Heart Defect Detection

Frank Dattalo
dattalo.2@osu.edu

Amy Miao
miao.106@osu.edu

Abstract

Traditionally, cardiovascular disease diagnosis requires highly trained professionals administering tests in order to properly detect anomalous heart sounds, which are an early indicator of heart disease. In this paper, we show how signal processing techniques and machine learning can be applied to provide free alternatives to otherwise expensive tests. Given raw audio of heartbeats, we attempted to predict whether a given subject has an anomalous heart sound. On our dataset, our model was able to predict with 85.5% accuracy and an F1 score of 64.2% whether a given subject has an anomalous heart sound. We believe that with further iteration, machine learning based methods will provide a free alternative to the traditional cardiovascular disease diagnosis methods.

1 Introduction

According to the World Health Organization, cardiovascular disease is the number one cause of death globally. [7] In 2017, cardiovascular disease was responsible for the death of an estimated 17.7 million people. [7] Furthermore, the World Health Organization estimates that 57 million people died in 2017; thus cardiovascular disease represented 31% of all deaths globally in 2017.

Cardiovascular disease represents a category of diseases in which a heart abnormality is present, the most common of which includes heart valve problems, arrhythmias, heart attacks, and strokes. [1] It is estimated that 80% of cardiovascular disease related deaths are due to heart attacks and strokes. [7] Additionally, the World Health Organization stresses that people who have a cardiovascular disease, or possess many high-risk factors such as hypertension, diabetes, or hyperlipidaemia, need early disease detection, health counseling, and medication for treatment to be effective. [7]

Two of the biggest barriers regarding treatment of cardiovascular diseases are monetary cost and symptom detection. The World Health Organization estimates that over 75% of global cardiovascular disease related deaths occur in low to middle income countries. Given a median household income of $59,039 [6] and mean out-of-pocket diagnosis cost of $5,541 [4] in the United States, it is clear to see that the unaffordable cost of diagnosis is a large barrier to treatment for many people affected by

cardiovascular disease. Furthermore, the Center of Disease Control estimates that only 27% of those with cardiovascular disease were aware of their symptoms early enough to seek treatment. [3]

Traditional approaches to initial discovery and diagnosis of cardiovascular diseases require trained professionals listening for anomalous heart sounds including heart murmurs and extrasystolic sounds as these sounds are often early indicators of heart diseases. [2] Electrocardiograph[1] approaches to anomaly detection use variance in peak R wave distance, or RR interval, to determine whether heart anomalies are present. [5] However, both aforementioned methods of cardiovascular disease detection are inaccessible to the low-income patients who need them most given each method's high monetary cost. Therefore, we would like to use signal processing and machine learning methods to provide a free alternative to the traditional and electrocardiograph methods of heart anomaly detection in order to reduce cardiovascular disease related deaths globally.

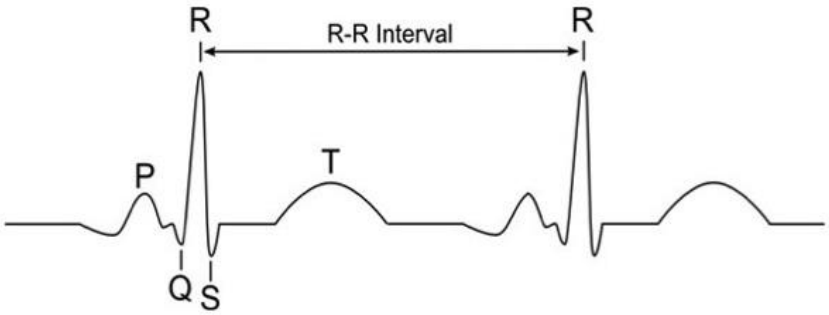

Figure 1: RR Interval of an ECG Signal

2 Data

For our research, we used the dataset collected for Bentley et. Al's Heartbeat Classification Challenge. [2] The dataset was gathered by trained professionals during a clinical trial using digital stethoscopes sampling at 4000 Hz. The original dataset contained audio files of normal heartbeats, heartbeats with murmurs, and heartbeats with extrasystolic sounds. For this paper, we will consider all abnormality under one category, and thus only consider whether a heartbeat is normal or anomalous. This dataset contains 461 raw audio files ranging in length from 1 second to 10 seconds. Of the original dataset, 141, or 31%, of the subjects bear heart abnormalities while the remaining 320, or 69%, do not.

3 Methods

In order to predict whether a given subject from our dataset has an anomalous heart sound, we first transformed the data through a series of steps. After all data transformation, we constructed a binary classifier to predict the labels for unseen subjects.

[1] Electrocardiograph (ECG) - machine which uses electrodes placed on skin to record electrical activity of heart over time

Filtering

First, we applied a low pass filter to each audio file at 195 Hz. Doing so eliminated external noise above this frequency such as talking and feedback from sphygmomanometers[2], which ultimately resulted in higher classification performance.

Normalizing

Next, we normalized each audio file to have an absolute peak amplitude of 1 by scaling each sample of the signal by the absolute maximum of the original signal. Normalization of signal amplitude allowed us to tune our feature extraction parameters independent of variability in signal amplitudes. Empirically, our results proved to be slightly better with normalization than without. The signal normalization computation is described in Equation 1 below.

$$S_{normalized} = \left\{ \frac{s_i}{\max|s_j| \in S} \middle| s_i \in S \right\} \tag{1}$$

S represents the original signal while $S_{normalized}$ represents the resulting normalized signal. The resulting value $S_{normalized}$ contains samples strictly within the interval [-1, +1].

Feature Extraction

After filtering and normalizing our data, we mapped each preprocessed audio file to a low-dimensional feature vector representative of the original file. First, we used moving window median outlier detection to find the number of outlier samples within a given signal. As the name suggests, moving median outlier detection uses the median of a moving window to determine whether any given point is an outlier with respect to the median measure. Using this technique, signals which make less sudden movements over time have a lower number of outliers with respect to signals which fluctuate more often. Thus, this technique aims to separate normal heartbeats which are more consistent over time from anomalous heartbeats which are less consistent over time. The moving median calculation is scaled by the length of the audio sample in order to provide an outlier calculation which is independent of sample length. Using Equations 3 and 4, the moving median computation calculation is expressed Equation 2 below.

$$m(S, w) = \frac{\sum_i^n false(3\sigma - M(S, w, i) \leq s_i \leq 3\sigma + M(S, w, i))}{\|s\|} \tag{2}$$

$$M(S, w, i) = median(S(w - i : i + w)) \tag{3}$$

$$false(x) = \begin{cases} x = true & \rightarrow 0 \\ x = false & \rightarrow 1 \end{cases} \tag{4}$$

[2] Sphygmomanometer - blood pressure meter

S represents the filtered and normalized audio signal, *w* represents the window size hyperparameter, and σ represents the standard deviation of the signal *S*. It can also be seen that the true window for the moving median calculation is $2w$ as the window centered evenly around any given sample *i*.

We then turned our attention to approximate entropy measures of the signal. Similar to moving median outlier detection, approximate entropy defines a measure of the regularity of a given signal. However, unlike moving median outlier detection, the approximate entropy metric does not consider regular fluctuations in a signal to be anomalous. Given a subsequence length and a signal, approximate entropy convolves all subsequences of the specified length with all other same-length subsequences within the signal. Each convolution measures the distance in an embedding space from one subsequence to the next. Equations 5 through 9 illustrates this computation below.

$$ApEn(S, w, t) = \Phi(S, w, t) - \Phi(S, w+1, t) \tag{5}$$

$$\Phi(S, w, t) = \frac{\sum_i^n \log\left(C(S, i, w, t)\right)}{n - w + 1} \tag{6}$$

$$C(S, i, w, t) = \frac{\sum_j^n true\left(dist\big(S(i: i+w), S(j: j+w)\big) < t\right)}{n - w + 1} \tag{7}$$

$$dis(s1, s2) = max|s1_k - s2_k| \tag{8}$$

$$true(x) = \begin{cases} x = true \ \rightarrow 1 \\ x = false \rightarrow 0 \end{cases} \tag{9}$$

Here *S* represents the signal, *n* is the length of the signal, *w* is the window length, and *t* is a threshold hyperparameter. It can be seen that sequences with low approximate entropy are naturally periodic.

Finally, as a proxy measure for the RR interval of an ECG signal, we used the variance in distance between local amplitude maxima beyond a hyperparameter threshold. By doing so, we were able to achieve a metric which mimicked that of the traditional electrocardiograph approaches. Figure 2 below illustrates the local maxima of an absolute audio signal. The peaks, marked with upside down triangles, were then used to compute the final variance in distance measure.

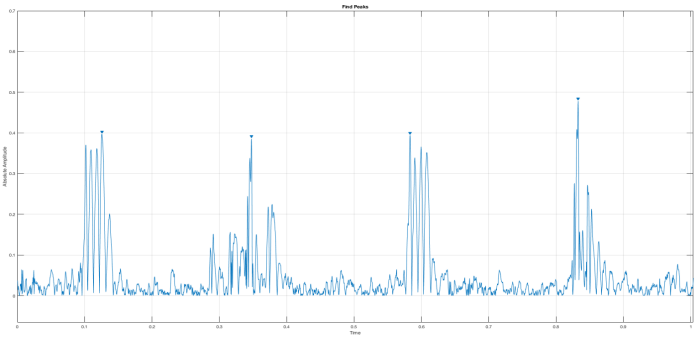

Figure 2: The Find Peaks Calculation of an Absolute Signal

Training

After mapping each preprocessed audio signal to a low dimensional feature vector, we began to train our classifier on 70% of the dataset. We selected a three layer feed forward neural network as our classifier as we believed that the nonlinear feature combinations which the network would learn were vital for anomaly detection. Then, we used cross validation with 15% of the dataset to tune our model's hidden layer size. Equations 10 through 13 below details the decision function learned by the three layer feed forward neural network.

$$neuralnet(X) = step(predict(X)) \tag{10}$$

$$step(x) = \begin{cases} x < 0.5 \rightarrow 0 \\ else \quad \rightarrow 1 \end{cases} \tag{11}$$

$$predict(X) = \sigma(\langle \sigma(\langle X, w_1 \rangle + b_1), w_2 \rangle + b_2) \tag{12}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{13}$$

Here, X represents the input feature vector, w_1 and w_2 represent the learned matrices which map from the previous layer to the next, and b_1 and b_2 represent the learned neuron bias vectors which are applied to the result of linear transformation with the respective matrices w_1 and w_2. Finally, it can be seen that our learned function is of the form $R^n \rightarrow \{0, 1\}$.

Testing

Finally, after training our classifier, we evaluated its performance against the remaining 15% of the dataset. As a human classifier control, we each predicted the label for 70 random samples. To evaluate the performance of all predictions, we used Accuracy, Precision, Recall, and F1 Score measures. Equations 14 through 17 below outline the derivation of these evaluation metrics.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1\ Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{17}$$

4 Results

Below are the results of our classifier. We have provided the confusion matrices in Figure 3, the receiver operating characteristics curves in Figure 4, and the table of resulting evaluation metrics in Table 1. It should be noted that a positive label signals that an anomalous heart sound is present. Additionally, we have benchmarked our neural network's ability against our own as a frame of reference.

Figure 3: Confusion Matrices for Training, Cross Validation, and Testing

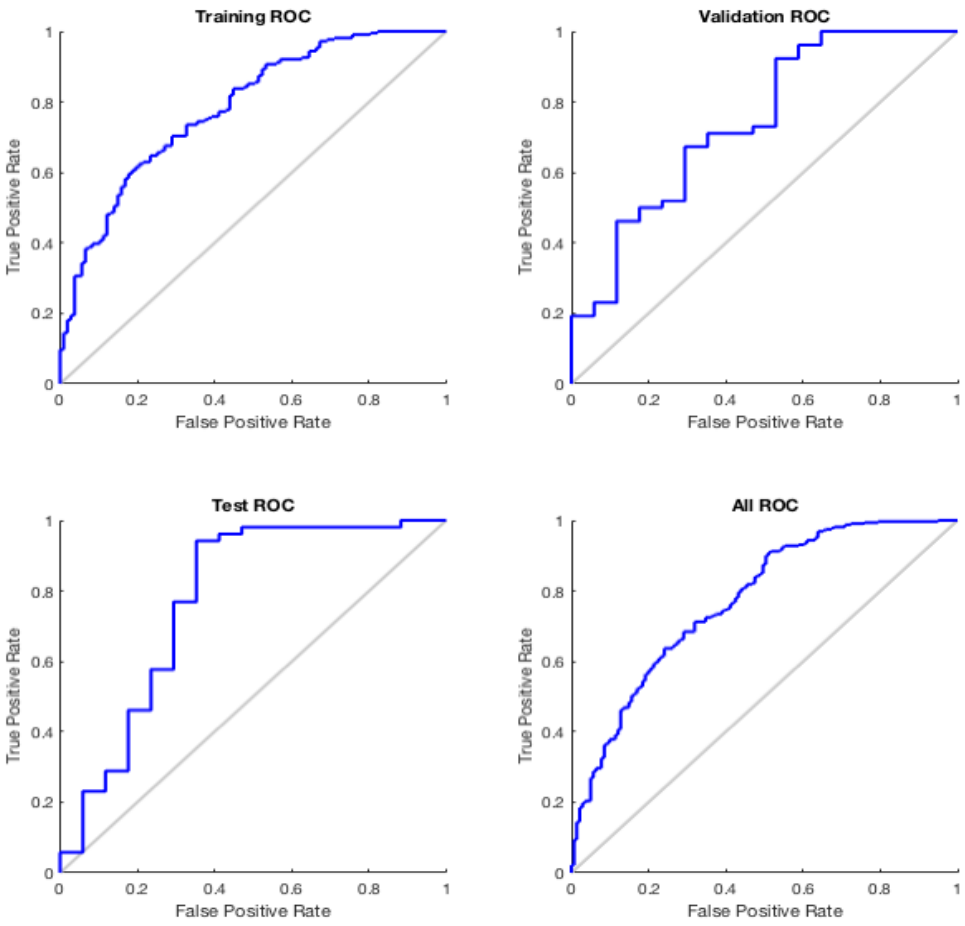

Figure 4: ROC curves for Training, Cross Validation, and Testing

Table 1: Evaluation Scores

Subject	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Neural Network	**85.5%**	**81.8%**	52.9%	**64.2%**
Amy	58.6%	35.1%	**72.2%**	47.3%
Frank	58.6%	33.3%	61.1%	43.1%

5 Discussion

As the results show, our neural network outperformed our own ability to classify heartbeats as anomalous or normal by a very large margin of 26.9%. Furthermore, the neural network surpassed our precision and F1 scores by a margin of 46.7% and 16.9% respectively. However, we were able to achieve a higher recall than our classifier at 72.2% by a margin of 19.3%. Ultimately, the neural network's predictions were very skewed towards a prediction of a normal label. This can be seen by a total of 92 abnormal

(20%) samples of the total 461 samples being mis-predicted as normal. We believe that we should collect more, preferably at least 10,000, samples of abnormal and normal heartbeats. Given a larger dataset of balanced labeling, we expect to observe a large increase in our neural network's recall metric.

Additionally, we recognize our low expertise in signal processing hindered our ability to construct and recognize meaningful features which differentiate between abnormal and normal heartbeats. During an initial model selection period, we observed our best results using neural networks. Given neural network's ability to ignore or combine multiple features in non-linear manners, we postulate that the non-performant results of other classifiers were a result of a shortsightedness in feature selection.

We expect that applying recurrent neural networks to the raw or downsampled signal might prove to be favorable. Given a larger dataset, enough computing power, and time, we believe that the learned classifier would surpass our current results. However, in the interest of time, we were unable to pursue this path as training locally was rather time consuming.

6 Conclusion

In this paper, we have shown how machine learning methods can be used to detect anomalous heart sounds, which are an early indicator of cardiovascular disease. Ultimately, we were able to create a classifier which would detect anomalous heart sounds using raw audio signals with an accuracy of 85.5%. We would like to arrange easy access to our classifier by providing a user interface over the internet. In doing so, we believe that its use would allow for earlier cardiovascular disease detection by removing the monetary barriers commonly associated with detection. We hope that the techniques which we applied here could be further iterated upon to provide a free alternative to cardiovascular disease detection and diagnosis in an effort to reduce disease related deaths globally.

References

[1] The American Heart Association. (2018, April 13). What is Cardiovascular Disease. Retrieved April 19, 2018, from http://www.heart.org/HEARTORG/Conditions/What-is-Cardiovascular-Disease_UCM_301852_Article.jsp

[2] Bentley, P. J., Nordehn, G., Coimbra, M., Mannor, S., & Getz, R. (2011, November 1). Classifying Heart Sounds Challenge. Retrieved April 13, 2018, from http://www.peterjbentley.com/heartchallenge/

[3] Center of Disease Control. (2017, November 28). Heart Disease. Retrieved April 13, 2018, from https://www.cdc.gov/heartdisease/facts.htm

[4] Dunlay, S. M., Shah, N. D., Shi, Q., Morlan, B., VanHouten, H., Long, K. H., & Roger, V. L. (2011, January 01). Lifetime Costs of Medical Care After Heart Failure Diagnosis. Retrieved April 13, 2018, from http://circoutcomes.ahajournals.org/content/4/1/68

[5] Jambukia, S. H., Dabhi, V. K., & Prajapati, H. B. (2015, July 23). Classification of ECG signals using machine learning techniques: A survey. Retrieved April 13, 2018, from https://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=7164783&tag=1

[6] US Census Bureau. (2017, September 12). Income, Poverty and Health Insurance Coverage in the U.S.: 2016. Retrieved April 19, 2018, from https://www.census.gov/newsroom/press-releases/2017/income-povery.html

[7] World Health Organization. (2017, September 29). Cardiovascular diseases (CVDs). Retrieved April 13, 2018, from http://www.who.int/cardiovascular_diseases/en/