Survivability of ICU Patients with Severe Sepsis/Septic Shock

Jeremy B. Crowe - crowe.jb@gmail.com

March 13, 2017

1 Domain Background

1.1 Overview

Sepsis is a serious, life-threatening condition that is partially responsible for 6% of all deaths in the United States [1]. A septic condition is highly variable in nature. A patient may arrive at the ICU with sepsis or may be diagnosed during the stay. There is no diagnostic test to confirm or deny the existence of such a condition in a patient. It is diagnosed based on the clinical judgment of the physician. Regardless of the cause, the speed at which treatment is administered and the types of treatments are critical in stabilizing a patients blood pressure, body temperature, eliminate infection, and keep cardiac output stable. Due to the variable, and difficult nature of the condition a machine learning approach may be able to aid in predicting the severity or survivability of the condition and the effectiveness of potential treatments.

A septic condition is caused by an overwhelming immune response to an infection. White blood cells release an array of chemicals to fight the infection which triggers systemic inflammation, vasodilation, increased permeability of vessels, and intracellular fluid build up. These "leaky vessels" deplete the body of coagulation factors. The increased fluid build-up and decreased blood pressure result in a lack of oxygenation of tissue, known as shock.

If sepsis is not treated quickly or with enough direct care, multiple organ dysfunction can occur which can result in kidney failure, liver failure, heart failure, acute respiratory distress, etc. The speed at which treatment is administered and types of treatments are directly correlated with the survivability of this severe condition. The speed of treatment has been found to be more important than the age of the patient [2]. Each hour of delay in antimicrobial administration over the initial 6 hours is associated with an average decrease in the survival rate of 7.6% [3].

Certain treatments may be highly effective, therefore it is important to pay special attention to several things: the type of infection and type of antibiotics used, the blood pressure and whether or not vasopressors were used, the correct amount of IV fluid provided to the patient. The correct type of antibiotic and appropriate administering of vasopressors can reduce the chance of organ failure and mortality greatly [4].

1.2 Problem

A predictive model would be highly useful for ICU staff and physicians. It would aid in prioritizing patients that are considered "high risk" and as a simple way to stratify patients in clinical research.

Upon patient admittance and initial blood and microbiology tests patient survivability will be classified over a 30 day period. While survival rates are available long after 30 days, this is the most critical period for surviving a condition such as sepsis/septic shock.

Test results and vitals from the first 24 hours in the ICU will be used for consideration. Several class features will be considered: gender, bacterial culture classification, antibiotic treatment, etc. Several continuous features will also be considered: blood urea nitrogen levels, age, blood PH, blood lactate levels, creatinine levels, blood pressure, etc.

The survivability of sepsis is strongly correlated with the speed of treatment, the age of the patient, type of infection, and the general state of the condition. The state of the condition can be measured in a variety of ways. Respiratory rate, blood pressure, creatinine (kidneys failing), and lactate levels(anaerobic cellular respiration) strongly suggest the current stage of the condition. Upon arrival, these measurements can be used to classify a patient into a binary survival group.

1.3   Metrics

The metric that will be used to gauge the prediction quality is the standard classification accuracy. Classification accuracy is the total number of true positives and true negatives out of all datapoints. It is defined as follows:

$$\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{total datapoints}}$$

This metric is used because the low survivability of sepsis. False positives and false negatives erode the usefulness of further patient studies.

Caveat: Accuracy is not enough in a real world application if the outcome affects human life. Accuracy weighs misclassifications evenly. This is useful when looking at a problem mathematically. Consider the following however: A patient is given a high probability that their cancer is in remission and no further treatment is given. It was a false negative classification and the patient dies. This is much more important to consider than a false positive. Therefore these results are only used in stratifying patients for further studies, not as a real time aid in patient treatment.

Many patients die of sepsis however, around 30%. Due to this high ratio and the intended problem of stratifying patients for future studies, an accuracy score is useful. If used in a different application, it may be useful to select a model with lower accuracy that ensures very low rates of false negatives, a more dangerous false prediction.

2   Datasets and Inputs

For this study, the MIMIC-III dataset(https://mimic.physionet.org/) will be used. This dataset is freely accessible and suggested for use by Udacity. More than 40,000 individual patient entries will be utilized. 1184 patients were admitted with sepsis and thousands more diagnosed by the end of their stay. Most of the focus will be put on the patients admitted with sepsis since there is an associated time of admittance that will aid in finding more effective treatments. Diagnosis data is tied to ICD-9 billing codes that do not have a time associated.

It was required to take a short course to gain access to this data by demonstrating some understanding of law and ethics in dealing with this de-identified dataset.

The data was downloaded locally as CSV files. The provided conversion scripts were modified to load the entries into the remote Amazon RDS Postgres database. This database and server is

secured and is HIPAA compliant. An Amazon compute EC2 instance was also spun up to do most of the heavy lifting. Many of the tables from the database will be used but the most critical are the following:

1. **admissions**: the admissions table provides basic patient information on admission including the initial diagnosis, sex, insurance type, etc.

2. **diagnoses_icd**: this table provides all diagnoses assigned to a patient during a hospital stay. Unfortunately, this is assigned at the end and cannot be given a time of diagnosis. A sequence is provided, however.

3. **microbiologyevents**: this table provides information about whether or not an infection is present, how it was obtained, and when.

4. **labevents**: this provides measurements such as blood pressure, and information from blood tests.

5. **prescriptions**: drugs given to patient along with the time and amount.

6. **patients**: this provides basic information about the patient including date of birth, date of death (if applicable), gender, etc.

In this supervised learning problem, the target attribute must be calculated from the provided information. The goal is to predict survivability within 30 days of admittance. Within the patient table exists a date of death (dod), if applicable, that can be used along with admittance date to calculate the boolean survivability target attribute.

3   Solution

The solution to is to categorizing patients by survivability at admittance. This is calculated by considering: speed of first antibiotic dose, the age of the patient, type of infection, respiratory rate, blood pressure, creatinine levels, and lactate levels. A supervised learning model will be trained off of this information along with the calculated target value, which is the patient survival over 30 days from admittance.

As the project progresses it may require me to include more complex information such as blood pressure measurements and categorical data about whether vasopressors were used to combat low bp and more. Multiple models using different techniques will be developed, analyzed and a comparative analysis between them will be done with the best performing and most generalizable solution being selected.

4   Benchmark

There are few studies into the predictability of patient survival from severe sepsis. There are also many factors that are important in how the prediction is made. I will be focusing on the following research: "Predicting survival of patients with sepsis by use of regression and neural network models" by J.R. Flanagan, et al.

"Survival after sepsis was predicted with an accuracy of 80% by the NN model, which used only information collected at the time of the diagnosis of sepsis. The development of multiple organ failure after the diagnosis of sepsis was predicted accurately (81.5%) with either the MLR or the NN model. Both the MLR and the NN methods depended on the interpretation of a likelihood quantity, requiring the choice of a threshold to make a survival prediction. "[5]

While this study was done in 1996, there have been few follow-up studies. The few follow-up studies had a similar success rate with slightly different attributes and test periods. Due to the detail of this study, most focus will be put on the work of Flanagan et al. for the benchmark.

5   Evaluation

The evaluation metric that will be used to quantify the performance of both the benchmark model and the solution model will be classification accuracy. The benchmark study by J.R. Flanagan, et al. details the classification accuracy of septic patient survival over a one month period. This is the approach that will be used in the solution model. The results of the solution model and that of the benchmark will finally be compared. Several different solution models will be designed and compared in order to see the difference in accuracy in a complex feature set such as this.

Using standard classification accuracy is a great metric for quantifying performance. While the feature set is highly complex, the target attribute is boolean and easily assessed. Therefore computing the model's accuracy at classifying patients by boolean survivability is accurate, simple, and understandable.

6   Design

Due to the nature of the problem and the relatively large number of features that will likely be involved, the results from the following models will be compared: multi-layer neural network, Adaboost with a decision stump as its base estimator, decision forests, and extreme gradient boosting.

The data is spread across multiple database tables. For most algorithms, a large single entry per patient will be formed from this data. For reusability, a new table with this joined information will be created.

1. Explore data structures:

    (a) Download the sample dataset, 100 patients, from physionet.org
    (b) Load data into Postgres database locally
    (c) Explore the data for useful trends, table structure, and layouts using DBVisualizer. Allows graphical exploration of data.

2. Set up full database:

    (a) Download entire compressed CSV dataset from physionet.org (30GB)
    (b) Spin up an Amazon RDS Postgres database and EC2 Compute instance.
    (c) Modify provided scripts from the MIMIC-III MIT Github repository to load data into the remote RDS Postgres database

(d) Set up docker with all necessary components including SQL alchemy for easy querying and object handling using its ORM.

3. Restructure data for use in ML model:

   (a) Create new DB sepsis table w/ one entry per patient hospital stay
   (b) Gather useful attributes across multiple tables by patient ID and/or hospital stay ID
   (c) Fields will contain the following (Likely will change): ID, Gender, Age (calculated), infection type, antibiotic treatment, time until first dose (calculated), blood lactate levels(first measurement), creatinine levels(first measurement), respiratory rate(first measurement), etc.

4. Develop ML models:

   (a) Develop and test multi-layer neural network in TensorFlow and/or Keras.
   (b) Develop and test extreme gradient boosting using XGBoost
   (c) Develop and test Adaboost w/ decision stumps as base estimator
   (d) Develop and test random forest

5. Run ML models remotely:

   (a) Spin up docker container on EC2 compute instance
   (b) Pull code from repo
   (c) Run ML models and save output logs

6. Compare results and tweak hyperparameters for better results

7. Analyze and write up results

References

[1] C. for Disease Control, "Sepsis data report." https://www.cdc.gov/sepsis/datareports/index.html, August 18, 2016. Accessed: 02-14-2017.

[2] J. Lundberg, "Septic shock: an analysis of outcomes for patients with onset on hospital wards versus intensive care units.." https://www.ncbi.nlm.nih.gov/pubmed/9635649, 1998. Accessed: 02-14-2017.

[3] A. Kumar, "Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock.." https://www.ncbi.nlm.nih.gov/pubmed/16625125, 2006. Accessed: 02-14-2017.

[4] M. Anthony C. Gordon, "Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock." http://jamanetwork.com/journals/jama/fullarticle/2540403, 2016. Accessed: 02-14-2017.

[5] J. Flanagan, "Predicting survival of patients with sepsis by use of regression and neural network models.." https://www.ncbi.nlm.nih.gov/pubmed/10156949, 1996. Accessed: 02-14-2017.