ECG Risk Stratification Using Multiple Instance Learning

Divya Shanmugam, Davis Blalock, John Guttag
{divyas, dblalock, guttag}@mit.edu

Goal

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

Challenges

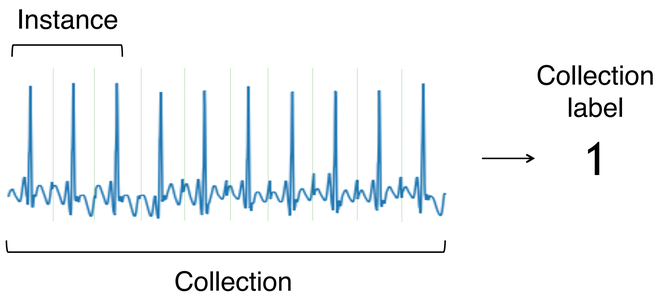
- Large number of heartbeats share one global label
- Most beats are not indicative of outcome

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

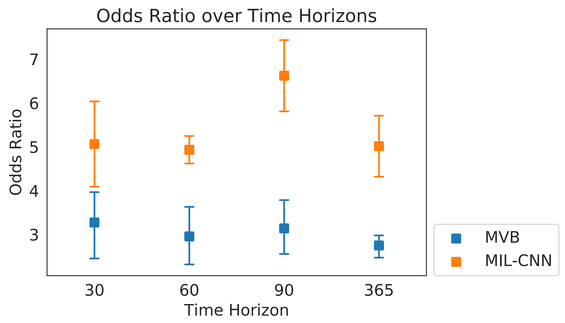


Takeaway

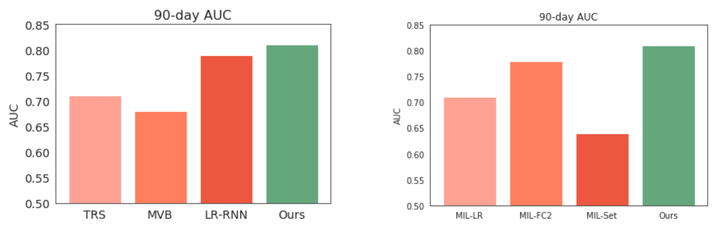
We propose a general-purpose method for incorporating very long time series into risk models and produce a state-of-the-art ECG-based risk score for cardiovascular death.

Results

- **Outperforms state-of-the-art ECG-based risk metric across multiple time horizons.**



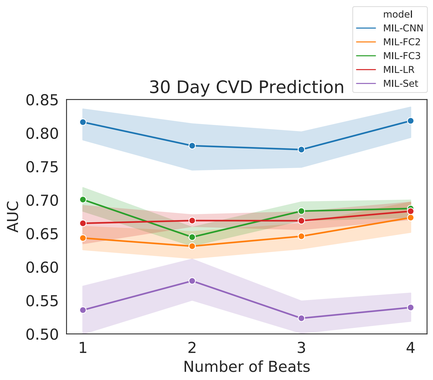
- **Outperforms existing CVD risk metrics and MIL methods.**



Robustness

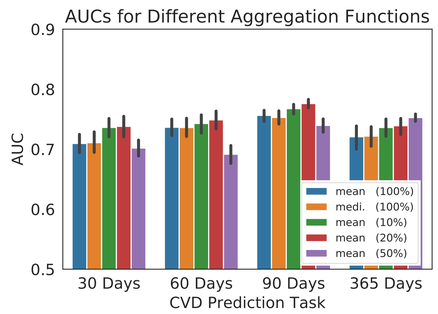
(1) Instance choice

MIL-CNN is robust to instance choice across 1-4 adjacent beats



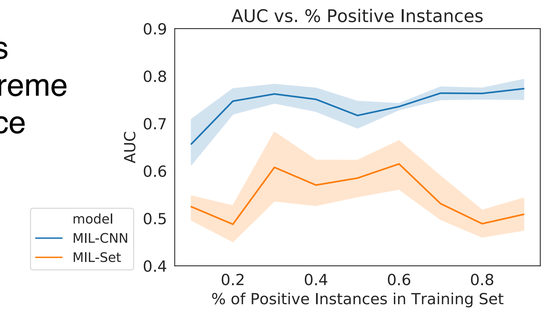
(2) Aggregation function

Choice of aggregation function has no significant effect on model performance



(3) Class imbalance

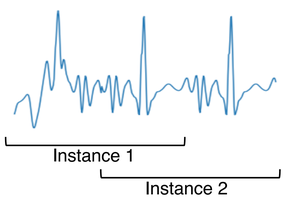
Model remains resilient to extreme class imbalance



Method

(1) Separation

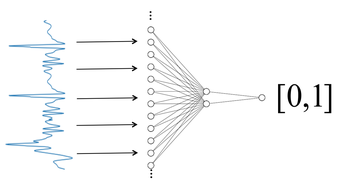
How do we separate collections into instances?



(2) Classification

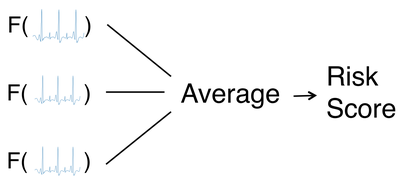
How do we map each instance to a collection label?

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



(3) Aggregation

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



Conclusion

- Method to include time series in risk models
- Strong performance despite training on soft labels
- Validates application of MIL to risk stratification

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