Predicting Sepsis

Capstone 1 – Project Proposal

Springboard Data Science Career Track

Aisling Casey, February 11th, 2021

Problem Statement:

Early intervention in sepsis patients is critical to their health outcome. Is it possible to identify sepsis in ICU patients 6 hours before the clinical diagnosis of sepsis?

Context:

Sepsis is a leading cause of death in US hospital patients. Sepsis occurs when.. "when the body's response to infection causes tissue damage, organ failure, or death"[1][3]. Prompt intervention in sepsis patients can improve the likelihood of their condition improving significantly, while unnecessary treatment in non-sepsis patients drains limited hospital resources.

Scope of Solution Space:

Quantifying the risk of sepsis & binary classification of sepsis every hour for patients in the ICU.

Criteria for Success:

Model produces high clinical utility score, relative to those given in Reyna et al [1]. The utility score rewards early prediction of sepsis and penalizes late or missed predictions. It also penalizes predictions of sepsis in non-sepsis patients.
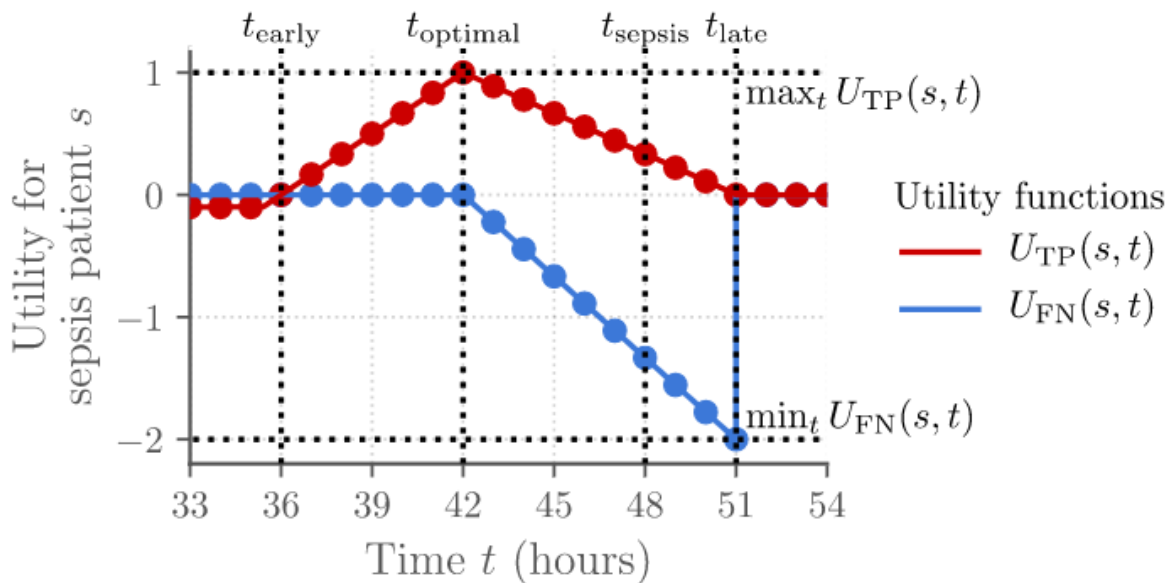

Figure 1: Utility function for a sepsis patient with time of sepsis at 48 hours. U_{TP} represents the utility score for a positive prediction at time t for a sepsis patient, while U_{FN} represents the utility score for a negative prediction for a sepsis patient. Image from Reyna et al 2019[1]

Constraints:

- Laboratory data not available hourly/not available for every patient.
- Do not have information on patients' comorbidities.
- Can only base model on datasets from two hospitals – may limit the applicability to other hospitals.

Stakeholders:

- **ICU Clinicians**: Real time, objective prognosis will enable doctors to deliver better healthcare & health outcomes.
- **Patients & Their Families**: Health outcomes of patient will directly impact their livelihood & their families'.
- **Hospitals:** Better health outcomes will lead to lower costs and greater profits.

Data Sources:

40,366 PSV patient data files (from 2 different hospitals). Each file contains 40 predictor variables, 1 target variable ('SepsisLabel'), with each row representing the data from one hour in the ICU. Data set can be accessed here[1].

Variable Type Column #	Vital Signs 1-8	Laboratory Values 9-34	Demographics 35-40	Sepsis Label 41
t_0	…	…	…	…
t_1	…	…	…	…
….	…	…	…	…
t_n	…	…	…	…

Figure 2: Data structure of one patient's dataset; there are as many rows as hours the patient spent in the ICU. There are 40,366 patient datasets in total.

Deliverables:

A model that provides the risk of sepsis & a classification of sepsis every hour for ICU patients. The model will be documented with the following:

1. Github repo for work completed on each step of the project
2. Slide Deck
3. Project Report

Citations

[1] Reyna, M., Josef, C., Jeter, R., Shashikumar, S., Moody, B., Westover, M. B., Sharma, A., Nemati, S., & Clifford, G. (2019). Early Prediction of Sepsis from Clinical Data -- the PhysioNet Computing in Cardiology Challenge 2019 (version 1.0.0). *PhysioNet*. https://doi.org/10.13026/v64v-d857.

[2] Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemati S, Clifford GD, Sharma A. Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge. Critical Care Medicine 48 2: 210-217 (2019). https://doi.org/10.1097/CCM.0000000000004145

[3] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):801-10. doi: 10.1001/jama.2016.0287. PMID: 26903338; PMCID: PMC4968574.