Severe sepsis mortality prediction with logistic regression over latent factors

Vicent J. Ribas[1], Alfredo Vellido[1], Juan Carlos Ruiz-Rodríguez[2] *MD* and Jordi Rello[3] *MD-PhD*

1- Dept. de Llenguatges i Sistemes Informàtics - Universitat Politècnica de Catalunya
Edifici Omega, Campus Nord, 08034, Barcelona - Spain

2- Vall d'Hebron Intensive Care Unit. Vall d'Hebron Research Institute. Universitat Autònoma de Barcelona
Passeig Vall d'Hebron 119-129, 08035, Barcelona - Spain

3- CIBERES (Centro de Investigacin Biomdica en Red de Enfermedades Respiratorias) and Vall d'Hebron Intensive Care Unit. Vall d'Hebron Research Institute. Universitat Autònoma de Barcelona
Passeig Vall d'Hebron 119-129, 08035, Barcelona - Spain

Abstract

Sepsis is one of the main causes of death for non-coronary ICU (Intensive Care Unit) patients and has become the tenth most common cause of death in western societies. This is a transversal condition affecting immunocompromised patients, critically ill patients, post-surgery patients, patients with AIDS, and the elderly. In western countries, septic patients account for as much as 25% of ICU bed utilization and the pathology affects 1% - 2% of all hospitalizations. Its mortality rates range from 12.8% for sepsis to 45.7% for septic shock.

The prediction of mortality caused by sepsis is, therefore, a relevant research challenge from a medical viewpoint. The clinical indicators currently in use for this type of prediction have been criticized for their poor prognostic significance. In this study, we redescribe sepsis indicators through latent model-based feature extraction, using Factor Analysis. These extracted indicators are then applied to the prediction of mortality caused by sepsis. The reported results show that the proposed method improves on the results obtained with the current standard mortality predictor, which is based on the APACHE II score.

Keywords: Sepsis, Mortality prediction, Factor analysis, Logistic regression

1. Introduction

Sepsis is a clinical syndrome defined by the presence of both infection and Systemic Inflammatory Response Syndrome (SIRS). This condition can lead to severe sepsis, which implies organ dysfunction, or to an even more severe state: septic shock (severe sepsis with hypotension refractory to fluid administration) and multiorganic failure [1, 2].

In western countries, septic patients account for as much as 25% of ICU bed utilization and the pathology occurs in 1% - 2% of all hospitalizations. The mortality rates of sepsis range from 12.8% for sepsis and 20.7% for severe sepsis, to up to 45.7% for septic shock [3].

This pathology has followed a clear upwards trend over the last 20 years, reaching 300,000 cases per year only in the United States. The number of sepsis cases in this country is projected to grow at a yearly rate between 1.5% and 8% as the population ages and treatment becomes more aggressive [4, 5].

The medical management of sepsis is therefore a serious challenge to healthcare systems as a whole. Recently, doubts have been raised about the usefulness of current diagnostic methods for this pathology [6]. This is due to their poor specificity and sensitivity results, as well as to their lack of prognostic significance. Raising to this challenge, the aim of this paper is to investigate new sets of descriptors that improve on current standards in terms of mortality prediction accuracy.

Here, we propose the use of a latent model-based feature extraction approach to obtain such new sets of descriptors, or prognostic factors. The experimental results reported in this study show them to be readily interpretable. Interpretability is, needless to say, a sensitive issue in the medical ambit, and one that should not be underestimated: lack of translation of the prognostic factors into usable clinical knowledge might render the proposed approach useless [7].

In the experiments of this study, the extracted factors are used to predict mortality through standard logistic regression (LR), a method commonly used in medical applications [8, 9] and widely trusted by clinicians. The prediction accuracy results herein reported improve on those obtained with current standard data descriptors and therefore provide support for the use of these new factors as risk-of-death predictors in ICU environments.

2. Related Work

The SIRS pathology is known to be a quite sensitive indicator of sepsis [10], but also one of poor specificity. Different studies have shown that the incidence of SIRS is quite high in critical patients in general. For example, Pittet *et al.* [11] present a SIRS incidence of up to 93% in critical care patients, while Rangel *et al.* show an incidence of 68% [10]. The latter study also shows that 25% of patients with SIRS developed a sepsis, 18% presented severe sepsis, and 4% of them, septic shock. Regardless of these incidence ratios, the early detection of patients with a higher risk of death remains a challenge.

The MEDS (Mortality in Emergency Department Sepsis) score is a collection of variables routinely recorded in the emergency departments (terminal illness, tachypnea/hypoxemia, septic shock, platelet count, age, lower respiration infection, bands, nursing home resident and mental status). It was shown in [12] to yield an area under the ROC (AUC) of 0.88 for the population under study: patients at the emergency department with SIRS (not taking into account those septic patients admitted in the emergency department who were not critical enough to be admitted in the ICU).

Since the publication in 1985 of the Organ System Failure (OSF) score by Knaus [13], which is a prognosis scale to evaluate and quantify the Multiple Organ Dysfunction Syndrome (MODS), alternative prognostic scores have been developed. They include the APACHE II (Acute Physiology and Chronic Health Evaluation II) score [14], the Multiple Organ Dysfunction Score (***MODS***) [15] and the SOFA (Sequential Organ Failure Assessment) score [16], and the LODS (Logistic Organ Dysfunction System) [17]. Two prognostic scores based on the PIRO model (Predisposition, insult/infection, response and organ dysfunction) have also been recently proposed: the SAPS3 PIRO score ([18]: AUC 0.77) and the PIRO score ([19]: AUC 0.70).

Machine learning methods have been used with varying success for the prediction of mortality caused by sepsis. A diagnostic system for septic shock based on ANNs (Radial Basis Functions and supervised Growing Neural Gas) was presented in [20], reporting an overall correct classification rate of 67.84%, with a specificity of 91.61% and an extremely poor sensitivity of 24.94%. Support Vector Machines (SVM) have also been used in this context. Tang *et al.* [21] presented a SVM-based system for sepsis and SIRS prediction from non-invasive cardiovascular spectrum analysis, reporting an accuracy

of 84.62%, with a rather low specificity of 62.50% and a high sensitivity of 94.44%.

3. Materials

This work is based on a prospective study approved by the Clinical Investigation Ethical Committee of the *Hospital Universitari del Vall d'Hebron* in Barcelona, Spain, and is based on a prospective database collected by the Research Group on Shock, Organic Disfunction and Ressuscitation of Vall d' Hebron's Intensive Care Unit. The database consisted of data collected in the ICU at this hospital between June 2007 and November 2008. During this period of time, 156 patients were admitted to the ICU (including medical and surgical patients) with severe sepsis.

The mean age of the patients in the analyzed database was 57.24 (with standard deviation ± 15.25) years, 41.03% of patients were female and the diagnosis on admission was 64.10% *medical* and 35.90% *surgical*. The origin of primary infection for the cases on the database was 49.36% pulmonary, 14.74% abdominal, 10.26% urinary, 7.05% skin/muscle, 2.56% central nervous system (CNS), 1.28% catheter related, 0.64% endovascular, 5.13% biliar, 2.56% mediastinum and 6.41% unknown.

The collected data show the worst values for all variables during the first 24 hours of evolution for severe sepsis. Organ dysfunction was evaluated by means of the SOFA score [16], which quantifies the dysfunction and failure of six organs/systems (Cardiovascular, Respiratory, CNS, Hepathic, Renal and Haematologic), as shown in Table 1, and scored from 0 (normal function) to 4 points (maximum failure). Severity was evaluated by means of the APACHE II score (for further reference, see [14]).

In the population under study, 56.41% received mechanical ventilation with a PaO_2/FiO_2 of 166 ± 100, 73% received vasoactive drugs, the platelet count was $1.84 \cdot 10^5/L \pm 1.36 \cdot 10^5/L$, the Lactate Levels were 3.40 ± 3.60 mmol/L, and the APACHE II score was 22.73 ± 8.53.

In 2004, the Surviving Sepsis Campaign (SSC) defined a set of guidelines for the management of severe sepsis and septic shock [22]. More specifically, these set of guidelines were proposed for both the first 6 hours of evolution and for the first 24 hours. Therefore, the compliance of the SSC bundles for the first 6 hours was 31.41%, out of which 77.56% had Haemocultures performed, 85.90% received antibiotics, 57.05% had their lactate monitored,

69.87% received Volume (i.e. Fluid Resuscitation) 18.59% received transfusions and 4.89% received Dobutamine. The $SVCO_2$ values were 45.53 ± 70.76 and the Haematocrit 26.53 ± 12.92 for the first 6 hours. The compliance of the first 24 hour SSC bundles was 51.92%, the glycaemia was < 150 mg/dL in 62.18% of cases and Plateau Pressure (PPlateau) < 30 cm H_2O in 44.23% of cases. The mortality rate intra-ICU for the our study population was 34%.

The variables used in our analyses are listed in Table 2.

4. Methods

4.1. Feature Extraction Methods

The most widely-used feature extraction methods are Principal Component Analysis (PCA) [23], Non-Negative Matrix Factorization (NMF) [24], and Factor Analysis (FA) [25]. PCA obtains new factors using the eigenvectors of the sample covariance matrix. This matrix presents the property that a sub-base made of the eigenvectors associated with the highest eigenvalues yields a reconstruction that minimizes the square error.

NMF is also a natural way of obtaining a meaningful base because the observations are all positive, and most are multinomially distributed. Provided that this factorization does not give a ranking of the elements of the base as in the case of PCA, an arbitrary dimension of the sub-base that spans the observation can be selected. The bases (factors) that are obtained with both methods span a subspace which reconstructs the original observation with an error.

The covariance matrix can be decomposed into the sum of two terms: the product of the base that we use in order to represent the observed data, plus an error term, in the form $\Sigma = \Lambda\Lambda^T + \Psi$.

In PCA and NMF, the covariance of the error term is a full matrix, which means that the factor base does not account for all the interactions between the observed variables. In other words, the error term still contains information about interactions or relations between these variables in addition to the specific information of each variable (diagonal term of Ψ).

To overcome this limitation, we propose the use of FA, which finds a decomposition of the covariance matrix $\Sigma = \Lambda\Lambda^T + \Psi$ such that Ψ is a diagonal matrix. This method selects the factors following a criterion based on the correlation between features of the observation vector. In our implementation, the model is estimated using maximum likelihood (ML), which explicitly assumes a Gaussian distribution for x. Nevertheless, and independently of

assumptions concerning data distribution, ML searches for a decomposition of Σ so that the error matrix Ψ has a diagonal structure. Therefore, the model generates the observation from a set of latent variables that are independent of the error term, and takes into account all the correlations between variables.

Sections 4.2 and 5.3 show that, although the variables $\{x, f\}$ in the analyzed data fail to pass a multivariate normality test, the covariance matrix of the residual error can be assumed to be diagonal.

4.2. Test of Multivariate Normality

The multivariate normality test was performed using the Mardia Test [26], which states that the skewness and kurtosis for multivariate Gaussian distributions with p degrees of freedom are: $\beta_{1,p} = 0$ $\beta_{2,p} = p(p+2)$. The null hypotheses for the skewness and kurtosis are defined as:

$H_{0,Skew}$ =multivariate skewness is consistent with a multivariate normal distribution.

$H_{0,Kurt}$ =multivariate kurtosis is consistent with a multivariate normal distribution.

4.3. Factor Analysis Through ML Estimation

The likelihood function for the sample X is:

$$L(X, \Theta) = \prod_{i=1}^{n} f(x_i, \Theta) \tag{1}$$

and, as a result, the log-likelihood function can be defined as:

$$l(X, \Theta) = \ln L(X, \Theta) = \sum_{i=1}^{n} \ln f(x_i, \Theta) \tag{2}$$

which has a support function:

$$\begin{aligned} F(\Lambda, \Psi; S) &= tr((\Lambda\Lambda^T + \Psi)^{-1} S) \\ &\quad - \ln \left|(\Lambda\Lambda^T + \Psi)^{-1} S\right| - p \end{aligned} \tag{3}$$

This is a linear function of the log-likelihood $l(X;\Sigma)$ and its minimum corresponds to a maximum in $l(X;\Sigma)$ [1]. The minimization of $F(\Lambda,\Psi;S)$ can be performed by an iterative Newton-Raphson method (resulting in an eigen value equation) by minimizing separately for Λ (over a fixed Ψ) and subsequently minimizing over Ψ.

4.4. Logistic Regression

Logistic regression studies binomially distributed variables of the form $C_i \sim B(n_i, p_i)$ where n_i and p_i correspond to the number of patients and the probability of exitus. In our study, C_i is a class label that takes the value 1 for survival and 0 for exitus. The logistic model proposes that, for each patient i, there is a set of explanatory variables that might inform the final probability. Thus, the model takes the form: $p_i = E(\frac{C_i}{n_i}|X_i)$, for each variable i (be it from the original set of variables listed in Table 2, or one of the extracted factors).

Here, the natural logs of the odds ratio for the unknown binomial probabilities are modeled as a linear function of X_i:

$$\log(\frac{p_i}{1-p_i}) = \beta_0 + B^T \cdot X_i, \tag{4}$$

where β_0 is the intercept and B is the vector of logistic regression coefficients. In this study, the intercept and regression coefficients were estimated by ML with a generalized linear model.

5. Results

5.1. Test of Multivariate Normality

The Mardia test with 35 degrees of freedom over the data matrix $X = \{x_1 \dots x_n\}$, where n is the number of cases, yielded $H_{0,Skew} = H_{0,Kurt} = 1$ and, therefore, the null hypothesis could be rejected for both the skewness and kurtosis with a p-value $<< 0.001$.

Also the Mardia test with 14 degrees of freedom over the factor matrix $FM = \{f_1 \dots f_n\}$ yielded $H_{0,Skew} = H_{0,Kurt} = 1$ and, therefore, the null hypothesis could be rejected for both skewness and kurtosis with a p-value $<< 0.001$. Although both matrices X and F fail to pass the gaussianity test, section 5.2 shows that it is reasonable to model X with 14 factors.

[1] It can be shown that this support function corresponds to the likelihood ratio for the hypotheses in 5.2

5.2. Selection of Latent Factors

Correlations between variables are likely to be found in the database and, as a result, the intrinsic dimensionality of the problem is bound to be lower than 35 (number of variables in Table 2). In order to determine the intrinsic data dimension, a statistical test based on the likelihood ratio was used. It contrasts the hypothesis that the covariance matrix can be decomposed into the sum of two components, where matrix Λ consists of $(d \times k)$, with $d = 35$ and k the number of factors of the model. For the problem at hand, the hypotheses to be tested are:

$H_0 : \Sigma = \Lambda\Lambda^T + \Psi$

$H_1 : \Sigma \neq \Lambda\Lambda^T + \Psi$

$$-2\log\lambda = nF(\Lambda, \Psi). \tag{5}$$

The statistic $-2\log\lambda$ has an asymptotic χ^2_s distribution under H_0, where s are the degrees of freedom and $n = 156$.

Using the Bartlett correction:

$$n' = n - 1 - \frac{1}{6}(2d + 5) - \frac{2}{3}k \tag{6}$$

for any specified number of factors k, we reject H_0 if the statistic $U_{k,n'}$ is greater than a threshold given by a chi-square value, i.e.

$$U_{k,n'} = n'F(\Lambda, \Psi) > \chi^2_{0.5((d-k)^2-(d+k))} \tag{7}$$

where

$$F(\Lambda, \Psi) = tr\Sigma^{-1}S - \ln\left|\Sigma^{-1}S\right| - d, \tag{8}$$

and S corresponds to the covariance matrix of X. When H_0 is true, this statistic has an asymptotic χ^2 distribution with $s = 0.5((d-k)^2 - (d+k))$ degrees of freedom.

From the results summarized in Table 3, we compare $U_{14,196} = 211.652$ with $\chi^2_{196} = 229.663$ (95% of significance) and, therefore, the analyzed data is consistent with a 14 factor model [26].

5.3. Diagnosis of the Factor Analysis Model

Given that the variables of the model do not follow a Gaussian distribution, we proceeded to test if Ψ was a diagonal matrix. Once the covariance of the residual error matrix was computed, the value of the diagonal elements was compared to the off-diagonal ones, for $i \in \{1 \dots d\}$. Specifically, the value of K so that

$$|\psi_{ii}| \geq K \sum_{j=1, j \neq i}^{d} |\psi_{ij}|$$

was calculated, turning out to be $K = 11.3$ for all ψ_{ii}. Because the maximum off-diagonal element is at least two orders of magnitude lower than any of the diagonal elements, diagonal dominance is clear and it can be assumed that all possible interactions between variables are accounted by the matrix Λ.

5.4. Factor Interpretation from a Clinical Viewpoint

As described in the previous subsections, the application of FA resulted in a consistent 14-factor model of the original data set. The cumulative proportion of total (standardized) sample variance explained by this model was found to be 79.12%.

Table 4 summarizes de matrix of loadings corresponding to the original variables listed in Table 2. Taking into consideration the highest factor loadings (in absolute value) for every given variable, these factors were mapped into different easily interpretable clinical descriptors, explained as follows:

- Factor 1: Related to cardiovascular function and, more specifically, to the cardiovascular SOFA score and vasoactive drugs c.f. table 1.
- Factor 2: Corresponds to haematologic function (haematologic SOFA score measured by platelet count).
- Factor 3: Corresponds to the use of Mechanical Ventilation) and the PPlateau.
- Factor 4: Also corresponds to respiratory function, Respiratory SOFA score (PaO_2/FiO_2 relation).
- Factor 5: Related to the microorganism producing the sepsis and wether this sepsis is polimicrobial or not.

- Factor 6: It corresponds to the 24 h. SSC bundles and glycaemic indexes.
- Factor 7: It corresponds to renal function measured by the SOFA score and the total SOFA score.
- Factor 8: Related to the hepathic function measured by the SOFA score.
- Factor 9: It corresponds to the administration of antibiotics and haemocultures taken during the first 6 h. of evolution.
- Factor 10: Relates to the number of dysfunctioning organs for a SOFA 1-2 and the total number of dysfunctioning organs.
- Factor 11: It corresponds to the CNS function measured in the SOFA score and the total number of dysfunctioning organs.
- Factor 12: Relates to the base pathology and the foci of sepsis.
- Factor 13: Corresponding to the global APACHE II score.
- Factor 14: Relates to the focus of sepsis, transfussions and hepathic SOFA score.

The factors obtained with this method are coherent with the SOFA score as a description and measure of organ failure and dysfunction [16], combined with the management guidelines defined by the Surviving Sepsis Campaign [22]. Therefore, it can be safely concluded that they are related to SOFA and the actions taken to mitigate this organ deterioration.

This is a result of particular interest. One of the main challenges in mortality prediction is that of producing flexible models that can robustely fit the observed data without the need for unnecesary contextual assumptions, and in the presence of subtle interactions between covariates. This happens because standard medical indicator-based models typically rely on hand-crafted parametric solutions to get around the problem [27]. One clear example of this is the categorization of the SOFA score prognostic indicators described in section 3. The obtained FA solution goes beyond this categorization while accounting for covariate interactions.

As mentioned in the introduction, the interpretability of results is paramount in real clinical applications [7]. The reported FA not only complies with this

requirement: it also provides a parsimonious data representation that can be used as a basis for mortality prediction related to the sepsis pathology.

5.5. Mortality prediction using logistic regression over 14 factors

We now progress to the task of mortality prediction itself, using the obtained 14-factor FA solution as starting point. The performance of the model was evaluated by bagging (bootstrap aggregating), as defined in [28]. One thousand new training sets consisting of 104 samples were generated by sampling uniformly and with replacement. For each new training set, approximately 52 training instances from the original database were left out. These left-out instances were used to evaluate the system performance (i.e. out-of-bag prediction). The model performance was averaged over the 1,000 bagged training sets.

Table 5 shows the coefficient estimates β, Z-Scores and p-values resulting from fitting a logistic regression model to the 14 factors (inputs) and the outcome in the ICU (output). The Z-Scores measure the effect of removing one factor from the model [29, 30]. A Z-score greater than 1.96 in absolute value is significant at the 5% level and provides a measure of the relevance for the prediction of a given factor.

As shown in table 5, factor 3, which is related to *Mechanical Ventilation* and *Pplateau*, shows the strongest effect together with factor 13, which is related to the APACHE II score. Factor 8 (Hepathic Function measured with the SOFA Score) and factor 10 (related to the number of Dysfunctional Organs) are also found to be relevant.

Table 6 shows the theoretical 95% confidence intervals for the LR coefficients and their corresponding odds-ratio. More specifically, $CI = \beta \pm 1.96\sigma$ and $OR = e^{CI}$.

It is worth noting at this stage that, with LR, the factors related to the *Surviving Sepsis Campaign* show no strong effect on mortality prediction. This result may be due to the low compliance with the *Surviving Sepsis Campaign Bundles* for the first 6 and 24 hours of evolution (31.41 % and 51.91 % respectively for the ICU under study). However, it is interesting to note that factor 9 (antibiotic administration and haemocultures) presents a higher impact than that of factor 6 (24 h. bundles with glycaemic indexes). For our ICU, 85.90 % of patients received antibiotics during the first 6 h of evolution and 77.56 % had haemocultures during the same period of time. In fact, timely administration of antibiotics and performance of haemocultures are considered critical to improving the prognosis of septic patients.

Regression on the 14 factors together with bagging resulted in an AUC of 0.75. It is also important to stress that, for the bagging experiment, LR for each coefficient varies within the 95% CI for each iteration, as shown in table 7. A decision threshold of $\gamma = 0.64$ was empirically selected (for the maximizization of the discrimination probability) to decide whether the patient survives. This resulted in a prediction error over the test data of 0.22, a sensitivity of 0.64, and a specificity of 0.84.

5.6. Comparison with Logistic Regression over a Selection of the Original Variables

Further experiments aimed to compare the predictive ability of the FA 14-factor solution with that of the original data attributes were carried out. For that, the most significant clinical attributes were selected in a backward feature selection process. The selected attributes were: the total number of dysfunctioning organs; the APACHE II score; and the worst lactate levels. The corresponding coefficients, z-scores and p-values for these three variables are presented in tables 8 and 9.

Regression on the most significant attributes together with bagging yielded an AUC of 0.71, a lower result than the one obtained with the FA solution. It is also important to note that for the bagging experiment, logistic regression for each coefficient varies within the 95% CI for each iteration, as shown in table 9. Following the procedure outlined in the previous subsection, a decision threshold of $\gamma = 0.64$ was empirically selected. This resulted in a prediction error over the test data of 0.25 (higher than the FA solution), a sensitivity of 0.72, and a specificity of 0.77.

5.7. Comparison with the APACHE II Score

The Risk-of-Death (ROD) formula based on the APACHE II score can be expressed as [14]:

$$\ln(\frac{ROD}{1-ROD}) = -3.517 + 0.146 \cdot A + \epsilon \tag{9}$$

Where A is the APACHE II score and ϵ is a correction factor depending on clinical traits at admission in the ICU. For instance, if the patient has undergone post-emergency surgery, ϵ is set to 0.613. The aplication of this formula with a threshold of $\gamma = -0.25$ to the population under study yields an error rate of 0.28 (higher than previous results), a sensitivity of 0.55 (very low) and a specificity of 0.82. The AUC was 0.70.

A previous study [31] presented very similar results to those reported here for a similar ICU. Furthermore, a recent study from 2009 [32] presented very similar results to those reported here for neurocritically ill patients (with a very low sensitivity of 0.47).

6. Conclusions

Sepsis is a prevalent pathology in the clinical ICU environment, and one with relatively high mortality levels associated. Its medical management is therefore both a sensitive issue and a serious challenge to healthcare systems.

The clinical indicators of sepsis currently in use are known to be of limited relevance as mortality predictors. In the assessment of ROD for critically ill patients, sensitivity is of paramount importance due to the fact that more aggressive treatment and therapeutic actions may result in better outcomes for high risk patients. As validated by the results reported in section 5.7 and similar ones reported in other studies [31], the ROD formula presented in [14] is poor in terms of sensitivity (i.e., it results in a high number of false positive cases). This is despite the fact that it is widely accepted in practice and yields acceptable accuracy results. Its poor sensitivity may be the result of its formula being based on clinical traits and the APACHE II score only.

In this paper we have put forward a new and simple method for the assessment of ROD in septic patients. It proposes a change of data representation in the form of feature extraction using FA, and uses LR over the resulting latent factors for the prediction itself. The main advantage of the proposed approach is that it removes collinearities and noisy inputs while keeping the method simple and fully interpretable from a clinical point of view. In other words, the strength of this study lies in the fact that it is possible to derive a prognostic score from a set of physiopathologic and therapeutic variables, which are available at the onset of severe sepsis.

The proposed method may be understood as a generalization of the ROD formula introduced in [14], where the ϵ corrective factor, which models clinical traits at admittance in the ICU, is accounted for by the latent-factor representation. It takes not only the contribution of the APACHE II score into consideration, but also other important clinical traits such as the number of dysfunctioning organs combined with the Sequential Organ Failure Assessment (SOFA), which also impacts on the mortality rates of Septic patients. The reported ROD assessment takes into consideration the Respiratory and Hepatic SOFA scores. It is precisely all the extra parameters considered in

our experiments the reason behind the significant improvement on sensitivity. This improvement is achieved while keeping model complexity under control and without compromising the interpretability of the results (given that all the parameters involved are routinely monitored in an ICU).

A word of caution must be given, though, as the system performance has only been evaluated in a single ICU and a limited population sample. For this reason, future work should lead towards a multi-centric prospective study, in order to validate the generalizability of the method.

7. Acknowledgements

This research was partially funded by Spanish MICINN R+D Plan TIN2009-13895-C02-01 project.

References

[1] 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; 20:864-874.

[2] Levy M.M, Fink M.P, Marshall J.C, Edward Angus A., Cook D., Cohen J., Opal S.M, Vincent J.L, Ramsay G. for the International Sepsis Definitions Conference (2003). "2001 SCCM/ESICM/ACCP/ATS/SIS International sepsis definitions conference". Intensive Care Med 2003; 29: 530-538.

[3] Esteban A.,Frutos-Vivar F., Ferguson N., Penuelas O., Lorente J.Al, Gordo F., Honrubia T., Algora A., Bustos A., Garcia G., Rodriguez I., Ruiz. R, "Sepsis incidence and outcome: contrasting the intensive care unit with the hospital ward". Crit Care Med 2007; 35(5):1284-1289.

[4] Angus D.C., Linde-Zwirble W.T, Lidicker J., Clermont Gl, Carcillo J., Pinski M.R., "Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care". Crit Care Med 2001; 29(7):1303-1310.

[5] Martin G.S., Mannino D.M., Eaton S., Ross M., "The epidemiology of sepsis in the United States from 1979 through 2000". N Eng J Med 2003; 348(16):1546-1554.

[6] Vincent J.L., "Dear SIRS, I'm sorry to say that I don't like you... Crit. Care. Med. 1997; 25(2):372-374.

[7] Lisboa P.J.G., Vellido A., Martn J.D. "Computational Intelligence in biomedicine: Some contributions", In Procs. of the 18^{th} European Symposium on Artificial Neural Networks (ESANN), 2010:429-438.

[8] Paliwal, M., Kumar. U.A., "Neural networks and statistical techniques: A review of applications". Expert Systems with Applications 2009; 36(1):2-17.

[9] Kurt, I., Ture, M., Kurum, A.T., "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease". Expert Systems with Applications 2008; 34(1): 366-374.

[10] Rangel-Frausto M.S, Pittet D., Costigan M., Hwang T., Davis C., Wenzel R.P., "The natural history of the sistemic inflammatory response syndrome (SIRS). A prospective study". JAMA 1995; 273: 117-123.

[11] Pittet D., Rangel-Frausto S., Li N., Tarara D., Costigan M., Remple L., Jebson P., Wenzel R.P, "Systemic inflammatory response syndrome, sepsis, severe sepsis and septic shock: incidence, morbidities and outcome in surgical ICU patients". Intensive Care Med 1995; 21: 302-309.

[12] Sankoff J.D., Goyal M., Gaieski D.F., Dietch K., Davis C.B., Sabel A.L., Haukoos J.S., "Validation of the Mortality in Emergency Department Sepsis (MEDS) score in patients with the systemic inflammatory response syndrome (SIRS)", Crit. Care. Med. 2008; 36(2):1-6.

[13] Knaus W.A, Draper E.A, Wagner D.P, Zimmerman J.E, "Prognosis in acute organ-system failure". Ann. Surg. 1985;202:685-93.

[14] Knaus W. A., Draper E. A., Wagner D. P., Zimmerman J. E., "APACHE II: A severity of disease classification system". Crit Care Med 1985; 13: 818-829

[15] Marshall J.C, Cook D.A, Christou N.V, Bernard G.R, Sprung C.L, Sibbald W.J, "Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome". Crit Care Med 1995;23:1638-52.

[16] Vincent J. L., Moreno R., Takala J., Willats S., De Mendoa A., Bruining H., Reinhart C. K., Suter P. M., Thijs L. G., "The SOFA (Sepsis Related Organ Failure Assessment) score to describe organ dysfunction/failure", Intensive Care Med 1996; 22:707-710.

[17] Le Gall J.R, Klar J., Lemeshow S, "How to assess organ dysfunction in the intensive care unit? The logistic organ dysfunction (LOD) system". Sepsis 1997;1:45-7.

[18] Moreno R.P, Metnitz B., Adler L., Hoechtl A., Baure P., Metnitz P.G.H, SAPS3 investigators, "Sepsis mortality prediction based on predisposition, infection and response". Intensive Care Med 2008; 34: 496-504.

[19] Rubulotta F., Marshall J.C, Ramsay G., Nelson D., Levy M., Williams M, "Predisposition, insult/infection, response and organ dysfunction: a new model for staging severe sepsis". Crit Care Med 2009; 37: 1329-1335.

[20] Brause R., Hamker F., Paetz J., Jain L.C. (ed.), "Septic Shock Diagnosis by Neural Networks and Rule Based Systems", Computational Intelligence Techniques in Medical Diagnosis and Prognosis, Springer Verlag, 2001, pp:323-356.

[21] Tang C.H.H., Middleton P.M., Savkin A.V., Chan G.S.H., Bishop S., Lovell N.H., "Non-invasive classification of severe sepsis and systemic inflammatory response syndrome using a nonlinear support vector machine: a preliminary study", Physiol. Meas. 2010; 31:775-793.

[22] Dellinger R. P., Carlet J. M., Masur H., Gerlach H., Calandra T., Cohen J., Gea-Banacloche J., Keh D., Marshall J. C., Parker M. R., Ramsay G., Zimmerman J. L., Vicent J. L., Levy M. M., "Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock". Intensive Care Med 2004; 30: 536-555.

[23] Jolliffe, I. T. , "Principal Component Analysis", Springer, 2002.

[24] Lee D.D., Seung S.H., "Learning the parts of objects by non-negative matrix factorization". Nature 1999; 6755(401):788-791.

[25] Rubin, D. B., Thayer, D. T. "EM Algorithms for ML factor analysis". Psychometrika 1982; 47:6976.

[26] Mardia K.V., "Measures of multivariate skewness and kurtosis with applications". Biometrika 1970; 3(57):519-530.

[27] Lisboa P.J.G., Vellido A., Tagliaferri R., Napolitano F., Ceccarelli M., Martín-Guerrero J.D., Biganzoli E. "Data Mining in Cancer Research", IEEE Computational Intelligence Magazine 2010 5(1):14-18.

[28] Breiman L., "Bagging predictors", Machine Learning 1996; 24(2):123-140.

[29] Johnson R.A., Wichern D.W. , "Applied Multivariate Statistical Analysis (6th Edition)",Prentice Hall,2007.

[30] Friedman J., Hastie T., Tibshirani R., "The Elements of Statistical Learning". Springer-Verlag (2008).

[31] Wong D.T., Crofts S.L., Gomez M., McGuire G.P., Byrick R.J., "Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients", Crit Care Med. 1995 Jul;23(7):1177-1183.

[32] Ying-Ying S., Xia L., Si-Jie L., Rong L., Jian-Ping D., Ling W., Gui-Hua C., Dong-Yu W., Jin-Xia G., "Predicting hospital mortality using APACHE II scores in neurocritically ill patients: a prospective study", J. Neurol. 2009 Apr; 256:1427-1433.

Table 1: List of SOFA scores, with their corresponding mean and (standard deviation) values.

Cardiovascular (CV)	2.980 (1.496)
Respiratory (RESP)	2.413 (1.062)
Central Nerv. Sys. (CNS)	0.419 (0.859)
Hepatic (HEPA)	0.387 (0.863)
Renal (REN)	0.968 (1.159)
Haematologic (HAEMATO)	0.877 (1.164)
Global SOFA score	7.948 (3.671)
Dysf. Organs (SOFA 1-2)	1.729 (1.124)
Failure Organs (SOFA 3-4)	1.152 (0.892)
Total Dysf. Organs	3.200 (1.406)

Table 2: List of variables used in this study.

Variable	Description
v1	Age
v2	Gender
v3	Sepsis Focus
v4	Germ Class
v5	Polimicrobial Infection
v6	Base Pathology
v7	Cardiovascular SOFA score
v8	Respiratory SOFA score
v9	CNS SOFA score
v10	Hepathic SOFA Score
v11	Renal SOFA Score
v12	Haematologic SOFA Score
v13	Total SOFA Score
v14	Dysfunctioning Organs for SOFA 1-2
v15	Dysfunctioning Organs for SOFA 3-4
v16	Total Number of Dysfunctioning Organs
v17	Mechanical Ventilation
v18	Oxygenation Index PaO_2/FiO_2
v19	Vasoactive Drugs
v20	Platelet Count
v21	APACHE II Score
v22	Surviving Sepsis Campaign Bundles 6h
v23	Haemocultures 6h
v24	Antibiotics 6h
v25	Volume 6h
v26	O_2 Central Venous Saturation 6h
v27	Haematocrit 6h
v28	Transfussions 6h
v29	Dobutamine 6h
v30	Surviving Sepsis Campaign Bundles 24h
v31	Glycaemia 24h
v32	PPlateau
v33	Worst Lactate
v34	O_2 Central Venous Saturation

Table 3: Results of goodness of fit test

k	$U_{k,s}$	s	p-value
12	331	241	$<< 0.001$
13	265.295	218	0.016
14	211.652	196	0.211
15	166.394	175	0.667

Table 4: Loadings Matrix: $|\Lambda(i,j)| >$ quantile 95 for Factor F_i are presented in bold.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
v1	.33	-.60	-.05	-.03	-.05	-.11	.58	-.12	.10	.09	.84	-.00	.17	.21
v2	.06	-.02	.18	.06	.12	-.04	.10	-.12	.13	-.02	-.10	.24	.12	.10
v3	.11	.00	-.05	-.28	-.03	.01	.03	.19	-.05	.06	.13	**-.43**	.16	**.26**
v4	-.16	-.06	.00	-.06	**.97**	-.06	.-.02	.-.01	-.07	-.02	-.06	-.01	-.07	-.08
v5	-.04	.00	.02	-.10	**.76**	-.15	-.07	-.03	-.12	-.03	.02	.06	-.01	.03
v6	-.11	.19	.02	.14	.00	-.05	-.00	-.07	-.01	-.04	.08	**.84**	-.08	.03
v7	**.95**	.06	.15	-.06	-.06	-.00	.08	.13	-.06	.02	-.00	-.07	.12	-.03
v8	.01	-.04	.36	**.89**	-.11	-.00	.01	.11	.03	.02	.11	.04	.02	-.07
v9	.02	.05	.03	-.02	-.03	-.02	.06	.01	.13	.11	**.96**	.01	.09	.13
v10	.08	.12	.03	.10	-.08	.12	.01	**.88**	.05	.22	.07	-.04	.14	**.29**
v11	.18	.12	.04	-.05	-.09	.02	**.91**	.07	-.06	.11	.04	-.00	.29	-.06
v12	.12	**.93**	.09	-.01	.02	-.02	.03	.16	-.03	.08	-.01	.04	.10	.23
v13	.57	.37	.23	.24	-.08	.06	**.38**	.34	.04	.16	.22	-.01	.21	.20
v14	.00	.27	-.12	.03	-.05	.03	.09	.14	.04	**.92**	.10	-.08	.08	.10
v15	.51	.27	.36	.19	-.14	.02	.30	.31	.08	-.37	.25	.04	.19	.01
v16	.34	.44	.06	.17	-.12	.05	.20	.30	.09	**.49**	**.33**	-.08	.20	.01
v17	.10	-.03	**.93**	.14	-.00	-.01	.08	.06	.12	-.09	.08	.02	.18	-.15
v18	.10	-.04	.03	**-.76**	.09	.02	.01	-.00	-.09	-.09	.10	-.24	-.15	-.08
v19	**.88**	.10	.13	-.08	-.13	.01	.06	.09	.11	-.01	-.01	-.06	.05	.02
v20	-.10	-.64	.12	.02	.09	.07	-.12	-.12	.01	-.21	-.08	-.17	-.14	.07
v21	.03	.16	.27	.21	-.01	.01	.35	.05	.09	.05	.21	.10	**.66**	.07
v22	.03	.06	.02	.07	.01	.06	-.01	.08	.48	.13	.02	.06	.04	.03
v23	.21	.04	-.02	-.03	-.07	-.03	.03	.04	**.66**	-.09	.09	.06	-.12	.06
v24	.04	-.01	.01	.06	-.14	.05	-.02	.06	**.68**	-.01	.02	-.08	.00	.01
v25	.52	.07	.06	-.06	.03	.11	-.02	.00	.41	-.00	-.04	-.03	.13	-.06
v26	-.00	.08	-.07	.02	.03	-.21	-.02	.18	.11	.14	.08	.06	-.11	-.00
v27	.04	-.15	.06	-.04	-.01	-.05	-.07	.03	.3	-.01	.05	.23	.05	-.22
v28	.01	.10	-.07	-.00	-.03	-.06	-.05	.09	.05	.05	.12	.00	.01	**.25**
v29	.09	.14	.21	-.15	.03	.05	.01	.06	.04	.06	-.01	.13	.09	.08
v30	.01	.02	-.05	-.03	-.10	**.73**	-.00	.11	.07	-.02	.02	-.03	-.00	.02
v31	.01	-.05	-.01	.01	-.10	**.98**	.02	.05	.07	.09	-.01	-.01	-.03	-.10
v32	-.22	.06	**-.52**	-.17	-.03	.06	.03	.13	.09	.09	.06	-.04	-.18	.03
v33	.25	.17	.22	.07	-.09	.05	.18	.15	.05	-.03	-.06	.02	**.48**	.08
v34	.09	.11	-.01	-.02	.01	.06	.05	**.40**	.12	-.02	-.03	-.01	.08	-.06

Table 5: Results for Logistic Regression over the 14 Factors (Z-score> 1.96 results in p-values< 0.05). Most relevant factors in bold lettering.

	β Coeff.	Z-Score	p-value
Intercept	**0.95**	**3.52**	**0.008**
F1	-0.392	-1.14	0.27
F2	-0.45	-1.84	0.53
F3	**-1.029**	**-3.72**	**0.005**
F4	-0.16	-0.66	0.45
F5	0.022	0.090	0.80
F6	0.47	0.58	0.18
F7	-0.61	-1.82	0.17
F8	**-0.61**	**-2.55**	**0.042**
F9	**0.22**	1.07	0.35
F10	**-0.50**	**-2.06**	**0.02**
F11	-0.31	-1.24	0.27
F12	-0.02	-0.09	0.52
F13	**-0.88**	**-3.11**	**<0.001**
F14	0.21	0.91	0.39

Table 6: 95% CI and Odds-Ratio

	CI-High	CI-Low	OR-High	OR-Low
Intercept	1.47	0.42	4.36	1.52
F3	-0.49	-1.57	0.61	0.21
F8	-0.14	-1.09	0.87	0.34
F10	-0.02	-0.98	0.98	0.37
F13	-0.32	-1.43	0.72	0.24

Table 7: β intervals for the 10-fold cross-validation experiment

Factor	MAX	MIN
Intercept	0.634	0.959
F3	-0.971	-0.509
F8	-0.790	-0.317
F10	-0.783	-0.361
F13	-0.711	-0.212

Table 8: Logistic Regression Results

	β Coeff.	z-score	p-value
Intercept	4.16	4.85	< 0.001
Num. Dysf. Org	-0.57	-2.05	0.03
APACHE II	-0.09	-2.15	0.03
Worst Lact.	-0.30	-2.49	< 0.001

Table 9: 95 % CI AND ODDS-RATIO

	CI-High	CI-Low	OR-High	OR-Low
Intercept	5.78	2.55	12.82	324
Dysf. Org	-0.04	-1.10	0.33	0.95
APACHE II	-0.01	-0.17	0.85	0.99
Worst Lact.	-0.10	-0.51	0.60	0.91