Heart Abnormality Classification Using Phonocardiogram (PCG) Signals

Md. Khayrul Bashar
Faculty of Gen. Educational Research
Dept. of Information Science
Ochanomizu University
Tokyo 112-8610, Japan
bashar.md.khayrul@ocha.ac.jp

Samarendra Dandapat
Dept. of EEE
Indian Institute of Technology
Guwahati
Guwahati, Assam 781-039, India
samaren@iitg.ac.in

Itsuo Kumazawa
Dept. of Information and
Communications Engineering
Tokyo Institute of Technology
Tokyo 226-8503, Japan
kumazawa@isl.titech.ac.jp

***Abstract*— Heart abnormality or disease is one of the leading causes of mortality worldwide. Sound signal produced by the mechanical activity of heart, known as phono-cardiogram (PCG), provides useful information about the heart's health. To increase discriminability among PCG signals of different normal and abnormal persons, an appropriate combination of signal features and classifiers is important. The segmentation of PCG signal, which requires corresponding ECG signal, is typically used for better prediction. But using ECG is generally expensive and time consuming. 781039In this paper, we therefore propose a segmentation free method to extract information from PCG signal. The signal is first preprocessed for DC removal and to limit the frequency to the required range. Four features (i.e. WPS, PS, FD, and SF) and four classifiers (i.e. LDA, ESVM, DT, and KNN) are then considered for the classification of heart murmur sound from PCG signals. A preliminary experiment with 56 signals showed the highest classification accuracy of 82.6%, obtained by simple statistical feature (SF) with ESVM classifier. On average, the best performing classifier was ESVM (accuracy: 77.17%), while the best feature was PS (accuracy: 75%). In addition, the PS feature showed stable and consistent performance irrespective of the classifiers used. Results also indicate the importance of combining multiple features and classifiers for better accuracy and reliability.**

Keywords—heart sound signal, feature-classifier combination, supervised classification

I. Introduction

According to the world health organization, cardiovascular disorders are a major burden worldwide, causing 30% of the death in the world [1]. Therefore, an early detection of the patients at risk, and the clear understanding of the disease mechanism is crucial. Typically, electrocardiogram (ECG) and cardiac MRI are used in the hospitals and clinics as routine modalities to intervene such disorders. However, there are expensive and medical quality devices are not portable. On the other hand, heart sound recording sensors like digital stethoscope is cheap and they can capture heart disorders as PCG signal before the symptoms actually appear [2].

Two clear sounds, namely a *lub* and a *dub*, are produced by a normal healthy heart. The *lub* sound is known as S1, while the *dub* sound is regarded as S2 [3]. The closure of atrioventricular valves produces S1 sound. In contrast, the S2 sound is produced when aortic and pulmonary valves are closed. Heart sound has two other components: S3 and S4. However, these are rarer heart sounds which are not normally audible and are sometimes visible on the graphical recording as PCG signal. Murmur sounds are high frequency sounds like a whooshing or swishing noise. They may be harmless or abnormal. Harmless murmur can be happened due to increased blood flows than normal through the heart such as during exercise, pregnancy, and rapid growth in children. Abnormal murmurs may be occurred due to congenital heart defects, heart valve diseases, and other symptoms such as shortness of breath, dinginess, bluish skin or a chronic cough. Different diseases may cause the audible murmur in different parts of the cardiac cycle. Murmurs are generally categorized into two classes, namely systolic and diastolic murmurs. Systolic murmurs occur between *lub* and *dub*, while diastolic murmurs happen between *dub* and *lub* sounds [3]. However, diagnosing heart sounds by using a stethoscope requires experience and skill and may include subjective error due to the human hearing limitation as well as the transient and non-stationary nature of PCG signal. An automated system is therefore necessary to objectively classify the *normal* and *murmur* heart sounds at low cost.

Over the past years, many investigations were done on heart classifying abnormal sound signals [2 - 4]. Some of the widely used feature extraction methods are FFT analysis, Gabor transform, discrete wavelet transform, ARMA models, mathematical simulation based methods, matching pursuit method, fractal based methods etc. However, most of the methods are based on segmentation of a heart signal into its component regions for each cardiac cycle, which requires simultaneous recording of ECG signal as a reference and further computation. ECG signals are sometimes impaired by heart diseases, which causes an error in the segmentation results [3]. Recently, a few authors started working on the segmentation-free methods. One such method used curve fitting technique, fractal dimension, and Mel frequency cepstrum coefficient (MFCC) and claimed to obtain 81 to 98% overall accuracy [2].

Data captured in the unconstraint environments contains different background noises. Heart sounds corresponding to the normal and pathologic heart conditions may be highly correlated. It is therefore challenging to discriminate between these two classes of signals using a single classifier. Combining multiple features and classification models is another way of looking into the problem, which can give more efficient representation for reliable and accurate classification. In this study, we therefore attempted to classify heart murmur sounds from their normal counterpart using several features and classifiers. Four well-known features, namely power spectrum (PS), multiresolution wavelet packet statistics (WPS), fractal dimension (FD), and simple statistical feature (SF) are investigated in association with four classifiers, i.e., the linear discriminant analysis (LDA), error correcting output coding support vector machine (ESVM), decision tree (DT), and K-nearest neighbor (KNN) to see the effects how they interact with each other.

II. Material and Methods

An overview of the proposed system is shown in Fig.1. The raw PCG signal is first preprocessed for noise and artifact reduction. Various feature extraction schemes (*SF*, *WPS*, FD, and PS) are then applied to each segment for PCG feature extraction. To compute *WPS*, PCG segments are first decomposed via wavelet packet transformation. Two sets of PCG signals (having 26 or 34 signals) corresponding to the "normal" and "murmurs" class are used in this study. Fig.1 shows the overview of our proposed system.

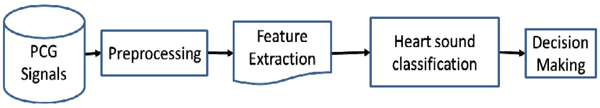


Fig. 1. Block diagram of the heart abnormality classification system

A. Signal Acquisition

The PCG signals were collected from an open access database [5]. Original dataset consists of four classes with 26 normal sound, 30 murmur sound, 13 Extra heart sound, and 40 artifacts. In this study, only *normal* and *murmur* signals have been considered. These signals were collected through mobile phone with sampling frequency of 44.1 KHz. The audio files are of varying lengths, between 1 second and 30 seconds.

In the ***normal*** category, there are normal, healthy heart sounds, which may contain a variety of background noises including occasional random noise corresponding to breathing, or brushing the microphone against clothing or skin. The collected data contains information from children and adults in calm or excited states, the heart rates in the data vary from 40 to 140 beats or higher per minute.

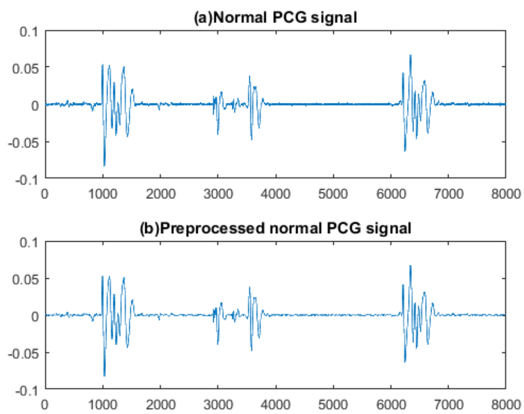


Fig. 2. Normal PCG signal (a) before (b) after processing

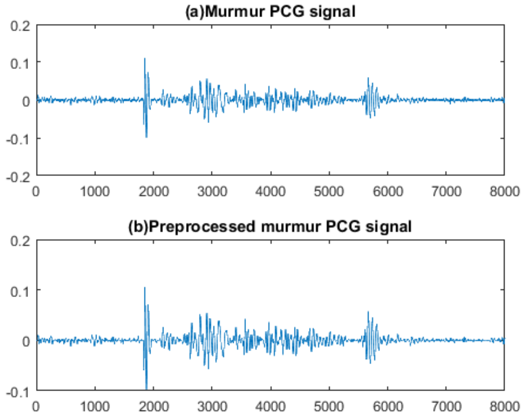


Fig. 3. Murmur PCG signal (a) and (b) after preprocessing

In contrast, the heart ***murmur*** sound contains whooshing, roaring, rumbling, or turbulent fluid noise that typically appears either as systolic or diastolic murmurs. These sounds can be a symptom of many types of heart disorders. However, they will still contain *lub-dub* patterns. One of the things that often confuses the non-medically trained people is to identify the exact location of murmurs. Figs. 2(a) and 3(a) show PCG signals corresponding to these two classes. These figures indicate how PCG amplitudes and frequencies vary in case of normal and pathologic conditions, respectively.

B. Preprocessing

PCG signals are usually noisy. Normal PCG signals may contain random noise corresponding to breathing, or brushing the microphone against clothing or skin. In contrast, murmur signals contain whooshing, roaring, turbulent fluid noises. Possible reduction of noises and artefacts was performed before feature extraction. Since PCG signals have varying lengths, all signals were standardized to 7 sec length starting from beginning by considering the significance of the problem. The DC component of signal is removed to eliminate the effect of microphone. Since the heartbeat frequency components are below 1000 Hz, a zero-phase lowpass filter having cutoff frequency of 900 Hz is applied to reduce high frequency noise from PCG signal. A median filter has also been applied to reduce impulsive noise. The original sampling rate (44.1 KHz) is then reduced to 2 KHz in order to enhance computational efficacy. Finally, the normalization was performed. Included are the figures of the signals from ***normal*** class (Fig. 2) and ***murmur*** class (Fig. 3) and their preprocessing results without normalization.

C. Feature Extraction

Appropriate and discriminatory feature extraction is an important step for the classification of signal events. Time, frequency, and joint time-frequency domain features have been successfully applied to many applications. Each PCG signal is divided into several segments, where segment width is fixed to 5000 samples by trial and error. The following features were considered in our study.

Statistical Features (SF): Signal shape information may be important to discriminate different abnormalities in PCG signals. They were supposed to have non-linear dynamic properties, so tried using higher order statistics like skewness, and kurtosis, but results are somewhat not convincing. Therefore, two simple statistical parameters, namely mean and standard deviation, of signal intensities were extracted in our study.

Wavelet Packet Statistics (WPS): They are statistical representation of terminal subbands of wavelet packet decomposition (WPD) of a discrete-time signal. WPD is a generalization of wavelet decomposition, which partition the frequency content of a signal into progressively finer equal-width intervals of the approximation and detail subbands and hence produces a richer signal description. In this study, Daubechies db5 wavelet is used for PCG signal decomposition. The following statistical descriptors were extracted from each PCG segment.

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_x\right)^2} \tag{2}$$

$$\varepsilon_x = -\sum_{i} x_i^2 \log\left(x_i^2\right) \tag{3}$$

In the above equations, μ_x, σ_x, ε_x represent the mean, standard deviation, and entropy for a subband coefficients, $X = \{x_i, i = 1, 2, ..., n\}$, where n is the length of X. The above three parameters from all subbands were concatenated to construct WPS feature vector. Please refer to [6, 15] for more information.

Power Spectral Density (PS): The power spectral density represents the distribution of signal power over the frequency. It is basically the magnitude squared of the discrete-time Fourier transform of a discrete-time signal with appropriate scaling. The power distribution across different frequencies of the *normal* and *murmur* PCG signals may contain discriminatory characteristics between these two classes. In this study, the concept of periodogram is applied to extract the PS feature from PCG signal. Please refer to [7] for detail computational procedure.

Fraction Dimension (FD): A fractal dimension is an index, which quantifies the complexity of self-similar patterns of a signal. This complexity can be defined as the ratio of the change in detail to the change in scale. Since heart murmur sound, which is the main symptom of heart abnormality, contains high frequency components and they are self-similar patterns like fractals, we choose FD as another important feature towards signal classification. In this study, we have chosen Katz's FD measure as given by

$$D = \frac{\log_{10}(L)}{\log_{10}(d)}, \tag{4}$$

where L is the sum of distances between successive points, and d is the diameter estimated as the distance between the first point of the sequence and the point of the sequence that provides the farthest distance. Katz proposed a generalized unit to make FD computation more correct. Please refer to [8] for more information.

D. Classification

Four classification algorithms, namely linear discriminant analysis (LDA), Error correcting output coding support vector machine (ESVM), decision tree (DT) and K-nearest neighbor (KNN) classifier, were considered in our study.

LDA: This is a simple and effective method for classification. This model assumes that each class (output) generates data (feature variables as inputs) using a multivariate Gaussian distribution. Mean and covariance matrix are computed from each class of training PCG signals. The posterior probabilities for a test signal are computed using trained model for each class. Finally, data classes are predicted by minimizing the expected classification cost. Please refer to [9] for more information about LDA model.

ESVM: This model minimizes the drawbacks of a standard multiclass SVM model by incorporating a coding and a decoding schemes with a binary SVM. The coding design determines the classes, while the decoding scheme combines classification results from binary SVM. We used non-linear SVM with radial basis function, i.e., RBF as kernel for ESVM classifier. Please refer to [6, 10 – 11] for the detail about ESVM.

DT: Decision tree is a flow-chart-like structure, where each internal node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node. There are many variations of the basic decision-tree algorithm. In this study, we adopted the one, implemented in MATLAB. Please refer to the [6, 12] for detail explanation.

KNN: It is a well-known non-parametric method, we adopted here for classification of PCG signals. For a test object (i.e. feature point), *k* closest feature points that might enclose training samples from the ***normal*** and ***murmur*** classes are identified as votes. The class membership is finally determined based on majority voting. In our experiment, we used k=7 for KNN implementation. Please refer to the [6, 13] for detail explanation.

III. Experimental Evaluation

Experiments were conducted on 56 PCG signals from a publicly available dataset, in which 26 signals were recorded from normal healthy subjects, while the rest 30 were from pathologic case representing heart murmur sound. In our study, we used first 7 sec recorded PCG signals for our analysis. This includes a total 280700 samples per PCG with the sample rate of 44.1 KHz. 60 % of the samples from each class (normal and murmur) was used for training, while the rest 40% was used for testing the classifier.

A. *Evaluation Criteria*

Experimental results have been analyzed using various evaluation parameters: (i) precision, (ii) sensitivity (true positive rate), (iii) specificity (or true negative rate), (iv) accuracy, and (v) F_{score} . The accuracy of a classifier is the percentage of the test set which is correctly classified by the classifier. The sensitivity and specificity are the proportions of positives/negatives that are correctly identified. F_{score} is the ratio between the geometric and arithmetic means of precision and sensitivity. Please refer to [14, 15] for more information on evaluation parameters.

B. *Results and Discussions*

Classification results are shown in Figs. 4, 5 and Table 1. Results showed that we can obtain more than 73 % average accuracy with PS feature with majority of the classifiers (ESVM, DT, and KNN) except LDA. This indicates the nonlinear nature of PCG signals, which cannot be well addressed by the linear classifier LDA. PS also showed stable performance, i.e., low variations in the Fscore values with all four classifier (Fig. 4). However, the local statistical feature (SF) produced the highest accuracy (82.6%) and highest Fscore (82.5%) with ESVM classifier. The ranking of the classifiers in terms of average (over four features) accuracy are: (i) ESVM, (ii) KNN, (iii) DT, and (iv) LDA (TABLE I).

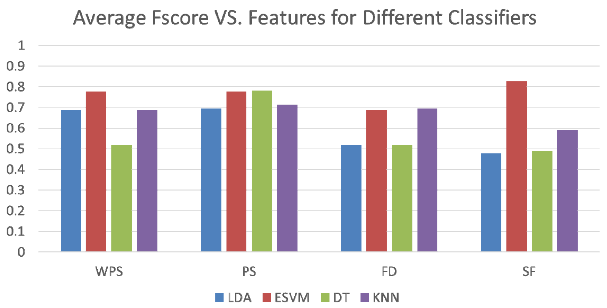


Fig. 4. Average Fscore for different classifiers against signal features: WPS, PS, FD, and SF. Blue, red, green, and purple represent LDA, ESVM, DT, and KNN classifiers, respectively.

Fig. 5 showed that most features (WPS, PS, and SF) performed better with the ESVM classifier (Column 3 of Table 1) except FD feature. Results showed the stable and consistent performances of the PS and WPS features, indicating the well capturing of heart conditions (i.e., “normal “and “pathologic”) in the frequency and time-frequency domain (Fig. 4 and Table 1).

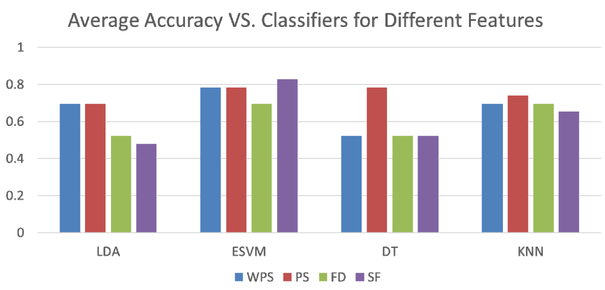


Fig. 5. Average Accuracy for different features against four classifiers: LDA, ESVM, DT, and KNN. Blue, red, green, and purple indicates four features (WPS, PS, FD, and SF. Blue, red, green), respectively.

TABLE I. COMPARATIVE PERFORMANCE

Feature Name	Average Accuracy				
	LDA	***ESVM***	***DT***	***KNN***	***Average***
WPS	0.6956	0.7826	0.5217	0.6956	**0.6739**
PS	0.6956	0.7826	0.7826	0.7391	**0.75**
FD	0.5217	0.6956	0.5217	0.6956	**0.6086**
SF	0.4782	0.8260	0.5217	0.6521	**0.6195**
Average	**0.5978**	**0.7717**	**0.5869**	**0.6956**	

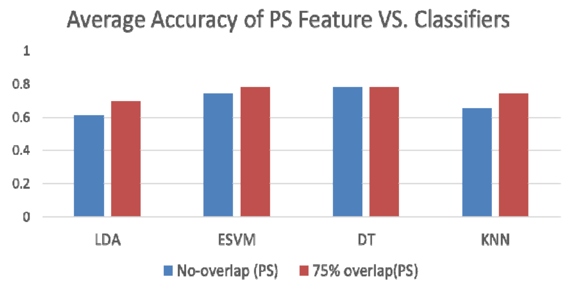
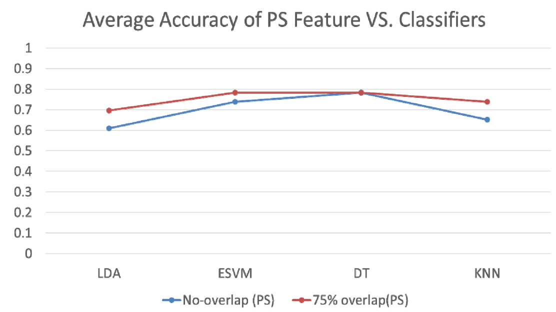


Fig. 6. Comparison of non-overlapping and overlapping sub-series based features for murmur classification. Blue and red lines indicate average *accuracy* in two cases, obtained by using PS features with four classifiers:



LDA, ESVM, DT, and KNN

Fig. 7 Comparison of full-series and sub-series based features for murmur classification. Blue and red bars indicate average Fscore in two cases , obtained by four classifiers: LDA, ESVM, DT, and KNN.

These results may include the effects, usually caused by the selection of mother wavelet, which was Daubechies d5 wavelet in this study. There are many mother wavelets in the orthogonal and bi-orthogonal family. An appropriate selection of it can further improve the performance of the classification.

Figs. 6 and 7 showed the effects of using non-overlapping and overlapping time window. Clearly, the overlapping sub-series processing (red line and bar) produced higher performance, i.e., higher average accuracy and average *Fscore*, compared to the case of using non-overlapping time-windows (blue bar and line) irrespective of the classifiers. A time-window of 5000 samples has been selected in this study. However, an optimization of the window length may further improve the detection results.

In the present study, we only included few features with few prediction models. More discriminatory features will be explored in future study including the practicability of other advanced learning algorithms, e.g., artificial neural network, and deep learning models.

IV. Conclusion

A preliminary study on the segmentation-free method for classifying of heart sound abnormality has been performed. Several time, frequency, and time-frequency domain features (SF, PS, WPS, and FD*)* were applied with four classifiers (LDA, ESVM, DT, and KNN). An experiment with 56 PCG records (26 normal cases and 30 murmur cases) from a heart sound database showed to obtain on average 82.6% classification accuracy by SF with the ESVM classifier. The best performing classifier with all four features was ESVM (on average 77.17% accuracy), while the best feature was PS (on average 75% accuracy). The ranking of classifiers in descending order of performance is (i) ESVM, (ii) KNN, (iii) DT, and (iv) LDA. Out of four features, the PS and WPS features showed consistent performances. With most of the classifiers, the sub-series based overlapping feature produced higher accuracy compared to the non-overlapping based sub-series. Future works will be performed with more classes, large databases, and deep learned features with a focus towards more clinical relevance and real-world applicability.

Acknowledgment

I would like to express my deepest appreciation to my co-authors Professor Samarendra Dandapat and Professor Itsuo Kumazawa, who contributed through their valuable discussions and advices. This work is partly supported by the program for leading graduate school at Ochanomizu University, Japan.

References

[1] A. Lyon, A. Minchole, J. P. Martinez, P. Laguna, and B. Rodriguez, "Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances," *J. R. Soc. Interface* 15: 20170821.

[2] M. Hamidi, H. Gassemian, and M. Imani, "Classification of Heart sound signal using curve fitting and fractal dimension," *Biomedical Signal Processing and Control*, vol. 39, pp. 351-359, 2018.

[3] S. K. Randhawa and M. Singh, "Classification of Heart Sound Signals Using Multi-modal Features," *Procedia Computer Science*, vol. 58, pp. 165 – 171, 2015.

[4] R. He , H. Zhang et al., "Classification of haert sound signals based on AR model," *Computing in Cardiology*, vol. 43, 2016. DOI:10.22489/CinC.2016.177-133.

[5] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor. The PASCAL Classifying Heart Sounds Challenge, available at http://www.peterjbentley.com/heartchallenge/index.html.

[6] M. K. Bashar and M. Murshed, "Texture Based Vein Biometrics for Human Identification: A Comparative Study," in *Proc IEEE SDIM-COMPSAC 2018*, July 23-27, 2018.

[7] Periodogram power spectral density estimate in MATLAB. [https://www.mathworks.com/help/signal/ref/periodogram.html]

[8] R. Esteller, "A Comparison of waveform fractal dimension algorithms", *IEEE Trans. Circuits and Systems*, Vol. 48, no. 2, February 2001.

[9] R. G. Brown and P. Y. C. Hwang, Introduction to Random Signals and Applied Kalman Filtering, 4ed, J. Wiley, NY, 2012.

[10] G. Armano, C. Chira, N. Hatami, "Error-Correcting Output Codes for Multi-Label Text Categorization," in *Proc. Third Italian Info. Retrieval Workshop (IIR)* , pp. 26 – 37, 2012.

[11] B. Li and C. Vogel, "Improving Multiclass Text Categorization with Error-Correcting Output Coding and Sub-class Partitions," *Canadian AI 2010*, vol. LNAI 6085, pp. 4–15, 2010.

[12] L. Rokach and O. Maimon, Data mining with decision trees: theory and applications, World Scientific Pub. Co. Inc. 2008. ISBN 978-9812771711.

[13] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175 – 185, 1992 (published online in February, 2012).

[14] J. Han, M. Kamber, and J. Pei, "Data mining: concepts and techniques, 3rd edn. Elsevier, Amsterdam ((2011).

[15] M. K. Bashar, F. Reza, Z. Idris, and H. Yoshida, "Epileptic seizure classification from intracranial EEG signals: A comparative study", in *Proc of IECBES, 2016*.