ECG Risk Stratification Using Multiple Instance Learning

Divya Shanmugam, Davis Blalock. Jen J. Gong, John Guttag
{divyas, dblalock, guttag}@mit.edu, {jengong}@berkeley.edu

Goal

Identify patients at highest risk for cardiovascular death based on their ECG signal to efficiently avert adverse outcomes.

Challenges

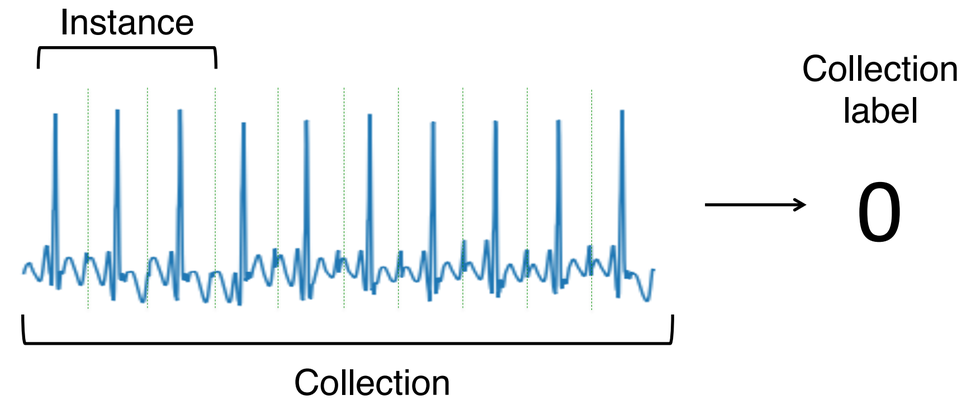
* Huge number of heartbeats share one global label
* Most beats are not indicative of outcome

How do we accurately classify signals, given no outcome annotations for heartbeats?

Background

Tackle problem using *multiple instance learning* (MIL)

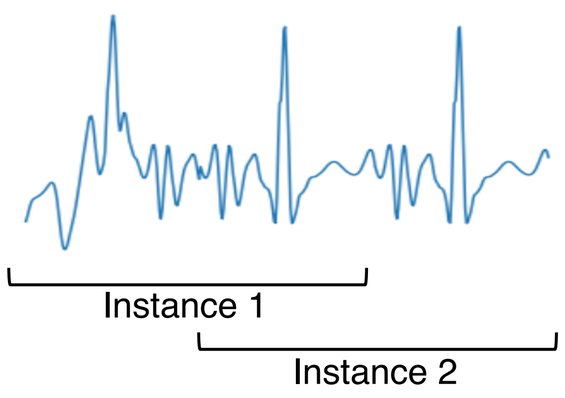
MIL: Classify collections of instances, given only collection-level labels



Method

(1) Separation

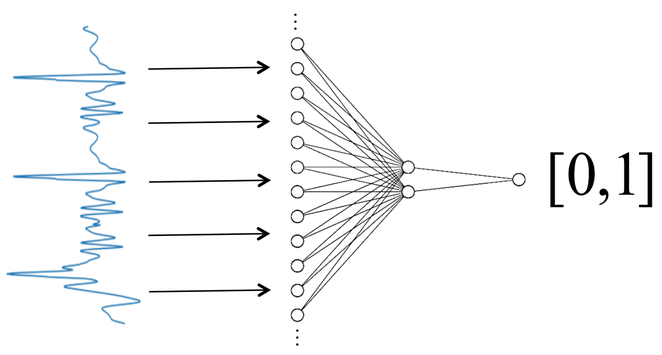
How do we separate collections into instances?



(2) Classification

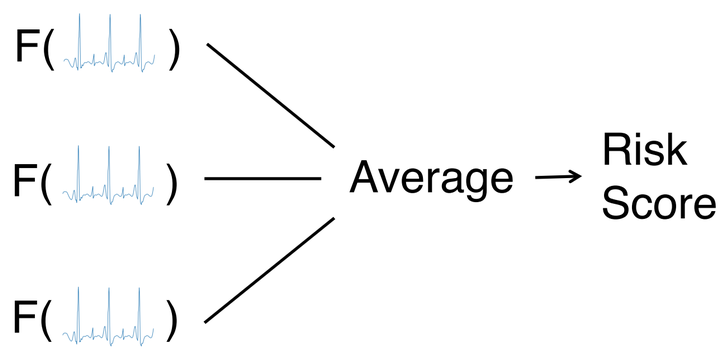
How do we map each instance to a collection label?

$$F(x) \rightarrow [0,1]$$



(3) Aggregation

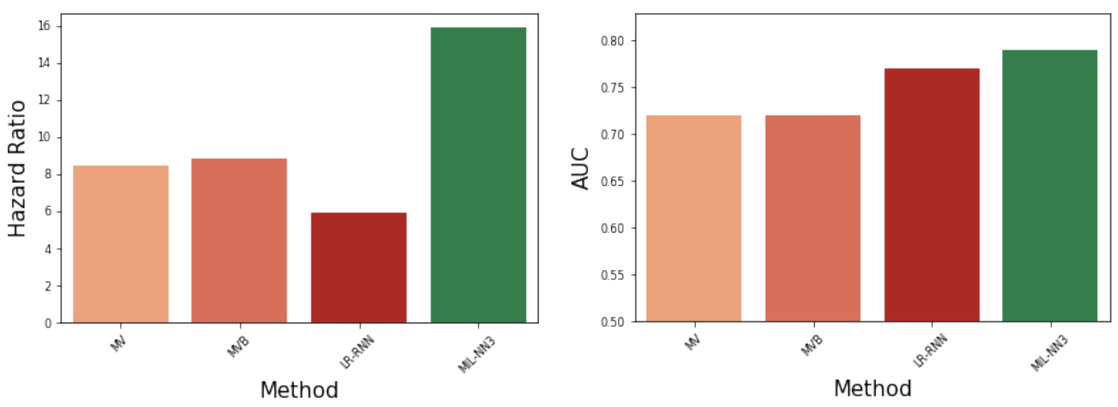
How do we assign a collection label based on instance predictions?



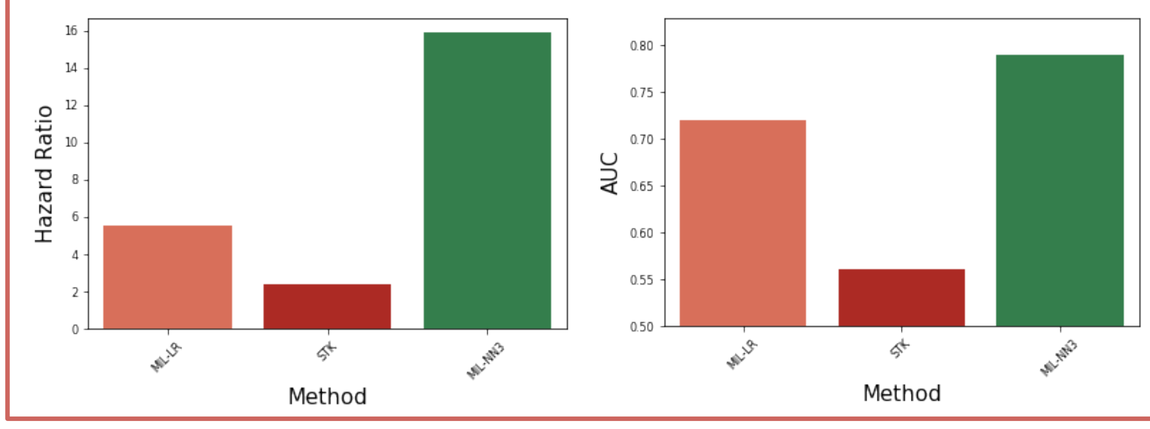
Results

1. Our method outperforms baseline methods.

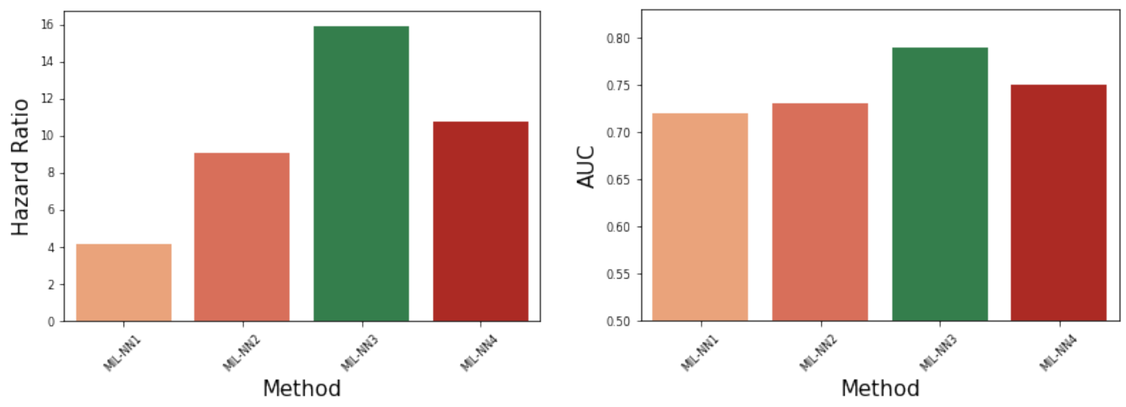
A. Existing ECG metrics



B. Existing MIL methods



2. The MIL approach reveals that 3 beats are more informative than 2, in contrast to a common assumption in ECG literature.



3. The proposed approach is more practical than existing MIL methods for datasets at this scale.

* 6000 patients, 1000 instances/patient
* Largest MIL benchmark dataset is two orders of magnitude smaller
* Kernel-based methods do not scale to this dataset size

Conclusions

* Powerful framework for risk stratification
  * Generalizes easily to other biometric signals
  * Strong performance despite training on weak labels
* Contributions
  * MIL Methods: Validation in new problem space
  * ECG Risk Metrics: *Learned* relationship between beats

Funding: This work is supported by Wistron and NSF Grant #1122374.