Development of a Model That Uses Data Obtained in the Admission to Predict One-Year Mortality in Patients with Sepsis in the Intensive Care Unit

Javier E. Garc á-Gallo[1], Nelson J. Fonseca-Ruiz[2], and John F. Duitama-Mu ñoz[1]
[1]Engineering and Software Investigation Group, Calle 70 No. 52-21,Universidad de Antioquia UdeA, Medell ń, Colombia
[2]Critical and Intensive Care Program, Calle 10A # 22-04, Universidad CES Medell ń, Colombia
Email: {jesteban.garcia, john.duitama}@udea.edu.co, nfonseca@ces.edu.co

***Abstract*—Sepsis is a syndrome that occurs with physiological and biochemical abnormalities induced by severe infection and carries a high mortality and morbidity, therefore the severity of its condition must be interpreted quickly. This study presents the development of a model for the one-year mortality prediction of the patients that are admitted in a ICU with a sepsis diagnosis. 5650 patients extracted from the MIMIC III database (divided in 70% for training and 30% for validation) were evaluated and predictors available from the ICU admission was used to develop a mortality prognosis prediction model based on Bayesian Additive Regression Trees (BART) methodology. Variable importance is also presented. In order to evaluate the predictive power of the model, we used the 1695 admissions of the validation subset, and obtained an area under the Receiver Operating Characteristic curve (AUROC) of 0.7354 (95% Confidence Interval (CI): [0.7118-0.7589]). The presented model outperform the results obtained with Sequential Organ Failure Assessment (SOFA), Oxford Acute Severity of Illness Score (OASIS) and Simplified Acute Physiology Score II (SAPSII) indicators on the same validation subset. Our approach demonstrates the importance of comorbidities for the long-term mortality in patients with sepsis in the ICU and shows that it is possible to obtain a model with adequate predictive capacity from the moment of the admission of a patient.**

***Index Terms*—bayesian additive regression trees, prognosis prediction, sepsis, intensive care unit**

I. Introduction

Sepsis is a syndrome of physiologic, pathologic, and biochemical abnormalities induced by infection. The reported incidence of sepsis is increasing, due aging populations with more comorbidities and greater recognition. Sepsis is a major health problem worldwide, associated with high mortality rates in all countries [1]-[3].

There is increasing awareness that patients who survive sepsis often have long-term physical, psychological, and cognitive disabilities with significant health care and social implications [4]-[7]. For these reasons, strategies that allow the early identification of a patient's poor prognosis are needed.

The model presented in this work is based on data that is available from the beginning of a sepsis related ICU stay, and perform the estimation of the one-year mortality prediction of patients admitted to an ICU with a sepsis diagnosis.

Bayesian Additive Regression Trees (BART) [8] were used to develop the model. BART is an ensemble model that consists of two parts: a sum-of-trees model and a regularization prior on the parameters of that model.

In this study 5650 admissions from MIMIC III (Medical Information Mart for Intensive Care) [9] database were used, which were divided into two groups: 70% for training and 30% for validation.

Model discrimination was examined using area under the receiver operator curve (AUROC), goodness of fit was evaluated by Pearson's chi-square and calibration was assessed using the Hosmer-Lemeshow test.

In order to evaluate the model, one-year mortality predictions were made based on three severity-of-disease classification systems: Sequential Organ Failure Assessment (SOFA) [10], Oxford Acute Severity of Illness Score (OASIS) [11] and Simplified Acute Physiology Score II (SAPSII) [12].

II. Methods

A model that allows the estimation of the one-year mortality of a sepsis patient from the beginning the stay, could be part of a useful tool that help to improve the prognosis of patients with sepsis admitted to an ICU.

A. Sepsis Criteria

Traditionally sepsis was considered a condition resulted from a host's systemic inflammatory response syndrome (SIRS) to infection. When organ dysfunction occurred, it was considered severe sepsis, a condition that, if aggravated, could turn into septic shock, defined as "sepsis-induced hypotension persisting despite adequate

Manuscript received August 21, 2018; revised December 17, 2018.

doi: 10.18178/ijpmbs.8.1.12-16

fluid resuscitation." [13], [14]. Table I presents the summary definitions.

TABLE I. DEFINITIONS FOR SYSTEMIC INFLAMMATORY RESPONSE SYNDROME (SIRS), SEPSIS, SEVERE SEPSIS, AND SEPTIC SHOCK.

Term	Definition
Systemic inflammatory response syndrome (SIRS)	Two or more of the following criteria: *Body temperature ≥38°C or <36°C *Heart rate > 90 bpm *Respiratory rate > 20 bpm or $PaCO_2$ < 32 mmHg *White blood cell count >12.0x10^9/L or <4.0x10^9/L or >10% immature band forms
Sepsis	SIRS + Infection
Severe sepsis	Sepsis + Organ Dysfunction
Septic Shock	Sepsis with arterial hypotension despite adequate fluid replacement

However, in The Third International Consensus Definitions for Sepsis and Septic Shock a task force developed new definitions that incorporate the current understanding of sepsis biology. Defining sepsis as a "life-threatening organ dysfunction caused by a dysregulated host response to infection"; and Organ dysfunction as an acute change in total Sequential Organ Failure Assessment (SOFA) score ≥ 2 points consequent to the infection [1]. In spite of the rigor of the methodology used by the consensus, currently, there remains some controversy around the new definitions, since the new definitions did not involve low or middle income countries, and SOFA is a score that is routinely calculated in some, but not all, ICUs [15]-[17]. Therefore, in order to follow the definition of the consensus, but without forgetting the doubts regarding the new way of doing the diagnosis, we used the Angus criteria [18] in this study to identify the patients with sepsis, which diagnoses organ dysfunction according to the codes in Table II.

TABLE II. INTERNATIONAL CLASSIFICATION OF DISEASES (NINTH REVISION) BASED CLASSIFICATION OF ACUTE ORGAN DYSFUNCTION.

Organ system	ICD-9-CM Code description	ICD-9-CM Code
Cardiovascular	Shock without trauma	785.5
	Hypotension	458
Respiratory	Mechanical ventilationa	96.7
Neurologic	Encephalopathy	348.3
	Transient organic psychosis	293
	Anoxic brain damage	348.1
Hematologic	Secondary thrombocytopenia	287.4
	Thrombocytopenia unspecified	287.5
	Other/unspecified coagulation defect	286.9
	Defibrination syndrome	286.6
Hepatic	Acute and subacute necrosis of liver	570
	Hepatic infarction	573.4
Renal	Acute renal failure	584

B. Working Dataset

For this study we used 5650 admissions (with a one-year mortality rate of 43.3%) from MIMIC III [9] (Medical Information Mart for Intensive Care), an open database, that provides demographic information, vital signs measures, laboratory test results, drug information, procedures, fluid balance, length of stay and mortality (both inside and outside the medical center).

As predictors we use the variables listed in Table III. These variables are available at the admission of the patients allowing the presented model to be applied from the beginning of the ICU stay.

The patients including in this cohort have a median age of 67.54 years, a median ICU stay of 5.6 days and a one-year mortality rate of 43.3%. The 5650 admissions were randomly divided into two groups: a train subset with 3955 admissions (70% of the working set), and a validation subset of 1695 admissions.

TABLE III. EXTRACTED DATA FROM EACH ADMISSION

Parameter	Unit
DATA TAKEN AT THE TIME OF ICU ADMISSION	
Gender	Female, Male
ICU Type	Medical, Scheduled Surgical, Unscheduled Surgical
Admission Type	Elective, Urgent, Emergency
Age	Years
Glasgow Coma Scale (GCS)	Integer 3-15
COMORBIDITIES	
Diabetes	Binary (Presence)
Immunosuppressive diseases	
AIDS	
Hypothyroidism	
Malignancy	
Metastatic Cancer	
Heart failure	
Pulmonary diseases	
Vascular diseases	
Coronary diseases	
Obesity	
Alcohol abuse	
Collagen diseases	
Drug abuse	
Malnutrition	

III. MODEL DEVELOPMENT

As other models, BART makes an inference about an unknown function f that maps a p dimensional vector of inputs $x = (x_1, \ldots, x_p)$ to Y. BART approximates Y by a sum of trees T each one of those have of a set of

parameter values associated with each of the b terminal nodes $M = \{\mu_1, \mu_2, \dots, \mu_b\}$.

$$Y = \sum_{j=1}^{m} g(x; T_j, M_j) + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2)$$

Where T_j is the jth tree consisting of a set of decision rules (associated with its interior nodes) and a set of terminal nodes, m is the total number of trees, and $\sigma^2$is the variance of noise.

In BART a prior is specified, a likelihood is defined using the data, and then a sequence of draws from the posterior using Markov chain Monte Carlo (MCMC) is obtained. For a complete illustration about the implementation, see [8].

BART can be used to identify which components are more important for explaining the variation of Y. Such variable selection information is model-free in the sense that it is not based on the usual assumption of an encompassing parametric model [8], [19].

When BART is being used in a binary classification problem, the output of interest is $Y(= 0 \; or \; 1)$, the probit model setup is used.

$$p(x) \equiv P[Y = 1|x] = \Phi[G(x)]$$

Where

$$G(x) = \sum_{j=1}^{m} g(x; T_j, M_j)$$

$\Phi[\cdot]$ is the standard normal cumulative distribution function and each classification probability $p(x)$ is obtained as a function of a sum of regression trees $G(x)$.

The performance of the developed model was evaluated using the area under the Receiver Operating Characteristic curve (AUROC) that is a common indicator of the goodness of a predictor in a binary classification task; The AUROC were compared with three commonly used severity-of-disease classification systems: SOFA [10], SAPS2 [12] and OASIS [11]. Goodness of fit and calibration of the developed model was assessed using Pearson's Chi-square Test and Hosmer-Lemeshow test respectively.

The Pearson's chi-square [20] measures the discrepancy between an observed and a predicted distribution. The test statistic for a binary problem is given by:

$$\chi^2 = \frac{(O_s - E_s)^2}{E_s} + \frac{(O_f - E_f)^2}{E_f}$$

Where O_s and O_f are the observed number of successes and failures, E_s and E_f the predicted successes and failures in all the observations.

The Hosmer–Lemeshow test [21] seeks to prove that a model fits the data, and it is a chi-square test conducted by sorting the n observations in the data set by estimated probability of success, dividing the sorted set into g groups and assessing the Hosmer–Lemeshow C statistic:

$$\hat{C}_g = \sum_{i=1}^{g} \left[\frac{(O_{s,i} - E_{s,i})^2}{E_{s,i}} + \frac{(O_{f,i} - E_{f,i})^2}{E_{f,i}} \right]$$

Where $O_{s,i}$ and $O_{f,i}$ are the observed number of successes and failures, $E_{s,i}$ and $E_{f,i}$ the predicted successes and failures in the ith group

The number of groups g is defined by the user, however, Paul et al. [21] made some recommendations to select the number of groups. Specifically for samples sizes between 1000 and 25000 g is given by:

$$g = max\left(10, min\left\{\frac{m}{2}, \frac{n-m}{2}, 2 + 8\left(\frac{n}{1000}\right)^2\right\}\right)$$

Where n is the number of observations, and m is the number of successes.

The chi-squared statistic for the binary classification problem and the $\hat{C}_g$ can be used to calculate a p-value with one and $g - 2$ degrees of freedom respectively. Additionally, observed versus predicted numbers of deaths were compared graphically within deciles of increasing probability of one-year mortality.

IV. Results

To estimate the one-year mortality at the admission moment of patients admitted in a ICU with a sepsis diagnostic, predictors reported in tables 3 were used. Categorical variables were binarized using one hot encoding. The BART model was implemented with the "bartMachine" R-package [22], in R software; for classification, this "bartMachine" implementation require two parameters, the number of trees m and the prior probability that E(Y|X) is between (-3,3),k; this parameter is associated with the shrinkage.

Many models with different values of k and m were evaluated over the train subset, and the parameter that presented a better performance were k=3 and m=50, with an AUROC on the 3955 admissions of the training subset of 0.7286. In order to evaluate the predictive power of the developed model, the 1695 admissions of the validation subset were used. An AUROC of 0.7348 (95% Confidence Interval (CI): [0.7111-0.7585]) were obtained.

The general effect on the model of each predictor was calculated using the inclusion proportions [22], which when scaled (so that the sum of the contributions of all the predictors is one) represents the variable importance (Fig. 1).

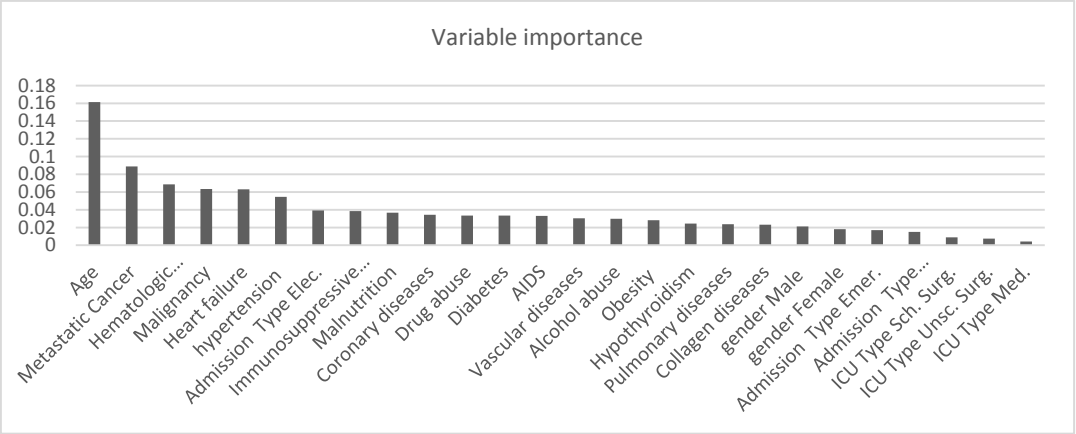

Figure 1. Variable importance measure.

Goodness of fit and calibration of the model was evaluated using Pearson's Chi-square Test and Hosmer–Lemeshow (HL) Test (with g=25) respectively. The p-value of the Pearson's Chi-square Test was 0.36 The HL p-value was 0.23 suggesting that the model produces probabilities that reflect the true mortality experience of the data, thus, the model is well calibrated.

Observed versus predicted of numbers of one-year deaths were compared graphically within deciles of increasing probability of the outcome based on the developed model (Fig. 2).

To benchmark the developed model, the AUROC of three severity-of-disease classification systems was calculated. The AUROC values for SAPS2, SOFA and OASIS scores on the validation subset were 0.724, 0.614 and 0.661 respectively.

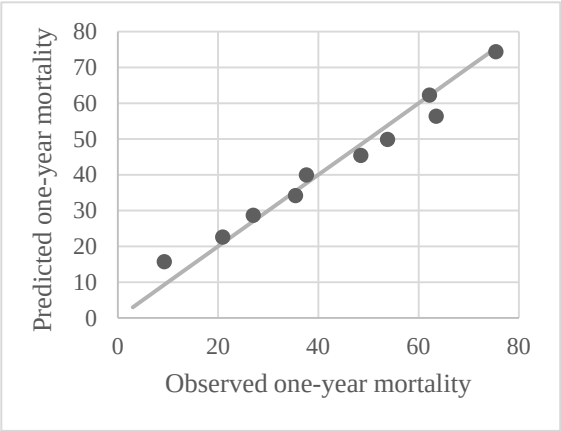

Figure 2. Comparison of observed versus predicted one-year mortality by deciles of estimated mortality. The solid line is the 45-degree line, and it represents a perfect match.

V. Conclusions

Sepsis is a syndrome that carries a high mortality, 43.3% for the admissions included in this study; therefore, tools that help clinicians to quickly predict a worse prognosis are needed.

AUROC evaluation over the validation data indicate that the developed model presents an adequate discrimination, similar to the SAPS2 performance and better than OASIS and SOFA; Which could lead to better management of illness within the ICU since the data on which the model is based could be taken at the admission moment, which means that from the beginning of a patient ICU stay the physician would have an estimate of the risk of one-mortality of the patient.

The presented model allows a faster stratification of patients than other usually used scoring systems, which are generally applied from the data of the first 24 hours.

As expected, older patients are at greater risk in consequence the most important parameter for the outcome is the age. Cancer in patients is also a strong indicator of worse prognosis. The ICU and admission types do not influence the outcome as much as the comorbidities.

The graph presented in Fig. 2 indicate that estimated and observed mortality pairs are similar and shows that the number of outcome events is indeed increasing along the probability deciles.

The main objective of this work is to present a model for the one-year mortality prediction of the patients that are admitted in a ICU with a sepsis diagnosis; and shows that the use of ensemble based algorithms (BART in this study) allows an adequate early mortality estimation.

Future works include the inclusion of later mortality estimates (with data from the first 24 hours of ICU stay), which is expected to improve the predictive capacity.

Acknowledgments

We would like to thank Dr. Leo A. Celi, MD for his support. This work if part of an ongoing doctorate thesis, which is possible thanks to the Scholarship Program No. 647 of COLCIENCIAS.

References

[1] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, *et al.*, "The third international consensus definitions for sepsis and septic shock (sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801-810, 2016.

[2] J. L. Vincent, J. C. Marshall, S. A. Ñamendys-Silva, B. Fran ̧ois, I. M. Loeches, J. Lipman, *et al.*, "Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit," *Lancet Respir. Med.*, vol. 2, no. 5, pp. 380-386, 2014.

[3] C. Fleischmann, A. Scherag, N. K. J. Adhikari, C. S. Hartog, T. Tsaganos, P. Schlattmann, *et al.*, "Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and

limitations," *Am. J. Respir. Crit. Care Med.*, vol. 193, no. 3, pp. 259-272, 2016.
[4] B. D. Winters, M. Eberlein, J. Leung, D. M. Needham, P. J. Pronovost, and J. E. Sevransky, "Long-term mortality and quality of life in sepsis: a systematic review," *Crit. Care Med.*, vol. 38, no. 5, pp. 1276-1283, 2010.
[5] S. Yende, S. Austin, A. Rhodes, S. Finfer, S. Opal, T. Thompson, *et al.*, "Long-term quality of life among survivors of severe sepsis: analyses of two international trials," *Crit. Care Med.*, vol. 44, no. 8, pp. 1461-1467, 2016.
[6] M. Shankar-Hari and G. D. Rubenfeld, "Understanding long-term outcomes following sepsis: Implications and challenges," *Curr. Infect. Dis. Rep.*, vol. 18, no. 11, p. 37, 2016.
[7] S. Yende and D. C. Angus, "Long-term outcomes from sepsis," *Curr. Infect. Dis. Rep.*, vol. 9, no. 5, pp. 382-386, 2007.
[8] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees," *Ann. Appl. Stat.*, vol. 4, no. 1, pp. 266-298, 2010.
[9] A. E. W. Johnson, T. J. Pollard, L. Shen, H. W. Li, M. L. Feng, M. Ghassemi, *et al.*, "MIMIC-III, a freely accessible critical care database.," *Sci. data*, vol. 3, 2016.
[10] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, *et al.*, "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," *Intensive Care Med.*, vol. 22, no. 7, pp. 707-710, 1996.
[11] A. E. W. Johnson, A. A. Kramer, and G. D. Clifford, "A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy," *Crit. Care Med.*, vol. 41, no. 7, pp. 1711-1718, 2013.
[12] J. R. Le Gall, S. Lemeshow, and F. Saulnier, "Simplified acute physiology score (SAPS II) based on a european/north american multicenter study," *JAMA*, vol. 270, pp. 2957-2963, 1993.
[13] S. R. Bennett, "Sepsis in the intensive care unit," *Surg.*, vol. 33, no. 11, pp. 565-571, 2015.
[14] W. J. Sibbald, G. Doig, and K. J. Inman, "Sepsis, SIRS and infection," *Intensive Care Med.*, vol. 21, no. 4, pp. 299-301, 1995.
[15] M. Singer, "The new definitions of SEPSIS and SEPTIC SHOCK: What do they give us? An answer," *Med. Intensiva*, vol. 41, no. 1, pp. 41-43, 2017.
[16] A. Rodríguez, I. Martín-Loeches, and J. C. Yébenes, "New definition of sepsis and septic shock: What does it give us?," *Med. Intensiva*, vol. 41, pp. 8-10, 2016.
[17] C. C. Medicine, E. Society, I. C. Medicine, S. O. F. Assessment, and Q. Sofa, " A ESPAÑOLA Debemos asumir la nueva definicio campo de la cirugı Should we assume the new definition of sepsis in the surgical field ?," 2017.
[18] D. C. Angus, W. T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, and M. R. Pinsky, "Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care," *Crit. Care Med.*, vol. 29, no. 7, pp. 1303-1310, 2001.
[19] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "Bayesian additive regression trees-based spam detection for enhanced email privacy," in *2008 Third International Conference on Availability, Reliability and Security*, Barcelona, 2018, pp. 1044-1051.
[20] S. E. Fienberg, "The use of chi-squared statistics for categorical data problems," *J. R. Stat. Soc. Ser. B*, pp. 54-64, 1979.
[21] P. Paul, M. L. Pennell, and S. Lemeshow, "Standardizing the power of the Hosmer—Lemeshow goodness of fit test in large data sets," *Stat. Med.*, vol. 32, no. 1, pp. 67-80, 2013.
[22] A. Kapelner and J. Bleich, "Bartmachine: Machine learning with bayesian additive regression trees," *Journal of Statistical Software*, arXiv:1312.2171, 2013.

Javier E. García-Gallo is a PhD student at the Universidad de Antioquia, Medellín-Colombia. He received his master degree in biomedical engineering from the Escuela de Ingeniería de Antioquia and the UNIVERSIDAD CES in 2013. His research interests include biomedical software development and health data analysis.

Nelson J. Fonseca-Ruiz is presently working as a critical care doctor at Clínica Central Fundadores in Medellín-Colombia, is a professor at the UNIVERSIDAD CES, Colombia. He received his master's degree in clinical Epidemiology from the UNIVERSIDAD CES in 2009. He graduated as Critical Care Specialist from the Universidad Pontificia Bolivariana in 2001, as Anesthesiologist from the Universidad de Antioquia in 1999 and as Doctor of Medicine from the Universidad de Antioquia in 1995. His research activities focus on Critical Care Medicine.

John F. Duitama-Muñoz is associated professor at the Universidad de Antioquia, Medellín-Colombia. He received his PhD degree in Informatics from the Institut National des Télécommunications and the Université d'Evry - France in 2005. He obtained a DEA (diplôme d'études approfondies) from the Université Paris I in 2000. He received his master degree in systems engineer from the Universidad Nacional of Colombia in 1998, and graduated as Systems Engineer from the Universidad de Antioquia in 1989.