Contact: Serge Assaad: serge.assaad@duke.edu
Mark Chen: mark.j.chen@duke.edu
Guangshen Ma: guangshen.ma@duke.edu

Automated Classification of Vascular Anomalies

Problem Statement

The goal of this project is to automate the diagnosis of vascular anomalies from Doppler Ultrasound data to both improve diagnostic accuracy and reduce physician time spent on simple diagnoses.

Data

- 38 patients, 5 Doppler ultrasound recordings per patient (provided by Duke Hospital)
- Binary patient labels (**healthy** = 0; **unhealthy** = 1)

Task

- Predict patient labels from given recordings

Figure 1: Example patient setup of ultrasound device and data collection

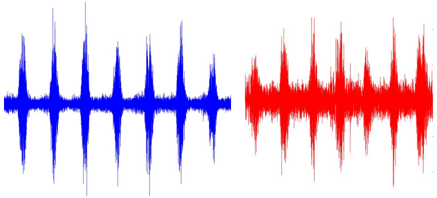
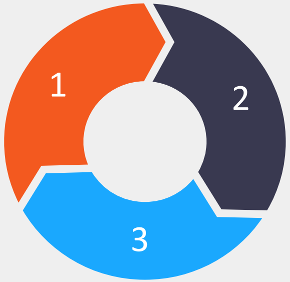


Figure 2: Sample audio signals from ultrasound machine. (a) Healthy signal and (b) Unhealthy signal

Methods

Feature extraction

Find key pieces of information from signals



Build model

Use features to discriminate between healthy and sick patients

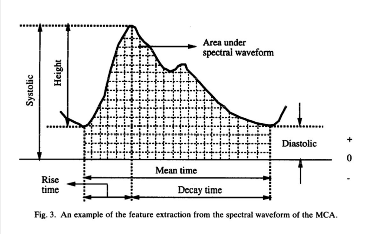
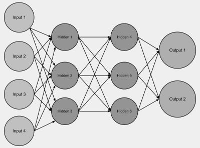
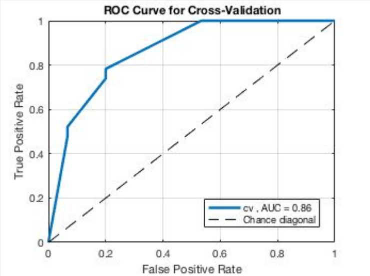


Fig. 3. An example of the feature extraction from the spectral waveform of the MCA.



Test model

Evaluate model performance using standard Machine Learning metrics



Results

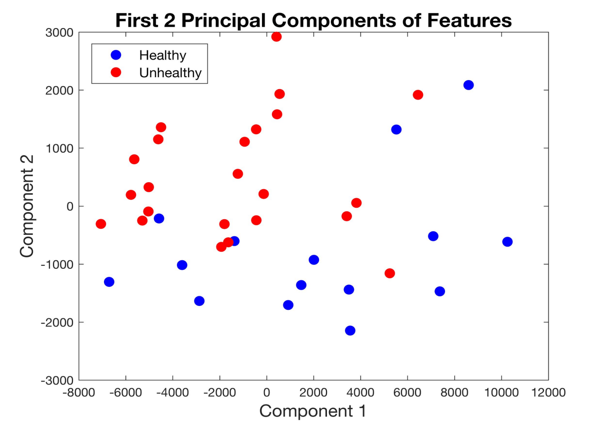


*Figure 3: Separation between healthy and unhealthy patients, as evidenced by first 2 principal components of 12-dimensional feature space (**97.3%** variance explained)*

List of Features:

Bandwidth (99% occupied), Mean Frequency, Peak Frequency (high), Peak Frequency (low), Total Harmonic Distortion, Power, Rise Time, Decay Time, Mean Time, Diastolic Level, Systolic Level, Total Peak Height,

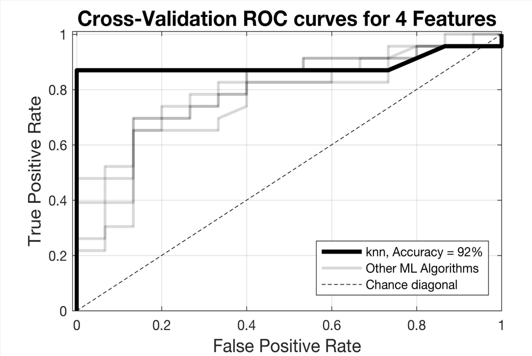


Figure 4: Cross-Validation ROC curves for 5 different ML algorithms (7 features selected to optimize AUC)

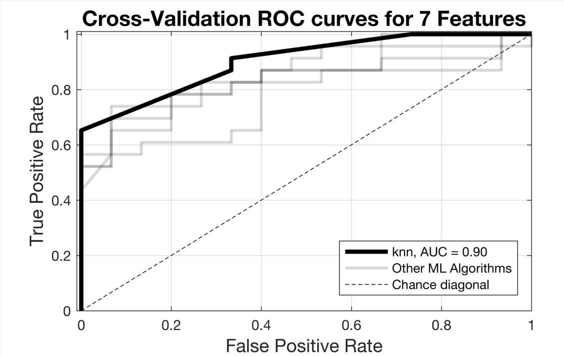


Figure 5: Cross-Validation ROC curves for 5 different ML algorithms (4 features selected to optimize %accuracy)

Table 1: Performance and model specs for different metrics

Metric	AUC	%accuracy
Performance	**AUC = 0.90**	**%accuracy = 92%**
Best features	Mean frequency, Peak frequency (high), Power, Total Harmonic Distortion, Normalized Systolic/Diastolic Difference, Systolic/Diastolic Ratio, Rise/Mean Time Ratio	Peak Frequency, Total Harmonic Distortion, Systolic/Diastolic Ratio, Rise/Mean Time Ratio
Best model	3-Nearest-Neighbor	4-Nearest-Neighbor

Conclusions

Best performance

- Best AUC performance: 3-Nearest-Neighbor classifier
- Best %accuracy performance: 4-Nearest-Neighbor classifier

Limitations

- Only 38 patients (model may not be general enough)
- Only used 10-fold cross-validation for performance metrics (not enough data for a test set to be meaningful)
- Selected features which optimized metrics for KNN (because of $O(2^n)$ runtime for brute-force feature selection)

Future Directions

Acquire more data

- Access databases of patient signals to make model more scalable
- Generate signals through ultrasound simulation (ANSYS,FIELD II)

Predict vessel shapes

- Create dictionary of signal models corresponding to vessel shapes
- Deduce vessel shape from audio signals using simulation data

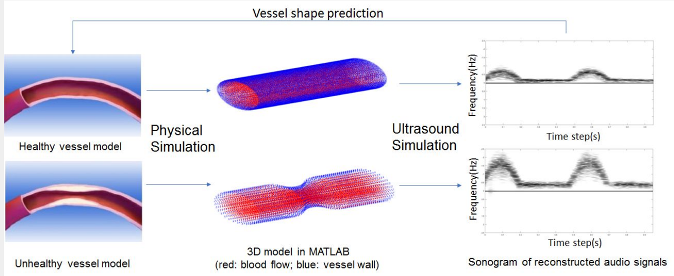


Figure 6: Ultrasound audio signal simulation process

Improve current methods

- Explore more complex features using **Mel Frequency Cepstral Coefficients (MFCC)** and **Topological Data Analysis (TDA)**
- Use more sophisticated ML techniques such as **dictionary learning** and **deep learning**