Classifying Heartbeat Anomalies

N. Chopp
Michigan Technological University,
Houghton, MI 49931
nichopp@mtu.edu

I. Cummings
Michigan Technological University,
Houghton, MI 49931
itcummin@mtu.edu

H. Sweidan
Michigan Technological University,
Houghton, MI 49931
hisweida@mtu.edu

A. Kern
Michigan Technological University,
Houghton, MI 49931
ankern@mtu.edu

ABSTRACT

Kaggle, Inc. hosted a data set called "Heartbeat Sounds" which is comprised of 176 audio samples of heartbeat sounds that were collected via iPhone app. Of the 176 samples, 124 of them are labeled as normal, murmur, extra heart sound, and artifact. The goal of this work was to develop a classification model to determine if the individual exhibits a normal heart sound, a heart sound with a murmur, an extra heart sound, or the recording contains an artifact. After preprocessing the data and extracting features, the data was split into a 75% training set and 25% testing set. Using 10 fold cross-validation, eight models were built using the training set. After the models were built, the testing set was used to determine the best model in terms of accuracy, sensitivity, and specificity. Based on the evaluation criteria, it was determined that the Support Vector Machine, Classification Tree, and Random Forest models were the best suited to classify heartbeat sounds given the features we chose. These classification models could be implemented in an app to check heartbeat recordings for abnormalities. This software could easily be used in real time at health clinics to assist physicians and medical staff with client diagnostics.

KEYWORDS

Heartbeat Sounds, Anomaly Detection, Classification, Feature Extraction, Cross-Validation, SVM, Classification Tree.

1 INTRODUCTION

Heart disease is the number one killer in America. Almost every minute, someone will die from a heart disease-related event [1]. It is crucial to be able to identify heart anomalies early so proper preventative and diagnostic medicine can be administered to the patient. In fact, proper use of an automated external defibrillator or the use of certain medicines may be dependent on early identification of heart anomalies.

1.1 Dataset

The Heartbeat Sounds data set, listed on the Kaggle website, contains 832 heartbeat audio file samples from two sources (data sets), A and B. A was collected from the general public via an iPhone app, and B was collected from a clinical trial in hospitals using a digital stethoscope [2]. Data set A contains recordings labeled as "normal", "murmur", "extra heart sound" (e.g. extrasystole), and "artifact." Data set B contains data from the classes "normal", "murmur", and "extrasystole". Due to the fact that these data sets were collected using different measurement methods and had different class outcomes, we did not combine the data sets together. For this project we focused on data set A, which contained 124 labeled samples and 52 unlabeled samples.

1.2 Project Goal

Given the data set, we set out to develop a multi-class classification model. Our goal was to locate the heart sounds in the audio files, extract features from those sounds and classify the heart sounds into one of the listed categories.

2 BACKGROUND

The first step in addressing this challenge was to understand the structure of the recordings and the characteristics of sound recordings in each class. According to Mayo Clinic, a normal resting heart rate for adults ranges from 60 to 100 beats per minute. However, since the data did not provide information about the individual (i.e.: age, emotional state, athleticism, etc.) [3] the recorded heart rate could vary between recordings and within individual recordings. It is also important to listen to the sounds between normal lub and dub pattern. "In healthy adults, there are two normal heart sounds (lub and dub). These are the first heart sound (S1) and second heart sound (S2)" [4]. If there are other heart sounds between S1 and S2, this could indicate a murmur or extrasystole condition. Table 1 describes the characteristic sounds of each heartbeat category.

Table 1: Heartbeat Sounds

Heartbeat Categories	Description
Normal	"…[N]ormal, healthy heart sounds" that have "a clear 'lub dub, lub dub' pattern" [2].
Murmur	"Sounds as though there is a 'whooshing, roaring, rumbling or turbulent fluid' noise in one of two temporal locations: 1) between 'lub' and 'dub', or 2) between 'dub' and 'lub'" [2].
Extra heart sound	"There is an additional sound, e.g. a 'lub-lub dub' or a 'lub dub-dub'" [2].
Artifact	"…[T]here are a wide range of different sounds, including feedback squeals and echoes, speech, music and noise. There are usually no discernable heart sounds" [2].

2.1 Internal Domain Features

We considered a number of different descriptive measures (listed in Table 2) to generate feature vectors for each recording that could be classified with standard classification algorithms. We intended to apply these measures to isolated samples of S1 and S2 from each recording. Below is a table of variables that were used in similar and related work, and were considered for this heartbeat classification project.

Table 2: Previously Used Variables in Related Work [5-15]

Domain	Techniques
Time	● Signal energy envelope. ● Number of zero crossings. ● Maxima and minima ● Distance from start to maximum and minimum ● Positive and Negative Area Under the Curve
Frequency	● Fast Fourier Transform (FFT) based features. ● Power Spectral Density (PSD) based features.
Time-Frequency	● Short Time Fourier Transform (STFT). ● Wavelet Transform (WT).

2.2 External Domain Features

In addition to features used in similar work, we considered measures that were developed for other domains of research. Various indices have been developed for ecology research that uses audio recordings. Historically, these indices may have been used to extract features from biotic sounds such as bird songs. The indices extracted from the heartbeat audio files were tested for their effect on prediction of heart anomalies. The indices we used and their descriptions were [16]:

1. Acoustic Complexity Index (ACI) - an estimate of the variability of intensity
2. Normalized Difference Soundscape Index (NDSI) - an estimate of the level of anthropogenic disturbances
3. Bioacoustic Index - the area under each curve for frequency bands within a given decibel range
4. Acoustic Diversity Index (ADI) – the proportion of the signals above a threshold falling into given spectrogram bins
5. Acoustic Evenness Index (AEI) - a measurement of evenness using the Gini index for binned sections of the spectrogram

2.3 Classification Methods

In this work, we considered eight different classification models. We applied Partial Least Squares Regression, Generalized Linear Model Regression, Linear Discriminant Analysis, Neural Networks, Support Vector Machine, Classification Tree, Random Forest, and Naïve Bayes in an attempt to classify the data. These models and the parameters used are described more thoroughly below.

3 METHODS

3.1 Data Preprocessing

Before features can be extracted from the audio files, pre-processing was required to address noise issues. The Kaggle website recommended that we low-pass filter the signals at 195 Hz, as all of the frequency content of interest resides within that range [2]. Figure 1 shows the heartbeat sound before and after the filter was applied.

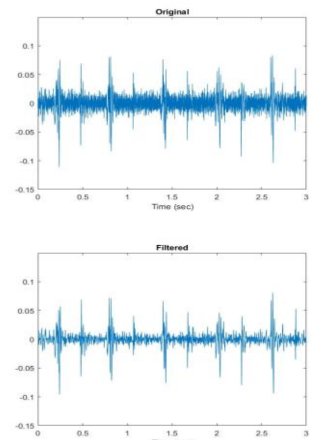


Figure 1: The impact of low-pass filtering the signal.

The feature extraction methods inspired by previous work were intended to be applied to isolated samples of S1 and S2. In an attempt to isolate these samples, the Continuous Wavelet Transform (CWT) was used, with the Morlet wavelet, to extract the location of the S1 and S2 sounds, as shown in Figure 2 [17]. Once the S1 and S2 sounds were located, the S1 systoles were then isolated from the time series by selecting the 2000 samples centered on the locations indicated by the CWT, and labeling the data in between S1 occurrences to be S2 occurrences. Given these isolated systole samples, a large number of features could be extracted.

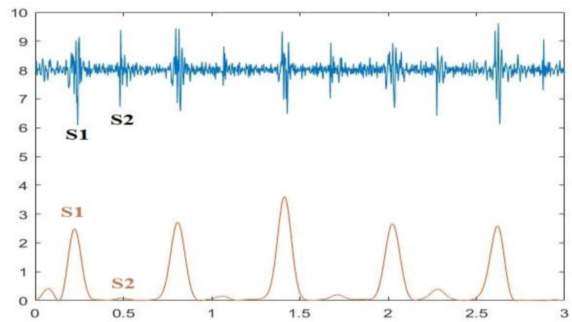


Figure 2: The use of CWT to extract S1 and S2 locations.

3.2 Feature Extraction

The method of isolating the S1 and S2 systoles using the CWT proposed above was not robust to high noise levels, and only worked on a portion of our data set. Therefore, we restricted our classification to the Acoustic Complexity Index, Normalized Difference Soundscape Index, Bioacoustic Index, Acoustic Diversity Index, and Acoustic Evenness Index, which were calculated using a sliding window method and did not rely on robust segmentation. These features were extracted from the data set after applying the 195 Hz lowpass filter. The output from the pre-processing and feature extraction phase was a flat file spreadsheet where the rows represented each sample and the columns represented a single feature or variable. Unfortunately, some of the features were not successfully extracted from a limited number of samples. This was handled using kNN to impute the missing values.

3.3 Data Exploration and Visualization

Figure 3 displays boxplots showing the distributions of the five extracted features. As you can see from the boxplots, the NDSI the medians are all relatively close (just above 0) and there is very little range in values from feature ADI.

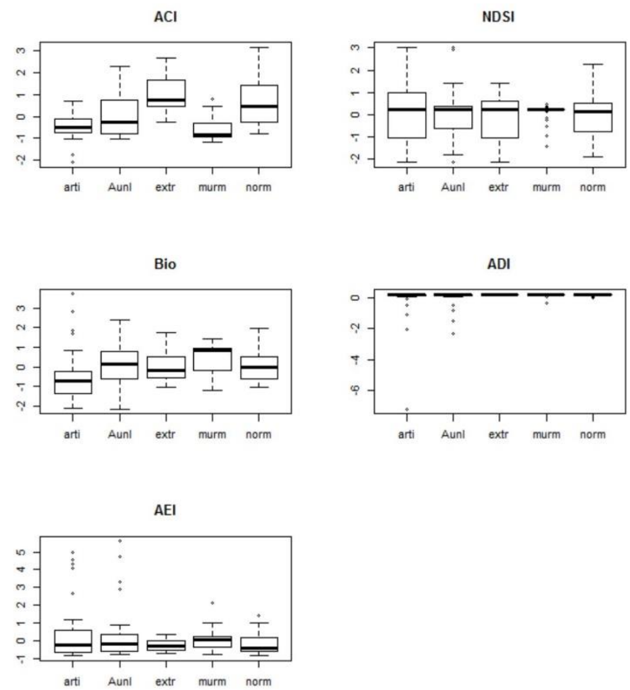


Figure 3: Boxplots of extracted features

Figure 4 shows histograms showing the distributions of the five extracted features. As you can see from the histograms, they are not normally distributed. ADI has is highly left skewed and AEI is highly right skewed.

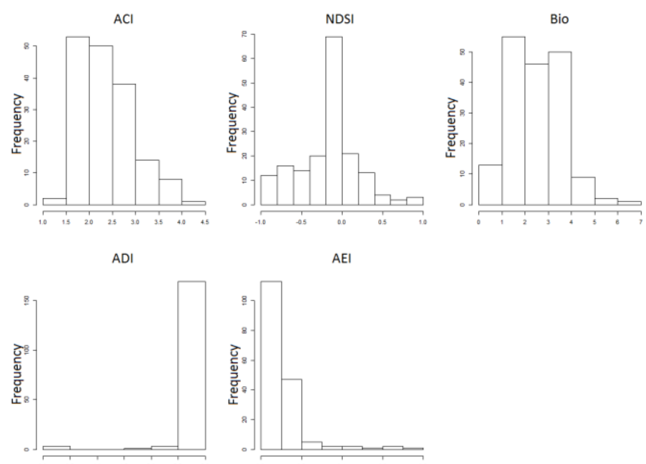


Figure 4: Histograms of extracted features.

3.4 Model Building

Given the goal of the project, we determined the best approach would be to train several different models on the labeled data available and select the highest performing model based on the criteria of: accuracy, sensitivity, and specificity. For building the models, the data was split into a training and testing set; 75% of the data was used for training and 25% was reserved for the model evaluation (testing). Due to unequal class balance, this split was stratified so that the appropriate distribution of classes were represented in the training and testing sets (see Figure 5). Within the training data, 10-fold cross validation was used to prevent model overfitting.

Figure 5: Data splitting.

The first model considered was *Partial Least Squares* (PLS) regression, which extends standard *Least Squares* (LS) regression to avoid the weakness in LS for correlated features [18]. "...PLS linear combinations of predictors are chosen to maximally summarize covariance with the response. This means that PLS finds components that maximally summarize the variation of the predictors while simultaneously requiring these components to have maximum correlation with the response" [18]. The model was allowed 1 to 5 components, and this parameter was selected using cross-validation [19]. A *Generalized Linear Model* (GLM) was also considered. The GLM picks the linear model coefficients that maximize the log-likelihood of the data given its class [20]. But, it also includes a regularization term [20]. The regularization term is parameterized (α) such that Lasso or Ridge regularization can be obtained, or some variant in between, and weighted with λ [20]. The values of λ that were tried were: 0.0100, 0.0575, 0.1050, 0.1525, & 0.2000, and α varied from 0 to 1 by 0.2. A *Linear Discriminant Analysis* (LDA) model was run, which assumes that the variables are Gaussian distributed with unit variance and some covariance [21]. The parameters of these distributions are calculated as sample means and covariances from the data, and elements are classified by selecting the class that maximizes a given sample's likelihood given its feature values [21]. No parameters are necessary for the LDA. A *Neural Network* model was trained as well. A neural network is composed of layers of nodes which each weight the data from all of the layers in the previous node and feed the result through an activation function to the nodes in the next layer. The weights are iteratively trained through a process called backpropagation. The nodes in the middle layers are called hidden nodes, and in the neural networks we used, there was only one layer of hidden nodes. The number of hidden nodes ranged from 1 to 10, and the weight decay parameter was allowed to take on values {0, 0.1, 1, 2} [18, 22]. The optimal number of hidden nodes and the optimal value of the weight decay parameter was found using cross-validation. The maximum number of back-propagation iterations was set at 50. We also considered a *Support Vector Machine* (SVM), using the *radial basis function* (RBF) kernel. An SVM projects the data into a hyper-dimensional feature space and finds a hyperplane that best separates the data in each class in the hyperspace. The cost parameter (C) for the SVM was chosen using cross-validation over model accuracy [18]. The σ-parameter was estimated from the data using the sigest() function in the kernlab package [22], and C was selected from the powers of two between 2^{-4} and 2^{16}, inclusive.

Table 3: Model evaluation over data set A (Accuracy, Sensitivity, Specificity)

Model	Overall Accuracy	Artifact Sens.	Artifact Spec.	Extrasystole Sens.	Extrasystole Spec.	Murmur Sens.	Murmur Spec.	Normal Sens.	Normal Spec.
Partial Least Squares	0.448	0.700	0.737	0.000	1.000	0.500	0.619	0.286	0.864
GLMN	0.448	0.600	0.789	0.250	0.960	0.500	0.667	0.286	0.818
Linear Discriminant Analysis	0.483	0.700	0.842	0.500	0.920	0.500	0.667	0.143	0.864
Neural Network	0.483	0.700	0.842	0.500	0.920	0.500	0.667	0.143	0.864
Support Vector Machine	0.552	1.000	0.789	0.500	0.840	0.500	0.810	0.000	0.955
Classification Tree	0.517	0.700	0.842	0.500	0.920	0.500	0.762	0.286	0.818
Random Forest	0.517	0.900	0.842	0.250	0.880	0.500	0.810	0.143	0.818
Naive Bayes	0.345	0.400	1.000	0.250	0.760	0.500	0.667	0.143	0.727

A *Classification Tree* model was considered as well, which is a sequential set of thresholds on features used to separate data into different classes. The model was allowed to cross-validate over 5 values of the tree complexity parameter (which were selected by the training algorithm). The CART algorithm was used [19], and the optimal tree of the 5 generated was selected. In addition to the classification tree, a *Random Forest* model was trained, which is essentially an ensemble of very short trees. The model was allowed to contain anywhere from 1 to 20 trees, but cross-validation against accuracy selected the one tree model. The final model considered was *Naïve Bayes*, which maximizes the posterior probability of the class given the features, by assuming all of the features are independent and estimating their distributions from the training data. No smoothing was done, and no kernel was used, so no parameters were necessary [23]. The appendix contains graphs of the example accuracy results over the tuning parameters for the neural network, classification tree, and SVM.

4 RESULTS

The models were designed to identify several different anomalies, so we considered sensitivity and specificity for each class, and overall accuracy. Using these measures, we selected the three highest performing models and determined the optimal model. Performance metrics from the test set can be found in Table 3 below. Based on the evaluation criteria, the three models that achieved a higher overall accuracy were: Support Vector Machine, Classification Tree, and the Random Forest. Also, a high specificity value achieved by these models (especially for the murmur and extrasystole cases) indicate their ability to dismiss the possibility of a heart disorder. In Table 4, the precision is evaluated for all models. The Classification Tree model seems to show the best tradeoff among all classes. A high precision value can indicate a lower number of false positives, resulting in a more exact model. The data set had a group of sample files that were not labeled. Several high performing models—based on the results from Tables 3 and 4—were used to predict the labels of these samples, shown in Table 5.

Table 4: Model evaluation over data set A (Precision)

Model	Artifact	Extra-systole	Murmur	Normal
Partial Least Squares	0.583	NaN	0.333	0.400
GLMN	0.600	0.500	0.364	0.333
Linear Discriminant Analysis	0.700	0.500	0.364	0.250
Neural Network	0.700	0.500	0.364	0.250
Support Vector Machine	0.714	0.333	0.500	0.000
Classification Tree	0.700	0.500	0.444	0.333
Random Forest	0.750	0.250	0.500	0.200
Naive Bayes	1.000	0.143	0.364	0.143

Table 5: Predicted classes for unlabeled samples

Class	LDA	NNET	SVM	Tree
Artifact	13	13	13	15
Extra Sound	7	7	18	6
Murmur	17	17	13	15
Normal	15	15	8	16

It was determined that the Classification Tree would be the optimal model based on its overall accuracy as well sensitivity and specificity. Below is the Classification Tree our model developed in order to classify heart sounds, see Figure 6.

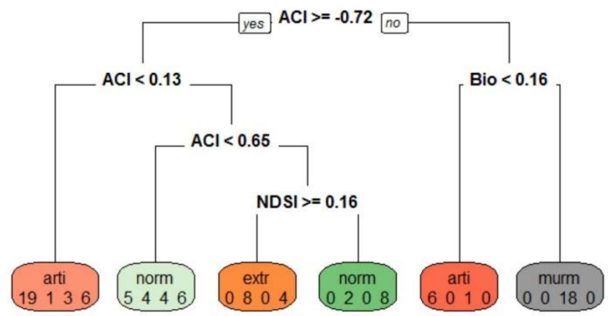
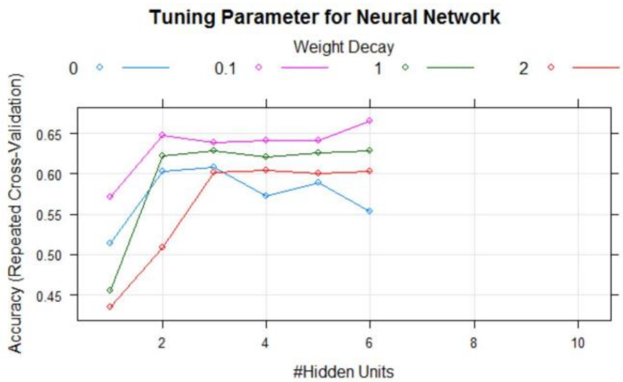


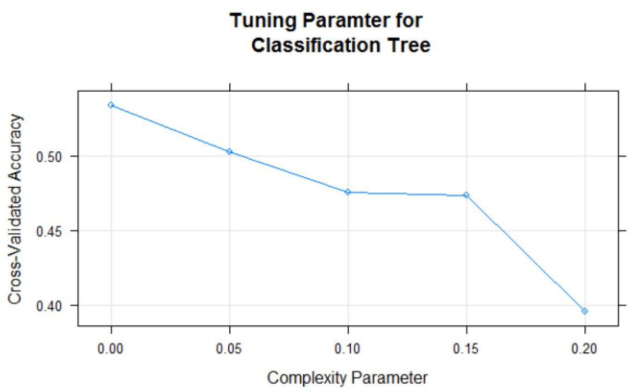
Figure 6: Classification Tree Model

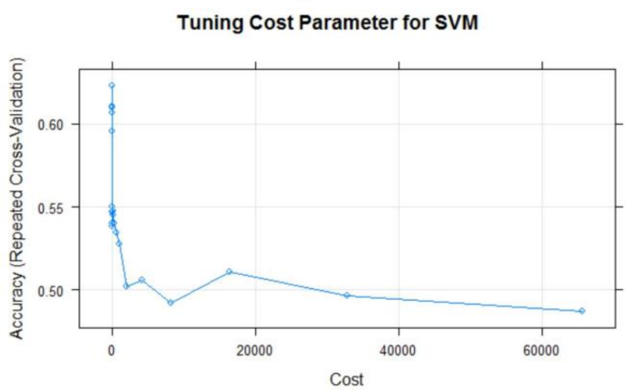
5 CONCLUSIONS

This work investigated the problem of heartbeat abnormality detection based on an audio recordings data set presented in a competition on the Kaggle website. From a set of four possible classes, the classification models were trained to predict the class of a testing heartbeat recording. The training relied on features that were extracted based on a set of indices historically used in the ecology research domain. Results show that three models achieved top performance based on the accuracy, sensitivity, specificity; the Support Vector Machine, Classification Tree, and the Random Forest. Based on our findings, we suggest using the Classification Tree as the optimal model to identify heart abnormalities due to its higher level of predictive capabilities and the interpretability of the model. It should be noted that our work outperformed that of the original contest winner according to their own metrics. Future work for this project would include further research into feature extraction in order to improve predictive power of these models. These models were built on a small number of features but more information could be extracted from the audio files. Furthermore, the same features could be extracted for data set B and then be tested for a statistical difference between data set A and data set B for each feature. If data set B is not considered to be different, then this would expand on the number of samples that could be used for model development. Outside of the model development phase, software could be developed that reads the audio files, automatically extracts the features, and then applies the predictive models to the features to classify the heartbeat to check for abnormalities. This software could easily be implemented in real time at health clinics for client diagnostics.

A APPENDIX: TUNING PARAMETERS







REFERENCES

[1] Heart Health. “Heart Disease Facts.” November, 2018. Retrieved from: http://www.your-heart-health.com/en-US/heart-disease-facts.html

[2] Kaggle. “Heartbeat Sounds.” Retrieved from: https://www.kaggle.com/kinguistics/heartbeat-sounds

[3] Mayo Clinic. “What’s a Normal Resting Heart Rate?” Retrieved from: http://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979

[4] Wikipedia. Heart Sounds. Retrieved from: https://en.wikipedia.org/wiki/Heart_sounds

[5] S. Erb, “Classification of vehicles based on acoustic features.” Master’s thesis, Begutachter: Univ.-Prof. Dipl.-Ing. Dr. Bernhard Rinner, 2007.

[6] German Gomez Herror, Atanas Gotchev, Ivaylo Christov, Karen Egiazarian, “Feature Extraction for Heartbeat Classification Using Independent Component Analysis and Matching Pursuits.” Retrieved from: https://pdfs.semanticscholar.org/0a98/1e6b5c31e9480784dcb6cdbee42d548f6a21.pdf. 2015

[7] H.-l. Wang, W. Yang, W.-d. Zhang, and Y. Jun, “Feature extraction of acoustic signal based on wavelet analysis.” In ICESSSYMPOSIA ’08: Proceedings of the 2008 International Conference on Embedded Software and Systems Symposia. Washington, DC, USA: IEEE Computer Society, 2008.

[8] G. P. Mazarakis and J. N. Avaritsiotis, “Vehicle classification in sensor networks using time-domain signal processing and neural networks.” Microprocess. Microsyst., 31(6) 381–392, 2007.

[9] M. Baljeet, N. Ioanis, H. Janelle, “Distributed classification of acoustic targets in wireless audio-sensor networks.” Computer Network, 52(13):2582–2593, 2008.

[10] S. S. Yang, Y. G. Kim1, H. Choi, “Vehicle identification using discrete spectrums in wireless sensor networks.” Journal Of Networks, 3(4):51–63, 2008.

[11] Varun Kumar Kakar and Manisha K, “Techniques of Acoustic Feature Extraction for Detection and Classification of Ground Vehicles,” International Journal of Emerging Technology and Advanced Engineering. Vol. 3, Issue 2, Feb. 2013.

[12] E.M. Munich, “Bayesian subspace methods for acoustic signature recognition of vehicles.” in Proceedings of the European Signal Processing Conference (EUSIPCO-04), Vienna, Austria, Sept. 2004.

[13] C. De Capua, A. Battaglia, A. Meduri and R. Morello, "A Patient-Adaptive ECG Measurement System for Fault-Tolerant Diagnoses of Heart Abnormalities," 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007, Warsaw, 2007, pp. 1-5. doi: 10.1109/IMTC.2007.379434

[14] Amir Averbuch, Valery A. Zheludev, Neta Rabin, Alon Schclar, “Wavelet-based acoustic detection of moving vehicles.” Multidim

[15] R. Besrour, Z. Lachiri and N. Ellouze, "ECG Beat Classifier Using Support Vector Machine," 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, 2008, pp. 1-5. doi: 10.1109/ICTTA.2008.4530053. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4530053&isnumber=4529902

[16] Cran.R Project. “An Introduction to the soundecology Package.” Retrieved from: https://cran.r-project.org/web/packages/soundecology/vignettes/intro.html

[17] Mathworks, Inc.. (2017). *Time-Frequency Analysis with the Continuous Wavelet Transform* [Online]. Available: https://www.mathworks.com/help/wavelet/examples/time-frequency-analysis-with-the-continuous-wavelet-transform.html

[18] M. Kuhn and K. Johnson, Applied Predictive Modeling, 1st ed. New York: Springer-Verlag, 2013.

[19] M. Kuhn. (2016). The caret Package [Online]. Available: http://topepo.github.io/caret/train-models-by-tag.html#support-vector-machines

[20] T. and J. Qian. (2014). Glmnet Vignette [Online]. Available: https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html

[21] J. Brownlee. (2016) Linear Discriminant Analysis for Machine Learning [Online]. Available: http://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/

[22] A. Karatzoglou, A. Smola, et al.. (2016). Package ‘kernlab’ [Online]. Available: https://cran.r-project.org/web/packages/kernlab/kernlab.pdf

[23] C. Roever, N. Raabe, et al.. (2015). Package ‘klaR’ [Online]. Available: https://cran.r-project.org/web/packages/klaR/klaR.pdf