Predictive Modeling Of Sepsis In Adult ICU Patients

Philip H. Schroeder[1], Roman Wang[2], Yasasvini Puligundla[3], Catherine Sun[4], Mawulolo Ameko[1],
Christopher C. Moore[5], Laura E. Barnes[1]
Department of Systems & Information Engineering[1], Department of Computer Science[2], Computer Engineering[3],
Department of Statistics[4], Department of Medicine[5],
University of Virginia, Charlottesville, VA

Introduction: Sepsis is a life-threatening syndrome, in which a dysregulated host response to infection causes organ dysfunction. Sepsis is responsible for the longest, most expensive hospital stays in the US, is a leading cause of in-hospital mortality, and, for surviving patients, leads to increased risk of permanent organ damage, cognitive impairment, and physical disability. Early and targeted treatment has been shown to significantly improve sepsis outcomes. Predictive models have been used to improve care in critical care settings, such as the Intensive Care Unit (ICU), and can potentially be used for earlier detection of patients at risk of becoming septic, allowing for earlier treatment. The majority of previous efforts to develop predictive models for sepsis in ICU patients have used an outdated definition of sepsis based on systemic inflammatory response syndrome (SIRS) criteria, which is not clinically validated and misses at least 1 out of 8 cases. The most recent definition of sepsis, Sepsis-3, considers patients septic if their Sequential Organ Failure Assessment (SOFA) score increases by two or more points, coincident with or consequent to the combination of a blood culture request and antibiotic administration. The objective of this study was to use the Sepsis-3 definition to build a multivariable logistic regression model that predicts onset of sepsis in adult ICU patients.

Materials and Methods: The data was extracted from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC)-III database through PostGreSQL queries. All patients extracted were ≥ 18 years old. Patients were labeled as having sepsis if they had an infection and their SOFA score increased ≥ 2 points within a window of 48 hours before and 24 hours after time of infection. Infection was defined as administration of antibiotics between 24 hours prior to and 72 hours after blood culture acquisition. Sepsis onset was defined as the earliest point of antibiotic administration or blood culture acquisition. A logistic regression model was built to predict sepsis onset, using clinical laboratory and vital sign data from the MIMIC-III database as predictors. For patients who did not develop sepsis, predictor values were selected from a random 48-hour time window during the patient's ICU stay. For those who developed sepsis, a random time was selected for the patient within 48 to 6 hours prior to onset of sepsis, and the predictor values closest to that time were extracted for that patient. Missing values for septic patients were imputed using a "carry-forward" method, where the last known predictor value was used to impute the missing value, capped at 72 hours prior to onset of sepsis. Predictors with ≥ 50% of values missing were eliminated. Then, patients with ≥ 50% of predictor values missing were eliminated. To reduce variables, the best-first search method was used to perform a greedy search algorithm using forward selection. Model selection was performed using Bayesian information criterion. The model was trained and validated using 10-fold cross-validation. For each cross-validation, the threshold was optimized with the selected validation set, where the optimal threshold was defined as that which maximized the F2 score.

Results and Discussion: From the 19,358 patients extracted from MIMIC-III, 4,915 patients were identified as developing sepsis. The final model included the variables shown in Table 1. The optimal threshold was calculated as .40. At a specificity of .896 (false positive rate of .104), the model achieved a sensitivity of .688, with an area under the receiver operating characteristic curve (AUC) of .792.

Table 1. Logistic regression model coefficients

Features	Coefficients
Intercept	-1.81
Respiratory rate	0.70
Anion gap	0.48
Chloride	-0.58
Oxygen saturation	-0.70
Magnesium	-0.30
Glasgow Coma Score	-1.72

Conclusions: In this project, the Sepsis-3 definition was used to build a logistic regression model that predicts sepsis onset in adult ICU patients. The model performed well, with a low false positive rate and high sensitivity. By using only six features and incorporating predictor values that were recorded up to 72 hours prior to sepsis onset, this logistic regression model shows promise for the development of transparent and interpretable predictive models that can successfully identify sepsis early and allow for more timely treatment. Finally, logistic regression provides for a suitable baseline model, with which other machine learning methods can be compared in future work.