R. Brause, E. Hanisch, J. Paetz, B. Arlt: Neuronal networks for sepsis prognosis - the MEDAN project,
Journal für Anästhesie und Intensivbehandlung (11)1-2004, pp.40-43

Neural Networks for Sepsis Prediction - the MEDAN-Project[1]

R. Brause, E. Hanisch, J. Paetz, B. Arlt

1. Introduction

Since the description of sepsis by Schottmüller in 1914, the amount on knowledge available on sepsis and its underlying pathophysiology has substantially increased. Epidemiologic examinations of abdominal septic shock patients show the potential for high risk posed by and the extensive therapy situation in the intensive care unit (ICU) (5). Unfortunately, until now it has not been possible to significantly reduce the mortality rate of septic shock, which is as high as 50-60% worldwide, although PROWESS' results (1) are encouraging. This paper summarizes the main results of the MEDAN project and their medical impacts. Several aspects are already published, see the references.

The heterogeneity of patient groups and the variations in therapy strategies is seen as one of the main problems for sepsis trials. In the MEDAN multi-center study of 71 intensive care units in Germany, a group of 382 patients made up exclusively of abdominal septic shock patients who met the consensus criteria for septic shock (3) was analysed. For use within scores or stand-alone experiments variables are often studied as isolated variables, not as a multidimensional whole, e.g. a recent study takes a look at the role thrombocytes play (15). To avoid this limitation, our study compares several established scores (SOFA, APACHE II, SAPS II, MODS) by a multi-dimensional neuronal network analysis.

For outcome prediction the data of 382 patients was analysed by using most of the commonly documented vital parameters and doses of medicine (metric variables). Data was collected in German hospitals from 1998 to 2001. The 382 handwritten patient records were transferred to an electronic database giving the amount of 2.5 million data entries. The metric data contained in the database is composed of daily measurements and doses of medicine. We used range and plausibility checks to allow no faulty data in the electronic database. 187 of the 382 patients are deceased (49 %).

2. The Neural Network Diagnosis

The MEDAN project followed the paradigm that an automatic, data-driven analysis and prediction should be done. We implemented this demand by using adaptive systems, especially artificial neural networks.

2.1 Networks and Scores

An artificial neural network can be seen as a net of information processing units which abstractions of biological neurons. Each formal neuron uses inputs $x_1,\dots,x_n$ and has an input-output function S_j. All output from the N neurons of one layer is combined to the final output of the network by weighting the different influences by weights w_j.

$$f(x_1,\dots,x_n) = \sum_{j=1}^{N} w_j S_j(x_1,\dots, x_n) \qquad (1)$$

In comparison to this, for computing a score we have first to determine in which interval a measured variable falls, then assign a score value to it and then to add all the values of the different variables together to the final score. We might model this by defining a function $S(x_i)$ to be one within an interval j if the measured value falls within the interval boarders α_j and β_j

$$S_j(x) = \begin{cases} 1 & \alpha_j \le x \le \beta_j \\ 0 & \text{else} \end{cases}$$

and assign as score value the weight w_j to it. As example the SOFA variable x_i ="Billirubin" with its associated score values w_{ji} is shown in Fig. 1.

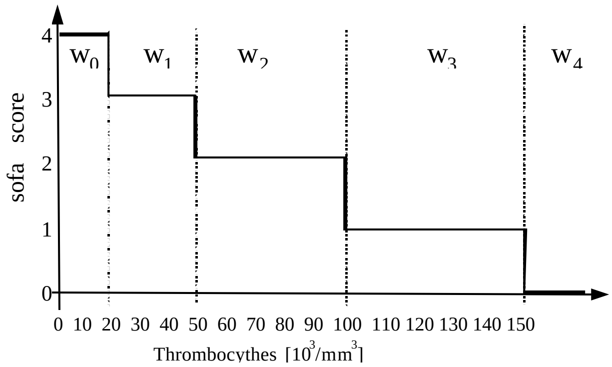

Fig. 1 The output function weights of the Thrombocythes SOFA variable

[1] This work was partially supported by the grant HA 1456/7-2 from the "Deutsche Forschungsgemeinschaft"

Then the final score value is computed as

$$\text{Score} = \sum_j \sum_i w_j S_{ji}(x_i) \quad (2)$$

using the variables x_i and the intervals j defined on each variable. Comparing the two expressions (1) and (2) we notice that they are formally equivalent under the condition that the input-output function $S_j(.)$ of the neural network is identical to the interval-shaping function $S_{ji}(.)$ of the scores. Thus, the resulting "score networks" are special cases of general artificial neural networks. This fact can be formulated also in other ways, see (13).

In difference to scores which are statically defined and do not change, neural network parameters like weights and parameters of S (e.g. location and width of intervals) are supposed to change. Special learning algorithms adapt the network diagnostic performance to become the maximum.

2.2 Results

For the diagnosis we trained a supervised neural network algorithm used in its modified, improved variant (12) – which uses the class information of the data in its adaptation process. Outcome labels are used {survived, deceased} as class information in the training procedure of the neural network. This kind of system adapts a non-linear classification to the data. For implementation details see (4).

The data of the 382 patients were analyzed for different tasks. The most important task was the prediction of death the most in advance as possible. This is discussed in detail in the next section.

Other results were:

- By our neural network analysis, we identified the systolic and diastolic blood pressure/thrombocytes system as the most relevant variables for outcome prediction (11).
- The metric variables hold most of the diagnostic information: After adding qualitative variables like treatment or medication the diagnosis augmented only slightly from AUC=0,90 to 0,92 for the subset of 138 patients.
- The diagnosis was impeded by unimportant variables like life variables (e.g. respiration) or medication (e.g. catecholamine)
- A diagnosis based only on the whole qualitative context like therapies or medication was not possible, because each patient had an unique combination of attribute values: there were no identical cases in the data base. Even the regrouping according to only one attribute did not solve the problem. For instance, the medically important attribute "reoperation" of 282 patients presented in Table 1 can hardly used for representative statistics or reliable prognosis.

Number of ReOp	Number Patients	Number deceased	percent deceased
1	69	32	46.38
2	19	15	78.,95
3	15	9	60.00
4	8	3	37.50
5	6	2	33.,33
6	5	4	80.,00
7	2	2	100.00
9	2	1	50.00
11	1	1	100.00
12	1	1	100.00

Table 1 The reoperations statistics

2.3 The Resulting Alarm System

It turned out that the diagnostic quality heavily depended on the time period analyzed. Using neural network results we have created an alarm system (11) (using 138 patients), here presented using the results of the extended group of 382 patients. An alarm message is given whenever input for the neural network generates high output for class "deceased". In Fig. 2 we see the resulting alarm percentage for the first three days, for the first and second half of ICU stay and for the last three days, indicated separately for patients who either died or survived.

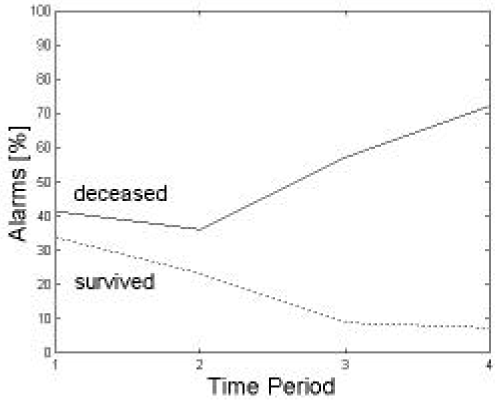

Fig. 2 Alarm rate in percent for 1) the first three days; 2) the first half of ICU stay; 3) the second half of ICU stay; 4) the last three days.

In the time periods 1, 2, 3 and 4 there are 34%, 23%, 9% and 7% alarms for surviving patients, respectively, and 41%, 36%, 57% and 72% alarms for deceased patients, respectively, (i.e. 1.2, 1.5, 6.5 and 9.9 times more alarms for deceased patients, respectively). Only alarms stemming from the last three days can be interpreted as false alarms with respect to outcome prediction; 7% were false alarms.

On the other days one cannot retrospectively examine if the alarms are due to critical or uncritical states. Patients may have a lot of critical or uncritical states, independent of their outcome. We did not interpret alarms for survived patients within the other time periods as false alarms, as surviving patients may get also trigger the alarm when they are critical during their stay in the ICU. For this reason alarms for surviving patients can be called "false alarms" only if they survived being in an alarm state. Since the absolute values for sensitivity and specifity depend on the diagnostic threshold, we will further compare the different diagnostic systems cf. scores by their corresponding ROC characteristics, i.e. their AUC value.

3. The MEDAN RRT Score

After training and testing the neural network we set up an alarm system (11) based on the resulting diagnostic rules. It is still available in the internet, see www.medan.de. Nevertheless, since internet access is not standard in ICUs, we decided to design a simpler version of the diagnostic system which is comparable to standard scores used in ICUs.

As described in section 2.1, there is a formal similarity between scores and special neural networks. Therefore we designed a score network and trained it for the best possible performance on the three already identified most important variables systolic blood pressure, diastolic blood pressure and number of thrombocythes. In difference to standard neural network learning algorithms, not only the number of the RBF neurons and width of the input receptive fields, i.e. the interval boarders on the three input variables, have to be determined, but as constraint the input fields (intervals) do not overlap and are separated by the interval boarders. We used an evolutionary algorithm for this task (13). After the generation of 360 mutations and selections a score network was obtained which did not change in the limit. This score network can be described like an ordinary score by a table, see Table 2.

Score	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
RR_{sys}	≤119	>119	>151	>221	>251	>265	-	-	-
RR_{dia}	≤ 42	> 42	> 47	> 49	> 64	> 83	> 117	> 121	> 126
Thromb.	≤112	>112	>202	>312	>371	>621	>770	-	-

Table 2 The new RRT score, from (13)

In order to compute the score, the three score values have to be summed up. The optimal threshold is $\theta = 6$: For a sum greater or equal to 6 the outcome prediction is favorable for the patient (85,7% correctly classified as “survived”), otherwise severe problems will arrive. In our experience, patients who stay several days below a score of 6 have a high probability to die. In Table 3, this is shown by the mortality associated to the score ranges.

Score	0..2	3..5	6..9	10..13
Mortality	98.41%	81.65%	13.68%	1.89%

Table 3 Mortality related to score ranges

How does the new score perform generally in comparison to the other standard scores used in ICUs? For a comparative analysis of the new MEDAN score we evaluated the following scores on our data.

a) SOFA (Sepsis-Related Organ Failure Assessment) (16): the SOFA score assesses organ malfunction (respiratory, cardiovascular, renal, coagulation, liver, neurological) each on a scale of 0 to 4 in whole-number values. The sum of these values for the individual organs is called the SOFA score. 10 variables and the Glasgow Coma Scale (GCS) (7) are needed to calculate the score.

b) APACHE II (Acute Physiological and Chronic Health Evaluation) (8): APACHE II is the score for outcome prognosis of ICU patients assessing acute disorders, age and overall health (on a scale of 0 to 71 of whole-number values). 13 variables and GCS are needed to calculate the score.

c) SAPS II (Simplified Acute Physiology Score) (9): The SAPS II score is another ICU score using 16 variables and GCS.

d) MODS (Multiple Organ Dysfunction Score) (10): The MODS score assesses organ states (respiratory, liver, renal, coagulation, heart, neurological) on a whole-number scale. It uses 6 variables and GCS.

For our analysis, a score was calculated every time when the necessary variables were given without considering the GCS. The GCS was not included in the scores since it was not always available for our data.

In Fig. 3 the diagnostic capabilities of the different scores are shown, compared by their AUC values for the first and last 3 days on the intensive care units.

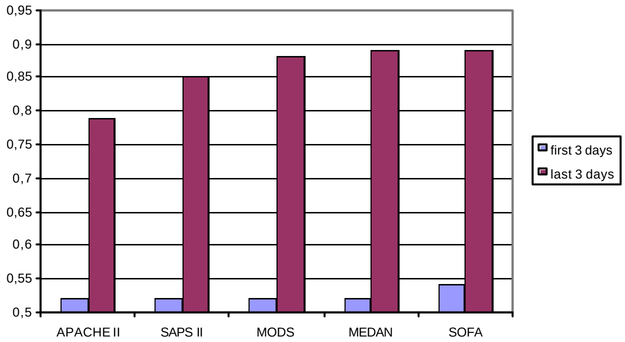

Fig. 3 The diagnostic AUC of different scores

We notice that the MEDAN RRT score is as performing as the best, by experience-evolved score, the SOFA score (13). We might interpret that we short-cut the evolution of the score by an evolutionary algorithm based on the available data instead of evaluating it with costly medical studies. Thus, it turned out that the resulting new score performs as well as the general neural network and the best score available, but it uses only three parameters for input.

The usage of the score is not limited to the last three days: A bad score indicates a bad situation for the patient "as if he or she is in the state of being in the last there days before death" whenever the score is computed. Comparisons of individual histories and computed scores showed good correlations for the whole ICU stay.

4. Discussion and Outlook

The objective of our MEDAN project was the creation of a reliable diagnosis and prediction for septic shock patients. For this task we compared the classification performance of a trained neural network and scores by ROC analysis. The analysis showed that the best score for abdominal septic shock outcome diagnosis is the SOFA score. Our system achieves a similar classification performance as with the SOFA score, but with fewer variables (three instead of ten). We argued that to create an alarm system with data from the last three ICU days is reliable and can be used throughout the whole patient's ICU stay. The system may be useful to warn physicians of critical patient conditions. If the patient is likely to die, then more alarms are given.

Additionally, we created a new score which performs as well as the neural network, but allows the computation of the outcome by hand. Since it uses only three variables in difference to the best score which uses ten and performs as well, the MEDAN score is suitable as cheap, reliable standard score on ICUs during the whole stay.

What are the clinical implications of this approach? If the prediction does not depend on the catecholamine level, do we have to revise all our therapies? It should be underlined that this is not at all the case. The MEDAN alarm system and the RRT score give only a prediction based on the most significant variables which were measured in the multicenter ICU stays. Here, we have to take into account several contextual restrictions:

- The prediction can be given without knowing the catecholamine level the respiration regime. However, we can assume that the catecholamines which were regularly given in the ICU are sufficient and adapted to the patients needs. Since this context was similar for all patients (“good practice”) it is no discriminative fact but already part of the diagnose “septic shock”.

- The discriminative variables are selected from the set of all variables which were sufficiently often measured. All variables like the concentrations of specific cytokines (14), the genetic disposition (6) and the inflammation state (2) (hyper/hypoinflammation) of the patients were not available to our analysis. Therefore, in other studies other well documented variables may give better results. However, because our predictive variables are very general and are influenced by a myriad of biochemical processes we do not attend a better performance by specific molecules.
- The prediction is based on the whole group of all patients with abdominal septic shock. If we have enough data to select subgroups of the patients, e.g. hyper/hypoinflammation responder, instead of averaging we might get better conditional results within each subgroup. This enrichment of the MEDAN RRT score is up to subsequent prospective studies which have to avoid the problem of missing values for the subgroups.

In April 2002 we started a prospective multi-center study to check the clinical usefulness of our alarm system. Everybody who is interested in participating is asked to contact the authors, see www.medan.de.

References

(1) Bernard GR, Vincent J-L, Laterre PF, et al. Efficacy and safety of recombinant human activated Protein C for severe sepsis. N. Engl. J. Med. 2001; 344:699–709

(2) Bone RC. Sir Isaac Newton, sepsis, SIRS and CARS. Crit.Care Med. 1996; 24:1125-1128

(3) Bone RC, Balk RA, FB Cerra, et al. American college of chest physicians/society of critical care medicine consensus conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit. Care Med. 1992; 20:864–875

(4) Brause R, Hamker F, Paetz J. Septic shock diagnosis by neural networks and rule based systems. In Schmitt M, et al., editors, Computational intelligence processing in medical diagnosis. Heidelberg, Physica, 2002; 323–356

(5) Hanisch E, Encke A. Intensive care management in abdominal surgical patients with septic complications. In Faist E, editor, Immunological screening and immunotherapy in critically ill patients with abdominal infections, Berlin, Springer-Verlag, 2001;71–138

(6) Janeway CA Jr., Medzhitov. Innate immune recognition. Annu Rev. Immunol. 2002; 20:197-216

(7) Jennett B, Teasdale G. Assessment of coma and impaired consciousness: A practical scale. Lancet 1974;1:81–84

(8) Knaus WA, Draper EA, Wagner DP, et al. APACHE II: A severity of disease classification system. Crit. Care Med. 1985;13(10):818–829

(9) Le Gall Jr, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a european / north american multicenter study. JAMA 1993;270:2957–2963

(10) Marshall JC, Cook DJ, Christou NV, et al. Multiple organ dysfunction score: A reliable descriptor of a complex clinical outcome. Crit Care Med. 1995;23(10): 1638–1652

(11) Paetz J, Arlt B. A neuro-fuzzy based alarm system for septic shock patients with a comparison to medical scores. Proc. of the 3rd Int. Symp. on Medical Data Analysis, 2002;42-52

(12) Paetz J. Metric rule generation with septic shock patient data. Proc. of the 1st IEEE Int. Conf. on Data Mining, 2001;637–638

(13) Paetz J. Evolving Score Neural Networks, Proc. of the 1st Indian Int. Conf. on Artificial Intelligence (IICAI), Hyderabad, India, 2003.

(14) Sfeir T, Saha DC, Astiz M, Rackow EC. Role of interleukin –10 in monocyte hyporesponsiveness associated with septic shock. Crit. Care Med. 2001; 29:129-133

(15) Vanderschueren S, De Weerdt A, Malbrain M, et al. Thrombocytopenia and prognosis in intensive care. Crit. Care Med. 2000;28:1871–1876

(16) Vincent JL, Moreno R, Takala J, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22:707–710

Authors: PD Dr. Rüdiger Brause, Prof. Dr. Dr. Ernst Hanisch, Dr. Jürgen Paetz, Björn Arlt

To contact the corresponding author:

PD Dr. R. Brause,
J.W.Goethe-University, FB 15,
Institut für Informatik,
D-60054 Frankfurt a.M., Germany

Tel. +(49)69-798-23977
Fax +(49)69-798-21228
Email: Info@Medan.de