The role of context in the prediction of acute hypotension in critical care

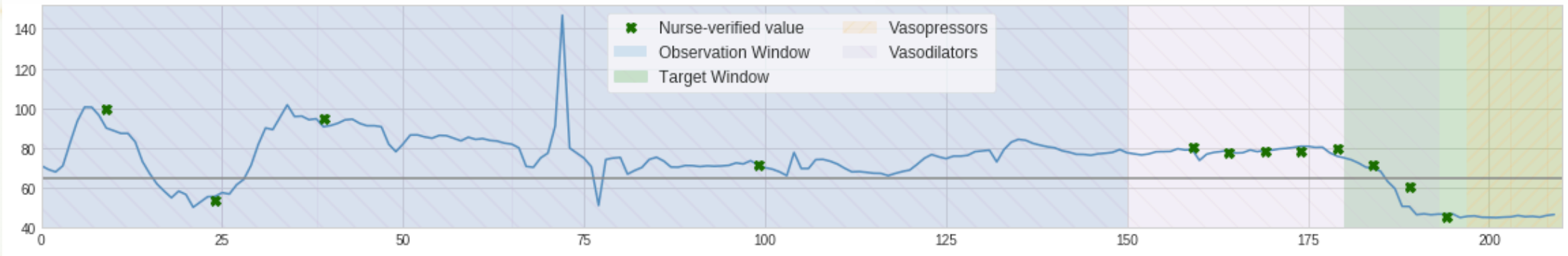
Niranjani Prasad *(Princeton University)* and Konstantina Palla *(Microsoft Research Cambridge)*

Motivation

- In this work, we explore how choices in data selection and algorithm design can confound ML for clinical decision-making, focusing on **acute hypotension events (AHEs)**
- An AHE—sustained low mean arterial pressure—can result from sepsis, neurogenic disorders, shock from sudden fluid loss, or directly from cardiovascular disorders. It is therefore associated with high mortality; timely prediction can allow clinicians to intervene before further decompensation.
- For AHE prediction in intensive care using four off-the-shelf classifiers, we consider the trade-offs inherent in cohort and feature selection, challenges in evaluation in this domain.

Data Selection and preprocessing

- 4,518 admits to ICU are filtered from **MIMIC III** and the corresponding waveform database.
- Each is classified as either patient admissions with one or more AHEs, or a control group.
- Fixed-length segments are extracted from each admission: target window for which we want to predict AHE risk, preceded by gap Δ, and observation window taken as classifier input.
- This observation window comprises key vitals, temporally aligned fluids or vasoactive drugs administered, and demographic data.

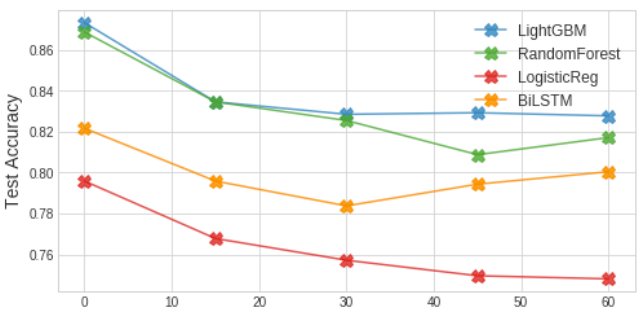


Experiments

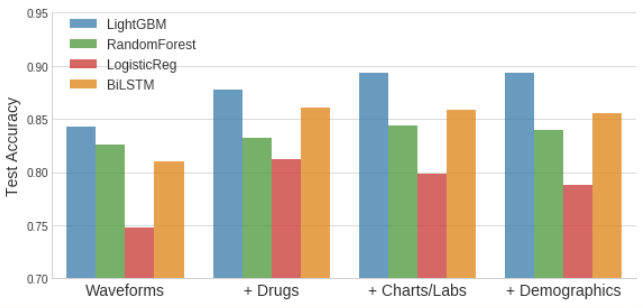
We compare the classification accuracy of four baseline models trained with different gaps Δ using waveform input alone, and with addition of clinical context, fixed Δ.

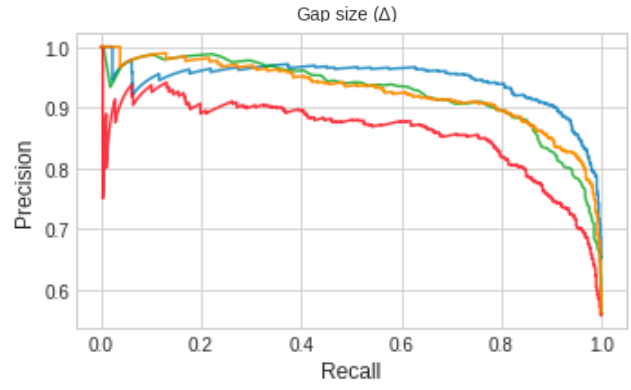
- Accuracy consistently decreases with large Δ but plateaus after $\Delta = 30$, suggesting that a reasonable estimate of AHE risk can be achieved with just a short segment of data from the start of an admission.
- Inclusion of drugs consistently improves test accuracy, while the gains from vitals measurements and lab test time series, as well as demographics, are more modest.

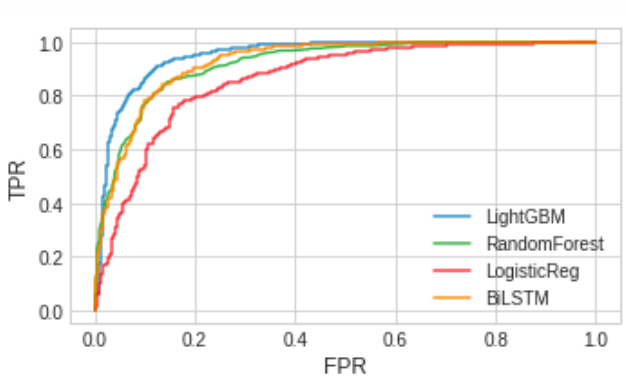
Effect of prediction gap size on classifier performance



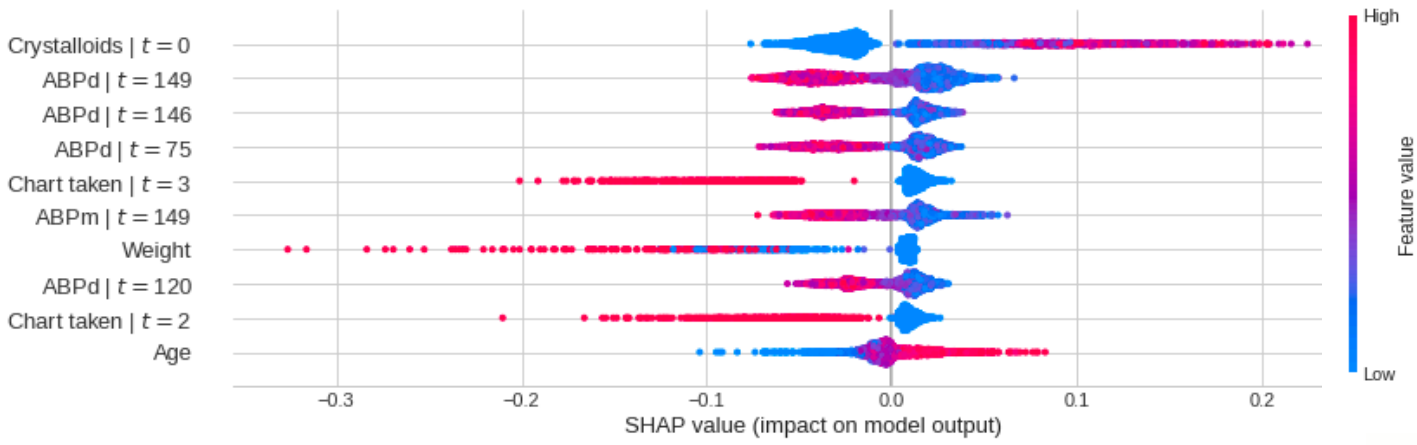
Effect of feature inclusion on classifier performance







Analysis



Shapley values show that influential features are dominated by recently measured ABP, as expected. However, the top feature is administration of crystalloids at start of observation. The model simply identifies patients already known to be at high risk and so for whom fluids are initiated early.

Errors are hence dominated by censoring from preventative fluids, or absence of fluids despite deterioration. Standard performance metrics therefore provide a limited view of model usefulness; evaluation in relation to the underlying clinical protocol determining data collection as well as decisions in data curation is crucial.