A study of early sepsis detection models based on multivariate medical time series

Aren Maes[0000−0003−1637−953X], Tom Van Steenkiste[0000−0002−3842−3151], Tom Dhaene[0000−0003−2899−4636], and Dirk Deschrijver[0000−0001−6600−1792]

Ghent University - imec, IDLab, Technologiepark-Zwijnaarde 126, 9052 Ghent, Belgium aren.maes@ugent.be - www.ugent.be/ea/idlab/en

Abstract. Sepsis is a life-threatening complication caused by the body's response to an infection. For that reason, it is important to have an accurate method to detect sepsis as early as possible. The features extracted from the used ICU data have missing values and non-uniform sampling frequencies, hence an advanced GP based interpolation method is proposed that increases the performance of the models. Additionally, this thesis abstract develops and compares different sepsis detection models based on real medical data [1]. The results show that accurate models can be developed to predict the occurrence of sepsis during an ICU stay.

Keywords: Sepsis · Early detection model · Medical time series · GP.

1 Introduction

Worldwide, more than 30 million people are affected by sepsis each year, of which 6 million people die as the mortality rate lies between 17% and 26% [2]. Hence, it is important to detect sepsis as early as possible. Current tests to diagnose sepsis are time consuming and often inaccurate. Therefore, an automatic detection system could be beneficial. In this work, electronic measurements of the patient's current condition are used to develop a machine learning model that can predict sepsis, allowing doctors to start a treatment as early as possible.

2 Prediction models

This work makes use of the MIMIC-III database from which 48063 samples with each 34 features are extracted. A 70-10-20 % ratio was used for the train, validation and test sets. The features are based on existing sepsis scores and use a combination of vital signs and laboratory results. To address missing values and different sampling frequencies of these features, a Gaussian Process based interpolation technique is introduced that fits the data. The interpolated points consist of a value and its uncertainty which is used to improve the models. Experimental results confirm the superiority of this technique to linear interpolation (see Figure 1).

Secondly, the performance of different models are compared to each other. The first three models (L1D, L3D and L5D) consist of one up to five LSTM layers, followed by a fully connected layer as the final layer of the model. The next three models (CL1D, CL3D and C2L3D) are similar to the previous ones except that they are now preceded by one or more 1D convolutional layers. These combined architectures take advantage of the automatic feature extraction property of CNNs and improve the model results.

3 Results

Figure 2 shows the performance of these models using the area under ROC metric. Comparing the L1D, L3D and L5D models to each other, a small improvement can be noted. The models with the extra convolutional layers perform better than without these extra layers but exceptions exists.

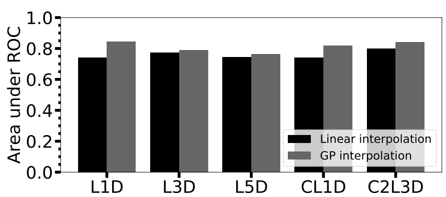

Fig. 1. Area under ROC performance of linear and GP interpolation.

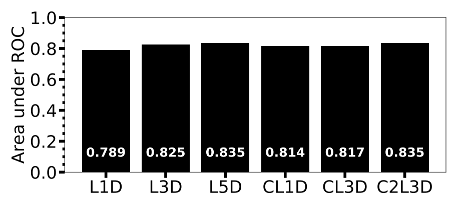

Fig. 2. Area under ROC performance of different models, tested on MIMIC-III.

4 Conclusion

In this work, an advanced interpolation method based on Gaussian processes was tested and compared to linear interpolation. The experiments showed that the GP interpolation method is superior to such a simpler method. Secondly, a combined architecture, called convolutional recurrent neural network, was developed. Experimental results showed that this architecture performs better in general than a regular RNN or CNN. Such architecture, with additional optimisations like hyperparameter tuning and dropout, can obtain an area under ROC of 0,918.

References

1. A. Maes, "Human health monitoring using machine learning and data analysis," Master's thesis, Ghent University, 2019. Via `https://lib.ugent.be/en/catalog/rug01:002785831?i=0&q=Human+health+monitoring+using+machine+learning+and+data+analysis`.
2. C. Fleischmann, A. Scherag, N. K. Adhikari, C. S. Hartog, T. Tsaganos, P. Schlattmann, D. C. Angus, and K. Reinhart, "Assessment of global incidence and mortality of hospital-treated sepsis. current estimates and limitations," *American journal of respiratory and critical care medicine*, vol. 193, no. 3, pp. 259–272, 2016.