Predicting Breathing Severity Scores for Asthma

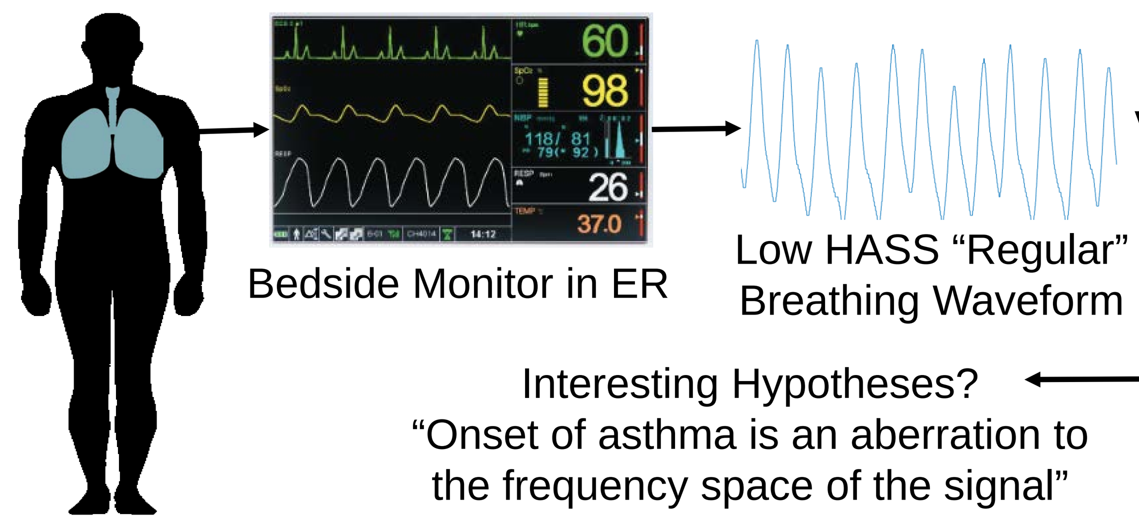
Sahil Loomba[1], Benjamin Matthews[2], Joseph Mooney[1], John Osborne[1]

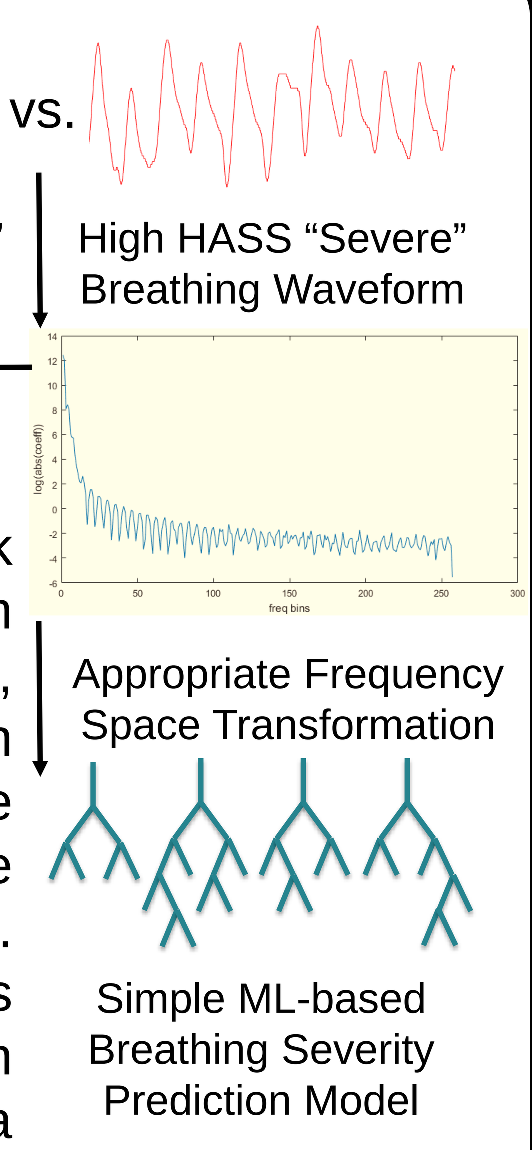
[1]Wyss Institute for Biologically Inspired Engineering at Harvard University, [2]Boston Children's Hospital

Abstract

An important component of Project Abbie is to develop a system that automatically and accurately monitors a subject's breathing severity. This could be helpful in diagnosing the onset of an asthmatic attack, and monitoring recovery post clinical interventions. We collect breathing waveform signals from bedside monitors for multiple patients admitted to the ER after an asthma attack, along with HASS breathing severity scores assigned by clinicians. Since these are close-to-periodic temporal signals, we use a litany of frequency representation techniques to extract appropriate features of the signal. We hypothesize that an increase in breathing severity is an aberration in the signal spectrum. More specifically, (1) we extract the multi-resolution wavelet spectrum followed by a multifractal analysis, (2) hypothesize and validate that an onset of asthma attack is the transition of the breathing waveform from mono to multifracticality. On top of this compact representation, we build simple machine learning models for a predictive score severity algorithm. We perform with a high 95% accuracy on the classification version of the problem, and with a low root-mean-squared-error of 1.18 on the regression version. We propose to expand our datasets with more supervised data collection, and transfer the bedside monitor model to waveforms recorded using sensor technology.

Concept





When a patient undergoing an asthmatic attack is admitted to the ER, the breathing waveform recorded by a bedside monitor can be extracted, transformed and analyzed to make a prediction algorithm of breathing severity, and to generate interesting hypotheses that provide some mechanistic insight into the onset of anaphylaxis. We use various frequency transform techniques to first establish the best feature representation of the breathing waveform. Nurses often assign a a "HASS" score, varying on a continuous scale from 5 (regular) to 13 (severe), which is then used as labels for a simple machine learning (ML) based prediction algorithm.

Methods

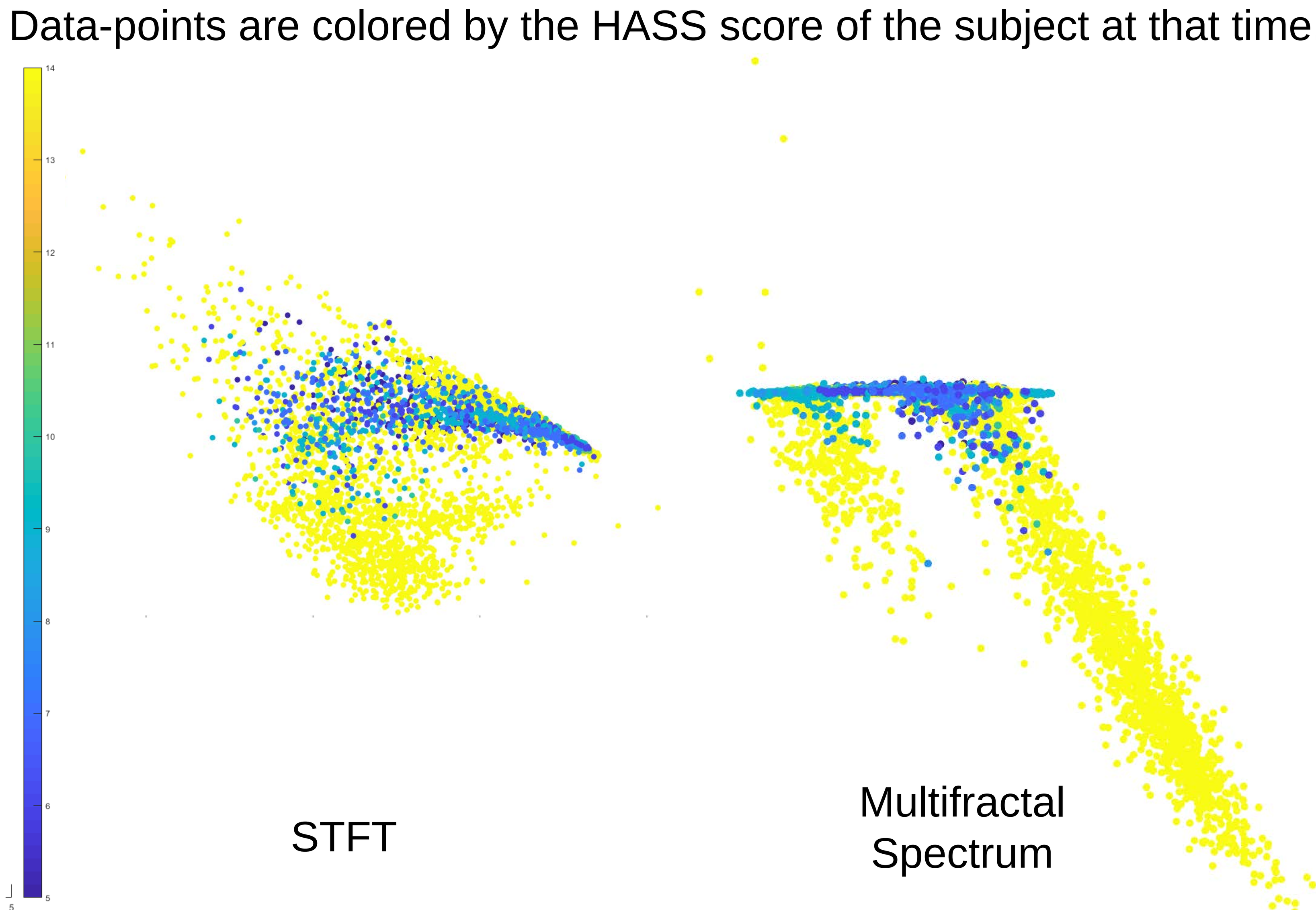
The representation of a waveform signal is essentially a periodic time-series that indicates the chest movement of the subjects, which leads us to believe that a frequency representation would suit this waveform. We try a litany of such transforms, each with an increasing level of complexity. But the most interesting representation is that of a fractal system. Signals that show self-similarity at different scales can be compactly represented in terms of its multifractal coefficients. Many natural objects (like snowflakes) and signals (like ECG) show such self-similarity. While a monofractal system can be represented by exactly one such coefficient, a multifractal system requires other higher-order terms. **We hypothesize that the onset of an asthmatic attack could be the transition of breathing from being a monofractal system to a multifractal one.**

Transformation	Usually applied to	Output	Notes
Fourier Series Fit (FS)	Periodic signal	Primary frequencies	Assumes one dominant frequency
Discrete Fourier Transform (DFT)	Aperiodic signal	Spectral coefficients	Most common frequency transform
Short-term Fourier Transform (STFT)	Noisy aperiodic signal	Averaged spectral coefficients	Signal is binned over time-windows
Discrete Wavelet Transform (DWT)	Fractal signals	Multiresolution spectral coefficients	Signal is binned over both time and frequency space
Multifractal Spectrum Extraction (MS)	DWT of a signal	Multifractal coefficients	Higher-order coefficients are 0 for monofractal systems

To build our ML-based model, we tried a library of techniques, but a simple Random Forest Regression (RFR) model for predicting continuous HASS scores, and Random Forest Classification (RFC) model for predicting a binarized HASS score "low" vs. "high", gave a strong performance and optimistic results.

Results

The methods are applied on data collected from recovery trajectories of 5 asthmatic subjects, sampled into a total of 6042 data-points. To visually discern the goodness of a representation space, PCA was used to reduce the dimensionality of STFT spectrum to 2. (Other frequency spaces are ignored here since they were less superior in our comparative analyses.) The multifractal spectrum is visualized by plotting the first and second multifractal coefficients on x and y axes. Data-points are colored by the HASS score of the subject at that time.



To build the ML-based algorithm, we do a 5-fold cross-validation. We present our result here for the best feature space: a composite of the original waveform shape descriptors, STFT spectrum, and the multifractal spectrum. **We obtained a low room-mean-squared error of 1.18 and a high classification accuracy of 95%.**

Conclusions

1. **Interpretability:** Looking at the visualization of the multifractal spectrum, which gives us merely 2 highly uncorrelated features, we observe a clear horizontal manifold where low-HASS patients lie (monofractal system) and two vertical manifolds where the high-HASS patients lie (multifractal system). **The fracticality hypothesis stands validated.** This provides us key mechanistic insight into breathing during an asthma attack.
2. **Quality:** With a simple ML-based algorithm, we were able to obtain high performance on predicting breathing severity. This showcases the goodness of the frequency transformed feature representations.
3. **Usability:** An RFR/RFC model would also be an easier model to embed into a real-time wearable monitoring device in the future.

Future Directions

1. **Model Expansion:** This study was performed on only 5 subjects with "clean" datasets. That is, it was ensured that the monitor sensors were correctly attached at all times, and the HASS scores were reliably assigned by nurses. We would like to expand our dataset to more subjects, to better capture the landscape of asthmatic breathing waveforms.
2. **Onto Wearables:** The real win of this algorithm would be for it to work on data collected through a wearable sensor, and not a bedside monitor. Although we expect the waveform characteristics to remain the same, we would like to transfer our model onto a wearable tech.
3. **Research:** Our hypothesis on asthma is fascinating, and needs a deeper study for a more complete understanding of why the system transitions from mono to multifracticality.

Acknowledgments

This project has been a highly collaborative endeavor across many people. We would like to thank the staff at Boston Children's Hospital for helping us with procuring and processing bedside monitor data. We are thankful to Christoph Kanzler for important early analysis and model building on this dataset, and to the rest of the Project Abbie team for constant encouragement and inputs to algorithm development.

We are grateful to the Wyss Institute and KeepSmilin4Abbie Foundation for their generous grant support.