Predicting sepsis with a recurrent neural network using the MIMIC III Database

Matthieu Scherpf[a,*], Felix Gräßer[a], Hagen Malberg[a], Sebastian Zaunseder[b]

[a] *Institute of Biomedical Engineering, TU Dresden, Germany*
[b] *Institute of Information Technology, FH Dortmund, Germany*

Abstract

Objective: Predicting sepsis onset with a recurrent neural network and performance comparison with InSight - a previously proposed algorithm for the prediction of sepsis onset.

Methodology: A retrospective analysis of adult patients admitted to the intensive care unit (from the MIMIC III database) who did not fall under the definition of sepsis at the time of admission. The area under the receiver operating characteristic (AUROC) measures the performance of the prediction task. We examine the sequence length given to the machine learning algorithms for different points in time before sepsis onset concerning the prediction performance. Additionally, the impact of sepsis onset's definition is investigated. We evaluate the model with a relatively large and thus more representative patient population compared to related works in the field.

Results: For a prediction 3 hours prior to sepsis onset, our network achieves an AUROC of 0.81 (95 % CI: 0.78-0.84). The InSight algorithm achieves an AUROC of 0.72 (95 % CI: 0.69-0.75). For a fixed sensitivity of 90 % our network reaches a specificity of 47.0 % (95 % CI: 43.1 %-50.8 %) compared to 31.1 % (95 % CI: 24.8 %-37.5 %) for InSight. In addition, we compare the performance for 6 and 12 hours prediction time for both approaches.

Conclusion: Our findings demonstrate that a recurrent neural network is su-

*Corresponding author
Email address: Matthieu.Scherpf@tu-dresden.de (Matthieu Scherpf)

perior to InSight considering the prediction performance. Most probably, the improvement results from the network's ability of revealing time dependencies. We show that the length of the look back has a significant impact on the performance of the classifier. We also demonstrate that for the correct detection of sepsis onset for a retrospective analysis, further research is necessary.
Keywords: sepsis; disease prediction; machine learning; clinical decision support systems; multivariate time-series data; temporal information extraction; prognostication

1. Introduction

To this day, sepsis exhibits a high mortality [1, 2, 3]. The multifactorial characteristic of the disease makes early diagnosis a challenging task for physicians. Additionally, the definition of sepsis exhibits a low specificity resulting in many patients that are wrongly identified as manifesting sepsis. In 1991, the first definition of sepsis and its different severity levels - severe sepsis and septic shock - was developed [4]. This definition was extended in 2001 to facilitate the bedside diagnosis of sepsis [5]. Finally, in 2016 the whole definition was renewed in order to clarify the state of sepsis and therefore to facilitate earlier recognition of sepsis [6]. Nevertheless, the definition of 2016 is criticized for its potential of leading to higher mortality due to the downgrading of the sepsis definition to infection and severe sepsis to sepsis [7, 8]. In our opinion, this definition could actually lead to delayed identification of sepsis as the definition of the term *sepsis* defines a more critical physiological status than before. With respect to this, we assume predicting sepsis with the first definition to be more challenging than with the latest one. Therefore, we use the first definition within this work, defining sepsis as the presence of the systemic inflammatory response syndrome (SIRS) and an infection at the same time (see figure 1). The database we use for our retrospective analysis, the Medical Information Mart for Intensive Care (MIMIC) III database [9], was recorded between 2001 and 2012.

Sepsis prediction is highly relevant though complicated, mostly due to a

Infection { e.g. bacteremia

+

SIRS is confirmed when at least two of the four symptoms are present {

temperature	<36 °C or >38 °C
heart rate	>90/min
respiratory rate (tachypnea) or $PaCO_2$ (hyperventilation)	>20/min or <32 mmHg
white blood cell count or immature band forms	<4 k/µl or >12 k/µl or >10 %

Figure 1: Definition of sepsis from Bone et al. [4] which is used in this paper

low specificity of usable physiological parameters. Several machine learning approaches were proposed in the literature which use vital signs to identify characteristic patterns leading to sepsis. Once identified, these patterns can then be used to predict sepsis onset. In 2016, Calvert et al. introduced the so called InSight algorithm. InSight uses 9 parameters - 8 routine vital signs and patient age [10]. InSight was evaluated with 1394 extracted admissions from the MIMIC II database [11]. A deep learning approach was applied by Kam et al. [12] for the same data set as proposed by Calvert et al. [10]. Desautels et al. applied a modified version of InSight to the MIMIC III database [13]. In this case, 22853 admissions were extracted from the database. Again, in 2018 Mao et al. introduced a revised version of InSight [16]. Nemati et al. [14] and Shashikumar et al. [15] as well used a machine learning approach to predict sepsis onset using the sepsis definition from 2016. Table 1 summarizes the results from previous works.

Previous works in most cases do not account for temporal developments. In this contribution we propose a recurrent neural network to exploit such information and compare it to the Insight algorithm. InSight is one of the first

Table 1: Summary of the results from related works on the prediction of sepsis onset; for explanation of the prediction time see fig. 2; the sepsis definitions from 1991, 2001 and 2016 are indicated as I, II and III, respectively; the first two definitions are considered to be essentially equivalent (see section 1); abbreviations: Sensitivity (sens.), specificity (spec.)

Author(s)	*Number of subjects (name of database/)*	*Sepsis definition*	*Length of look back in hours*	*Prediction time in hours*	*AUROC*	*Sens.*	*Spec.*
Calvert et al. [10]	1394 (MIMIC II)	I/II	5 hrs.	3 hrs.	0.92	0.90	0.81
Mao et al. [16]	90353 (Dataset from University of San Francisco (UCSF))	I/II	3 hrs.	0 hrs.	0.84	0.80	0.75
Kam et al. [12]	approx. 6362 (MIMIC II)	I/II	5 hrs.	3 hrs.	0.93	0.91	0.94
Desautels et al. [13]	22853 (MIMIC III)	III	2 hrs.	4 hrs.	0.74	0.80	0.54
Nemati et al. [14]	approx. 69000 (Emory Cohort and MIMIC III)	III	6 hrs.	4 hrs.	0.85	0.85	0.67
Shashikumar et al. [15]	242 (Emory affiliated hospital)	III	N.a.	4 hrs.	0.78	0.85	0.55

developed machine learning algorithms for the prediction of sepsis onset, well described and uses the first sepsis definition. We evaluate the machine learning approach using roughly 30000 admissions extracted from the MIMIC III database. Hence, the underlying population is considered to be more representative than the one used by Calvert et al. We measure the performance by the area under the receiver operating characteristic (AUROC) and calculate the specificity and sensitivity for chosen values. Additionally, we investigate the impact of the length of the look back, i.e. the length of the sequence of values used for sepsis onset prediction. We also investigate the implementation of the gold standard proposed by Calvert et al.

2. Methodology

We propose a recurrent neural network based approach for the prediction of sepsis onset. As the goal is to distinguish patients who obtained sepsis at any point in time during their stay in the intensive care unit from those who did not, we defined two classes: the *sepsis-class* and *non-sepsis-class*. The point in time from where sepsis is to be predicted is defined by the difference of *sepsis onset* and the *prediction time* (see fig. 2).

2.1. Gold standard and definition of sepsis onset

We defined the gold standard as proposed by Calvert et al. [10]. A graphical illustration is shown in figure 3. The gold standard consists of two criteria. The first one determines if the patient manifested sepsis. This is identified by the international classification of disease (ICD) codes, delivered by the MIMIC III database. Additionally, a second criterion is necessary to determine the point in time of the sepsis diagnoses, as the aim is to predict the sepsis onset. Hence, in this work we determine the sepsis onset by the related ICD-Codes and the 5-hour-SIRS-interval. We provide a descriptive example in figure 2 where the SIRS can be confirmed for at least 5 hours, as more than 1 parameter is higher than the threshold (see fig. 1). Therefore, sepsis onset occurs at the 119th

Table 2: Admissions according to the admission inclusion chart (see fig.4), the evaluated quantities of interpolations (*Ints.*) for the 5h-SIRS-interval and the prediction time (PT)

Ints.	*Y1*	*X1*	*3hrs. PT*		*6hrs. PT*		*12hrs. PT*	
			Y2	*X2*	*Y2*	*X2*	*Y2*	*X2*
0	32790	2724	31575	1509	31444	1378	31238	1172
1	33143	3077	31498	1432	31356	1290	31151	1085
2	33424	3358	31375	1309	31241	1175	31058	992
3	33708	3642	31293	1227	31136	1070	30946	880
4	33903	3837	31179	1113	31032	966	30831	765
5	33985	3919	31116	1050	30967	901	30754	688

hour. Here, a difficulty can be observed: the sepsis onset depends on the used quantity of interpolations. This fact is not mentioned in previous papers. Here, we investigate the different quantities of accepted interpolations as they have a large impact on the point in time of the detection of sepsis onset (see section 4).

2.2. Data collections and inclusion criteria

The MIMIC III database was recorded between 2001 and 2012, in the Beth Israel Deaconess Medical Center in Boston, Massachusetts. We use the most recent Version (v1.4) for this work. The database contains 58976 admissions of 46520 patients. The criteria shown in figure 4 were applied to filter out patients showing a sufficient minimum amount of collected data. The composition of the final data collections are presented in table 2. As we evaluated 6 different quantities of interpolations (0/1/2/3/4/5) for the detection of sepsis onset and 3 different prediction times (3/6/12 hours) we created 18 different data collections. The associated quantities of admissions are encoded with Y2/X2.

2.3. Extracted parameters from the MIMIC III database and data preprocessing

The parameters that we extracted from the database were chosen based on the paper of Calvert et al. [10] and the SIRS parameters (see fig. 1). For

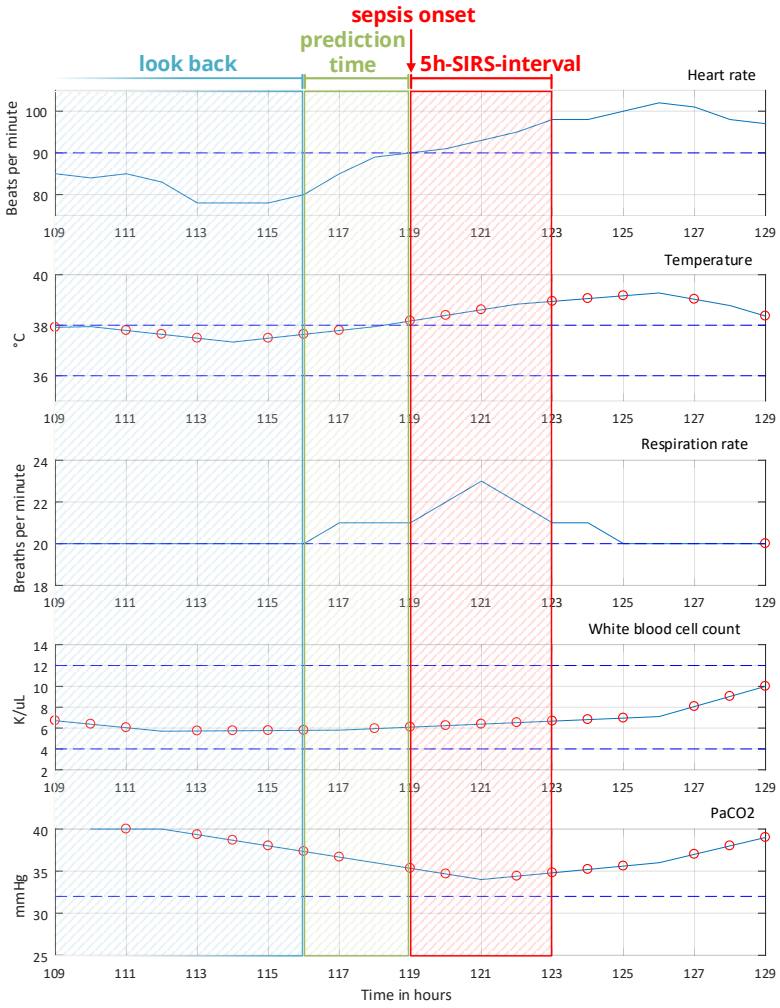

Figure 2: Patient with sepsis onset at the 119th hour of his/her ICU stay; The look back is the sequence of values that is used to predict if there will occur a sepsis onset or not - hence, if a specific look back is classified as belonging to the *sepsis-class* or *non-sepsis-class*; we used 5/10/15/20 hours of look back; the prediction time represents the duration between sepsis onset and the latest values of the look back - here 3 hours; we evaluated 3/6/12 hours for the prediction time

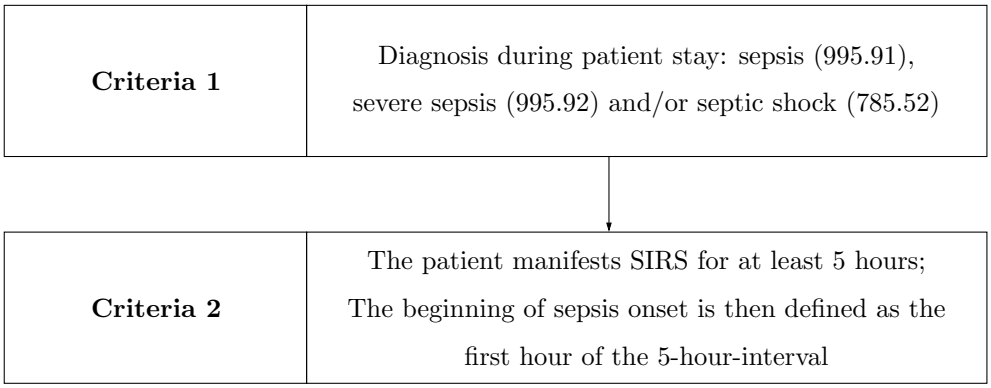

Figure 3: Gold standard used by Calvert et al. [10]; In brackets: ICD-9-Codes

the purpose of reproducibility, we provide the *ITEMIDs*, which indicate the underlying measurement from the MIMIC III database in table 3.

We calculated the mean value for every one-hour-interval for each extracted parameter. We use two different strategies for the imputation of missing values. For the implementation of the gold standard linear interpolation and "carry forward/backward" extrapolation is used - extrapolating the last or first available value forward or backwards, respectively. For the classification task "carry forward" interpolation and no extrapolation was performed.

2.4. Evaluation strategy

We used 4-fold-stratified-cross-validation to evaluate the implemented method. The *stratified* validation method was implemented to take the different class proportions into account. For the training procedure of the RNN, the training data was additionally split into training and validation data resulting in $\frac{9}{16}$ training data, $\frac{3}{16}$ validation data and $\frac{1}{4}$ test data for each of the four cross-validation runs.

As we evaluate the quantity of interpolations for the detection of sepsis onset (0/1/2/3/4/5), the prediction time (3/6/12 hours) and the length of the look back (5/10/15/20 hours) 72 cycles of 4-fold-cross-validation were performed.

The extraction of the look back for the *sepsis-class* is shown in fig. 2. For the *non-sepsis-class* a sequence with the length according to the look back is

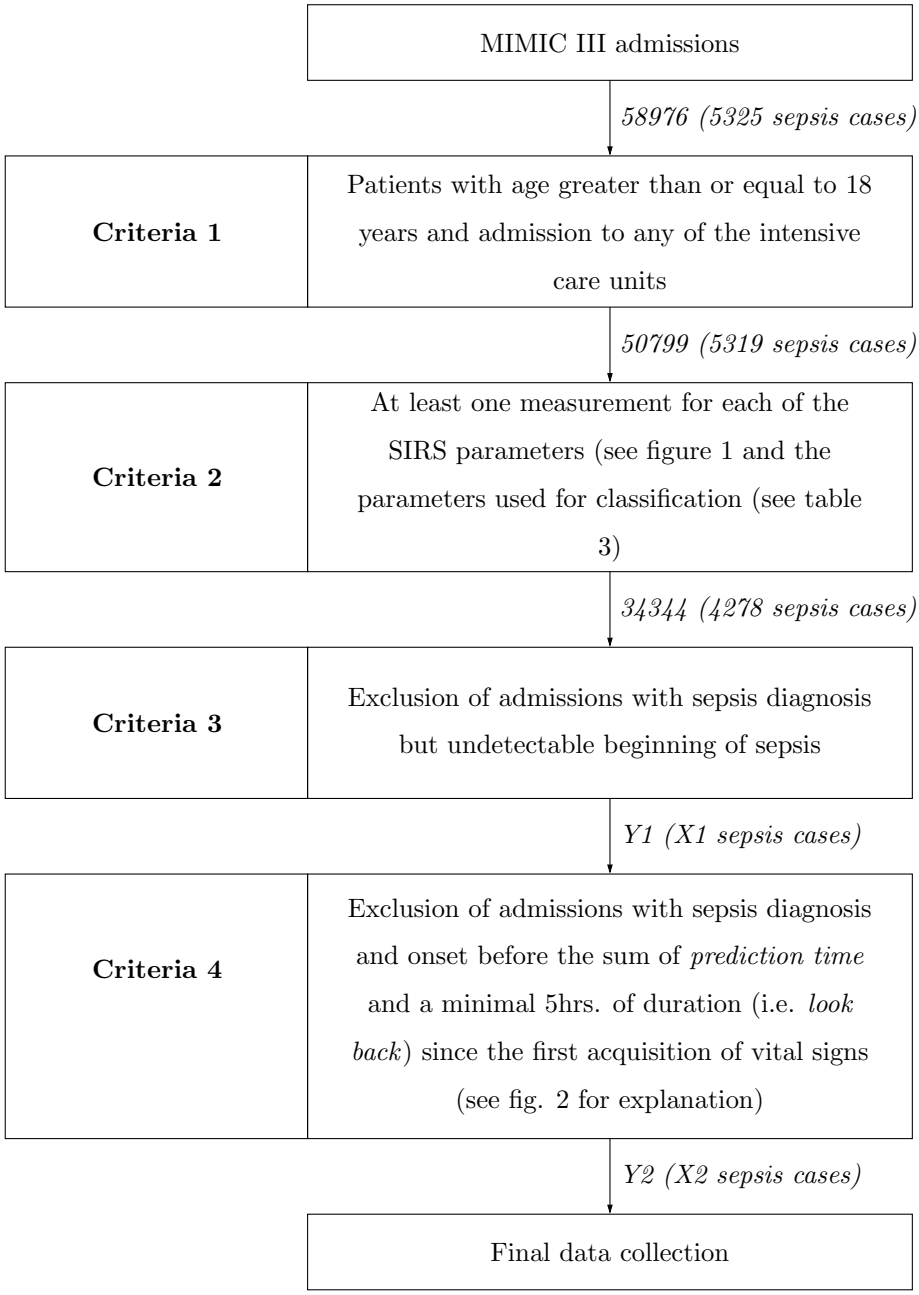

Figure 4: Admission inclusion chart; for the explanation of Y1/X1 and Y2/X2 see sec. 2.2; for the concrete quantities of Y1/X1 and Y2/X2 see tab. 2

Table 3: Extracted parameters from the MIMIC III database. The parameters marked in green represent SIRS parameters and those marked in blue represent parameters used for prediction. The ITEMID represents the identification number of a measurement in the MIMIC III database

Parameters	*ITEMIDs*		
patient age	-		
systolic blood pressure	220050, 225309, 220179, 51, 455, 6701		
diastolic blood pressure	220051, 220180, 225310, 8368, 8441, 8555		
pH value	50820		
blood oxygen saturation (SO_2)	220227, 220277, 834, 646		
temperature	223762, 223761, 676, 678		
heart rate	220045, 211		
respiratory rate	220210, 224422, 224689, 224690, 618, 651, 615, 614		
CO_2 partial pressure ($PaCO_2$)	220235, 778		
white blood cell count	51301		

randomly picked from the whole duration of the admission. We do this random selection once for each *non-sepsis-class* admission.

2.5. Implementation of the InSight algorithm

To rank the results of the developed classifier in this paper, we implemented the InSight algorithm introduced by Calvert et al. [10]. This algorithm extracts 101 features from the look back.

For the extracted look back of each admission, the mean (M_i) and the difference (D_i), between the first and the last value of each parameter of the look back, is extracted. The features then consist of the conditional probabilities $P_{M_i}(s = 1|M_i)$, $P_{D_i}(s = 1|D_i)$, $P_{D_{ij}}(s = 1|D_i, D_j)$ and $P_{D_{ijk}}(s = 1|D_i, D_j, D_k)$, whereby the indices i, j, k represent the parameters and $s = 1$ expresses the fact that the observation leads to sepsis. Hence, the features are probabilities to suffer from sepsis according to the found value of a parameter or parameter combination. P_{M_i} is calculated for all the 9 parameters (see tab. 3), whereas P_{D_i}, $P_{D_{ij}}$ and $P_{D_{ijk}}$ are calculated for all parameters except the patient age. $P_{D_{ij}}$ and $P_{D_{ijk}}$ take combinations of different parameters, and therefore correlations between each of them, into account. Thus, $9P_{M_i}$ values, $8P_{D_i}$ values, $28P_{D_{ij}}$ values and $56P_{D_{ijk}}$ values exist.

For each look back, a score is calculated by

$$Score = a\sum_{i\in A} P_{M_i} + b\sum_{i\in B} P_{D_i} + c\sum_{(i,j)\in C} P_{D_{ij}} + d\sum_{(i,j,k)\in D} P_{D_{ijk}}. \qquad (1)$$

$A - D$ allow the sums to be compactly written by representing the several sets of features. The variables $a - d$ are used as calibration constants in terms of the maximization of the area under the receiver operating characteristic (AUROC) for the training set.

To allow for the implementation of the InSight algorithm despite the occurrence of missing values, they were replaced by the mean of the look back of the corresponding parameter.

2.6. Implementation of the recurrent neural network

In this paper we propose a recurrent neural network (RNN) for the prediction of sepsis onset. The aim is to better exploit time-dependent patterns within the data that finally are followed by sepsis onset, and thus to show that the implementation of the neural network as a support system for the clinicians is a promising approach. Although the network exhibits a black box character, its implementation still remains reasonable as it serves as a support system and the final decision is made by the attending physician. Consequently, such a system is alerting the physician if there is a remarkable deterioration of the patient's health condition that indicates a beginning of sepsis.

The RNN consists of 2 hidden layers with 40 neurons each. We use a gated recurrent unit (GRU) [18] as the hidden layer architecture. The network is optimized on binary cross-entropy cost function, which represents a standard approach for dichotomous classification tasks. As optimizer we use the so called Adam algorithm, which typically yields solid results [19, 20].

The features that we use for the RNN consist of the normalized parameter values of each hour of the look back. The normalization is necessary, as the activations of the neurons in the network should lie between 0 and 1 [21], whereby an activation of 0 implies that there is no information.

As previously mentioned, the data sets contain missing parameter values. For the InSight algorithm, these were replaced by the mean of each parameter. This is not necessary for the RNN, as the input data is normalized and missing data can be padded by 0.

3. Results

Figure 5 shows the results for the AUROC with 95 % confidence intervals (CI) when the detection of sepsis onset is based on no accepted interpolations. The RNN shows an overall higher performance than the InSight algorithm with a maximum AUROC of 0.81 (95 %: 0.79-0.83) and 0.72 (95 %: 0.69-0.75), respectively. In contrast to the InSight algorithm, the RNN benefits from the

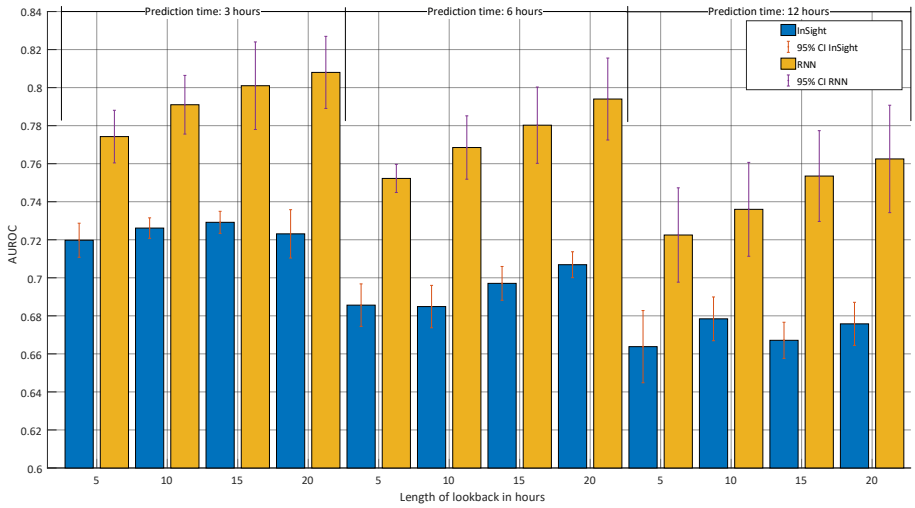

Figure 5: AUROC for 5/10/15/20 hours of look back and 3/6/12 hours prediction time and *0 accepted interpolations* for the 5h-SIRS-interval

elongated look back. The performance decreases with increasing prediction time for both methods.

We also computed the results for 1 to 5 accepted interpolations. In figure 6 we provide the results for 5 accepted interpolations for the 5h-SIRS-interval as the discrepancy between 0 and 5 accepted interpolations is the most significant. For this setup, the InSight algorithm also benefits from the elongated look back whereas it must be said that the confidence intervals of the InSight results have enlarged. In addition, the increase of the AUROC for the RNN, induced by the elongation of the look back, is greater than for the previous case with no accepted interpolations.

In table 4 the results for a look back of 20 hours are summarized. We also indicate the specificity for a fixed 90 % sensitivity for a better assessment of the results. The specificities of the RNN are significantly higher than for the InSight algorithm. This finding is consistent with the computed AUROCs as the AUROC relates to sensitivity and specificity.

Table 4: Results for 20 hours of look back for the RNN and InSight; for the calculation of the specificity the sensitivity was fixed to 90 %; 95 % confidence interval (CI) with lower and upper bound in brackets; prediction time (PT) is indicated in hours

Computed AUROCs

PT	*RNN (CI)*	*InSight (CI)*
0 accepted interpolations		
3	0.81 (0.79, 0.83)	0.72 (0.71, 0.74)
6	0.79 (0.77, 0.82)	0.71 (0.70, 0.71)
12	0.76 (0.73, 0.79)	0.68 (0.66, 0.69)
5 accepted interpolations		
3	0.81 (0.78, 0.84)	0.72 (0.69, 0.75)
6	0.80 (0.79, 0.83)	0.72 (0.67, 0.76)
12	0.79 (0.76, 0.82)	0.71 (0.64, 0.77)

Computed Specificities in %

PT	*RNN (CI)*	*InSight (CI)*
0 accepted interpolations		
3	46.9 (39.9, 53.9)	31.4 (30.4, 32.4)
6	45.3 (38.1, 52.6)	29.1 (26.6, 31.5)
12	38.8 (31.7, 45.8)	23.9 (21.6, 26.2)
5 accepted interpolations		
3	47.0 (43.1, 50.8)	31.1 (24.8, 37.5)
6	44.9 (35.3, 54.6)	32.5 (24.4, 40.6)
12	46.3 (40.5, 52.1)	34.1 (27.4, 40.7)

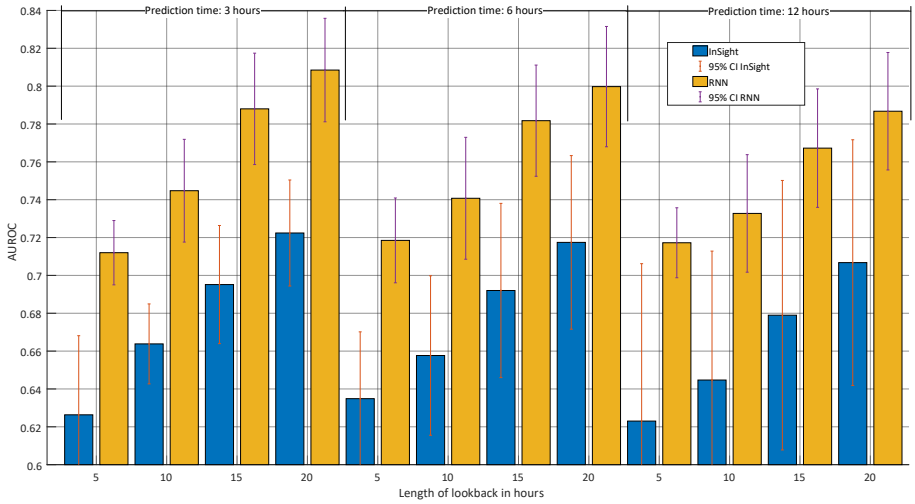

Figure 6: AUROC for 5/10/15/20 hours of look back and 3/6/12 hours prediction time and *5 accepted interpolations* for the 5h-SIRS-interval

4. Discussion

4.1. Results evaluation and comparison

The RNN exhibits a higher AUROC compared to InSight in all computed scenarios. This demonstrates that the application of such networks can improve the prediction of sepsis in intensive care, and thus potentially further reduce sepsis mortality.

We show that the performance of the RNN can be significantly increased for the underlying problem if the sequences of vital signs given to the machine learning algorithm is enlarged. We observed this effect especially for the RNN but also for InSight.

The AUROC, sensitivity and specificity is lower compared to related works using the same sepsis definition [10, 12]. However, the amount of data used to train, validate and test the RNN is greater and therefore we believe our classifier to have an increased generalizability. In comparison with works using a comparable amount of data but the latest sepsis definition, our machine learning approach shows superior performance (see table 1 - [13, 15]). Nemati et al.

[14] achieved better results, although it can be said that they worked with a database twice as large. Concerning Mao et al. [16] we compare to the results for classification of sepsis onset without the features used for the gold standard definition as we assume these results to be more meaningful. Their method also achieves higher performance. However, it must be said that they were classifying sepsis onset which we assume to be an easier task compared to a prediction as in the latter case less information is available.

Our finding of an improved prediction based on a longer look back might strengthen the knowledge about sepsis. It seems that the symptoms and related vital sign patterns of sepsis appear quite early. Machine learning algorithms are capable of detecting such complex interdependencies between different physiological parameters.

Further investigation on the detection of sepsis onset is a necessity. This applies for the definition of sepsis onset used in this work but also to the new definition from 2016. Both are partly based on physiological parameters acquired in the laboratory at time intervals far greater than one hour. It remains unclear how to handle missing values for the hourly detection of sepsis onset. We suggest a linear interpolation and "carry forward/backward" extrapolation whereas it is not exactly clear how to determine the best amount of accepted interpolations. An additional variation of the length of the SIRS-interval could also be investigated.

4.2. Limitations

Apart from the classifier, the data basis itself and the related implementation of the gold standard affect the performance of the implemented machine learning method. Here, we define the gold standard by the two criteria shown in figure 3. Especially the second criterion affects the detection of sepsis onset and therefore the prediction performance. In certain cases, patients might be in a sepsis-like state over longer periods, but "borderline" parameters violate the criterion 2 intermittently. Consequently, the look back itself will hold sepsis-like parameters which will ease the prediction. This problem raises the

question if we detect the sepsis onset correctly. In figure 7 the proportion of sepsis admissions with manifested sepsis onsets depending on the point in time when they occur after admission can be observed. Apparently, the discrepancy between 0 and 5 accepted interpolated SIRS-hours (INT 0/INT 5) for the 5h-SIRS-interval is high. That means when, for example, accepting 0 interpolated SIRS-hours, about 50 % of the sepsis cases manifested sepsis between 0 and 10 hours after admission - in contrast to about 78 % when accepting 5 interpolated SIRS-hours. The question concerning the correct sepsis onset detection is difficult to answer, as several factors are relevant. For example, Werdan et al. [1] state that significant discrepancies between Germany and the USA can be observed when analyzing epidemiological data for sepsis, although the population characteristics should be comparable. Possible reasons could be a different quality of documentation and/or a financial incentive by the cost units. The MIMIC III database is recorded in one hospital in a 11 years period. Therefore, the database itself is possibly biased as it consists of people mostly from a specific country and region. One general advantage of neural networks to mention in this context is the possibility of fine-tuning them to slightly different data belonging to the same problem. This seems to be an appropriate method to counter such a bias but is rarely used in combination with recurrent neural networks so far [21].

Apart from that, the black box character of the RNN can be problematic, but as we intend its implementation strictly as a support or early warning system and not as a decision system this deficit seems reasonable. The classification performance may not be good enough for an implementation in clinical practice yet, because with a specificity of 47 % the classifier would issue a false-alarm for roughly every second patient in the ICU although identifying nearly all sepsis cases correctly. This high rate of false-positive alarms potentially leads to increasing alert fatigue what will negatively influence the acceptance of such a classifier as a support system [22].

As we developed the proposed classifier based on the sepsis definition from [4] future efforts will be made for a comparison with the latest sepsis definition

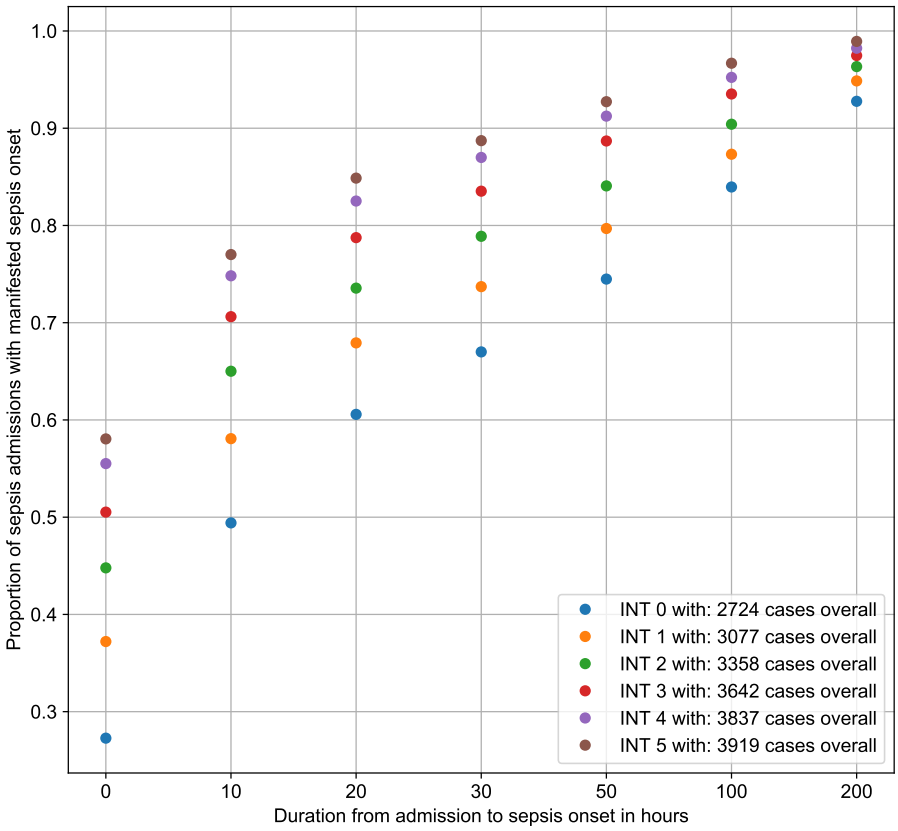

Figure 7: Proportion of patients with detected sepsis onset from all sepsis patients depending on the time from admission when the sepsis onset is detected; 6 variations of the accepted amount of interpolations (defined by INT0...5) for the 5h-SIRS-interval were evaluated; an interpolated SIRS-hour is defined as an hour where the SIRS is manifested but less than 2 criteria are fulfilled by non-interpolated parameter values

from 2016 [6]. Overall, the objective of sepsis prediction will surely benefit from the Physionet challenge in 2019 which addresses sepsis prediction based on the latest sepsis definition [23].

We use the InSight algorithm proposed by Calvert et al. from 2016 [10] and therefore we do not compare our classifier to the most recent version of InSight which is based on gradient-boosted decision trees. We will address this lack of comparability in future works.

As the classifier is specifically trained with mostly dynamic data acquired from the ICU, its implementation in an non-ICU environment remains problematic. Less timely data would be available, potentially having a negative influence on the classification performance and leading to an increased rate of false alarms. Most importantly, the classifier would likely alert later as a re-evaluation for the risk of sepsis can only be done at each time when there is new data presented to the classifier. This is a general problem in the practical implementation of machine learning methods with dynamic data from the electronic health record [22].

5. Conclusion

Within this study we demonstrated that a recurrent neural network can reliably predict sepsis onset. This machine learning approach outperforms the InSight algorithm developed by Calvert et al. [10]. Our findings emphasize the value of temporal information and a gradual development of sepsis. They also show that further research is necessary to determine the correct sepsis onset detection as it varies depending on the amount of accepted interpolations. This does not only count for the definition of Calvert et al. related to the sepsis definition from 2001 [5] but also for the definition of 2016 from Singer et al. [6] as it also relies on laboratory measurements.

Financial support

None received.

Conflict of interest

None declared.

References

[1] K. Werdan, U. Müller-Werdan, H.-P. Schuster, F. M. Brunkhorst (Eds.), Sepsis und MODS, 5th Edition, Springer Science and Business Media, 2016. doi:10.1007/978-3-662-45148-9.

[2] C. Fleischmann, D. O. Thomas-Rueddel, M. Hartmann, C. S. Hartog, T. Welte, S. Heublein, U. Dennler, K. Reinhart, Hospital incidence and mortality rates of sepsis, Deutsches Arzteblatt international 113 (10) (2016) 159–166. doi:10.3238/arztebl.2016.0159.

[3] N. S. Ward, M. M. Levy, Sepsis: Definitions, Pathophysiology and the Challenge of Bedside Management, Respiratory Medicine, Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-48470-9.

[4] R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus, R. M. Schein, W. J. Sibbald, Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis, in: Chest, Vol. 101, 1992, pp. 1644–1655. doi:10.1378/chest.101.6.1644.

[5] M. M. Levy, M. P. Fink, J. C. Marshall, E. Abraham, D. Angus, D. Cook, J. Cohen, S. M. Opal, J.-L. Vincent, G. Ramsay, 2001 sccm/esicm/accp/ats/sis international sepsis definitions conference, Vol. 31, pp. 1250–1256. doi:10.1097/01.CCM.0000050454.01978.3B.

[6] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J.-L. Vincent, D. C. Angus, The third international consensus definitions for sepsis and septic shock (sepsis-3), JAMA 315 (8) (2016) 801–810. doi:10.1001/jama.2016.0287.

[7] S. Q. Simpson, New sepsis criteria: A change we should not make, Chest 149 (5) (2016) 1117–1118. doi:10.1016/j.chest.2016.02.653.

[8] M. Sartelli, Y. Kluger, L. Ansaloni, T. C. Hardcastle, J. Rello, R. R. Watkins, M. Bassetti, E. Giamarellou, F. Coccolini, F. M. Abu-Zidan, A. K. Adesunkanmi, G. Augustin, G. L. Baiocchi, M. Bala, O. Baraket, M. A. Beltran, A. C. Jusoh, Z. Demetrashvili, B. de Simone, H. P. de Souza, Y. Cui, R. J. Davies, S. Dhingra, J. J. Diaz, S. Di Saverio, A. Dogjani, M. M. Elmangory, M. A. Enani, P. Ferrada, G. P. Fraga, S. Frattima, W. Ghnnam, C. A. Gomes, S. S. Kanj, A. Karamarkovic, J. Kenig, F. Khamis, V. Khokha, K. Koike, K. Y. Y. Kok, A. Isik, F. M. Labricciosa, R. Latifi, J. G. Lee, A. Litvin, G. M. Machain, R. Manzano-Nunez, P. Major, S. Marwah, M. McFarlane, Z. A. Memish, C. Mesina, E. E. Moore, F. A. Moore, N. Naidoo, I. Negoi, R. Ofori-Asenso, I. Olaoye, C. A. Ordoñez, M. Ouadii, C. Paolillo, E. Picetti, T. Pintar, A. Ponce-de Leon, G. Pupelis, T. Reis, B. Sakakushev, H. S. Kafil, N. Sato, J. N. Shah, B. Siribumrungwong, P. Talving, C. Tranà, J. Ulrych, K.-C. Yuan, F. Catena, Raising concerns about the sepsis-3 definitions, World journal of emergency surgery : WJES 13 (2018) 6. doi:10.1186/s13017-018-0165-6.

[9] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, R. G. Mark, Mimic-iii, a freely accessible critical care database, Scientific data 3 (2016) 160035. doi:10.1038/sdata.2016.35.

[10] J. S. Calvert, D. A. Price, U. K. Chettipally, C. W. Barton, M. D. Feldman, J. L. Hoffman, M. Jay, R. Das, A computational approach to early sepsis detection, Computers in biology and medicine 74 (2016) 69–73. doi:10.1016/j.compbiomed.2016.05.003.

[11] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, R. G. Mark, Multiparameter intelligent monitoring in intensive care ii: a public-access intensive care unit database, Critical care medicine 39 (5) (2011) 952–960. doi:10.1097/CCM.0b013e31820a92c6.

[12] H. J. Kam, H. Y. Kim, Learning representations for the early detection of sepsis with deep neural networks, Computers in biology and medicine 89 (2017) 248–255. doi:10.1016/j.compbiomed.2017.08.015.

[13] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, D. J. Wales, R. Das, Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach, JMIR medical informatics 4 (3) (2016) e28. doi:10.2196/medinform.5909.

[14] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, T. G. Buchman, An interpretable machine learning model for accurate prediction of sepsis in the icu, Critical care medicine 46 (4) (2018) 547–553. doi:10.1097/CCM.0000000000002936.

[15] S. P. Shashikumar, M. D. Stanley, I. Sadiq, Q. Li, A. Holder, G. D. Clifford, S. Nemati, Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics, Journal of electrocardiology 50 (6) (2017) 739–743. doi:10.1016/j.jelectrocard.2017.08.013.

[16] Q. Mao, M. Jay, J. L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Y. Zhou, R. Das, Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and icu, BMJ open 8 (1) (2018) e017833. doi:10.1136/bmjopen-2017-017833.

[17] J. Calvert, N. Saber, J. Hoffman, R. Das, Machine-learning-based laboratory developed test for the diagnosis of sepsis in high-risk patients, Diagnostics (Basel, Switzerland) 9 (1) (2019). doi:10.3390/diagnostics9010020.

[18] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling.
URL `http://arxiv.org/pdf/1412.3555v1`

[19] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization.
URL http://arxiv.org/pdf/1412.6980v9

[20] S. Ruder, An overview of gradient descent optimization algorithms.
URL http://arxiv.org/pdf/1609.04747v2

[21] F. Chollet, Deep learning with Python, Safari Tech Books Online, Manning, Shelter Island, NY, 2018.
URL http://proquest.safaribooksonline.com/9781617294433

[22] A. M. Harrison, O. Gajic, B. W. Pickering, V. Herasevich, Development and implementation of sepsis alert systems, Clinics in chest medicine 37 (2) (2016) 219–229. doi:10.1016/j.ccm.2016.01.004.

[23] Physionet, The physionet/computing in cardiology challenge 2019: Early prediction of sepsis from clinical data (2019).
URL https://physionet.org/challenge/2019/