The Importance of Baseline Models in Sepsis Prediction

The University of Texas at Austin
Electrical and Computer Engineering
Cockrell School of Engineering

Christopher Snyder[1]. Jared Ucherek[2]. Sriram Vishwanath[3].

[1] Department of Biomedical Engineering, University of Texas at Austin, christopher.g.snyder@utexas.edu
[2] Department of Electrical and Computer Engineering, University of Texas at Austin, jared.ucherek@utexas.edu
[3] Department of Electrical and Computer Engineering, University of Texas at Austin, Professor, sriram@austin.utexas.edu

Summary

Reliable sepsis prediction for at-risk patients could drastically change the likelihood of mortality in the ICU. Emerging applications of Deep Learning in the medical field are making inroads in sepsis prediction and other challenging predictive problems. However, most metric comparison do not reveal how effective these models would be in realistic scenarios. We ultimately strive for our solutions to have the largest impact possible, and showing testbed performance gains only count when the models can be applied to realistic clinical settings.

Methods

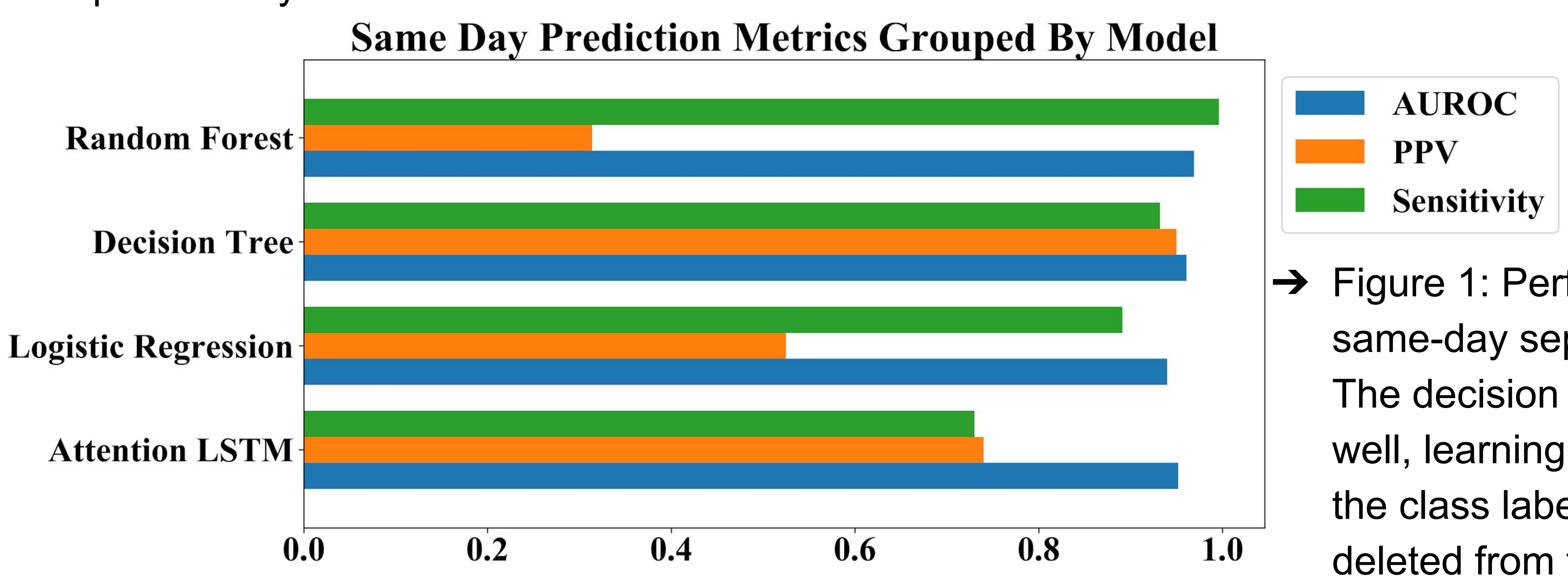
- Sepsis prediction was on the MIMIC III dataset [3] and patients are sepsis positive with SIRS criteria more than 2 (Heart Rate > 90; Respiratory Rate > 20; White Blood Cell Count > 12,000 or < 4,000; Temperature (F) > 100.4 or < 96.8).
- We closely followed the work by Kaji et al.[1] and their Github repo [2] to process the raw data into file checkpoints for training, validation, and testing datasets.

Approach

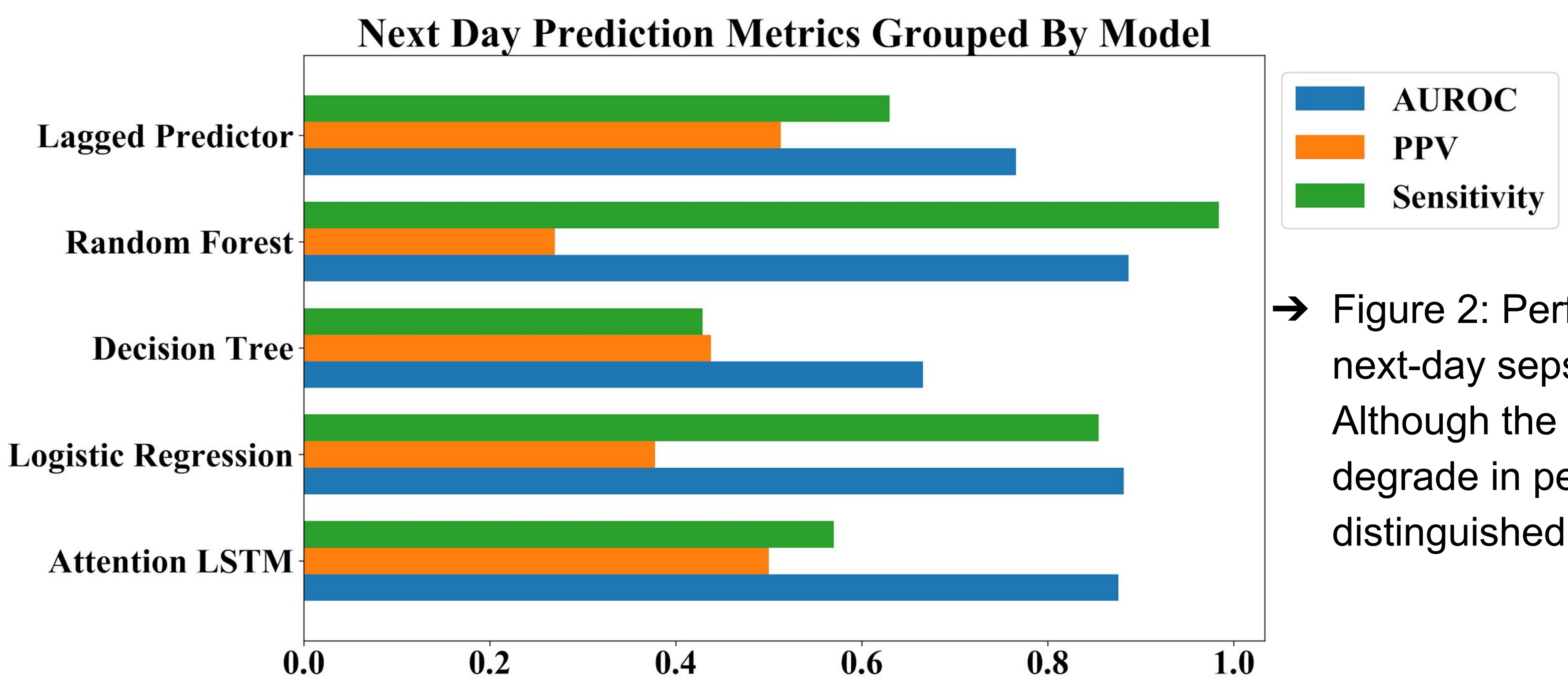
- To qualify how well current models predict sepsis began by comparing their performance to several baseline models.
- We hypothesized that large performance gaps would be present in various metrics between the two groups.
- We were surprised by the comparison, and try to explain the various metrics, advantages, and disadvantages for these predictors.

Results

- After directly comparing various models on the dataset, we found no significant performance advantage for the AttentionLSTM developed by Kaji et al. Upon inspection, the baseline AUROC metric presented in many sepsis predictors would not offer powerful predictive power for novel cases of sepsis being developed during patient stays.



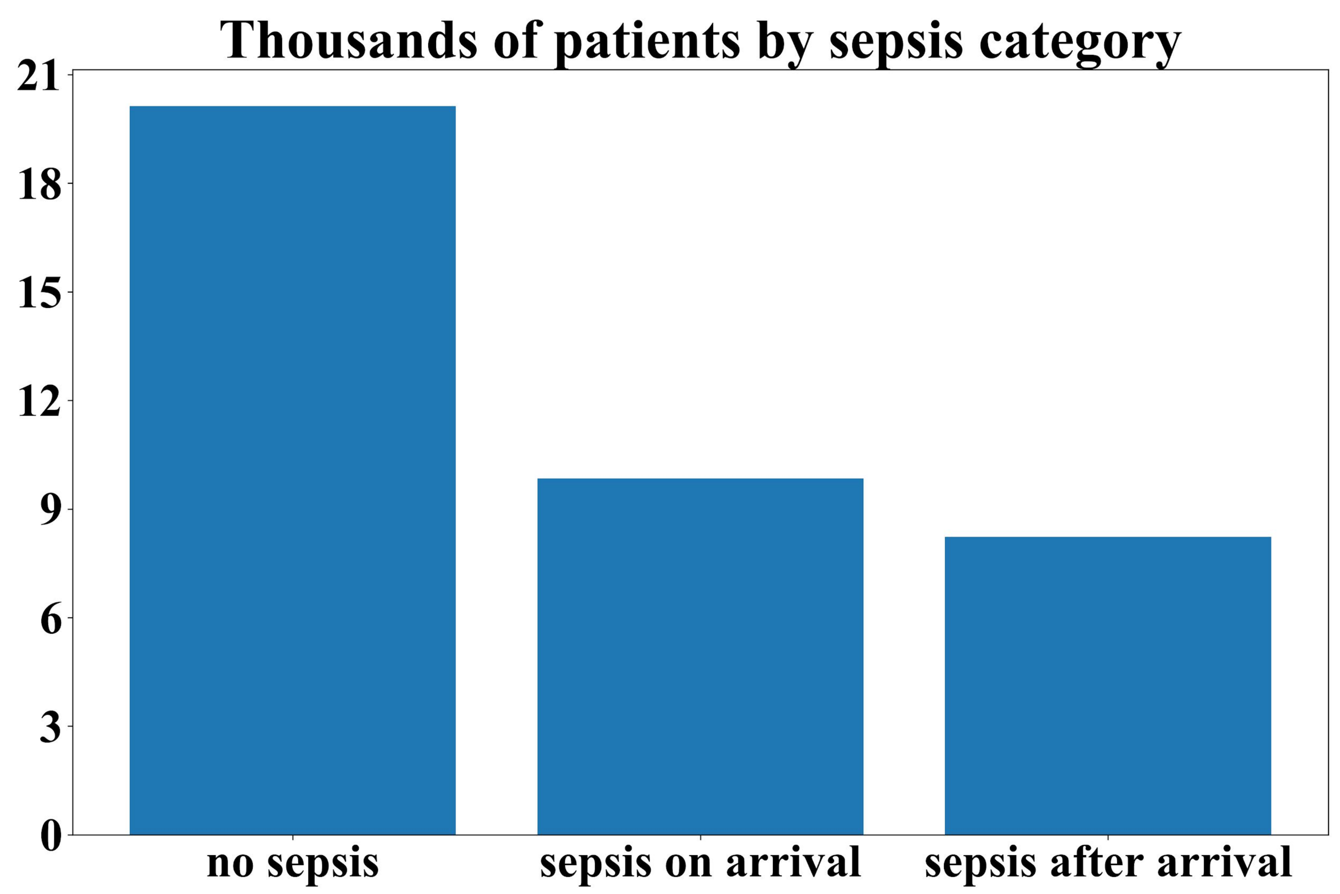
➔ Figure 1: Performance comparison for same-day sepsis prediction in patients. The decision tree performs exceptionally well, learning the rule based generation of the class label with the data that was not deleted from the dataset (heart rate, white blood cell count, and temperature).



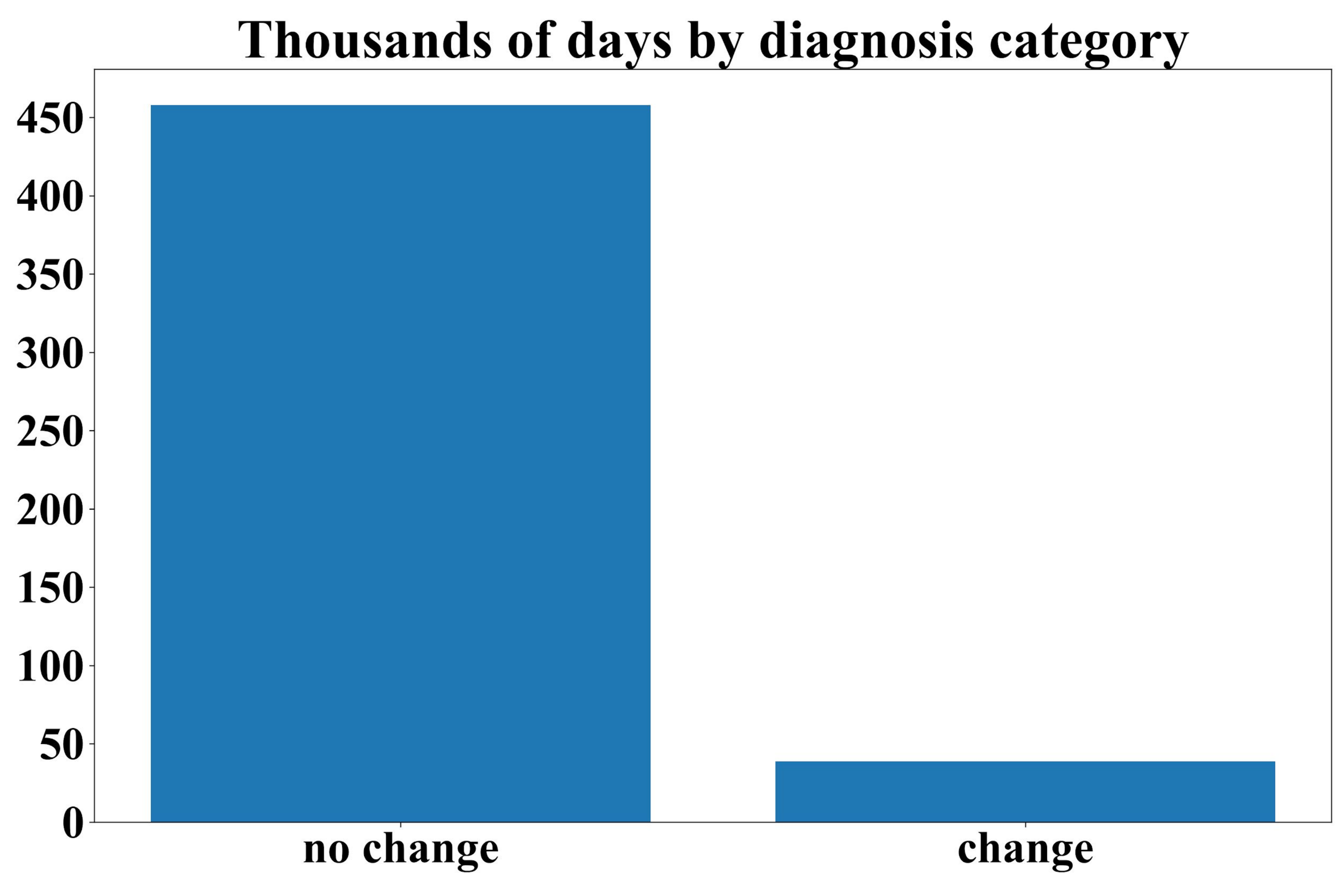
➔ Figure 2: Performance comparison for next-day sepsis prediction in patients. Although the time-independent models degrade in performance, there is still no distinguished results in this modeling

Dicussion

- There are several dynamics occurring in the sepsis diagnostics that convey why the current metrics fail to quickly convey helpful models for ultimate predictive task.
- For applications of black box clinical models, we suggest at the very least making standard practice the reporting of performance values with base reference models.



➔ Figure 4: Patient histogram of first day sepsis occurrence. The large plurality of patients are never diagnosed with sepsis. Additionally, most cases of sepsis are recorded by patients that arrive at the hospital with sepsis (within the first 24 hours of stay).



➔ Figure 5: Predicting tomorrow with today's diagnosis (Lagged Predictor) performs exceptionally well, mainly due to how infrequently the sepsis status switches states in patients. Refocusing the predictions for only critical patients or reconfiguring the metric to reward correctly diagnosing a changing sepsis state would greatly benefit the functionality of these classifiers.

[1] D.A. Kaji, J.R. Zech, J.S. Kim, et al., An attention based deep learning model of clinical events in the intensive care unit, PLoS One 14 (2019), e0211057. https ://www.ncbi.nlm.nih.gov/pubmed/30759094.
[2] deepak-kaji. (2018, October 29). deepak-kaji/mimic-lstm: Initial Release (Version 1.0). Zenodo. Doi: 10.5281/zenodo.1473691
[3] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). Doi: 10.1038/sdata.2016.35.