

**The Importance of Baseline Models in Sepsis Prediction***Christopher Snyder<sup>1</sup>, Jared Ucherek<sup>1</sup>, Sriram Vishwanath<sup>1</sup>*<sup>1</sup>*The University of Texas at Austin*

**Background.** Reliable sepsis prediction is just one of many machine learning problems where proper modeling could drastically change the likelihood of mortality in the ICU. Emerging applications of Deep Learning (DL) in the medical field are making inroads in sepsis prediction and other challenging predictive problems. This general-purpose approach has produced breakthroughs in a variety of different fields, proving itself effective in various diagnostic tasks.<sup>[5]</sup> Although DL offers substantial opportunities to improve clinical outcomes and influence health<sup>[6]</sup>, it comes with some key limitations with implications for how we evaluate models. In general, the neural networks used to model relationships in expressive data are uninterpretable. Our shallow sense of understanding results in slower iterative performance gains and less interpretable output for clinicians. We ultimately strive for our solutions to have the largest impact possible, and tested performance gains only count when they can be appropriately translated to new and diverse clinical settings. For applications of black box clinical models, we suggest making standard practice the reporting of performance values with base reference models.

**Approach.** Our goal was to find a published clinical DL model whose performance we would try to approximate by training simpler models. We hypothesized that the deep network would surpass the performance of these models, and we planned to use these performance gaps to qualitatively compare advantages and disadvantages of the various models. To ensure the same training data was used, we focused on publications using publicly available, clinical data with clearly-documented code. We settled on recent work by Kaji et al.<sup>[1]</sup> to model sepsis prediction on the MIMIC III dataset<sup>[3]</sup> with attention-based LSTM. We utilized scripts from their Github repo<sup>[2]</sup> to preprocess the raw data into file checkpoints for training, validation, and testing datasets. These files were subsequently fed into their model for training and testing without any further processing.

**Methods.** We developed additional preprocessing scripts to load these checkpoints, align the labels for the corresponding predictive task, flatten the dataset, and remove the extra padding. In principle, the best sepsis prediction models would learn time-dependent covariates and contextualize predictions based on patient history. However, we removed these time-series aspects of the data to create simple and instructive reference models. For analysis, we trained several time-independent classification models, listed in Table 1. Additionally, for next-day predictions, we created a Lagged Predictor that predicts the next-day label only using the current-day label. The performance metrics of our models are compared to the final reported numbers in the original paper.

**Results.** We found no significant performance advantage for the AttentionLSTM developed by Kaji et al. Using the full feature set, our findings show a very similar performance between the base models and the AttentionLSTM. Every trained model performed better than the baseline scores provided by the lagged predictor.

**Table 1:** Five classification models comparing same-day and next-day performance metrics

Models		Same-day sepsis prediction			Next-day sepsis prediction		
		AUROC	PPV	Sensitivity	AUROC	PPV	Sensitivity
Time Dependent	Attention LSTM (Kaji et al.)	0.952	0.740	0.73	0.876	0.500	0.570
Time Independent	Logistic Regression	0.940	0.525	0.891	0.882	0.378	0.855
	Decision Tree	0.961	<b>0.950</b>	0.932	0.666	0.438	0.429
	Random Forest	<b>0.969</b>	0.314	<b>0.996</b>	<b>0.887</b>	0.270	<b>0.984</b>
	Lagged Predictor	-	-	-	0.766	<b>0.513</b>	0.630

**Conclusion.** AttentionLSTM models are among the most innovative and complicated time series models, possessing both intricate LSTM gates and high-class attention layers beyond a simpler RNN model. For sepsis prediction, this model is a sensible and robust choice. However, its good performance does not instruct us on how to model better in the future. Models succeed or fail for reasons that we should know. Additional analysis into which model performs better may contribute to a speculative understanding of a model’s behavior, which we can build on as we continue modeling similar clinical problems. We showed that for this particular dataset, the complicated components of the AttentionLSTM were completely superfluous. Ignoring the temporal structure of this particular sepsis prediction problem led to higher performance. While it is unclear whether these observations hold on other sepsis prediction datasets, higher standards should be sought after when constructing models for high-risk clinical decision making.

**References.**

- [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 (2) (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.
- [4] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011. <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- [5] Schaffter T, Buist DSM, Lee CI, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Netw Open. 2020;3(3):e200265. doi:10.1001/jamanetworkopen.2020.0265
- [6] Matheny ME, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine. JAMA. 2020;323(6):509–510. doi:10.1001/jama.2019.21579