A Novel Approach for Classification of Normal/Abnormal Phonocardiogram Recordings using Temporal Signal Analysis and Machine Learning

Sachin Vernekar[1], Saurabh Nair[2] , Deepu Vijaysenan[1], Rohit Ranjan[3]

[1] NIT Surathkal, Surathkal, India
[2] VJTI, Mumbai, India
[3] Simon Fraser University, British Columbia, Canada

Abstract

This paper discusses a novel approach used for classification of phonocardiogram (PCG) excerpts into normal and abnormal classes as a part of Physionet 2016 challenge [10]. The dataset used for the competition comprises of cardiac abnormalities such as mitral valve prolapse (MVP), benign murmurs, aortic diseases, coronary artery disease, miscellaneous pathological conditions etc. [3], We present the approach used for classification from a general machine learning application standpoint, giving details on feature extraction, type of classifiers used comparing their performances individually and in combination. We propose a technique which leverages previous research on feature extraction with a novel approach to modeling temporal dynamics of the signal using Markov chain analysis [7, 9]. These newly introduced Markov features along with other statistical and frequency domain features, trained over an ensemble of artificial neural networks and gradient boosting trees, with bagging, gave us an accuracy of 82% on the validation dataset provided in the competition and was consistent with the test data with the best result of 78%.

1. Introduction

This work describes a novel approach designed for Physionet 2016 Challenge Classification of Normal/Abnormal Heart Sound Recordings. The objective here is to classify Phonocardiogram (PCG) recordings into normal and abnormal categories. A comprehensive detail on the database, explaining how the PCG signals were collected and the type of abnormalities found are discussed in the paper mentioned in the reference section [3].

Usually, statistical features such as means, standard deviations of systole, diastole intervals, and signal complexity features are used as features for classification and these are enough to give decent results. But these features fail to completely capture the temporal information of the signal. This could be very important since it represents how each heart beat changes over time.

To capture the temporal dynamics, we take PCG signal beat by beat and assign each beat a symbol/category based on different thresholds set on features (ratios of systole intervals to RR interval, diastole intervals to RR interval, beat energy, the power of frequency component above 200 Hz). Thus, a sequence of symbols for the entire signal is obtained. We then extract features out of this sequence for classification. One of the ways we employed is to create a Markov chain with symbols being the states of the matrix and the resulting transition probabilities are used as features. These features along with marginal probabilities of states and rest of the acoustic features like sample entropy, instantaneous frequency analysis etc. are used to train an ensemble/bag of 4 class boosted tree classifiers [8, 13] and 4 artificial neural networks [12].

2. Preprocessing

The training data consists of PCG signals of varying length, anywhere between 5s to just over 120s all sampled at 2000 Hz. For training, all the signals were re-sampled to 1000 Hz and features were extracted. Since PCG recordings were collected under uncontrolled environments, they were corrupted by various noise sources. Hence signals were filtered with a band-pass Butterworth filter of the frequency range, 25 Hz to 400 Hz to remove high-frequency noise as well as artifacts such as baseline wandering. The signal spikes were then removed using Schmidt spike removal technique [4] and the signal was normalized to zero mean and unit variance. Reference annotations for four heart sound states (S1, systole, S2, diastole), for each heart beat, were then obtained for the pre-processed signals using Springer's segmentation algorithm [5] which is a state of the art solution for heart beat segmentation.

3. Feature Extraction

The time duration generated by segmentation algorithm for the four heart sound states: S1, systole, S2, diastole were used to get few selected, typical statistical features such as mean and standard deviations of each state lengths, powers, ratio of each state length to RR interval, ratio of systole to diastole intervals, ratio of amplitude of systole to diastole, number of zero crossings, etc. This comprised a total of 38 features.

From frequency domain analysis, FFT statistics, Autoregressive Moving-Average (ARMA) features [1], Mel-frequency Ceptral Coefficients (MFCC) features [2], sample entropy [6], music features [1], octave band features [1] etc. were extracted and it comprised a total of 102 features.

The above features even though proved to be statistically significant in classification don't really capture the temporal dynamics of the signal, which is to say, how a beat differed from other beats in terms of energy of the beat, energy of states of individual beats, duration of beat etc, Hence a Markov chain analysis [7, 9] was done to capture this, which we found to be useful to improve the accuracy of the results.

3.1. Markov Chain Analysis

We noticed that when we do beat by beat analysis, there is a significant variance in the following features, namely: the energy of beat, power in the region above 200 Hz, systole to RR ratio, diastole to RR ratio. The first 2 of the 4 features account for variation in amplitude component of the signal, and the remaining 2 account for the time component variations.

First, we list out the values of these 4 features for all the beats in all the recordings, then for each of the features, 2 thresholds were selected, which divides the training data at 33 percentile and 66 percentile points rendering them into 3 parts [7, 9]. Thus a total of 8 thresholds for 4 features in combination, give us 12 different regions in 4-dimensional space of 4 features. These 12 regions are used as 12 states of a Markov chain.

For each signal, each beat is classified into one of the 12 states, giving a sequence of states, labeled A through L, an example of which is given in Figure 1. A first order transition probability matrix is constructed for the sequence obtained, and these transition probabilities along with marginal probabilities are used as features making a total of 156 features, of which 40 were selected based on feature importance analysis using XGBoost [8, 13] and some manually curated methods.

Finally combining all features together we obtained a total of 180 features, of which 72 features were eliminated based on the feature importance obtained through random Forest and XGBoost [8, 13] feature importance analysis, giving a final total of 108 features. Approximate, normalized feature importance of individual feature types are summarized in the Table 1.

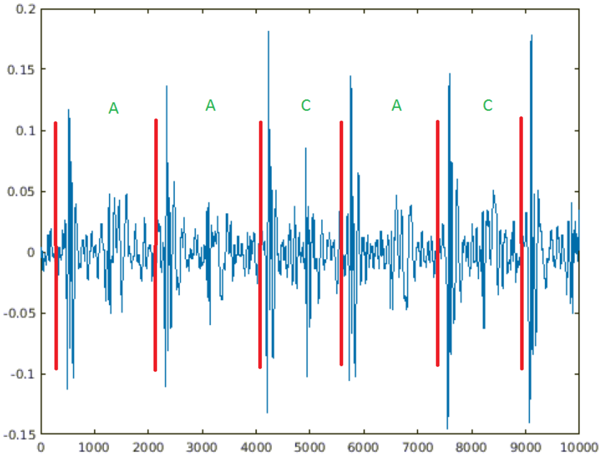

Figure 1. Sequence of states generated in Markov analysis of PCG.

Table 1. Features extracted and importance

Feature	Importance
ARMA	14%
MFCC	12%
FFT Statistics	8%
Music features	9%
Octave band features	12%
Markov features	12%
Rest of the features	33%

4. Training

Training process briefly describes training data used, way it is stratified, analysis of feature space and classification techniques employed.

4.1. Training Data

Training data [3] consists of 3153 recordings of which 242, labeled extremely noisy to be correctly classified are eliminated as no special method is employed to detect such a noise, some of which don't even contain PCG signals. 296 of the rest that is marked for validation by Physionet are taken as a validation set and are removed from the training set. Thus, the training data comprises of 2615 recordings, containing 454 abnormal recordings (class 1) and 2161 normal recordings (class 0), roughly a class ratio of 1:4, and the validation set comprises of 296 samples with a class ratio of 1:1. Test data is not revealed to the participants.

4.2. Experimentation

The normalized training data was further analyzed using Linear Discriminant Analysis (LDA) [14], Principal Component Analysis (PCA) [14] and t-Distributed Stochastic Neighbor Embedding (t-SNE) [15] to get a visual representation of the training to aid in selection of the classifier. As you can clearly see from the Figure 2, even though, classes seem to have fairly well separated, a simple linear classifier is highly unlikely to work. This is because classes don't seem to have a linear boundary, hence needing to go for non-linear classifiers with higher VC-dimensions.

Figure 2. 1-D plot of LDA, 2-D plots of PCA and t-SNE respectively.

In order to span through multiple hypothesis spaces, We tried quite a few classifiers, such as, XGBOOST [8, 13] which is an AdaBoost based gradient boosting tree classifier, single hidden layer artificial neural networks (ANN) with 10 neurons (Keras [12]), Support Vector Machines (SVM) with linear kernel (sklearn [14]) and random forest (sklearn [14]). Since the training data is highly class imbalanced, we train with sample weights of 0.8 for class 1 and 0.2 for class 0 samples. The results of which on validation data are summarized in Table 2.

The final model is a weighted ensemble of 4 XGBOOST and 4 ANN classifiers trained with feature bagging, manually tuning the parameters to get optimized ones. Figure 3 gives flow chart of the methodology used for classification.

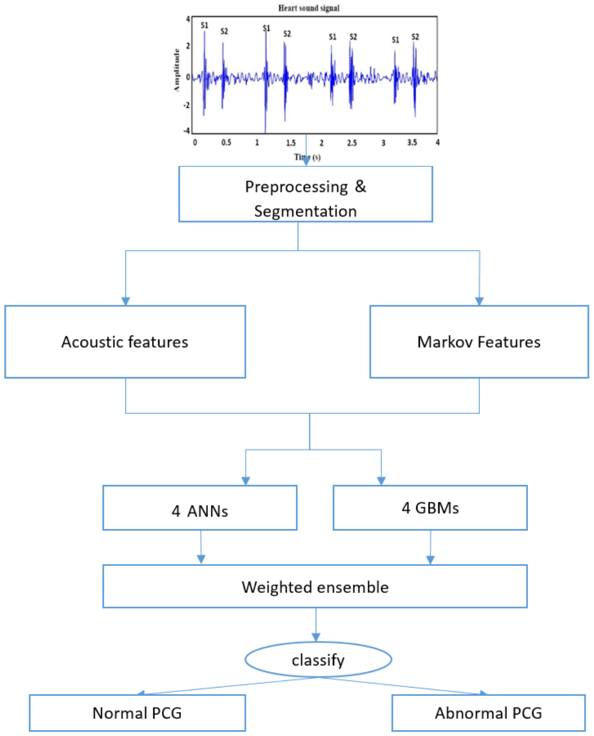

Figure 3. Algorithm flow.

Table 2. Classifiers and accuracy.

Classifier	validation accuracy
ANN	76.2%
Gradient Boosting	78.4%
SVM	71.1%
Random Forest	73%

5. Results

The scoring function used for the competition treats extremely noisy signal to be a separate class, but since our analysis doesn't do any such special handling for noise classification, the validation score will be reported as per the below scoring function.

$$Sensitivity, Se = \frac{100 * TP}{TP + FP}$$

$$Specificity, Sp = \frac{100 * TN}{TN + FN}$$

where TP, TN, FP, FN stand for true positive, true negative, false positive and false negative respectively.

$$score = \frac{Se + Sp}{2}$$

The overall results are summarized in Table 3.

Table 3. Final results: Val Set stands for validation set and Feats stands for features

Data	Se	Sp	Score
Val Set(with Markov Feats)	79.2%	84.3%	81.75%
Val Set(without Markov Feats)	72.9%	80.9%	76.3%
Test Set	71.6%	82.7%	77.2%

6. Conclusion

The experimentation with newly introduced Markov Features proves to be successful, improving the result by as much as 5% on the validation data. The validation scores are consistent with the test scores, hence over-fitting is minimal. The overall accuracy could be improved with better insight into the cause of abnormalities, type of abnormalities, and even if an abnormality is observed, knowing what region of PCG is subjected the same would be helpful. Moreover, noise, impossible to separate in the frequency domain, makes it very hard to classify, hence should be dealt in a better way.

The link for code location is mentioned in Reference [11].

References

[1] Schmidt, S.E.; Holst-Hansen, C.; Hansen, J.; Toft, E.; Struijk, J.J. Acoustic features for the identification of coronary artery disease. IEEE Trans Biomed Eng 2015, 62, 2611-2619.

[2] Quiceno-Manrique, A.F.; Godino-Llorente, J.I.; Blanco-Velasco, M.; Castellanos-Dominguez, G. Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiograph signals. Ann Biomed Eng 2010, 38, 118-137.

[3] Liu, Chengyu; Springer, David; Li, Qiao; Moody, Benjamin; Juan, Ricardo; Chorro, Francisco; Castells, Francisco; Roig, Jose; Silva, Ikaro; Johnson, Alistair; Syed, Zeeshan; Schmidt, Samuel; Papadaniil, Chrysa; Hadjileontiadis, Leontios; Naseri, Hosein; Moukadem, Ali ; Dieterlen, Alain; Brandt, Christian ; Tang, Hong; Samieinasab, Maryam ; Samieinasab, Mohammad ; Sameni, Reza; Mark, Roger; Clifford, Gari D., An open access database for the evaluation of heart sound algorithms, Volume 37 Number 9, September 2016.

[4] S. E. Schmidt et al., "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," Physiol. Meas., vol. 31, no. 4, pp. 513-29, Apr. 2010.

[5] D. Springer et al., Logistic Regression-HSMM-based Heart Sound Segmentation, IEEE Trans. Biomed. Eng., In Press, 2015.

[6] Richman, J. S. and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol 2000; 278(6):H2039-H2049; http://ajpheart.physiology.org/cgi/content/abstract/278/6/H2039

[7] Sebastian Garcia, Modeling the network behavior of malware to block malicious patterns. The Stratosphere Project: A behavioral IPS, DOI: 10.13140/RG.2.1.3784.7765 Conference: Virus Bulletin Conference, At Prague.

[8] T. Chen and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. ArXiv e-prints.

[9] https://stratosphereips.org/stratosphere-ips-generation-of-the-behavioral-models.html

[10] https://physionet.org/challenge/2016/

[11] https://github.com/sachinvernekar/pcgClassify

[12] https://github.com/fchollet/keras

[13] https://github.com/dmlc/xgboost/tree/master/python-package

[14] http://scikit-learn.org/stable/

[15] https://github.com/danielfrg/tsne

Address for correspondence:

Rohit Ranjan
Dept. of Computer Science, Simon Fraser University, British Columbia, Canada.
rranjan@sfu.ca

Sachin Vernekar
Dept. of Electronics and Communication Engineering, NITK, Surathkal, India.
sachinvernekar50@gmail.com

Saurabh Nair
Dept. of Electronics and Telecommunication Engineering, VJTI, Mumbai, India.
saurabh.nk@gmail.com

Dr. Deepu Vijayasenan Dept. of Electronics and Communication Engineering, NITK, Surathkal, India.
deepu.senan@gmail.com