Omar Battisha
Siva Bhavani
1/24/2020
MACS 33002: Project Proposal

Predicting Infection Patient Mortality from Vital Signs

Overview

Our goal is to use longitudinal data of an infected patient's vital signs from the first 24 hours of their hospitalization, to predict whether or not they will die within 30 days. We plan to base our mortality prediction model on a Recurrent Neural Network (RNN) with Long Short Term Memory Units (LSTM).

Introduction

Sepsis is defined as a dysregulated response to infection that results in life-threatening organ dysfunction often leading to death. Sepsis is responsible for over 20 billion dollars in healthcare spending and is the cause of over 250,000 deaths in the U.S. every year. Several methods have been used to prognosticate mortality in sepsis including eCART, SOFA and MEDS scores. However, most of these prognostic models use static measurements of vitals or laboratory data. Static measurements may not adequately capture the dynamic nature of sepsis. Instead, leveraging the widely available longitudinal data in the electronic health record (EHR) could result in much more sophisticated and accurate prognostic models. There is a paucity of research on utilizing this longitudinal data for prognosticating patients with infection.

Patient Population

Patients with infection were identified using the criteria published by Rhee et al., which require a blood culture order and at least four consecutive days of antibiotics (or antibiotics continued until one day prior to death or discharge), with the first day of antibiotics required to be a parenteral agent given within 48 hours before or after the blood culture order. In order to create a more homogenous population, the cohort was narrowed to patients who met Rhee criteria for infection

in the emergency department and received antibiotics within 24 hours of presentation, with presentation defined as time of first measured vital signs.

Significance

Patients have vastly different responses to infection, and very different outcomes. Current prognostication tools lack accuracy because they use static measurements (such as maximum heart rate and lowest blood pressure) to make predictions about mortality. Developing a model with longitudinal vital signs could better predict prognosis in patients with infection, and could serve as a tool for clinicians to make decisions in managing these complex patients.

Dataset

The derivation dataset will be gathered from 2,582 de-identified adult patients admitted to a tertiary medical center. There will be no identifying names, dates of birth, or hospitalization dates. The data will consist of time-varying and time-invariant data from the first 24 hours of hospitalization (see Data Structure below). The outcome our model will predict is 30-day in hospital mortality. We will validate the model in a separate test dataset with a similar number of de-identified adult patients.

Data structure

Patient	Hour of hospitalization	HR	RR	sBP	dBP	Temp	O2
1	0						
1	1						
1	2						
1	3						
1	4						
1	5						
1	6						

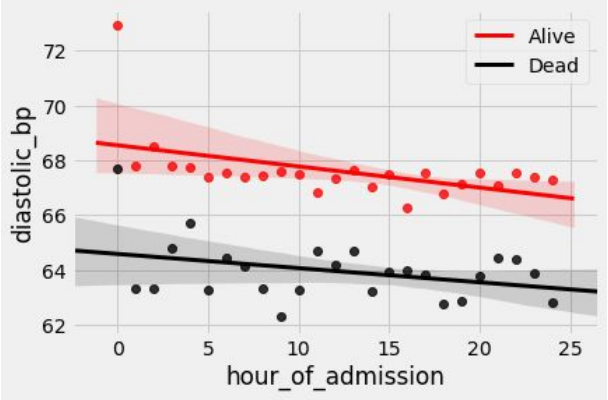
1	7						
1	8						
1	9						
1	10						
1	11						
1	12						
1	13						
1	14						
1	...						
1	24						

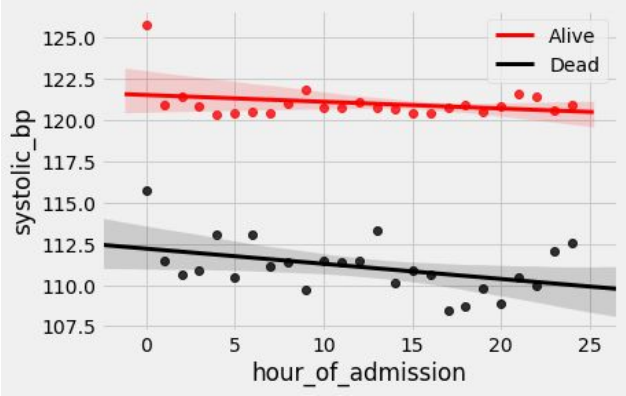
Definition of abbreviations: HR - heart rate, RR - respiratory rate, sBP - systolic blood pressure, dBP - diastolic blood pressure, Temp - temperature, O2 - oxygen saturation

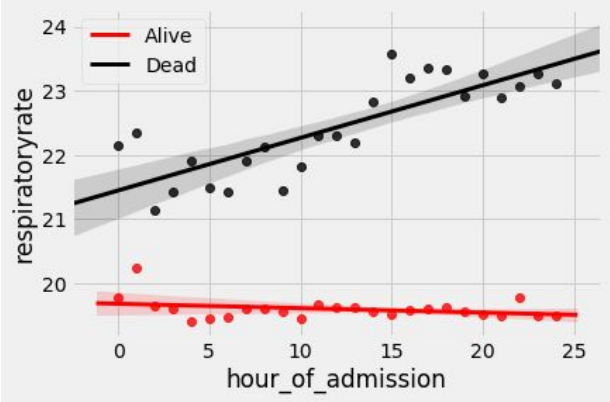
In addition to these time-varying measurements, patients will have the following information: age, presence of chronic diseases (heart failure, lung disease, liver disease, renal disease, diabetes mellitus, hypertension and cancer) and the outcome (mortality in 30 days).

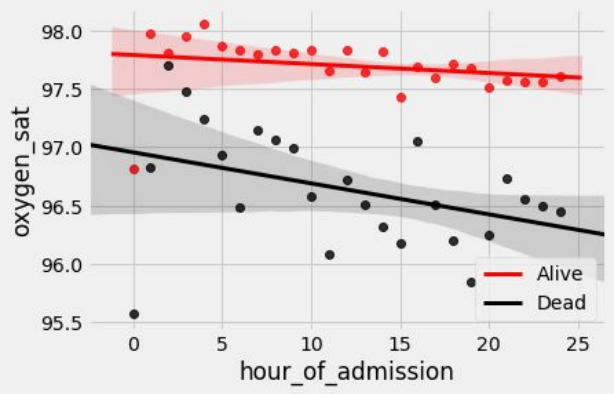
Initial Observations in the Data:

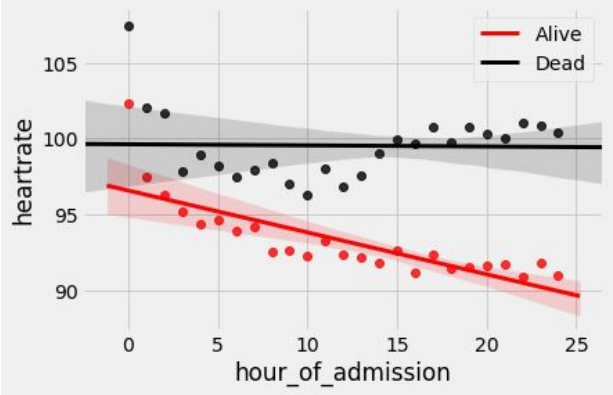
As an initial test to determine if our vital signs parameters could even be used to predict patient mortality, we've charted averages of each vital sign over time for both mortal patients and survivors, and attached 95% confidence intervals. The graphs for respiratory rate, blood pressure, heart rate and oxygen saturation are shown below.











As our charts show, not only do mortal patients and survivors have disparate vital sign values, but they also tend to have differing vital sign trends in the first 24 hours. For example, while patients who end up surviving tend to have lower, and slightly decreasing respiratory rates in the first 24 hours, mortal patients tend to have higher and strongly increasing respiratory rates in the first 24 hours. Importantly, the confidence interval shading for the graphs show that the differences between the vital signs for the survivors and the mortal patients are unlikely to be due

to chance, and therefore give us confidence in using vital signs as parameters for determining patient mortality.

Methods

In order to effectively take advantage of the longitudinal data that we have, we required a Machine Learning model that had a "memory" of sorts; it needed to be able to analyze each hourly data point in the context of the data points that preceded it to effectively make a prediction. Furthermore, that "memory" had to be able to take into account multiple preceding predictions, as our initial observations show that the trends present in the vital signs are best understood within the context of many preceding points. The model we believed fit these criteria best was a Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM). While RNNs allow us to take into account previous data points, LSTMs allow us to consider a larger trailing string of data points and effectively decide which data points in that string are of relevance, making their combination an optimal application for the task at hand. As such, our current aim is to develop a RNN with LSTM model that we can train with our current data set of Infected patients, and then apply to another set of Sepsis patients.

Where we need further direction is in determining how to use a RNN with LSTM to determine a binary output (alive or dead) based on the pattern generated from the entire sequence of longitudinal vital signs data points. While our research into previous experiments performed using a similar method has shown that using RNN with LSTMs in this manner is possible, we have not (as of yet) been able to identify the technical specifications of how this has been done.