Machine Learning and Statistical Techniques to Predict Sepsis: Unifying Previous Work

Nathaniel D. Bastian, Paul M. Grffin, Shravan Kethireddy, Joseph J. Klobusicky, Alexander Murph, Alexander Robinson, Nathan C. Ryan

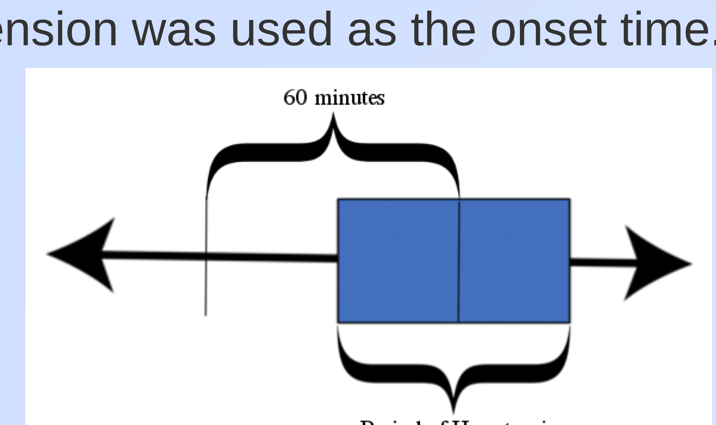
Introduction

Sepsis is a syndrome of dysregulated inflammation caused by infection. In 2010 it was that 11th leading cause of death in the United States [1] and, in 2011, it was the single most expensive diagnosis treated in hospitals in the United States [2]. Septic shock is a subset of sepsis in which the complications induced by the presence of sepsis lead to an increase in the mortality of the patient.

Methods

Finding a Septic Shock Diagnosis

Following a method developed by Ho et. al. [3] for determining Septic Shock, we classified periods of hypotension with a given patient as any period of time where hypostolic blood pressure was below 90. Using the center time of each of these periods of hypotension, we summed up all the fluids given to the patient between this time and an hour before. If the amount of fluids in this timespan exceeded 600ml, the patient was considered to have hypotension despite fluid intake. As such, if the patient was also determined to have Sepsis, we labeled the patient as positive for Septic Shock. The start time of that patient's period of hypotension was used as the onset time.

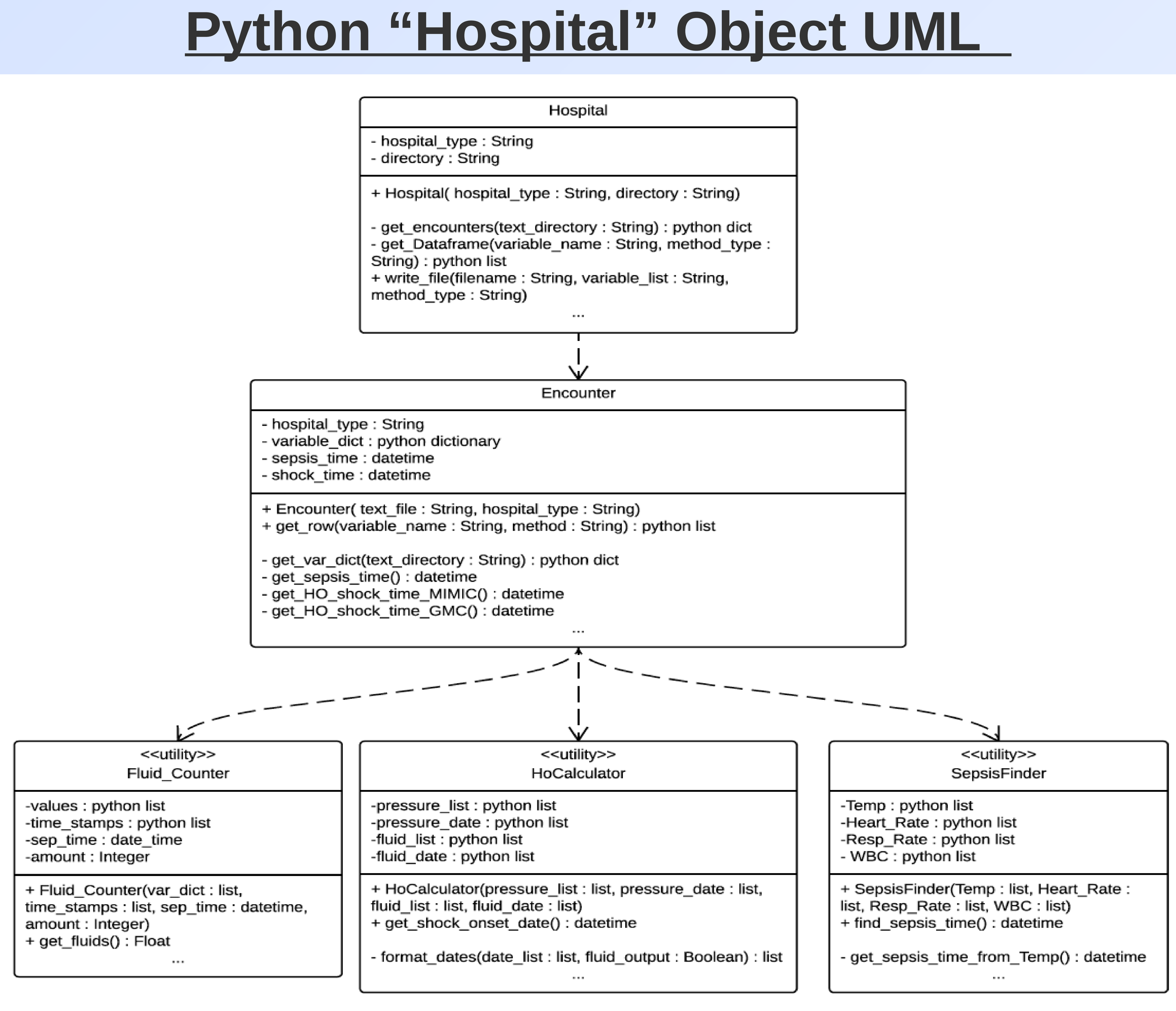


To find a patient's time of Septic Shock onset, to choose data points closest to this time, and to merge data from multiple sources into a single dataset, we used an object-oriented approach in Python. After the proper data was collected and written to a csv file, the cleaning, processing, and modeling-making was done in R.

Predicting Septic Shock

To do our predictive analysis, we used four existing mathematical models: a Logistic Regression model, a Decision Tree model, a Support Vector Machine, and an Artificial Neural Network. For each model, we used the following features for prediction: **Heart Rate, Respiratory Rate, SOFA and SAPS scores, White Blood Cell count, Temperature, and Blood Pressure.**

Python "Hospital" Object UML



Data

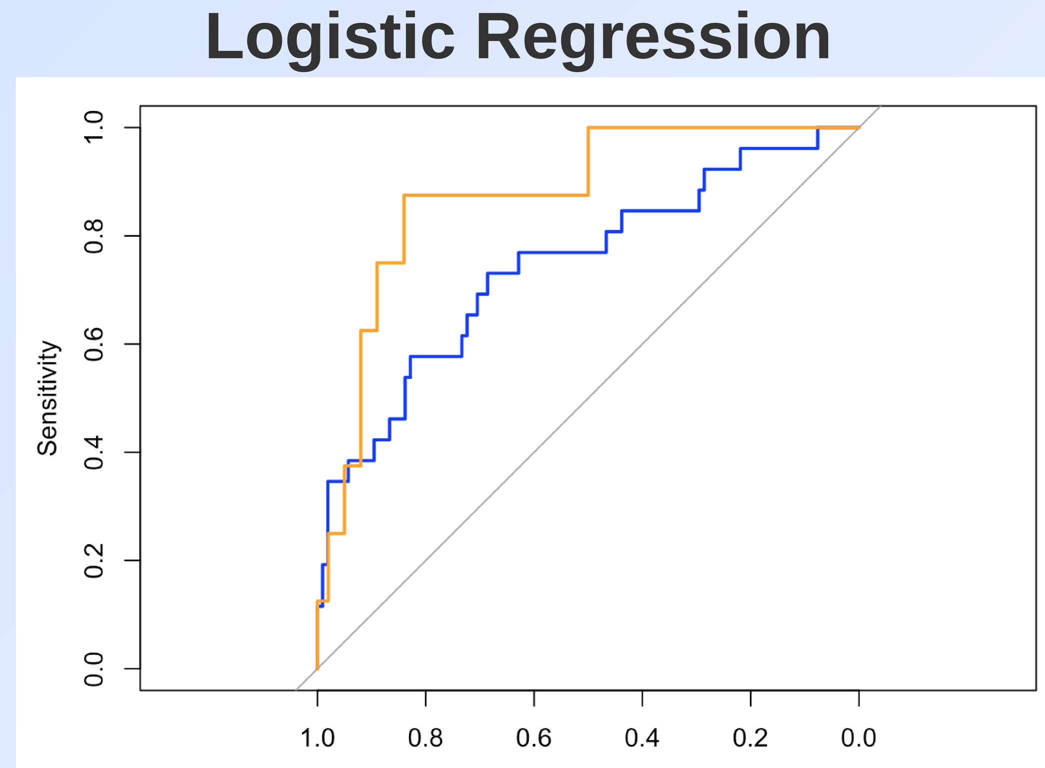
Two datasets were used for comparative analysis. First, we used the publically available Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database, which was gathered from Boston's Beth Israel Deaconess Medical Center [4]. Secondly, we analyzed medical data collected at the Geisinger Health System (GHS).

Dataset:	MIMIC-II	GMC
Data amount	>30,000	>18,000
# of ICU	27,542	17,864
Length of Stay	3 – 18 days	1 – 7 days
Resp. Average Samples/ Frequency	257.47 / 20 - 58 min	44.64 / 3 – 4 min
HR Average Samples/ Frequency	264.28 /27 - 67 min	42.26 /2.5 – 3.5 min
BP Average Samples/ Frequency	121.81 /61- 125 min	45.68 /2.6 – 3 min
Temp Average Samples/ Frequency	62.69 /120-280 min	25.29 /5.8 – 6 min
SpO2 Average Samples/ Frequency	259.6 /28-68 min	38.73 /3.3 – 3.6 min
WBC Average Samples/ Frequency	13.52/490-134 3min	4.37 /33.4 – 37.5 min

Results

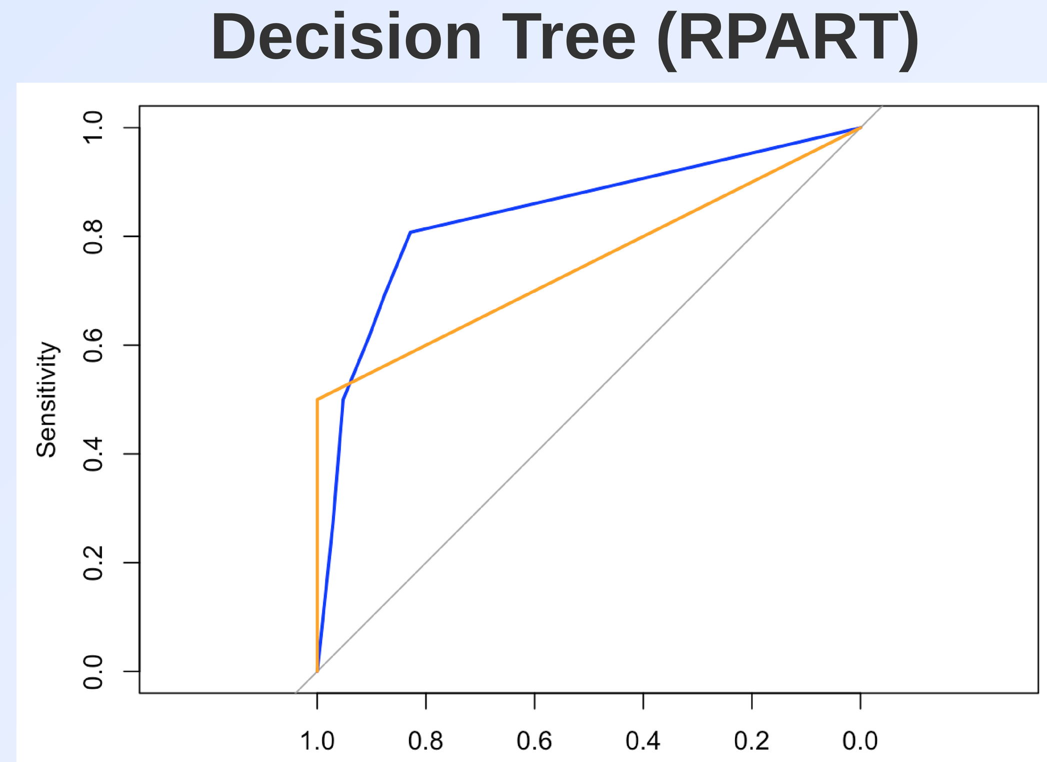
Each model yielded comparable results for the two data sets. To measure the performance of our models, we created for each a Receiver Operating Characteristic (ROC) curve.

Logistic Regression



GMC AUC: .875, MIMIC AUC: .7451

Decision Tree (RPART)



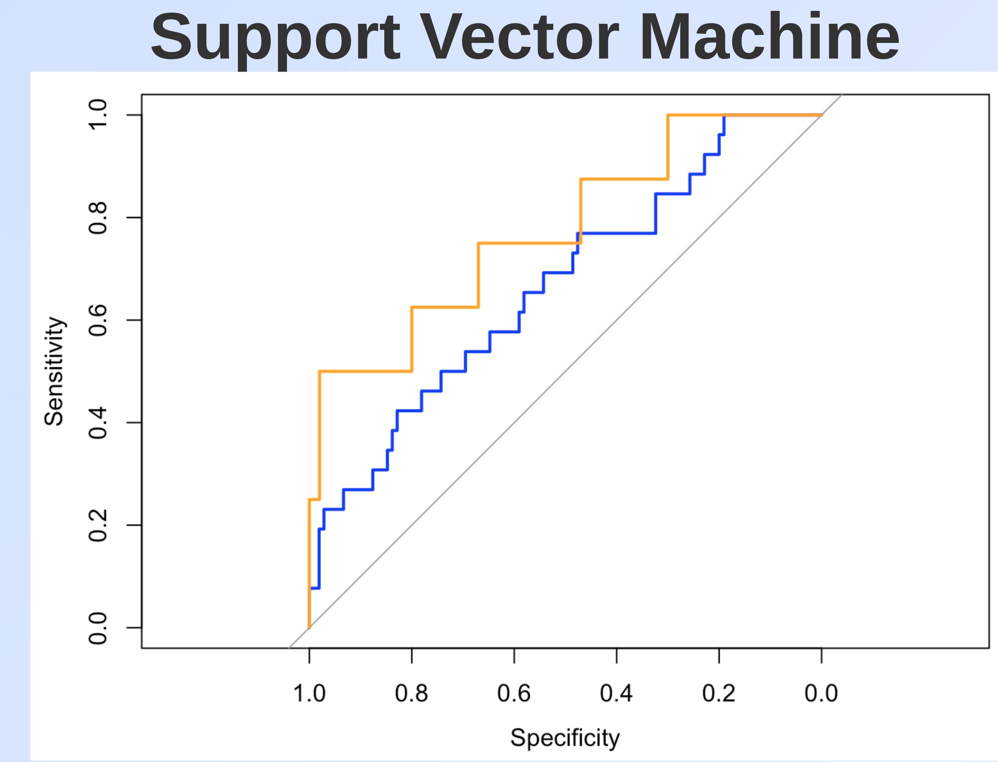
GMC AUC: .75, MIMIC AUC: .8408

Acknowledgements

We would like to thank the Bucknell-Geisinger Research Initiative for making this project possible, and Bucknell's Provost Office for providing funding for this trip.

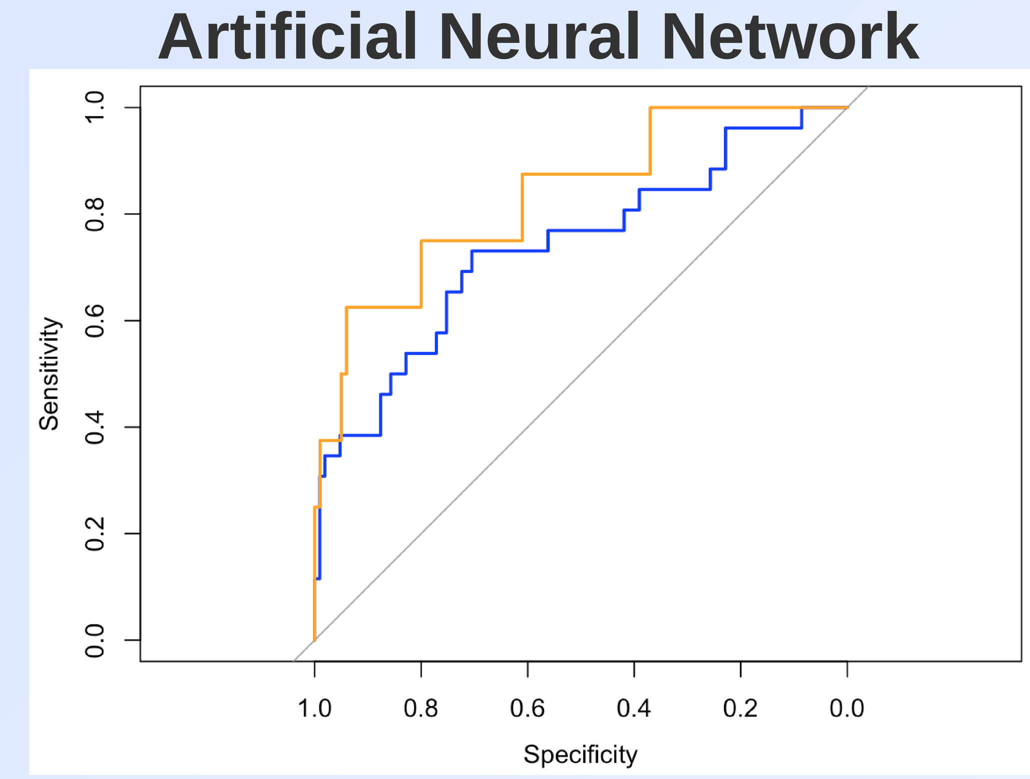
Results (Continued)

Support Vector Machine



GMC AUC: .775, MIMIC AUC: .6656

Artificial Neural Network



GMC AUC: .8325, MIMIC AUC: .7385

Conclusion

We replicated models in the literature on the same publicly available data set on which they were developed. After getting these baseline results, which were similar to those in the literature, we applied the same models (with the same features and definitions) to privately held data. For a fixed model, its performance on the publicly available data and the privately held data were comparable. The quality of performance was especially good considering that the models were developed on demographic features of the individuals and only seven additional features. In some cases, our models were trained on larger data sets and also outperformed models in the literature.

The greatest challenge in the development of these models is the determination of which individuals have the right kind and right amount of data to be included in the models. In particular, identifying which patients had sepsis and when they had it (and septic shock if applicable) was essential for our experiments. The fact that in some cases we improved on existing results we think can be attributed to the choice that we made to identify patients as having sepsis in a way that is not based on the ICD-9 the patients have had.

For our next step, we will be examining how we might examine the paths of how patients move through the hospital, and use predictive techniques like Markov Chains on the GMC data.

References

[1] L Hoyert, *et al.* *Nat. vil.* **21** (2012)

[2] C Torio, et al. Nat. inp. **160** (2013)

[3] J Ho, et al. Sep. sho. **1** (2014)

[4] M Saeed, et al. Mult. intel. **5** (2011)