Classification of Normal and Abnormal Heart Sound Recordings through Robust Feature Selection

Chetanya Puri[1], Arijit Ukil[1], Soma Bandyopadhyay[1], Rituraj Singh[1], Arpan Pal[1], Ayan Mukherjee[2], Debayan Mukherjee[3]

[1]TCS Research & Innovation, Tata Consultancy Services, Kolkata, India
[2]Indian Institute of Technology, Kharagpur, India
[3]Indian Statistical Institute, Kolkata, India

Abstract

We aim to develop a reliable and robust algorithm that accurately analyses a single short PCG recording (10-60s) from a single precordial location to determine the presence of heart abnormality for the Physionet/ Computing-in-Cardiology 2016 challenge.

We extract timing information for the fundamental Heart Sounds i.e. S1 and S2 using Hidden Markov Model based Springer's improved version of Schmidt's method. These values are then used to generate statistical features set in temporal, frequency, time-frequency and wavelet domain. We choose the optimal feature set out of the pool of overall 54 features using mutual information based minimum Redundancy Maximum Relevance (mRMR) technique. In order to cope with bad signals, we also check the signal quality of the PCG signal. Signals are rejected for further normal abnormal classification when the outside/background noise has rendered them useless for processing. Then, non-linear radial basis function based Support Vector Machine (SVM) classifier along with ensemble based methods is used to train with the reduced optimal feature sets, on a balanced training set chosen from the group of all PCG datasets.

Our algorithm is tested with hidden Physionet Challenge 2016 datasets and performance achieved is: Sensitivity (Se) = 0.7749, Specificity (Sp) = 0.7891 and Overall Score calculated as mean (Se, Sp) = 0.7820

1. Introduction

We aim to develop a reliable algorithm that accurately analyses a single short PCG recording (10-60s) from a single precordial location to determine the presence of heart abnormality.

We have considered features based on the statistical parameters derived from the analysis of the PCG recordings in the time and time-frequency domains. We segmented each PCG sequence into four states, namely, first heart sound (S1) and second heart sound (S2), Systole and Diastole using Hidden Markov Model based Springer's improved version of Schmidt's method [8]. Different features like standard deviation of S2 intervals, mean value of ratio between wavelet coefficient energy of Diastole period and RR in each heart beat are derived and overall 54 features are selected that represent a broad spectrum of energy, frequency, and temporal signature of typical PCG signal. Further, using minimum Redundancy Maximum Relevance (mRMR) algorithm, top 5 features are identified from the initial pool of 54 features. Then, non-linear radial basis function based Support Vector Machine (SVM) classifier is trained with the reduced set of optimal feature set, where balanced training set of 630 normal and 630 abnormal PCG datasets are considered. We also used ensemble based supervised learning with bagging and boosting techniques to handle the skewed class distributions. The trained SVM classifier and ensemble based classifier have been tested with the validation datasets as well as uploaded to test with hidden PCG datasets [1].

It is also noted that detection of pathological conditions based on machine learning techniques is flawed by a) variety of datasets b) presence of outside noise in the datasets and poses a greater challenge to detection of abnormality. We also detect signal quality to handle this problem.

2. State-of-the-art

Several articles are available in the literature on the analysis of the PCG signal. The segmentation of the heart sound is the first step to analyze the signal and for the identification of systolic and diastolic regions allowing for the detection of pathological events. Various existing literature are available for the segmentation of the PCG signal. Authors in [7], have applied an ensemble empirical mode decomposition (EEMD) based method combined with kurtosis features to detect the location of first (s1) and second (s2) heart sound. In [8] authors have proposed a duration-dependent hidden Markov model (DHMM) based methodology for the robust segmentation of heart sounds. They identified the events, the amplitude of the signal envelope and a predefined model structure. Authors of [9] proposed a logistic based HSMM for the identification of positions of heart sounds. These extensive works on the detection of heart sound location enable us to start the exploration for the classification of normal and abnormal PCG signals.

Second most problem in the analysis of signal is the noise. Noise is the major challenge for the information extraction in various bio-medical signals as in Photoplethysmogram (PPG), Electrocardiogram (ECG), Arterial Blood Pressure (ABP) signals [2]. Significant amount of noise occurs due to several factors such as motion artifacts, environmental conditions, device calibration etc., which leads to corrupt segments in the signal. Henceforth, it's a prime need to preprocess the data to remove the ambient noise from the signal.

Several signal processing and machine learning techniques exits in the literature to denoise the corrupt part of the signal. In [2] we attempted to denoise the PPG signals using template based matching using adaptive use DTW to derive accurate heart rate of the signal. In [5], we showed the effectiveness of removal of motion artifacts in PPG signals which directly enhanced the accuracy in arrhythmia condition detection, specifically to reduce the false negative alarms. From our previous experience of the signal analysis, we also observed that the morphology of the signal plays a vital role in information extraction of the signal used for medical analysis. Considering this motive, we exploited the morphology of the PPG signal (with/without motion artifacts) using basic feature derivation of maxima and minima to get the exact onset time of the signals [4]. Therefore, the preprocessing the signal is a key necessity to extract the desired information from the signal.

3. Proposed algorithm

The datasets used were provided as part of the Physionet/Computing-in-Cardiology Challenge 2016 [1]. The datasets given for training consisted of 3153 recordings. Out of the given 3153 recordings, 2302 recordings annotated as normal were clean, 186 normal recordings were noisy. Similarly, 572 abnormal recordings were annotated as clean and 93 as noisy.

Given a PCG time series, a set of 54 features is generated. These features pertain to time domain, frequency domain and wavelet domain. Our Abnormality classification algorithm primarily consists of four stages. We first pre-process the data by resampling the signal from 2000 Hz to a lower frequency of 1000 Hz and then find statistical features like kurtosis and trimmed mean in the frequency domain that help finding signal quality. If the signal is found out to be noisy, we discard it for further processing. If not, it is then fed to the second stage of segmenting the data. The timing information of the Fundamental Heart Sounds (FHSs) and the systole and diastole regions in the PCG signal is found using Hidden Markov Model based Springer's improved version of Schmidt's method [8]. This is known as segmenting the signal into fundamental components. Once the timing information is obtained, feature derivation is done. ***Feature Derivation***: We derive domain dependent features such as duration between the RR beats, S1 interval, systole interval, S2 interval, diastole interval, etc. Domain-independent features such as Wavelet Packet (WP) energy coefficients with 'db3' as the mother wavelet and the mean amplitude ratios of the inter-FHS intervals were also extracted. Statistical moments such as mean, standard deviation (SD) were then calculated as different features of the PCG signal. A list of features with their definitions is given in Table 1.

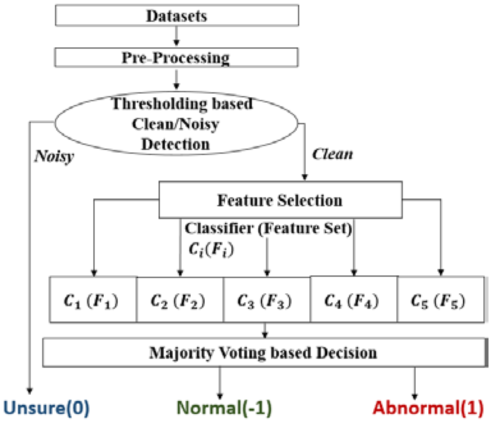

Figure 1: Flowchart of our detection scheme

3.1. Importance of noise elimination

With the advent of non-invasive smart devices for bio-medical applications, automating disease diagnosis has now become feasible. However, presence of ambient noise and motion artifacts poses a challenge in correctly identifying the condition of patient, even to a medical expert. Denoising [12,10] can increase the disease detection accuracy for signals like PPG, ABP. Non-stationary sound signals like PCG are highly prone to the background noises, even in clinical settings. We estimate signal quality of PCG signals based on auto-thresholding of certain features extracted from PCG waveform.

3.3. Balancing training data with respect to normal and abnormal labels for unbiased estimation

Due to the skewed distribution of normal and abnormal classes with only 19% of the clean training data being abnormal, training on such imbalanced data biases the classifier towards one class [15]. We randomly undersample from the normal set and make the distribution of the classes equal. 630 datasets from normal waveform were taken and 630 from abnormal for training with 50-50 class balance.

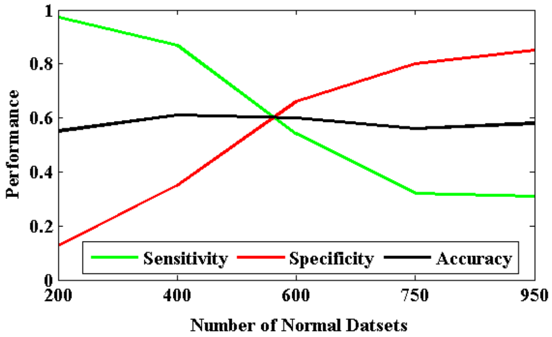

Figure 2: Variation of Performance with balancing of the majority class (Normals) towards minority class (Abnormals).

3.4. Progression towards optimal feature set generation

For a feature set $F=\{f_i\ ,\ i:1:N\}$, with N number of features and c class labels. If the mutual information I [3] between the feature f_i and a class label is high, the feature can be said to having maximum relevance. Mathematically, $D(F,c)=\frac{1}{|F|}\sum_{f_i} I\ (f_i,c)$.

However, it has been observed that in feature selection, combining many good features does not always increase the classifier's performance. Authors in [13] have put it simply, that "the best k features are not necessarily the k best features". It is easily possible that the n best features might have high inter-dependence making them redundant. The redundancy between two features can be formulated as $R=\frac{1}{|\mathrm{F}|^2}\sum_{f_i,f_j} I\left(f_i,f_j\right)$.

We have used mRMR [14] which maximizes the relevance of a feature set while minimizing the redundancy among them. Mathematically, $max_F\ \ \phi(D,R),\phi=D-R$

3.5. Classifier for normal abnormal Separation

Out of the total 54 features, the fittest five features were selected with the mRMR (minimum Redundancy Maximum Relevancy) algorithm. We choose non-linear SVM classifier and 'RBF' kernel due to its good generalization ability to train with the modified feature set.

The trained SVM classifier has been tested with the validation datasets and Sensitivity (Se), Specificity (Sp) and Overall Score (OS) are {0.68, 0.71, 0.69} and {0.70, 0.70, 0.70} for top five and seven featuers respectively. Due to insignificant improvement in case of higher dimensional feature space, we considered top five MRMR features, particularly, for avoiding classifier's over-fitting to minimize the curse of dimensionality.

We also selected a subset of features (F_1,F_2,F_3,F_4,F_5) from a total set of 54 features, F ranked by mRMR and use the i_{th} feature set F_i with i_{th} classifier chosen from $C_i=C_1,C_2,C_3,C_4,C_5$. A description of these features and classifiers is given in Table 1 and 2 respectively.

We use Ensemble based supervised learning technique with bootstrap aggregation (bagging) to train multiple weak learners and combine them into one strong decision maker.

We also considered ensemble based learning with "RUSBOOST" (Random Under Sampling) due to its ability to alleviate class imbalance. RUS boost is said to perform good when the class distribution is skewed.

Based on the output of each of the above classifier, a decision is taken through majority voting if the recording is normal or abnormal. Choosing such a majority voting decision logic based on different classifiers trained on different features strengthens the decision making. We next present our results for individual classifications and majority voting classifier in Table 3.

Table 1. Different Features used and their definitions

Feature	***Definition***
f1, f2, f3, f4	Mean value of RR intervals, systole intervals, diastole intervals, SD of diastole intervals respectively
f5, f6	Mean of the mean absolute amplitude ratios between systole and S1 period and diastole and S2 period in each heart beat for F5 and F6 respectively
f7	SD of ratio between WP Coefficient Energy sums of S1 period and RR in each heart beat
f8, f9	Mean of ratio between WP Coefficient Energy of S2 period and RR and Diastole period and RR in each heart beat for F8 and F9 respectively.
f10, f11, f12, f13, f14	SD of ratio between WP Coefficient Energy of S1-S2 period, Systole-Diastole Period, S1-Systole period, S1-Diastole Period and S2-Diastole period in each heart beat respectively.
f15, f16	SD of RR intervals and Fisher's Kurtosis [6] of the frequency spectrum respectively.
f17,f18	Frequency at which 80% of the power is contained and Spectral roll off frequency respectively

Table 2. Classifiers with Features sets and Training Data ratio

Method (C_i)	Dataset Balancing (Normal-Abnormal ratio)	Features Set (F_i)
C_1 = SVM	Unbalanced with 3:1 ratio	F_1 = {f1, f2, f3, f5, f6, f7, f8, f11, f13, f14}
C_2 = RUS Boost	Unbalanced with 5:1 ratio	F_2 = {f1, f2, f3, f5, f6, f7, f8, f11, f13, f14}
C_3 = SVM	Balanced with 1:1 ratio	F_3 = {f2, f4, f8, f9, f10, f12}
C_4 = Ensemble bagging	Unbalanced with 2:1 ratio	F_4 = {f1, f15, f9, f16, f17, f18}
C_5 = Ensemble bagging	Unbalanced with 2:1 ratio	F_5 = F={f1, f2, f3, f4, f5,…..f53, f54}

Table 3. Results on Validation Datasets (above) and Hidden datasets (below) with different methods (C_i)

Method (C_i)	Performance Score		
	Sensitivity	Specificity	OS
C_1	0.77	0.55	0.66
	0.76	0.76	0.76
C_2	0.79	0.44	0.61
	0.73	0.71	0.72
C_3	0.72	0.61	0.67
	0.78	0.77	0.77
C_4	0.64	0.72	0.68
	0.59	0.84	0.71
C_5	0.66	0.76	0.71
	0.67	0.84	0.76
Majority voting decision C = $\{C_1, C_2, C_3, C_4, C_5\}$	**0.77**	**0.66**	**0.72**
	0.77	**0.79**	**0.78**

4. Results

We eliminate the noisy signals unfit for classification and use different classifiers on optimally generated feature sets and achieve performance of Sensitivity (Se) = 0.7749, Specificity (Sp) = 0.79 and Overall Score (OS) calculated as mean(Se, Sp) = 0.7820 through majority voting on the hidden datasets during the Physionet Challenge 2016.

We also observed that the balancing of the classes for training the classifier is as important as the choice of classifier. As is evident from the results, choosing a much more balanced dataset drives the performance gains.

References

[1] Liu C Y, Springer D B, Li Q, Moody B, Juan R A, Chorro F J, Castells F, Riog J M, Silva I, Johnson A E W, Syed Z, Schmidt S E, Panadaniil C D, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab M R, Sameni R, Mark R G and Clifford G D 2016 An open access database for the evaluation of heart sound algorithms Physiol Meas 37 (in press).

[2] Ukil A, Bandyoapdhyay S, Puri C, Pal A. IoT Healthcare Analytics: The Importance of Anomaly Detection. IEEE 30th International Conference on Advanced Information Networking and Applications 2016; Mar 23:994-997.

[3] Bandyopadhyay S, Ukil A, Puri C, Singh R, Bose T, Pal A. SensIPro: Smart sensor analytics for Internet of things. In2016 IEEE Symposium on Computers and Communication (ISCC) 2016 Jun 27 (pp. 415-421). IEEE

[4] Bandyoapdhyay S, Ukil A, Puri C, Singh R, Pal A, Mandana K, Murthy C. An Unsupervised Learning for Robust Cardiac Feature Derivation from PPG Signals. In2016 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2016. IEEE (in press).

[5] Puri C, Ukil A, Bandyopadhyay S, Singh R, Pal A, Mandana K. iCarMa: Inexpensive Cardiac Arrhythmia Management--An IoT Healthcare Analytics Solution. InProceedings of the First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems 2016 Jun 30 (pp. 3-8). ACM.

[6] Ukil A, Bandyopadhyay S, Pal A. Sensitivity inspector: Detecting privacy in smart energy applications. In2014 IEEE Symposium on Computers and Communications (ISCC) 2014 Jun 23 (pp. 1-6). IEEE.

[7] Papadaniil CD, Hadjileontiadis LJ. Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features. IEEE journal of biomedical and health informatics. 2014 Jul;18(4):1138-52.

[8] Schmidt SE, Holst-Hansen C, Graff C, Toft E, Struijk JJ. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. Physiological Measurement. 2010 Mar 5;31(4):513.

[9] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering. 2016 Apr;63(4):822-32.

[10] Ukil A, Bandyopadhyay S, Puri C, Singh R, Pal A, MandanaK. CardioFit: Affordable Cardiac Healthcare Analytics for Clinical Utility Enhancement," In 2016 1st EAI International Workshop on Personal, Pervasive and mobile Health, 2016. Springer. (in press).

[12] Ukil A, Bandyopadhyay S, Puri C, Pal A. Heart-trend: An affordable heart condition monitoring system exploiting morphological pattern. In2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016 Mar 20 (pp. 6260-6264). IEEE.

[13] Cover TM. The best two independent measurements are not the two best. IEEE Transactions on Systems, Man, and Cybernetics. 1974 Jan(1):116-7.

[14] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence. 2005 Aug;27(8):1226-38.

[15] Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intelligent data analysis. 2002;6:429-49.

Address for correspondence.
Chetanya Puri
4B-39, 1B, Ecospace, TCS, New Town, Kolkata- 700156, India
chetanya.puri@tcs.com