CSE6250 Project Proposal, Team 2

Samuel J. Callister, B.S.[1], **Imre Patyi, Ph.D.**[1], **Timothy Whittemore, M.A.**[1]
[1]**Georgia Institute of Technology, Atlanta, Georgia, USA**

Abstract (A1)[1]

We propose to use the MIMIC III dataset with a view towards producing a benchmark and a machine learning system to facilitate the prediction of sepsis and its time of onset.

Motivation

(A1). Following recommendations issued in 2016 by a task force of specialists, sepsis is currently defined as "a life threatening organ dysfunction caused by a dysregulated host response to infection."[11] It is a global public health emergency, representing one of the largest causes of death worldwide.[3] Early detection and intervention is critical, with studies consistently demonstrating increases in hospital mortality for each hourly delay in treatment: from 1.8%[13] to as much as 7.6%.[2] However, over-triage of patients who may not have sepsis itself incurs potential harms and costs.[3,13] Meanwhile, clinical decision support (CDS) systems with high false positive rates contribute to alert fatigue and disuse of the CDS.[7] The need for both fast and accurate detection systems is widely-recognized.[2,4,5,7,10,12,13]

Summary of challenges and current approaches

(A2). Detecting sepsis is challenging for several reasons, with inter-related sources of difficulty including the following: lack of clear definitions, multiple phenotypes, inability to identify pathogenic organisms in sufficient time, disagreements concerning clinical criteria, and reliance on clinical surrogates for organ dysfunction. First, although considered an improvement over the previous 2001 definition, the Sepsis 3 definition fails to draw distinctions between either the type of infection or the host response.[3] Second, sepsis is a heterogeneous syndrome, currently believed to encompass a number of subtypes based on the combination of pathogenic organisms and patient characteristics.[3] For example, the immune response of septic patients ranges from "an exuberant pro-inflammatory cascade to a profoundly immunosuppressed phenotype."[3] Another example is that there are significant differences in incidence and mortality based on whether sepsis is present on admission (POA) or whether sepsis develops in the hospital (NPOA): approximately 85% of cases are POA, with a mortality rate of 12%, whereas 15% of cases are NPOA but carry a mortality rate of 35%.[10] Third, identification of pathogenic organisms that may be present can take days using conventional technology — moreover, a significant number of patients never have positive cultures.[3] Fourth, there is disagreement about what constiutes shock: whereas the Sepsis 3 criteria for septic shock require all three of hypotension, vasopressors, and elevated lactate, many clinicians continue to believe that the criteria should consider either the combination of hypotension and vasopressors OR elevated lactate.[3] Finally, there is no tissue diagnostic or reliable seriological test for sepsis.[14] Instead, clinical identification is based on surrogates for organ failure including: serum creatinine, serum bilirubin, blood pressure, PaO_2/FiO_2 ratio, Glasgow coma scale, platelet count, and respiratory rate. These surrogates are not sufficient to create a gold standard for sepsis.[3] Furthermore, adjustments may be necessary to account for comorbidities that could otherwise lead to misclassification: for example, chronic renal disease leads to high levels of creatinine; end-stage liver disease affects serum lactate, bilirubin, and platelet counts; and medications such as beta-blockers affect heart-rate.[7]

Research efforts thus far have considered multiple aspects of sepsis prediction, including: early prediction of sepsis,[1,4–7,9,12] prediction of septic shock,[2] modeling distinct subgroups of septic patients,[10] and developing models for dealing with gaps in data needed to predict sepsis.[8] Of the studies on early prediction of sepsis, a small number have included an implementation phase in hospital settings, after which they observed specific changes in outcomes.[7,9] The remainder consist of purely retrospective studies for which researchers assessed models based on metrics such as area under curve (AUC) as well as comparison against other scoring systems, including SIRS, MEWS, and qSOFA.[4–6,12]

Models used in early prediction of sepsis include linear regression models[1,6,10] and generalized linear models,[12] deep learning models,[4,6] support vector machines (SVM),[5] and analysis of medical record text.[5] Among the first group,

[1]Reference to CSE6250 Projects, § II, A.

Desautels et al.[1] studied the performance of the InSight model on patients selected from the MIMIC III database. InSight is a type of regression model with handcrafted features for predicting sepsis.[6] It achieved AUC values as high as 0.8799, versus 0.77 for qSOFA. One of the notable features of their case selection process was omission of data from one EHR system, both because it provided less clinical detail than the other represented system and because the researchers suspected under-reporting of negative cultures. The second point is interesting in light of the observation that a significant number of sepsis patients never have positive cultures.[3]

Researchers have recently compared deep learning aproaches against InSight.[6] For patients selected from the MIMIC II database, InSight achieved AUC of 0.887 (similar to Desautels et al.) and was outperformed by both multilayer perceptron models (AUC of 0.915) and long short-term memory models (AUC of 0.929). In addition to outscoring InSight, one of the major advantages of deep learning approaches is that they can learn the important features independently, with no specialized knowledge provided. By contrast, InSight's handcrafted features depend very much on domain knowledge. However, the "blackbox" nature of deep learning is a major limitation in the domain of medicine, which stresses the importance of being able to interpret models and explain the causes leading to the results.[6]

Researchers have also reported significant improvements in predictions by combining the results of text analysis with other clinical features as inputs to an SVM model.[5] Specifically, analysis of chief complaints and free-text nursing assessments boosted SVM performance by more than 26%, as measured by AUC. Interestingly, the most predictive words selected from bag of words analysis included: cellulitis, sore throat, dysuria, pneumonia, cyst and infection.

Lastly, the two studies with implementation phases are notable because they provide "real-world" examples of the benefits of early detection. In one case, sepsis mortality decreased by 53%, although no significant change in length of hospital stay was observed.[7] In the other case, sepsis mortality decreased by 60.24%, with reduction in length of stay by 9.55% and decrease in sepsis-related 30-day readmission by 50.14%.[9] The researchers did not discuss the details of their machine learning algorithm, however.

Formulation of research problem

(A3). We will develop a model to predict the probability that a patient develops sepsis for a fixed time range during an ICU stay. We aim to take data processing steps outlined as follows.

(A3), (A4), (A8).
(i) Data gathering from the MIMIC III database — this involves identifying a set of qualifying hospital stays, labeling patients as septic or not (a classification), and determining onset of sepsis. For the latter two tasks, we will adopt the approach taken by Desautels et al.,[1] which involves computing SOFA scores on an hourly basis per patient. Based upon our literature review, we expect a lower bound of just under 2000 sepsis cases from the MIMIC III database, accounting for slightly less than 10% prevalence among ICU stays.[1] We plan to do clustering on a projection of the data, where we map to the space of such variables as age, vital measurements, and number of data points available. We plan to prune clusters with very low frequency of sepsis to obtain a sufficiently balanced dataset.

(ii) Extraction, transformation, and loading — we will construct features from the direct measurements of vitals, the history of lab results, and medications. We intend to aggregate features by hour and make available to the learning algorithms only those that fall into the admissible time window. In a manner similar to the labeling procedure, we will develop a score that tracks deviation of vitals and measurements from the normal range in an absolute sense, updated hourly. (The scores themselves may have a different probability distribution than their differences; a differenced feature set may be differently fit even by a linear predictor.) The idea is that as patients slide towards sepsis, they accumulate more and more aggregate deviations from the normal range. We hypothesize that this feature will boost performance of the learning algorithm.

(iii) Prediction — we intend to solve a machine learning problem that predicts the probabilty that a patient develops sepsis within a fixed time range. (A8) We plan to employ soft classification algorithms such as gradient boosted trees, logistic regression, and random forest that can output label probabilities. (A4) To measure the performance of the prediction algorithms we plan to use the area under the receiver operating characteristic curve (AUROC) and the area under the precision versus recall curve (APR). We will also attempt to put a cost as a dollar amount on false positives, true positives, etc., and obtain the total cost of this prediction for discrete values of the probability threshold.

(iv) Monitoring and alarm — developing a model to predict sepsis probabilities at various hours ahead of the observation window will enable us to build an alarm system. We can put another cost as a dollar amount on true sepsis alarms and on false sepsis alarms as well as evaluating the potential for "alarm fatigue." These costs will inform our choice of threshold on the sepsis probability at or above which to trigger an alarm. We will then compare the rate of false alarms against scoring criteria such as qSOFA.

Analytic Infrastructure

(A5) We will leverage local Spark clusters to filter and transform data. If local processing power or memory is not sufficient, we will setup a Spark cluster on AWS. Various Python packages such as numpy and scikit-learn will be used to implement necessary algorithms and models.

Key Data

(A6) We will divide the MIMIC III dataset into sepsis patients and non-sepsis patients. We will rely on the SOFA criteria to determine if a patient has sepsis or not. A patient will be considered septic if his or her SOFA score rises at least 2 points within our prediction window. All other patients will be considered non-septic.

(A7) We have not yet built our sepsis labeling pipeline, but we have found some initial statistics about our patient population. Of the 46,520 total patients, 39,364 were found to have suspicion of infection during their ICU stays as indicated by the administration of antibiotics or a culture taken. A large portion of the patients have a suspicion of infection and from this populace we will find our sepsis cases.

Table 1: Initial Population Statistics

Characteristic		Suspicion of Infection	No Suspicion of Infection
Gender	Female	17,608	2,791
	Male	21,756	4,365
Age	0-17	6,772	1,200
	18-29	1,608	381
	30-39	1,732	354
	40-49	3,415	747
	50-59	5,523	1,148
	60-69	6,586	1,254
	70+	13,728	2,072
Percentage of Patients		84.6%	15.4%

Timeline (A9)

Date	Milestone
10/14	***Submit proposal document***
10/18	Check and clean data, phenotyping
10/25	Initial model implementation and predictions
11/01	Evaluate model results and compare with published literature
11/08	Refine features and/or machine learning program as needed
11/11	***Submit project draft***
11/15	Write and document a clean implementation of data preparation and machine learning algorithms
11/22	Try to make up for any slippage in the schedule
11/29	Make the final version of project code; make video or audio presentation
12/06	Complete final report
12/09	***Submit final report (paper, video or audio presentation, data, and code)***

References

1. T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, D. J. Wales, and R. Das. Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Med Inform*, 4(3):e28, 30 Sept. 2016.
2. K. E. Henry, D. N. Hager, P. J. Pronovost, and S. Saria. A targeted real-time early warning score (trewscore) for septic shock. *Science translational medicine*, 7(299):299ra122–299ra122, 2015.
3. Coopersmith, C. M., De Backer, D., Deutschman, C. S., Ferrer, R., Lat, I., Machado, F. R., ... and Evans, L. E. Surviving sepsis campaign: research priorities for sepsis and septic shock. *Intensive care medicine*, 44(9), 1400–1426. Sep. 2018.
4. Futoma, J., Hariharan, S., and Heller, K. Learning to detect sepsis with a multitask gaussian process rnn classifier. *arXiv preprint arXiv:1706.04152*, 13 Jun. 2017.
5. Horng, S., Sontag, D. A., Halpern, Y., Jernite, Y., Shapiro, N. I., and Nathanson, L. A. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PloS one*, 12(4), e0174708. 6 Apr. 2017.
6. Kam, H. J., and Kim, H. Y. Learning representations for the early detection of sepsis with deep neural networks. *Computers in biology and medicine*, 89, 248–255. 1 Oct. 2017.
7. Manaktala, S., and Claypool, S. R. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. *Journ Am Med Inform Assoc*, 24(1), 88–95. 25 May. 2016.
8. Mayhew, M. B., Petersen, B. K., Sales, A. P., Greene, J. D., Liu, V. X., and Wasson, T. S. Flexible, cluster-based analysis of the electronic medical record of sepsis with composite mixture models. *Journ Biomed Inform*, 78, 33–42. 1 Feb. 2018.
9. McCoy, A., and Das, R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual*, 6(2), e000158. 1 Oct. 2017.
10. Rothman, M., Levy, M., Dellinger, R. P., Jones, S. L., Fogerty, R. L., Voelker, K. G., ... and Beals IV, J. Sepsis as 2 problems: Identifying sepsis at admission and predicting onset in the hospital using an electronic medical record-based acuity score. *Journ Crit Care*, 38, 237–244. 1 Apr. 2017.
11. Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., ... and Angus, D. C. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, 315(8), 801–810. 23 Feb. 2016.
12. Guirgis, F. W., Jones, L., Esma, R., Weiss, A., McCurdy, K., Ferreira, J., ... and Gerdik, C. Managing sepsis: electronic recognition, rapid response teams, and standardized care save lives. *Journ Crit Care*, 40, 296–302. 1 Aug. 2017.
13. Liu, V. X., Fielding-Singh, V., Greene, J. D., Baker, J. M., Iwashyna, T. J., Bhattacharya, J., and Escobar, G. J. The timing of early antibiotics and hospital mortality in sepsis. *Am Journ Resp Crit Care Med*, 196(7), 856–863. 1 Oct. 2017.
14. Macdonald, S. P., Williams, J. M., Shetty, A., Bellomo, R., Finfer, S., Shapiro, N., and Keijzers, G. Sepsis in the emergency department– Part 1: definitions and outcomes. *Emergy Med Austral*, 29(6), 619–625. Dec. 2017.