UNIVERSITY OF VIRGINIA DATA SCIENCE INSTITUTE



INTRODUCTION

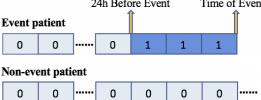
This project focuses on top-10 features comparison of acutely ill patients in the cardiac ward at the UVA Health System. Around 5% Patients in the cardiac ward deteriorate and are transferred to the Intensive Care Unit (ICU). Failure of responding to deterioration puts the patients at great risk.

My goals are to explore and visualize the top-10 features distribution: (1) between patients with events and patients without events (2) between correctly-predicted event patients and incorrectly-predicted event patients.

DATA

DATA DESCRIPTION

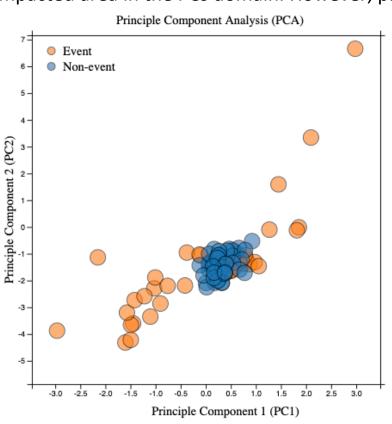
The data includes vital signs, lab results and ECG-derived data from 8,105 acute care patient admissions. A subset is used here. The top-10 features are selected based on feature importance derived from random forest.

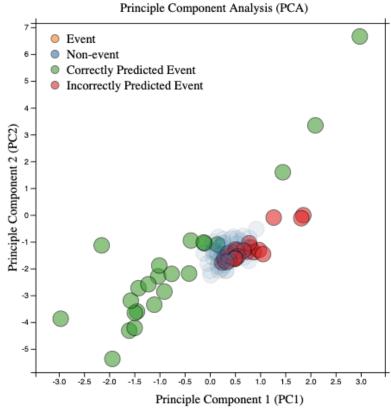


If a patient has never been transferred to ICU during stay, the records were all labeled as "0"; Otherwise, the records were labeled as "1" 24 hours prior to transfer.

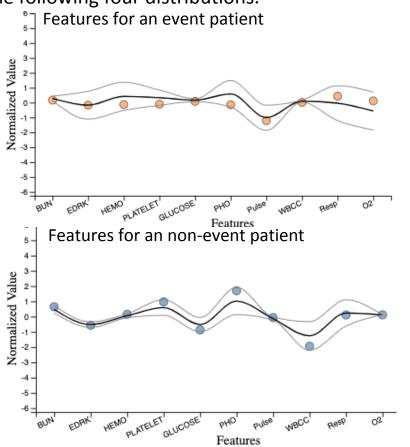
RESULTS

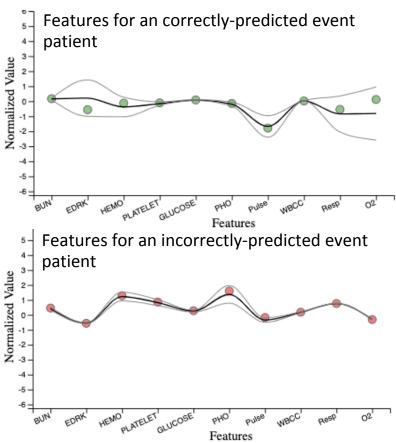
The top two principle components (PCs) are also used for comparison. Patients without events locate in a compacted area in the PCs domain. However, patients with events spread out across whole range of PCs.





 Mean and standard deviation (SD) were calculated for four categories, which are different from each other. For a new record, the potential to be an event or not could be obtained by comparing the feature distribution with the following four distributions.





CONCLUSION

- PCA can be an effective way to separate patients who have events and those who do not have events.
- Mean and SDs are different for the four situations, which can help roughly define a record as an event or not.