

# A Recurrent Neural Network for the Prediction of Vital Sign Evolution and Sepsis in ICU

Benjamin Roussel<sup>1</sup>, Joachim Behar<sup>2</sup>, Julien Oster<sup>1</sup>

<sup>1</sup> IADI, U1254, INSERM, Université de Lorraine, Nancy, France

<sup>2</sup> Faculty of Biomedical Engineering, Technion, Israel Institute of Technology, Haifa, Israel

## Abstract

*Sepsis is a life-threatening reaction to an infection, responsible for 6 million deaths globally each year. Moreover, this condition is one of the major cost to healthcare. Our aim is to develop a new technique for the early detection of the sepsis onset. Such an early detection would allow for the improvement of sepsis outcome.*

*Our technique is based on the assumption that accurate and early prediction of sepsis requires to be able to predict the evolution of the vital signs. This idea was translated in the use of a recurrent neural network, a Long Short-term memory (LSTM) network, which was trained to accomplish two tasks: the prediction of (i) sepsis and (ii) the vital signs at time  $t+6$ . We assume that the use of this auxiliary task allows for a better training of the network given the low prevalence of sepsis. The network consists in three modules: (i) an embedding module aiming at providing a compact representation of the inputs, (ii) a recurrent module with three LSTMs layers with highway connection between each layer (iii) the prediction modules consisting in linear layers for the prediction of two tasks.*

*The network achieved a final utility score of 0.309 on the full hidden test set (0.387 on the test set A, 0.365 on the set B, and -0.148 on the set C). The team name was "IADI".*

*Further improvements are required before transferring such an approach into clinical practice.*

## 1. Introduction

Sepsis is a life-threatening reaction to an infection [1], and is one of the leading causes of mortality and morbidity in Intensive Care Units (ICU) [2, 3]. Moreover, this condition is one of the major cost to healthcare [4]. It has recently been shown that the outcome of sepsis is highly dependent on the timing of the antibiotics treatment [5]. There is therefore a need to develop a technique that is able to detect the apparition of Sepsis as early as possible in order to speed up the antibiotics treatment and improve the outcome of this infection. Unfortunately, there exists

actually no technique that have been clinically validated to predict the apparition of sepsis in ICU patients. Artificial Intelligence has been touted as being the key for solving the problem of Sepsis prediction in patients, who are continuously monitored and for whom huge amount of data (and therefore information) remained (mostly) untapped.

The 2019 Physionet/Computing in Cardiology challenge aims at developing new techniques for the early detection of the sepsis onset [6]. Such early detections would allow for the improvement of sepsis outcome.

## 2. Methods

### 2.1. Challenge data

The training set consists in the collection of data from 40,336 individual patients collected in three different hospital systems. Outcome for these patients were provided on an hourly basis, and was computed according to the Sepsis-3 guidelines, i.e. a two point change in the Sequential Organ Failure Assessment (SOFA) score and clinical suspicion of infection[1].

For each patient, 40 clinical features are provided as a sequence of measurements with a new sample every hour. The data were divided into 3 major co-variables defined as follow: 8 vital signs (Heart Rate, Blood Pressure,...), 26 laboratory values, and 6 demographics (Age, Gender, ICU length of stay,...). Laboratory values are highly sparse values, as such co-variables are rarely measured.

### 2.2. Data preprocessing

First, the input values were linearly normalized between 0 and 1 using extrema values in the training set. Missing values (NaNs) were replaced by -1, in order to distinguish missing values from an extremum value.

We decided to apply recurrent neural networks (RNN) in order to encode the evolution of the patient health. Even though using RNN, the complexity of the time dependencies involved in the sepsis prediction process might be

complicated to catch. We therefore decided to add a 7-window long history of vital data to the input. At each time step, the network is not only informed of the value monitored at time  $t$  but is also reminded of the past 7 values of the considered feature. By doing so, we expect the temporal dependencies to be treated not only by the RNN but also by the linear layers. Furthermore, the RNN is also expected to capture more complex time dependencies.

Given the sparsity of laboratory values, a binary mask of presence for these features is added to the input. The mask contains 0 if a feature is missing and 1 if the value is monitored. We believe that the presence and presence rate of those features play a significant role in determining whether a patient is susceptible or not of developing a sepsis, since the measurement of a given lab value is highly informative of the patient health deterioration.

The different features are then concatenated into a 122-component vectors which is inputted in our model.

Finally, the labels were extended so to predict the onset of sepsis with different delays ranging from 1 to 6 hours. The labels were shifted by 6 hours and, for each time-step, a vector composed of the next 6 labels was used as the target instead of only considering the  $t+6$  label of interest. A flowchart of the label extension is given in figure 1. We assumed it would be easier to predict the onset of sepsis with a smaller delay, since the deterioration of the patient health is more pronounced. We also assumed that training the network on short delays will facilitate the 6-step prediction. Shorter and longer time horizon were also tried but none showed significant performances improvements.

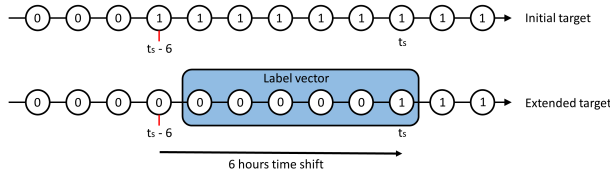


Figure 1: Flowchart explaining the label extension for prediction of sepsis from time  $t+1$  to  $t+6$ .

### 2.3. Model architecture

The proposed model is inspired from a model recently proposed for the detection of acute kidney injury[7].

The proposed neural network can be divided into 3 sub-modules (see figure 2): (i) an embedding module, composed of several linear layers followed by RELU activation functions, whose role is to find a compact and lower-dimensional representation of the input. (ii) This compact representation is then fed into a recurrent module containing 3 LSTM layers with skip connections between each of them, and a post-processing linear layer. This module is supposed to model the historical dependencies of

time-series. (iii) The output of the recurrent module is then fed to two prediction modules. (a) A label prediction module aims at predicting the next 6 hours probabilities of developing a sepsis through linear layers. The output of this module go through a cumulative distribution layer to assure monotony and is then compared to the ground truth label vector using the weighted binary cross-entropy loss function (loss  $l1$ ). The weights are chosen to compensate for the class imbalance between the number of positive and null labels. (b) An auxiliary task consisting in the numerical-vital-features prediction module aims at predicting the vital data 6 hours ahead. We hypothesised this multi-task approach results in better generalization and more-robust representations, especially under class imbalance. The output of the vital features prediction module are compared with the values at time  $t+6$  when provided. The L2 loss is used (loss  $l2$ ). The final loss is the weighted sum of the two former losses using a parameter  $\alpha$  such as:

$$Loss = \alpha \times l1 + (1 - \alpha) \times l2$$

Once the network trained, only the last component of the label-prediction output is kept (corresponding at a  $t+6$  prediction). A softmax is applied to get the probability of the onset of sepsis at time  $t+1$ .

The training set was divided into 4 random sets, each containing the same amount of patients developing a sepsis. 3 sets are used as training sets and the other as validation set. The validation sets were then used in order to calibrate the output and to determin the probability threshold leading to the highest Utility score. This process yield to a threshold of 0.57: if the resulting probability is greater than the threshold, then the patient is considered at risk and the associated output label will be 1.

Finally, the 4 trained networks were applied as an ensemble neural network, the probability outputs were averaged before applying the threshold decision.

### 2.4. Training

To asses the performance of our model, the whole 4-fold cross validation process was repeated 10 times by changing the random splitting. The average and standard deviation were computed for the Area under the receiver operating curve (AUROC), the area under the precision recall curve (AUPRC), accuracy, F1 score, and finally the Utility score developed for this challenge. To demonstrate the utility of the auxiliary prediction module, two models were evaluated and compared : the first without auxiliary task ( $\alpha = 1$ ) and the second with the auxiliary task ( $\alpha = 0.75$ ).

The network was implemented with the Pytorch library and trained using the Adam optimizer. The learning rate was set to 0.0001, with a batch size of 256. A weight decay of  $10^{-5}$  was used as L2 penalty. A grid search was

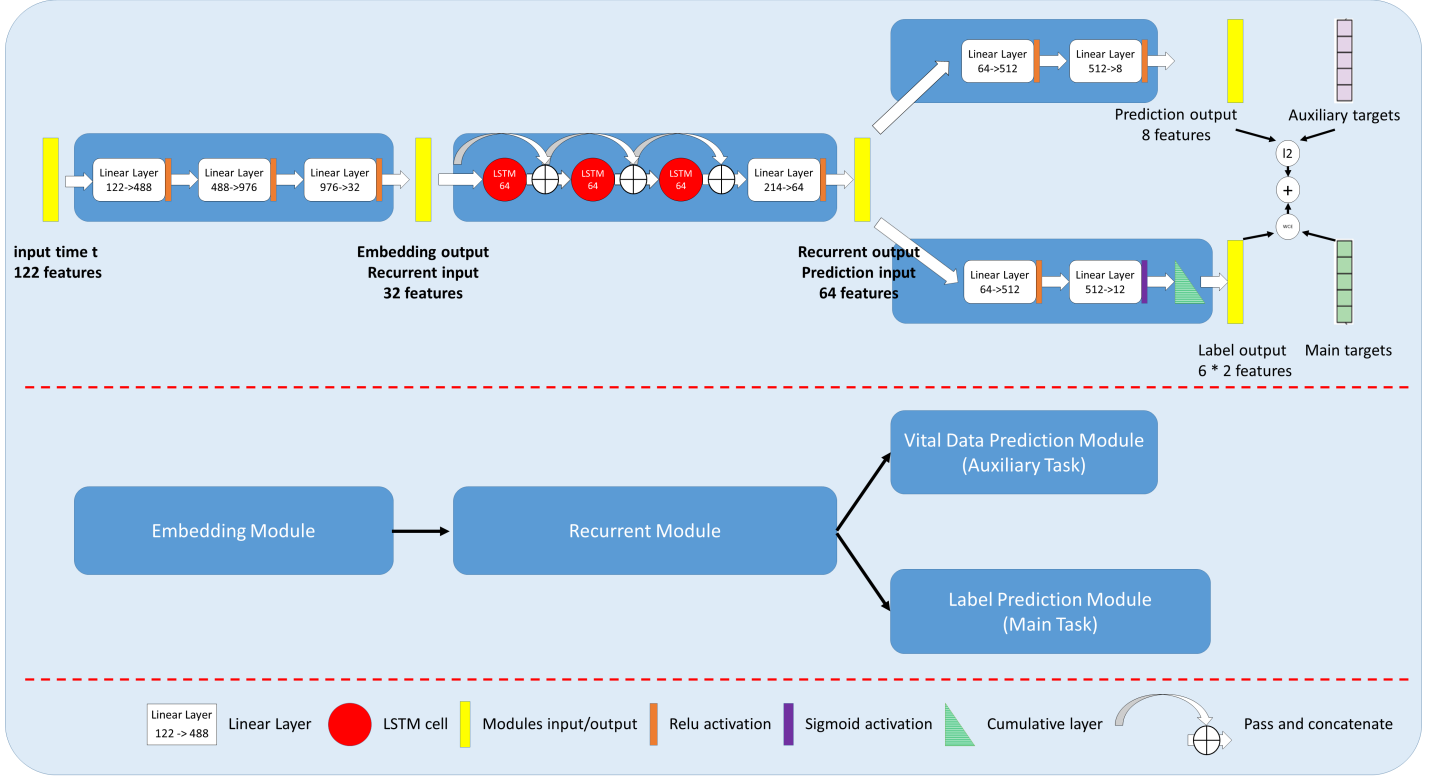


Figure 2: Architecture of the proposed network.

performed to find the hyper-parameters presented here. Dropouts and batch normalization were also considered but yielded no improvements.

### 3. Results

Table 1 assembles the experimental results of the repeated cross-fold validation. The mean Utility scores achieved during the cross-fold validation by the two configurations are respectively  $0.386 \pm 0.005$  for the network without auxiliary task, and  $0.393 \pm 0.005$  for the network with the auxiliary task.

The submitted model (network with the auxiliary task) obtained a utility score of 0.309 on the full hidden test set (0.387 on the test set A, 0.365 on test set B and -0.148 on the test set B). The result on the test set A is comparable to the result obtained during the repeated cross-fold validation, but not for the results on test set C.

### 4. Discussion

The benefit of introducing an auxiliary task to guide the training of the RNN towards the right direction can be discussed. Even though a higher Utility score is reached through the use of the auxiliary task, the other four metrics

(AUROC, AUPRC, Accuracy and F1 score) seemed to be suffering from adding this extra task. The Utility score being the one to maximize in the scope of this challenge, the auxiliary task was naturally added to our submission.

The Utility score introduced for this challenge is quite difficult to interpret in a clinical context. AUPRC and F1 are more easily interpretable metrics. The Utility score is rewarding TP more than FP, which explains the Utility score to be around 0.4 with the F1 score being only around 0.1.

With a mean Utility score of 0.393 on the publicly available dataset, but also a low F1 score (0.12) and AUPRC (0.09), the proposed architecture does not seem to be adequate for its application in clinical routine. The network achieves a Sensitivity of 66%, meaning that one third of sepsis events are currently missed, but is also associated with high level of false positives, which would only deepen the problem of false alarms in ICU.

One interesting avenue of research would consist in the use of high time resolution Heart Rate or Blood Pressure or even continuously measured Electrocardiogram (ECG) or Photoplethysmogram (PPG) data [8]. These additional features could provide the information needed to make a step further towards a clinically usable early-sepsis-predictor model.

	AUROC	AUPRC	Accuracy	F1	Utility
$\alpha = 1$	$0.798 \pm 0.005$	$0.095 \pm 0.002$	$0.854 \pm 0.009$	$0.123 \pm 0.004$	$0.386 \pm 0.005$
$\alpha = 0.75$	$0.793 \pm 0.005$	$0.088 \pm 0.003$	$0.836 \pm 0.012$	$0.117 \pm 0.005$	$0.393 \pm 0.005$

Table 1: Results of the repeated cross-fold validation.

Another important problematic with ICU data, and especially laboratory values comes from their sparsity and the fact that very few measurements are performed. The fact that a measurement is performed is also informative of the patient health. Several variants of recurrent neural networks have recently been proposed to overcome this problem, and should be considered for the detection of Sepsis [9, 10].

## 5. Conclusion

We introduced a new Recurrent Neural Network model for the prediction of Sepsis in ICU patients. The model was trained for the detection but also with an auxiliary task consisting in the prediction of the vital sign evolution. The obtained results do not allow to envision its application in clinical practice with a low sensitivity and high false positive rate. A model, which better incorporates the sparse laboratory values, should yield improved performance.

## Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan XP GPU used for this research.

The authors would also like to acknowledge the Region Grand Est and the Doctoral School "IAEM" from the University of Lorraine for funding Benjamin Roussel's PhD.

## References

- [1] 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.
- [2] Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, Kumar A, Sevransky JE, Sprung CL, Nunnally ME, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive care medicine* 2017;43(3):304–377.
- [3] Soong J, Soni N. Sepsis: recognition and treatment. *Clinical medicine* 2012;12(3):276–280.
- [4] 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. *Critical care medicine* 2018;46(12):1889.
- [5] Frost R, Newsham H, Parmar S, Gonzalez-Ruiz A. Impact

of delayed antimicrobial therapy in septic ICU patients. *Critical Care* 2010;14(2):P20.

- [6] Reyna M, Josef C, Jeter R, Shashikumar S, Brandon-Westover M, Nemati S, Clifford G, Sharma A. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *Critical Care Medicine* In Press 2019;.
- [7] Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572(7767):116.
- [8] Shashikumar S, Stanley D, Sadiq I, Li Q, Holder A, Clifford G, Nemati S. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *Journal of electrocardiology* 2017;50(6):739–743.
- [9] Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Scientific reports* 2018;8(1):6085.
- [10] Rubanova Y, Chen RT, Duvenaud D. Latent odes for irregularly-sampled time series. *arXiv preprint arXiv:1907.03907* 2019;.

Address for correspondence:

Julien Oster  
Inserm, U1254 IADI, Nancy, France  
julien.oster@inserm.fr