Learning from Few Subjects with Large Amounts of Voice Monitoring Data

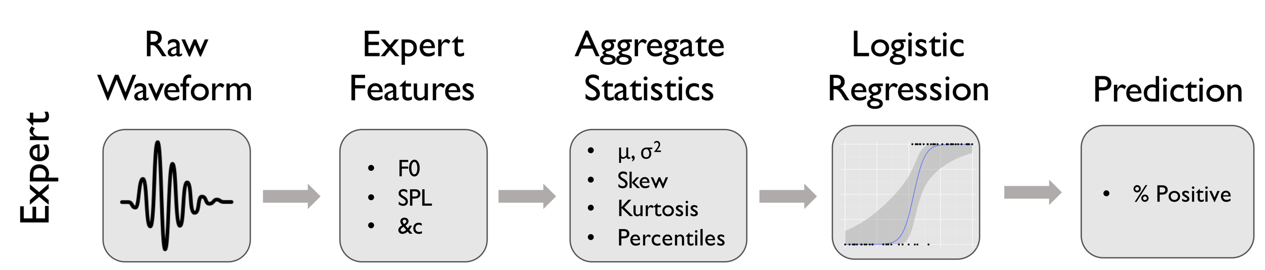
Jose Javier González Ortiz

John V. Guttag
Robert E. Hillman
Daryush D. Mehta
Jarrad H. Van Stan
Marzyeh Ghassemi

Medical Time Series

Challenges

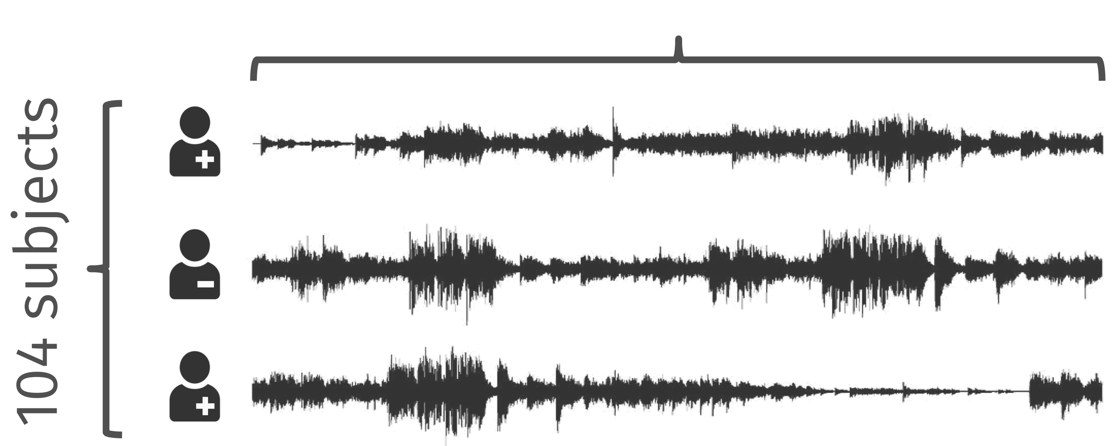
- Often few subjects and large amounts of data
  → Easy to overfit to subject-specific traits
- No obvious mapping from signal to features
  → Feature engineering is labor intensive
- Usually, we only have subject-level labels
  → In many cases, no way of getting annotations

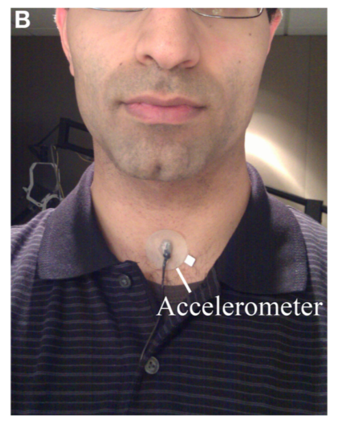


Example: Voice Monitoring

- Voice disorders affect 7% of the US population
- Data collected with neck-placed accelerometer
- 52 patients with vocal fold nodules & 52 controls

1 week = ~4 billion samples/subject

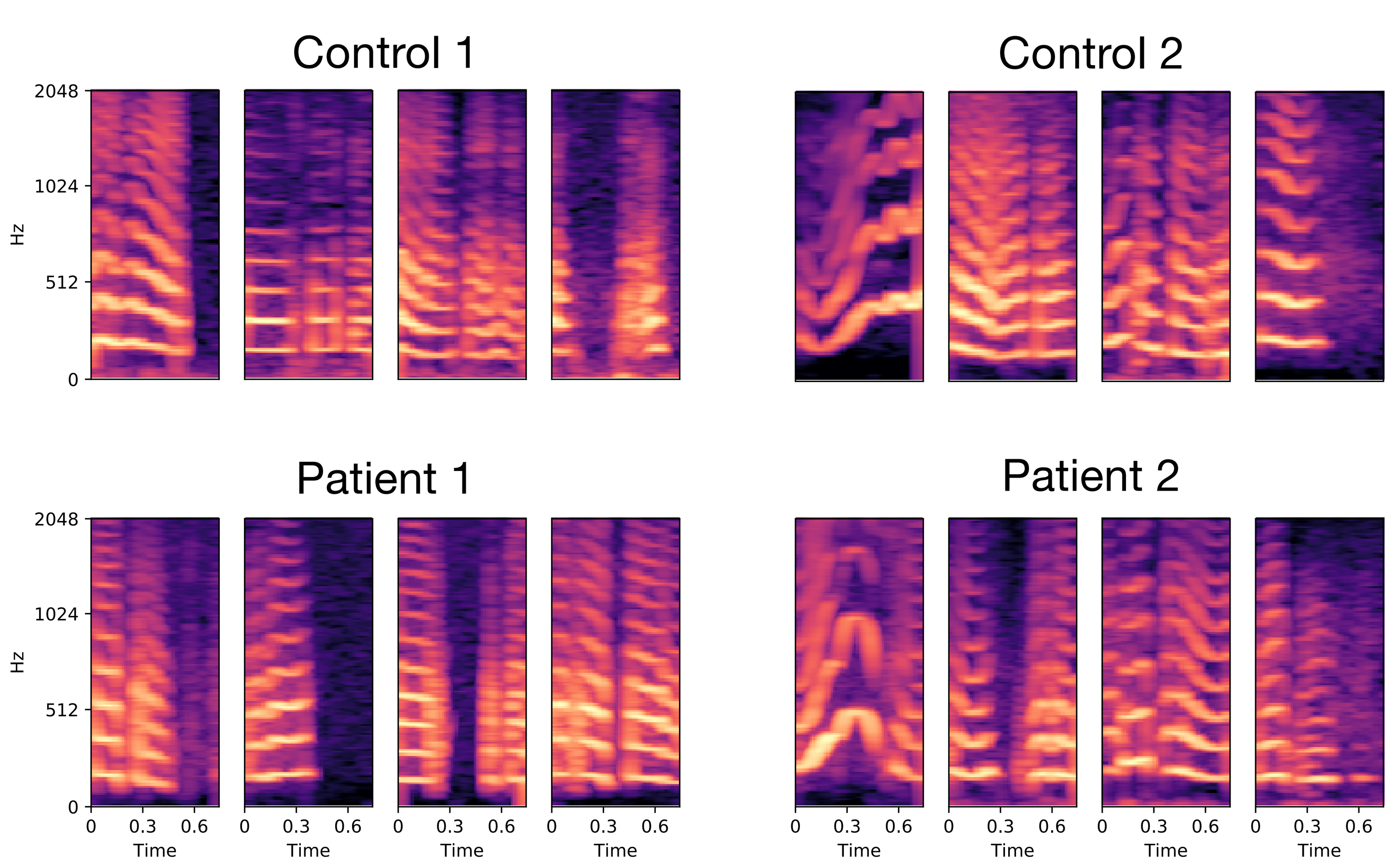




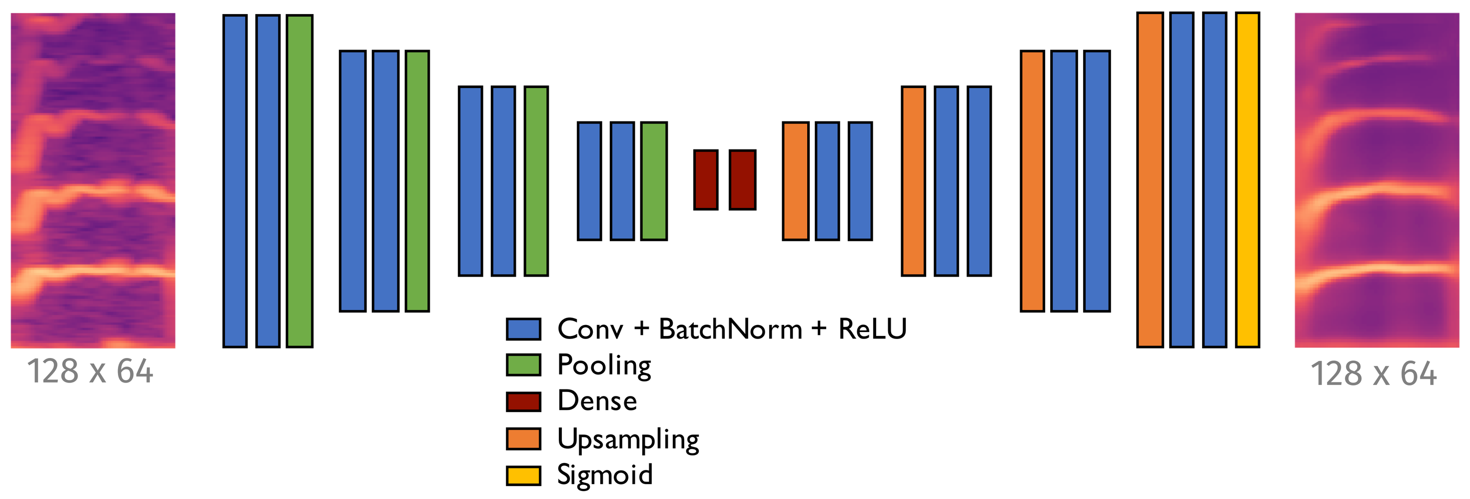
We automatically learn useful features for large time series data, reducing the need for laborious task-specific feature engineering.

Method

- Segment signal into fixed length windows
- Compute time-frequency representation

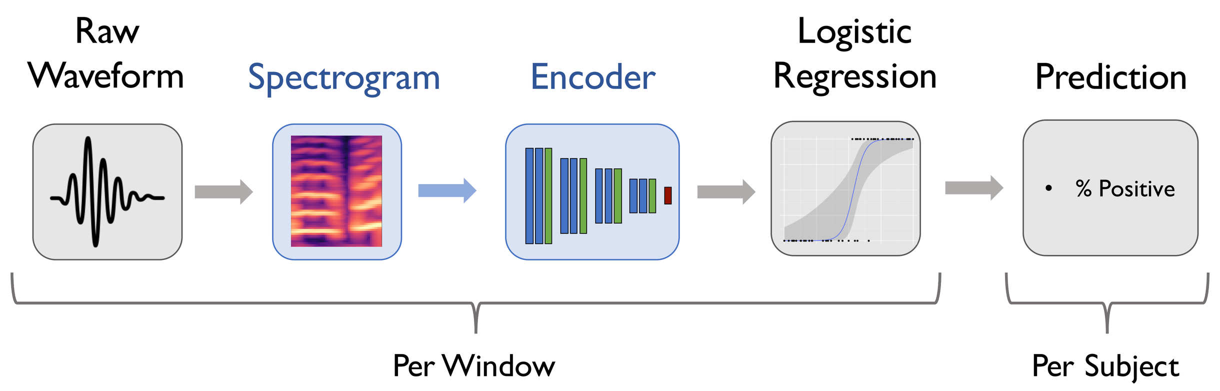


- Unsupervised feature extraction using autoencoder



Classification Results

Task: *Classifying between patients and controls*



- LR model on learned features with subject labels
- Aggregate prediction using % positive windows
- Previous work relied on expert driven features[1]

		AUC	Accuracy
Expert	Train	0.70 ± 0.05	0.71 ± 0.04
	Test	0.68 ± 0.05	0.69 ± 0.04
Ours	Train	0.73 ± 0.06	0.72 ± 0.04
	Test	0.69 ± 0.07	0.70 ± 0.05

Comparable performance **without** task-specific feature engineering!

Voice Usage Results

Task: *Does the amount of vocalization impact patients & controls differently?*

- We answer that using the same learned features
- Predict recent voice utilization from windows
- Statistically significant difference between predictions for patients and controls (p = .04)

[1]. Marzyeh Ghassemi et al. Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: initial results for vocal fold nodules. IEEE Trans. Biomed Engineering

This research program is supported by the Voice Health Institute, the NIH National Institute on Deafness and Other Communication Disorders under Grant P50 02032, and by "la Caixa" Foundation Fellowship.

josejg@mit.edu