Anomaly Detection Semi-supervised Framework for Sepsis Treatment.

Ines Krissaane[1], Richard Wilkinson [1], Kingsley Hampton [2], Jumanah Alshenaifi[3]

[1] School of Mathematics and Statistics, University of Sheffield, Sheffield, UK
[2] IICD Department, University of Sheffield, Sheffield, UK
[3] MD Andreson UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA

Abstract

Sepsis is one of the leading causes of morbidity and mortality in hospitals. Early diagnosis could substantially improve the patient outcomes and reduce the mortality rate. In this paper we propose a machine learning approach for anomaly detection to aid the early detection of sepsis. Using the medical data of over 40,000 patients [1], we use both unsupervised and supervised methods to extract relevant features from the data, and then use standard classification approaches to predict sepsis six hours before clinical diagnosis occurs. To extract features, we used the reconstruction error of an auto-encoding neural network trained on control patients free of sepsis, and used random forest classifiers to learn the most important features for the classification of patients. We then combined the features from both of these approaches with a variety of standard classification models. Cross-validation as well as the asymmetric utility function designed for this challenge are used to evaluate the resulting models. We obtained a utility function score for the full unseen dataset of 0.177 (Team Kriss); achieved with a logistic regression classifier. All the implementation is publicly available at `https://github.com/ineskris/SepsisChallenge-Cinc2019`.

1. Introduction

Sepsis in one of the highest leading causes of hospitalized patients mortality worldwide [2]. Machine learning techniques have been used previously to improve the understanding sepsis [3] and to improve sepsis prediction [4]. As a contribution to the Physionet/Computing in Cardiology Challenge 2019 [1], this paper focuses on the implementation of a modelling pipeline using ML techniques to detect adverse events [5] as sepsis.

1.1. Sepsis condition

Sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated body's response to infection [6]. Sepsis follows a continuum through severe sepsis to septic shock, starting with the systemic inflammatory response syndrome (SIRS). The early detection of sepsis is highly valuable and may potentially save a patient's life [7], reduce medical complications, and reduce the cost to the healthcare system [8]. By interrogating a large dataset of hospitalized patients, our aim is to identify which clinical measurements can best be used to predict sepsis, and moreover, to train models to predict the probability a given patient will develop sepsis.

1.2. Anomaly detection

Sepsis detection can be seen as an anomaly detection problem where you may think about an individual who developed sepsis as a patient with specific abnormal representative features, i.e., for each patient, we aim to detect a change from their usual clinical measurements. We use data provided as part of the Computing in Cardiology Challenge 2019 [1] which contains clinical data from 40,336 ICU patients, of which 2,932 develop sepsis. The aim is to distinguish between the usual variation in patient data that occurs in ill patients who don't have sepsis, with the change in measurements that occur in patients who develop sepsis.

Our approach is to use a selection of standard classification methods, trained using carefully learned features based on two machine learning methods, random forests (RF) and auto-encoding neural networks (AENN), both of which have proven to be effective at extracting important features, and at identifying irregularities in data.

1.2.1. Supervised learning problem

A random forest [9] consists of multiple random decision trees, where each individual tree gives a class prediction. To classify a new instance, the forest chooses the classification having the most votes over all the trees in the forest. As a classifier, random forest performs an implicit feature selection, using a small subset of variables for the classification. And Gini importance can be used to indicate

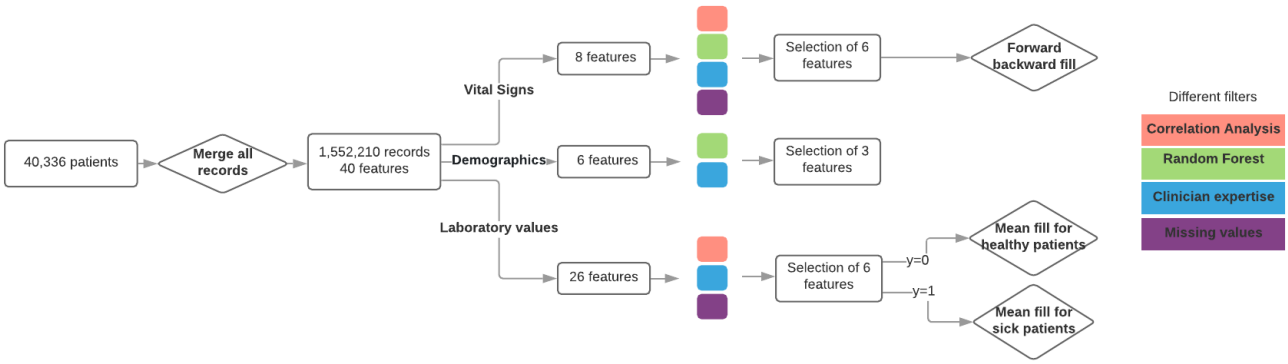


Figure 1: Data processing pipeline and feature selection

how often a particular feature was selected for a split [10].

1.2.2. Unsupervised learning problem

Sepsis is detected in 7% of the patients in our data. Since abnormality is a low probability event, data from the abnormal class are rare and detection can be targeted as an unsupervised learning problem. Unsupervised anomaly detection techniques detect anomalies in an unlabelled data set under the assumption that the majority of the instances in the data set are healthy (normal) patients, as in this case. Autoencoders [11] are a type of artificial neural network which learn to copy its inputs to its outputs, via a hidden layer of limited size. The representation of an input in its hidden layer can then be seen as an efficient compression of the data. We can use them in an anomaly detection scenario [12] considering a neural network trained only on the *normal* patients. For a new instance, we expect the reconstruction error to be higher for *abnormal* patients.

2. Materials and Methods

2.1. Data pre-processing and correlation analysis

For each patient, we used 40 clinical measurements over time, including demographics, vital signs, and laboratory values from where we eliminated variable which had greater than 80% missing values, except *Bilirubin direct*, *Lactate*, *PTT*, *Creatinine*, *WBC*, *Glucose* which are known as significant variables for detecting sepsis. In addition, the variables *Unit1* and *Unit2* are equally distributed across the whole dataset, and are uncorrelated with the sepsis label y, and were therefore removed. *SBP* and *DBP* are highly correlated with *MAP* via the relationship MAP = (SBP + 2*DBP) / 3, and they were discard for the RF feature selection but kept in the final prediction. For the laboratory values remaining variables, we impute any missing values with the mean obtained from the healthy group for this variable if the patient is healthy and conversely if the patient is sick. For testing on the unseen dataset, we impute the mean between both groups if there are missing values. For the vital signs variables, we impute any missing values by filling gap with the non missing values forward or backward along a Series. We merged all our patient files and add an extra column *ID* for the patient identifier. The variable *O2Sat* corresponds to the pulse oximetry percentage, and would be controlled for ventillated patients. Our methods find that *O2Sat* is not appearing as a predictive variable (See Fig. 3), and we did not include this variable in the analysis. For the demographic variables, we select *Age*, *Gender* and *ICULOS*. The data processing pipeline is described in Fig. 1. In the following, we kept 5500 files for final evaluation including 500 patients with sepsis.

2.2. Random forest importance features computation

We want to evaluate the importance of the remaining features in the prediction of sepsis. We split the dataset (of 40,336 - 5,500 = 34,836 patients) into training and test sets in proportion 70:30. Using the training data, we perform a 5-fold cross-validation. The Gini importance was used to calculate feature importance from the RF [10].

2.3. Auto-encoding neural networks

We train a dense feed-forward neural network autoencoder using only the patients who do not have sepsis. During training, we minimize the L_2 reconstruction error (RE), which is the mean squared distance between input and output : $L(X, X') = \|X - X'\|_2^2$. We use the selected features from the RF as well as the laboratory variables as inputs, and the neural network architecture used,

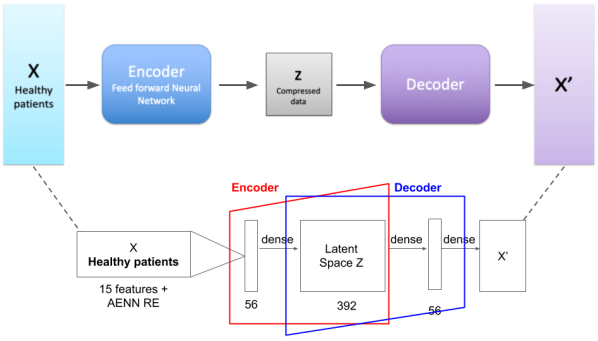


Figure 2: Auto-encoder Neural Network architecture

composed of only dense layers (7/56/392) is described in Fig. 2. We use the auto-encoder by passing patient data through it and evaluating the reconstruction error for each row measurement. This will be an additional feature for the classification.

2.4. Classification and Cross-Validation

Six standard classifiers were used in this study (see Table 1) with and without the RE's feature. Five-fold cross validation was used to evaluate each classifier, performance evaluated with the expected utility (the Computing in Cardiology Challenge defined an asymmetric utility function which heavily penalizes false negative predictions).

Data processing, feature extraction, and classification were all performed using Python 3.6. We used Keras in Tensorflow for the AENN and the library Scikit-Learn for classification. All the implementation is available in : `https://github.com/ineskris/SepsisChallenge-Cinc2019`.

3. Results

3.1. Feature importance

Fig. 3 shows the relative importance of each feature as estimated by the random forests and the Gini importance. As might be expected, the variable *Age* is the most significant in line with previous work [13][14]. Sepsis is equally prevalent in both genders, and so as expected the Gender variable is judged to be of low importance. The variable *O2Sat* will not be use in the AENN.

3.2. Reconstruction error

Fig. 4.a shows the observed distribution of reconstruction errors evaluated using a set of 1,000 patients (from the test set), half of who have sepsis. Observe that the

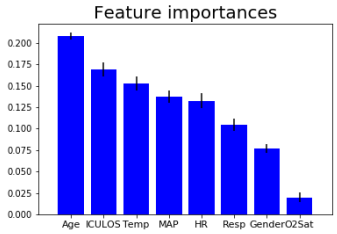


Figure 3: Feature importance according to the random forest.

RE for those with sepsis is higher on average than for those without sepsis. Precision and recall are commonly used to evaluate the accuracy of anomaly detection problem. In Fig. 4.b, the precision/recall shows how the trade off between missing an abnormal patient and the cost of falsely flagging a patient who is healthy according to different threshold for the RE. Based on the utility function provided by the challenge [1], we want to make the maximization of the recall prevail. Taking a $RE = 4000$ makes sense (Fig. 4) for this particular dataset.

3.3. Classification Model

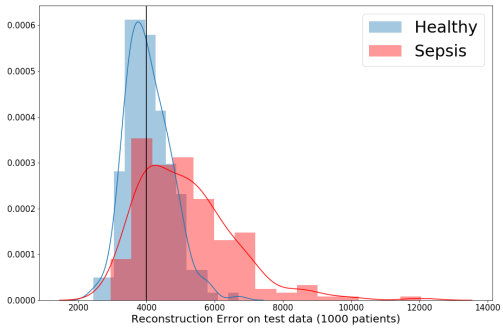
Table 1 highlights the positive impact of the RE on the prediction accuracy for all classifiers. Using the expected utility function from the challenge, logistic regression was found to be the classifier with the highest expected utility (using the *class weight* module to emphasis on the abnormal classes).

Table 1: Machine learning model utility score with or without the RE's feature on the validation dataset

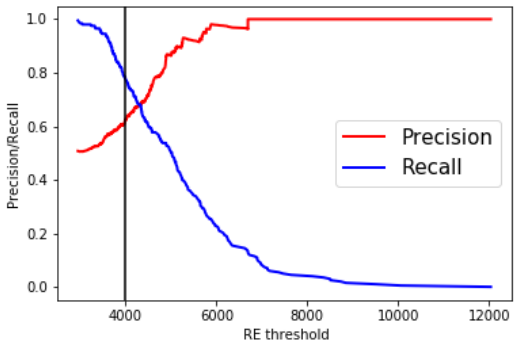
Classifier	Without RE	With RE
Logistic Regression	0.61	0.67
AdaBoost	0.56	0.62
Gradient Boosting	0.56	0.62
RandomForest	0.57	0.61
Decision Tree	0.54	0.56
KNeighbors	0.55	0.56

4. Conclusion

Anomaly detection methods enabled us to identify from a patient's data unusual patterns which do not conform to expected behaviour. In this paper, we used features extracted from the RF and a new feature, the RE obtained from the AENN, to predict sepsis with a classical logistic regression model. The method was validated on PhysioNet Challenge 2019 dataset and the results are encour-



(a) Distribution of the RE for the test set using the AENN



(b) Precision/Recall for different RE threshold values

Figure 4: Reconstruction error (RE) for the Auto encoder Neural Network (AENN) as a significant variable

aging, suggesting that the early prediction of sepsis is an achievable task (Utility Score = 0.177, Team *Kriss*). Ideas develop in this paper were used for the Hackathon (Team *Sepsyd*) and the utility score obtained was 0.329.

References

[1] Reyna M, Josef C, Jeter R, Shashikumar SP, Westover M, Nemati S, Clifford GD, Sharma A. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. Crit Care Med 2019 ;In press.

[2] Herrán-Monge R, Muriel-Bombín A, García-García MM, Merino-García PA, Martínez-Barrios M, Andaluz D, Ballesteros JC, Domínguez-Berrot AM, Moradillo-Gonzalez S, Macías S, Álvarez-Martínez B, Fernández-Calavia MJ, Tarancón C, Villar J, Blanco J. Epidemiology and changes in mortality of sepsis after the implementation of surviving sepsis campaign guidelines. J Intensive Care Med September 2019;34(9):740–750.

[3] Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res November 2017;4(1):e000234.

[4] Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med April 2018;46(4):547–553.

[5] Shamout FE, Zhu T, Sharma P, Watkinson PJ, Clifton DA. Deep interpretable early warning system for the detection of clinical deterioration. IEEE J Biomed Health Inform September 2019;.

[6] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA February 2016;315(8):801–810.

[7] Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, Gurka D, Kumar A, Cheang M. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med June 2006;34(6):1589–1596.

[8] Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the united States-An analysis based on timing of diagnosis and severity level. Crit Care Med December 2018;46(12):1889–1897.

[9] Breiman L. Random forests. Mach Learn October 2001; 45(1):5–32.

[10] Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics July 2009;10:213.

[11] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature October 1986;323(6088):533–536.

[12] Borghesi A, Bartolini A, Lombardi M, Milano M, Benini L. Anomaly detection using autoencoders in high performance computing systems November 2018;.

[13] Kotsanas D, Al-Souffi MH, Waxman BP, King RWF, Polkinghorne KR, Woolley IJ. Adherence to guidelines for prevention of postsplenectomy sepsis. age and sex are risk factors: a five-year retrospective review. ANZ J Surg July 2006;76(7):542–547.

[14] Martin GS, Mannino DM, Moss M. The effect of age on the development and outcome of adult sepsis. Crit Care Med January 2006;34(1):15–21.

Address for correspondence:

Ines Krissaane
SoMas, The University of Sheffield, UK
ikrissaane1@sheffield.ac.uk