Automated Diagnosis of Vascular Anomalies

Serge Assaad, Mark Chen, Guangshen Ma
PI: Wilkins Aquino, Ph.D., Leila Mureebe, M.D., Kyle Burris
Duke University, Durham, NC

Introduction
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 or unhealthy = 1)

Task
- Predict patient labels from given recordings

Results

Feature Extraction
- Examine both temporal and frequency features
- Use MATLAB to implement feature extraction algorithms

Model Training
- Used various feature subsets to train Machine Learning (ML) models
- Trained different ML models, including: K-Nearest-Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Gradient Boosting, and Naïve Bayes

Model Testing
- Performed 10-fold Cross-Validation (CV) on entire data set to generate a Receiver Operating Characteristic (ROC) curve
- Selected optimal classifier using 2 performance metrics:
  - Area Under Curve (AUC) (best: AUC = 1, worst: AUC = 0.5)
  - Percent accuracy (best: accuracy = 100%, worst: accuracy = 50%)

Table 1: List of 12 features extracted from signals
<table>
<thead>
<tr>
<th>Temporal</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rise time</td>
<td>Bandwidth (99% occupied)</td>
</tr>
<tr>
<td>Decay time</td>
<td>Mean frequency</td>
</tr>
<tr>
<td>Mean time</td>
<td>Peak frequency (high)</td>
</tr>
<tr>
<td>Diastolic level</td>
<td>Peak frequency (low)</td>
</tr>
<tr>
<td>Systolic level</td>
<td>Total Harmonic Distortion</td>
</tr>
<tr>
<td>Total peak height</td>
<td>Power</td>
</tr>
</tbody>
</table>

Figure 3: Main temporal features

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

Figure 4: Separation between healthy and unhealthy patients, as evidenced by first 2 principal components of 12-dimensional feature space (97.2% variance explained)

Table 2: Performance and model specs for different metrics
<table>
<thead>
<tr>
<th>Metric</th>
<th>AUC</th>
<th>%accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>AUC = 0.90</td>
<td>%accuracy = 92%</td>
</tr>
<tr>
<td>Best features</td>
<td>Mean frequency, Peak frequency (high), Power, Total Harmonic Distortion, RI, SDR, RMTR</td>
<td>Peak Frequency, Total Harmonic Distortion, SDR, RMTR</td>
</tr>
<tr>
<td>Best model</td>
<td>3-Nearest-Neighbor</td>
<td>4-Nearest-Neighbor</td>
</tr>
</tbody>
</table>

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 more 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

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

Build user-friendly product
- Design a Graphical User Interface for integration into health systems

Acknowledgements
We thank Dr. Wilkins Aquino, Dr. Leila Mureebe, and Kyle Burris for their stellar guidance and constant support.
We are also immensely grateful to Dr. Paul Bendich, Dr. Ashlee Valente, Ariel Dawn, and Kathy Peterson for making the Data+ program possible.
We would also like to show our gratitude to Dr. Chris Tralie and Dr. Kyle Bradbury for their helpful insights.