beatDB: A Large Scale Waveform Feature Repository

Franck Dernoncourt[1], Kalyan Veeramachaneni[1], Una-May O'Reilly[1]

[1] Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
[francky, kalyan, unamay,]@csail.mit.edu

Abstract

A great majority of the effort is spent assembling the data and formulating the features, while, rather ironically, the model building exercise takes relatively less time.

beatDB aims at radically shrinking the time of large scale investigations by judiciously pre-computing beat features which are likely to be frequently used.

In this poster we present beatDB structure and use beatDB for a concrete research study: predicting acute hypotensive event with blood pressure.

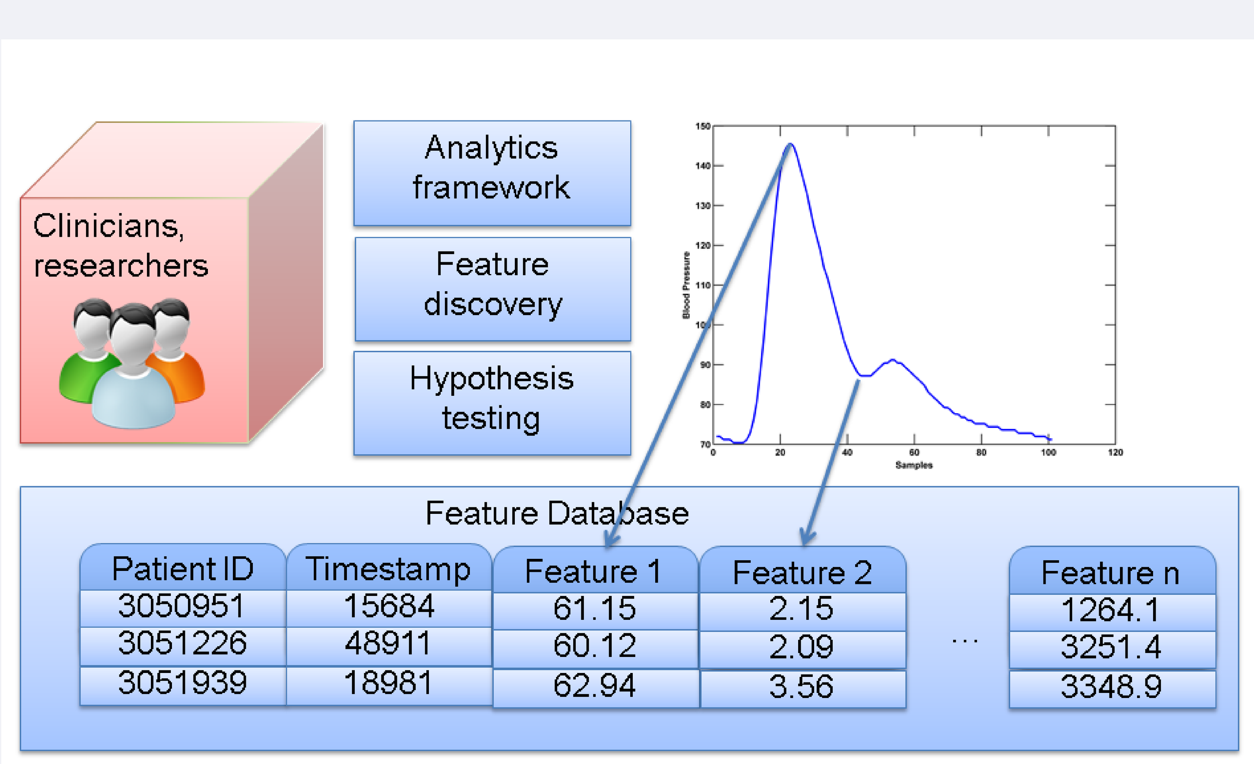
1. Introduction / Motivation

For typical physiological waveform studies, researchers usually follow this process:

- define a study group within which they designate case and controls.
- extract the group's waveforms
- filter the signals
- pre process them
- extract features
- iteratively execute evaluating and interpreting a pre-selected machine learning algorithm with metrics such as area under the curve and analyses such as variable sensitivity.

A typical study, even with modest quantities of patients, can therefore take 6 to 12 months. beatDB aims at taking care of the data wrangling part and subsequently cutting down significantly the time required to perform the study.

2. beatDB structure

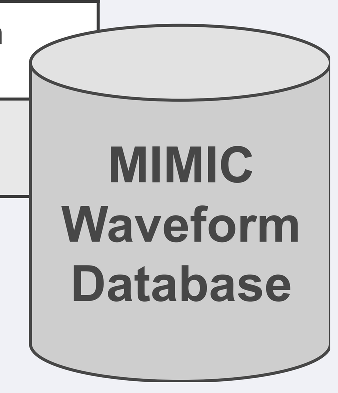


Features :

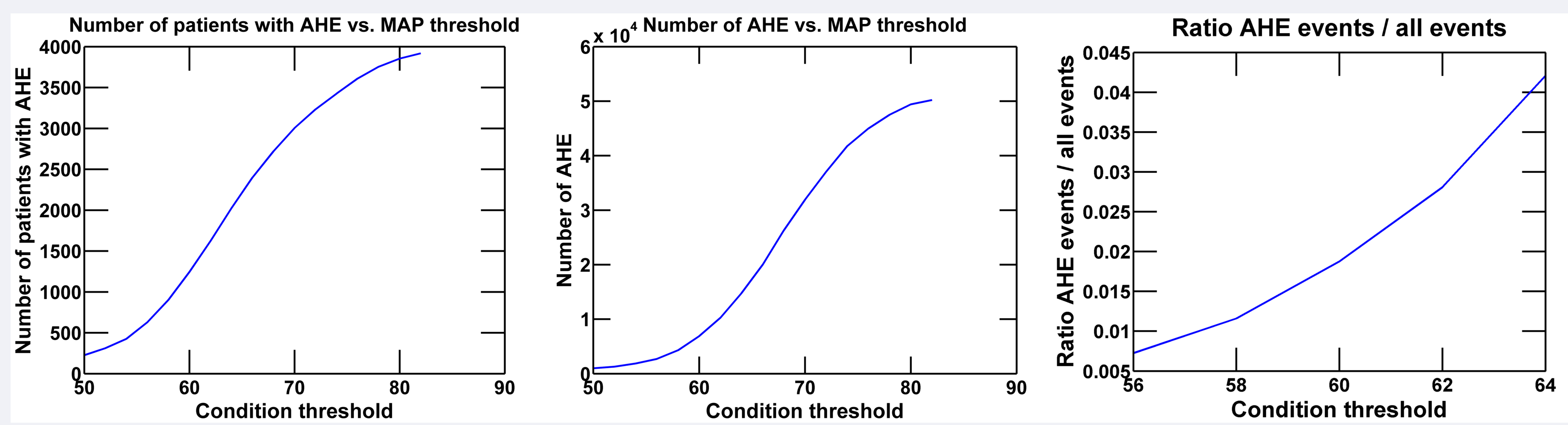
- Mean of signal
- Root-mean-square level (RMS) of signal
- Standard deviation of signal
- Kurtosis of signal
- Skewness of signal
- Systolic blood pressure (Max ABP)
- Diastolic blood pressure (Min ABP)
- Pulse pressure
- Duration of each beat
- Duration of systole
- Duration of diastole
- Pressure area during systole
- Crest factor (Peak-to-average ratio)
- Mean arterial pressure (MAP)

3. Results

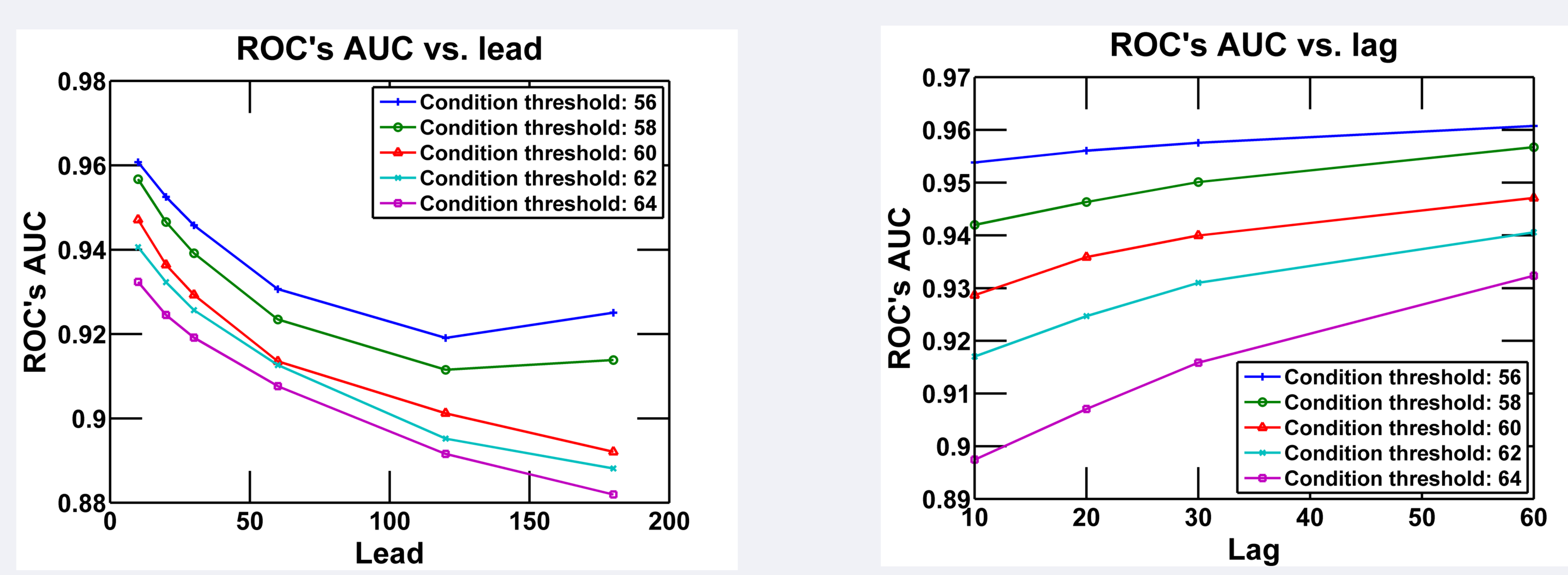
Feature	Value
Total size	4 Terra Bytes
Waveform types	22
Signal sampling frequency	125 samples/ sec
Number of samples	500m
Number of patients	6190



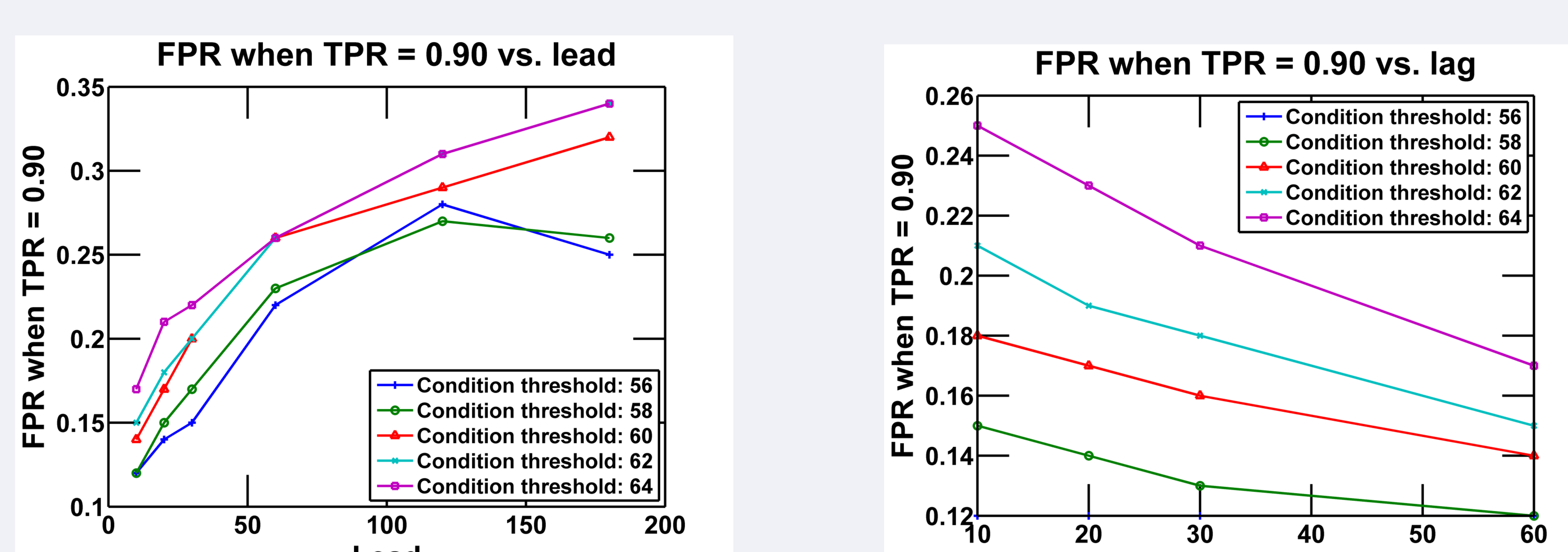
Parameter names	Parameter choice
Condition definition: window duration	30 minutes
Condition definition: variable	minute average MAP (mean arterial pressure)
Condition definition: variable threshold	[56, 58, 60, 62, 64] mmHg
Condition definition: frequency	90% minimum
Prediction definition: lag	[10, 20, 30, 60] minutes
Prediction definition: lead	[10, 20, 30, 60, 120, 180] minutes
Prediction definition: covariates	root-mean-square, kurtosis, skewness, systolic blood pressure, diastolic blood pressure, pulse pressure, beat duration, systole duration, diastole duration, pressure area during systole, standard deviation, crest factor, mean, mean arterial pressure.
Sub-aggregation window	1 minute
Sub-aggregation function	Mean
Aggregation window	Prediction definition: lag
Aggregation functions	mean, standard deviation, kurtosis, skew, trend
Machine learning algorithm	logistic regression
Hypothesis evaluation metric	AUC (area under the ROC curve)



(Left): Number of patients with AHE cases as we change the threshold. (Center) Total number of AHE cases present as we change the threshold for the condition. (Right): Balance between the AHE and non AHE cases as the thresholds are varied



Discriminatory strength in terms of Area Under the Curve for the classifiers for different condition threshold with prediction lead and lag sweeps.



False positive rates achieved for a true positive of 90% as we changed the prediction problem parameters and the thresholds for AHE determination

5. Conclusions

beatDB contains many features characterizing blood pressure beats.

It allowed us to investigate how the definition of an acute hypotensive event impacts the number of event in our dataset, and run 120 experiments to predict those events.

The successful use of beatDB shows that the feature repository along with the aggregation framework can drastically cut down the time required for such studies.

References

JH Henriques and TR Rocha. Prediction of acute hypotensive episodes using neural network multi-models. In Computers in Cardiology, 2009, pages 549–552. IEEE, 2009.

[2] JX Sun, AT Reisner, and RG Mark. A signal abnormality index for arterial blood pressure waveforms. In Computers in Cardiology, 2006, pages 13–16. IEEE, 2006.

NIPS 2013, Lake Tahoe, Nevada, United States