33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# Severe Sepsis Mortality Prediction with Relevance Vector Machines

Vicent J. Ribas
SOCO Research Group
Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Barcelona - Spain.

Juan Carlos Ruiz-Rodríguez *MD*
Servei de Medicina Intensiva
SODIR Research Group
Hospital Universitari Vall d'Hebrón
Vall d'Hebrón Institut de Recerca
Autonomous University of Barcelona.

Anna Wojdel
Research Associate
Llenguatges i Sistemes Informàtics
Sabirmedical S.L.
Barcelona - Spain.

Jesús Caballero López *MD*,
and Adolf Ruiz-Sanmartín *MD*
Servei de Medicina Intensiva
SODIR Research Group
Hospital Universitari Vall d'Hebrón
Vall d'Hebrón Institut de Recerca
Autonomous University of Barcelona.

Jordi Rello *MD-PhD*
Servei de Medicina Invensiva
CRIPS Research Group
Hospital Universitari Vall d'Hebrón
Autonomous University of Barcelona
Centro de Investigación Biomédica en Red
de Enfermedades Respiratorias (CIBERES).

Alfredo Vellido
SOCO Research Group
Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Barcelona - Spain.

***Abstract*—Sepsis is a transversal pathology and one of the main causes of death at the Intensive Care Unit (ICU). It has in fact become the tenth most common cause of death in western societies . Its mortality rates can reach up to 45.7% for septic shock, its most acute manifestation. For these reasons, the prediction of the mortality caused by sepsis is an open and relevant medical research challenge. This problem requires prediction methods that are robust and accurate, but also readily interpretable. This is paramount if they are to be used in the demanding context of real-time decision making at the ICU. In this brief paper, such a method is presented. It is based on a variant of the well-known support vector machine (SVM) model and provides an automated ranking of relevance of the mortality predictors. The reported results show that it outperforms in terms of accuracy alternative techniques currently in use, while simultaneously assessing the relative impact of individual pathology indicators.**

## I. 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 multiorgan 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].

The medical management of sepsis is therefore a serious challenge to healthcare systems and the prediction of the mortality caused by the acutest forms of the pathology is a relevant medical research challenge. Not a closed one, though, because, although SIRS yields a good sensitivity in the prediction, it lacks specificity [4].

This problem still requires the investigation on prediction methods that are robust and accurate, but also readily interpretable. Interpretability is paramount if they are to be used by medical experts in the demanding context of real-time decision making at the ICU. We present such a method in this paper. It is based on a variant of the well-known and widely used support vector machines (SVM), namely, the relevance vector machine (RVM) [5]. It shares with SVM the good performance in terms of accuracy, but it also provides an automated ranking of relevance of the mortality predictors. This is not just an *a posteriori* ranking, but part of the model *training* itself. That is, as the model is built, the adaptive weights corresponding to irrelevant features (mortality predictors) are updated as to minimize the impact of these features on the prediction, in an automatic procedure.

In this paper, the performance of RVM as an ICU Sepsis Mortality Predictor is compared to that of alternative techniques currently in use for ICU-related prediction, such as shrinkage methods for logistic regression and a risk-of-death (ROD) formula based on the standard APACHE II score [6]. The proposed model is shown to outperform these techniques, while simultaneously assessing the relative impact of individual indicators of the pathology on the prediction. Interestingly, only a reduced number of these indicators, which are also readily interpretable, are shown to have an impact on mortality prediction. We believe that this is a result that should help to simplify the decision making process at the ICU.

## II. Materials

This work is based on a prospective observational cohort study of adult patients with severe sepsis. The study was conducted at the Critical Care Department of the Vall d'

Hebron University Hospital (Barcelona, Spain), and it was approved by the Research Ethics Committee of the Hospital. The database consists of data from patients with severe sepsis, collected at the ICU by the Research Group in Shock, Organic Dysfunction and Resuscitation (SODIR),between June, 2007 and December, 2010. During this period, 354 patients with severe sepsis (medical and surgical patients) were admitted in the ICU.

The mean age of the patients in the database was 57.08 (with standard deviation $\pm 16.65$) years; 40% of patients were female and the diagnosis on admission was 56.15% *medical* and 44.85% *surgical*. The origin of primary infection for the cases on the database was 40.24% pulmonary, 23.17% abdominal, 10.75% urinary, 7.21% skin/muscle, 4.88% central nervous system (CNS), 1.55% catheter related, 1.00% endovascular, 2.22% biliar, 4.99% mediastinum and 3.99% 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 through the SOFA score system [7], which objectively measures organ dysfunction for 6 organs/systems, the details of which are provided in Table I. Severity was evaluated by means of the APACHE II score (for further reference, see [6]). The APACHE II score was $23.03 \pm 9.62$ for the population under study.

TABLE I
LIST OF SOFA SCORES, WITH THEIR CORRESPONDING MEAN AND STANDARD DEVIATION VALUES FOR THE POPULATION UNDER STUDY (SCORING ORGAN DYSFUNCTION).

| | |
|---|---|
| Cardiovascular (CV) | 2.86 (1.62) |
| Respiratory (RESP) | 2.31 (1.15) |
| Central Nerv. Sys. (CNS) | 0.48 (1.00) |
| Hepatic (HEPA) | 0.48 (0.92) |
| Renal (REN) | 1.06 (1.20) |
| Haematologic (HAEMATO) | 0.78 (1.14) |
| Global SOFA score | 7.94 (3.86) |
| Dysf. Organs (SOFA 1-2) | 1.68 (1.09) |
| Failure Organs (SOFA 3-4) | 1.51 (1.02) |
| Total Dysf. Organs | 3.18 (1.32) |

In 2004, the Surviving Sepsis Campaign (SSC) defined a set of guidelines for the management of severe sepsis and septic shock [8]. More specifically, these set of guidelines were proposed for both the first 6 hours of evolution (resuscitation) and the first 24 hours (treatment). The compliance of the SSC bundles for the first 6 hours was 28.64%, out of which 77.89% had haemocultures performed, 83.41% received antibiotics, 57.05% had their lactate monitored, 59.04% received volume (i.e. fluid resuscitation) 16.33% received transfusions and 5.02% received dobutamine. The $SvcO_2$ values were $34.97 \pm 36.60$ and the haematocrit $28.07 \pm 12.48$ for the first 6 hours. The compliance of the first 24 hour SSC bundles was 49.75%, the glycaemia was $< 150$ mg/dL in 56.28% of cases and plateau pressure (PPlateau) $< 30$ cm $H_2O$ in 44.23% of cases. The mortality rate intra-ICU for our study population was 29.44%.

The specific set of 34 features used for the mortality prediction analyses in this study are listed in Table II.

TABLE II
LIST OF FEATURES 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 | Hepatic 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 | Transfusions 6h |
| v29 | Dobutamine 6h |
| v30 | Surviving Sepsis Campaign Bundles 24h |
| v31 | Glycaemia 24h |
| v32 | PPlateau |
| v33 | Worst Lactate |
| v34 | $O_2$ Central Venous Saturation |

## III. METHODS

### A. Relevance Vector Machines

The general regression problem posed by RVM can be written as [9], [5]:

$$y = w^T \psi(x), \tag{1}$$

where $\psi(x)$ is a basis function. In order to estimate the weights $w$ from our training examples, it is assumed that each target $t_i$ in the training sample (valued 1 for survival and -1 for exitus in the current study) represents the true model $y_i$ contaminated by i.i.d Gaussian noise $\epsilon_i \sim N(0, \sigma^2)$, so that, $\forall i$ [10]:

$$t_i = w^T \psi(x_i) + \epsilon_i \tag{2}$$

Therefore,

$$p(t_i \mid x_i, w, \sigma^2) \sim N(y_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2}\left(t_i - w^T\psi(x_i)\right)^2\right) \tag{3}$$

For the $N$ training points,

$$p(t \mid x_i, w, \sigma^2) = \prod_{i=1}^{N} N(w^T \psi(x_i), \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{1}{2\sigma^2}\|t - \Psi w\|\right), \tag{4}$$

where $t$ is the vector of training targets $t_i$, the $N \times M$ matrix $\Psi$ is built so that the $i^{th}$ row represents vector $\psi(x_i)$.

The growth of weights $w$ can be constrained by defining an *explicit* prior probability distribution on $w$. Therefore, assuming a Gaussian distribution on $w$, and defining $S = sI$ as the hyperparameter matrix where $I$ is $N \times N$ identity matrix and $S = [s_1, \ldots, s_N]$ is a vector where each $s_i$ describes the inverse variance for each $w_i$. For each weight, the hyperparameter $s_i$ modifies the strength of the prior.

The posterior probability over the unknown parameters is defined as:

$$p(w, s, \sigma^2 \mid t) = p\left(w \mid t, s, \sigma^2\right) p\left(s, \sigma^2 \mid t\right)$$

$$p\left(w \mid t, s, \sigma^2\right) = \frac{|\Sigma|^{1/2}}{(2\pi)^{N/2}} \exp\left(\frac{-1}{2}\left(w - \mu\right)^T \Sigma^{-1} \left(w - \mu\right)\right), \tag{5}$$

where $\Sigma = \left(\frac{1}{\sigma^2}\Psi^T\Psi + S\right)^{-1}$ and $\mu = \frac{1}{\sigma^2}\Sigma\Psi t$. To estimate $\mu$ and $\Sigma$, we need to maximize the evidence:

$$p(t \mid s, \sigma^2) = \int p\left(t \mid w, \sigma^{-2}\right) p\left(w \mid s\right) dw \tag{6}$$

Assuming uniform hyperpriors and expanding eq.6, it is possible to calculate the following marginal likelihood function:

$$\ln p(t \mid s, \sigma^{-2}) = \frac{1}{2}\sum_{i=1}^{M} \ln s_i - \frac{N}{2}\left(\ln \sigma^{-2} + \ln(2\pi)\right) - \frac{1}{2}\left(\sigma^{-2} t^T t - \mu^T \Sigma^{-1} \mu + \ln|\Sigma|\right), \tag{7}$$

which has to be maximized w.r.t. $\sigma^{-2}$ and $s$.

It is important to note that during the iterative process of eq. 7 maximization, some $s_i$ may tend towards infinity, which entails $\lim_{s_i \to \infty} \Sigma = 0$ and $\lim_{s_i \to \infty} \mu = 0$. In this situation, some $w_i$ will take values close to zero, which means that the adaptive effect of the hyperparameters will effectively *switch off* those input features that are deemed to be irrelevant for the prediction. This is, in fact, a form of *soft* feature selection, or, more precisely, a form of *automatic relevance determination*.

## IV. Results

### A. Mortality Prediction with RVM

The model performance was evaluated by means of 10-Fold Cross-Validation. The RVM yielded an accuracy of mortality prediction of 0.80 as measured by the area under the ROC plot (AUC); a prediction error of 0.24; a sensitivity (proportion of correctly predicted survivors out of all survivors) of 0.66; and a specificity (proportion of correctly predicted exitus out of all exitus) of 0.80.

Beyond classification accuracy, and as described in the previous section, RVM performs soft feature selection through automatic feature relevance determination. The following relevance vector (with the weights associated to each input feature) was obtained:

- Number of dysfunctioning organs ($w_1 = -0.039$)
- Mechanical Ventilation ($w_2 = -0.101$)
- APACHE II ($w_3 = -0.337$)
- Resuscitation Bundles (6h) ($w_4 = 0.037$)

The coefficients corresponding to the rest of features were set to values close to zero as part of the training process. This effectively reduces the complexity of the prediction procedure (34 features reduced to just 4) and improves its interpretability. Given that a linear basis function has been used to estimate the relevance vector, it becomes apparent that the negative weights (number of dysfunctioning organs, mechanical ventilation, APACHE II) are related to a higher mortality risk (note again that we have coded survival as 1 and exitus as -1), whereas the SSC bundles (resuscitation bundles) are associated to a protective effect (i.e. antibiotics administration, performance of haemocultures, administration of volume and vasoactive drugs and so on). In fact, timely administration of antibiotics and performance of haemocultures are considered critical to improving the prognosis of septic patients. Equally important is the knowledge of which features are deemed *not* to be relevant by RVM.

### B. Comparison with Shrinkage Feature Selection Methods for Logistic Regression

The predictive ability of the RVM was then compared to that of other well established shrinkage methods for logistic regression. More particularly, we have tested the performance against Ridge Regression, the Lasso and Logistic Regression. The latter using a subset of features selected in a backward process by removing those coefficient yielding the lowest z-scores [11]. The selected features and coefficients for each method were:

- Ridge Regression:
  - Number of dysfunctioning organs for SOFA 3-4 ($w_1 = -0.021$)
  - APACHE II ($w_2 = -0.127$)
  - Worst Lactate ($w_3 = -0.126$).
- Lasso:
  - Age ($w_1 = 0.007$)
  - Germ Class ($w_2 = 0.005$)
  - $PaO_2/FiO_2$ ($w_3 = 0.001$)
  - APACHE II ($w_4 = -0.006$)
  - $SvcO_2$ 6h ($w_5 = -0.001$)
  - Haematocrit 6h ($w_6 = 0.009$)
  - Worst Lactate ($w_7 = -0.023$)
  - $SvcO_2$ ($w_8 = -0.006$).
- Logistic Regression with backward feature selection:
  - Intercept ($w_1 = 4.16$)
  - Number of Dysfunctioning Organs ($w_1 = -0.57$)
  - APACHE II ($w_2 = -0.09$)
  - Worst Lactate ($w_3 = -0.30$)

The three shrinkage methods evaluated in this section agreed in detecting as prognostic factors the Severity measured by the APACHE II score and acidosis measured by the

lactate levels. Apart from that, it becomes apparent that organ dysfunction and mechanical ventilation or other parameters related to it like $PaO_2/FiO_2$ also play a role in the prognosis of Sepsis. Table III shows the results of AUC, Error Rate, Sensitivity and Specificity for each method.

TABLE III
RESULTS FOR SHRINKAGE METHODS

| Method | AUC | Error Rate | Sens. | Spec. |
|---|---|---|---|---|
| RVM | 0.80 | 0.24 | 0.66 | 0.80 |
| Logistic | 0.77 | 0.27 | 0.66 | 0.76 |
| Ridge | 0.69 | 0.28 | 0.67 | 0.73 |
| Lasso | 0.70 | 0.32 | 0.67 | 0.68 |

### C. Comparison with the APACHE II Mortality Score

The Risk-of-Death (ROD) formula based on the APACHE II score can be expressed as [6]:

$$\ln(\frac{ROD}{1-ROD}) = -3.517 + 0.146 \cdot A + \epsilon \qquad (8)$$

where $A$ is the APACHE II score and $\epsilon$ is a correction factor depending on clinical traits at admission in the ICU. For instance, if the patient has undergone post-emergency surgery, $\epsilon$ is set to $0.613$. The application of this ROD formula to the population under study yields an error rate of 0.28 (higher than RVM), a sensitivity of 0.55 (very low) and a specificity of 0.80. The AUC was 0.70 (lower than RVM). Previous studies [12] reported very similar results.

## V. CONCLUSIONS

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 IV-C and similar ones reported in other studies [12], the ROD formula presented in [6] is poor in terms of sensitivity (i.e., it results in a high number of false negative cases). This is despite the fact that APACHE is widely accepted in practice and yields acceptable accuracy results. Its poor sensitivity may be the result of its formula being based on non-sepsis specific clinical traits and the APACHE II score only.

In this paper we have put forward an RVM-based method for the prediction of ROD in septic patients. It has been shown to produce accurate results, more particularly in terms of specificity, while improving the interpretability and actionability of the results through an embedded feature relevance determination process. This method has proven to be superior in terms of accuracy (error rate, specificity and AUC) than other well established shrinkage methods (Lasso and Ridge). Specifically from a medical viewpoint, the strength of this study lies in the fact that it shows that it is possible to derive a reliable prognostic score from a parsimonious set of physiopathologic and therapeutic variables, which are available at the onset of severe sepsis for medical experts at the ICU.

The proposed method may be understood as a generalization of the ROD formula introduced in [6], where the $\epsilon$ corrective factor, which models clinical traits at admittance in the ICU, is accounted for by the rest of attributes obtained with RVM. It takes not only the contribution of the APACHE II score into consideration, but also other important life-threatening clinical traits such as the number of dysfunctioning organs combined with mechanical ventilation (RVM) or worst lactate levels (Shrinkage methods). The prognosis indicator is also balanced with important procedures to overcome sepsis such as the administration of volume, antibiotics, vasoactive drugs and the performance of haemocultures (i.e. SSC Resuscitation Bundles).

The performance of the proposed method has been evaluated in a single ICU and a limited population sample. Future work should lead towards a multi-centric prospective study, in order to validate its generalizability.

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

[5] Tipping M., "Sparse Bayesian Learning and the Relevance Vector Machine". Journal of Machine Learning Research 2001; 1:211-244.

[6] 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.

[7] Vincent J. L., Moreno R., Takala J., Willats S., De Mendoca 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.

[8] 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.

[9] Shölkopf B., Smola A.J, "Learning with Kernels", The MIT Press, 2002.

[10] Fletcher T. "Relevance Vector Machines Explained", http://www.tristanfletcher.co.uk/RVM%20Explained.pdf.

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

[12] 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.