Classifier and Predictor

We used a machine learning ensemble technique random forest for conducting our studies. The random forest is trained on approximately 90% of the available data. The other 10% is used as a blind test set for testing the classifier and arriving at a cross-validation score to make sure that the classifier generalizes the given dataset rather than overfit it.

A window size (w) of six (6) is used in our early prediction of sepsis method. Initially, we use classification if the window size is less than six. If the results are showing the onset of sepsis, we use zero fillings to create a window of data and verify the same using early predictor as shown in the block diagram in Figure 2.

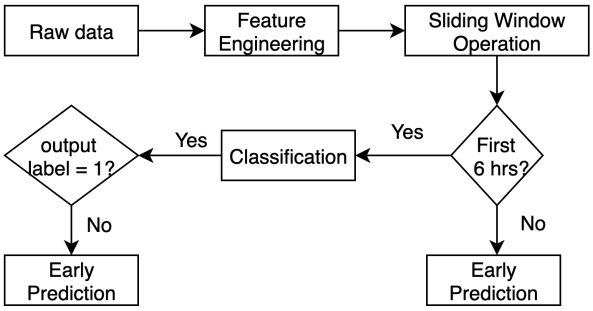

Figure 2: Overall architecture

4. Results and Discussions

This approach produced a satisfactory utility score in a held-out dataset and a completely new dataset taken from an entirely different hospital's electronic health records.

We performed 10-fold cross-validation on two training sets from two different hospitals to see if the random forest is learning. The 10-fold training accuracy on set A is 68.26% (standard deviation (SD) = 0.003), training area under receiver operating characteristic (AUROC) is 0.6945 (SD = 0.001) and training F-score is 0.0756 (SD = 0.0006). This model yielded a test accuracy on the training set B of 75.99%, a test AUROC of 0.6435 and a test F-score of 0.0581. Performing 10-fold cross-validation on the training set B yielded a train accuracy on train set B of 72.32% (SD = 0.005), a train AUROC of 0.7349 (SD = 0.002) and a train F-score of 0.0594 (SD = 0.0008). This model yielded a test accuracy on the training set A of 63.20%, a test AUROC of 0.6212 and a test F-score of 0.067.

4.1. Utility function

The utility function [11] scores a classifier giving a positive or negative score for the predictions done for each patient. The utility function calculates a score that rewards the early prediction between the 12th and 3rd hour and penalizes if the classifier fails to do so. Also, it penalizes false alarms a little less when compared to missing a true case as described in [11].

The above utility score is measured on the combined classifier and predictor system over a held-out data set. Below tables describe measures on three held out test sets and on the combined test set. The measures also include the area under the precision-recall curve (AUPRC).

Table 1: Utility scores on test sets and full test set

Test Set A	Test Set B	Test Set C	Full Test Set
0.249	0.152	-0.327	0.142

Table 2: Various measures on test sets

Measures	AUROC	AUPRC	Accuracy (%)	F-measure
Test Set A	0.721	0.045	85.5	0.113
Test Set B	0.736	0.03	86.6	0.08
Test Set C	0.669	0.012	74.2	0.033

The above measures were taken from the official ranking published after the challenge by Physionet. Our team named 'Tricog' finished 58th in the competition with a utility score of 0.142 on the full test set.

5. Conclusion

The suggested approach produced a satisfactory performance on the dataset provided. Though the dataset contains the majority of values missing, with our approach sepsis condition can be detected approximately six hours before its clinical onset.

6. Future directions

The proposed approach uses a basic machine learning ensemble technique with data preprocessing. There is a huge scope of improvement since the clinical time series can be modeled better using deep learning models, with the availability of data. Further research can be done on developing a recurrent neural network (RNN) based classifiers to accurately predict the onset of sepsis earlier from an approximately fixed window size. Later, a study can be done by varying the window sizes to figure out the optimal window size to predict the onset of sepsis.

7. Conflict of interest statement

The authors Manmay Nakhashi, Anoop Toffy, Achuth P V, Lingaselvan Palanichamy and Vikas C M are employed at Tricog Health India Pvt. Ltd.

Acknowledgments

We acknowledge the steering committee for CinC 2019 for allowing us to present our work as a part of the conference. This work was supported by Tricog Health India Pvt Ltd.

References

[1] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):801–10.

[2] Torio CM, Andrews RM. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: statistical brief# 160 [online] Available at: https://www.ncbi.nlm.nih.gov/books/NBK169005/ [Accessed 1 Oct. 2019].

[3] Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term Cognitive Impairment and Functional Disability Among Survivors of Severe Sepsis. JAMA. 2010 Oct 27;304(16):1787–94.

[4] Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, et al. Early detection of sepsis utilizing deep learning on electronic health record event sequences. CoRR. 2019;abs/1906.02956.

[5] Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. Chest. 1992 Jun 1;101(6):1644–55.

[6] Vincent J-L, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996 Jul;22(7):707–10.

[7] Angus DC, van der Poll T. Severe Sepsis and Septic Shock. N Engl J Med. 2013;369(9):840–51.

[8] Futoma J, Hariharan S, Sendak M, Brajer N, Clement M, Bedoya A, et al. An Improved Multi-Output Gaussian Process RNN with Real-Time Validation for Early Sepsis Detection. In: MLHC. 2017.

[9] Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One. 2017;12(4):1–16.

[10] Futoma J, Hariharan S, Heller K. Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier. arXiv e-prints. 2017 Jun;arXiv:1706.04152.

[11] Reyna M, Josef C, Jeter R, Shashikumar S, Westover M, Nemati S, Clifford GD, Sharma A. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. Critical Care Medicine. 2019 Aug.

[12] Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):762–74.

[13] Shankar-Hari M, Phillips GS, Levy ML, Seymour CW, Liu VX, Deutschman CS, et al. 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 Feb 23;315(8):775–87.

Address for correspondence:

Tricog Health India Pvt LTD
94/H 9th Cross, RMV Extn,
Bangalore - 560080, Karnataka, India
anoop@tricog.com