THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Diagnosing Heart Abnormality from PCG Signals using K-Means Clustering

Senthil Kumar T. K.
Assistant Professor, Department of ECE, Rajalakshmi Institute of Technology, Chennai, India
Arun A.
UG Scholars, Department of ECE, Rajalakshmi Institute of Technology, Chennai, India
Jagannathan PV.
UG Scholars, Department of ECE, Rajalakshmi Institute of Technology, Chennai, India
Lakshminarayanan S.
UG Scholars, Department of ECE, Rajalakshmi Institute of Technology, Chennai, India

Abstract:
Diagnosingcardiovascular diseases are now a days getting very critical , though there may be several classical methods like electrocardiography and ultrasound imaging to identify the abnormality in the functioning of heart , processing the PCG signals gives a lot of value added information in classifying the murmurs separately from S1(lub) and S2(dub). It is a cheap and non-invasive method which provides better information regarding the mechanism of heart valves and hemodynamics.It has been known that the presence of the heart murmurs in one's heart sound indicates that there is a potential heart problem. Thus, the goal of this paper is to develop a technique for detecting and classifying murmurs. Such a technique can be used as part of an automatic heart diagnostic system. Initially we developed an algorithm to detect S1 and S2 heart sounds, we extracted several features from the PCG signals and tested it with pathological and non-pathological heart sounds. The k-Means clustering concept was implemented , which is used to classify the signals based on the obtained features.The obtained results had an overall efficiency of 86.67 % and sensitivity of 92.857 % from a total of 52 PCG signals that were obtained from clinical database. The algorithm was implemented in Matlab programming language version R2013b.

Keywords*: murmurs, hemodynamics, diagnosing, k-means and pathological*

1. Introduction

Cardiovascular diseases (CVD) are the leading cause of death worldwide. Obesity, irregular lifestyles are the main cause of CVD.Most devices for cardiac remote monitoring are based on ECG, however, as opposed to cardiac auscultation, it doesn’t provide information on heart valves or hemodynamics, important elements for heart diseases detection, especially valvular disorders. Additionally, cardiac auscultation is non-invasive, low cost, reliable and easy to perform [2].There are advanced imaging techniques like EKG, MRI and CT , though they provide more direct evidence but require expensive equipment, specialized technicians to operate, experienced cardiologists to interpret the results, high maintenance cost, a permanent place to be installed and generally require more resources to function properly[1]. But processing of the PCG signals does not require much capital, the heart sounds can be obtained from patients using a i-stethoscope or by any other digital stethoscope. So by generating a computer aided system that can assist a physician in identifying the abnormalities of heart at a higher accuracy rate we can eradicate CVD to a much extent. In further discussion a automated tele-diagnosing system can be developed which can be included in the home care unit, so it serves well for elderly people who can’t come to the hospital for regular check-up.

2. Methodology

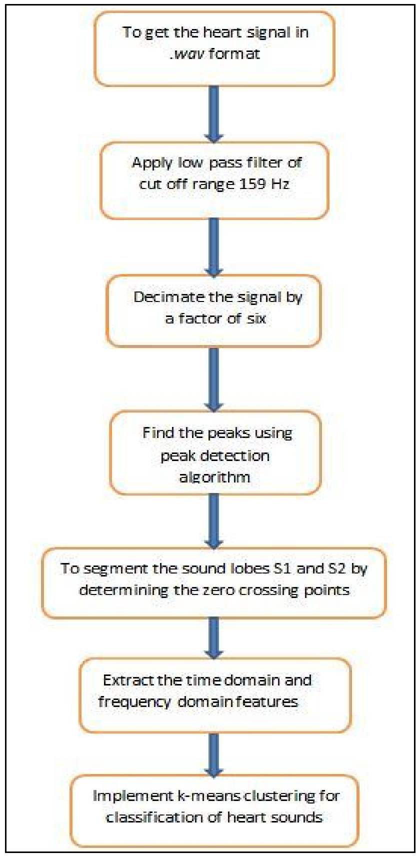

Figure 1: Flow chart of the proposed methodology

After getting the raw heart signal, we have to pre-process the signal. The pre-processing involves the steps of filtering, decimating etc. Then from the pre-processed signal we calculate the peaks of sound lobes and differentiate between S1(lub) and S2(dub). After which the sound lobes is to be segmented to determine the features. Several time domain and frequency domain features are calculated for the segmented signal. The extracted features are given through k-means clustering algorithm for classification purposes.

2.1. Data Acquisition and Preprocessing

The data was obtained from clinical database from a total of 52 patients which had both pathological and non-pathological diseases. The average age group of the patients was 37 years.

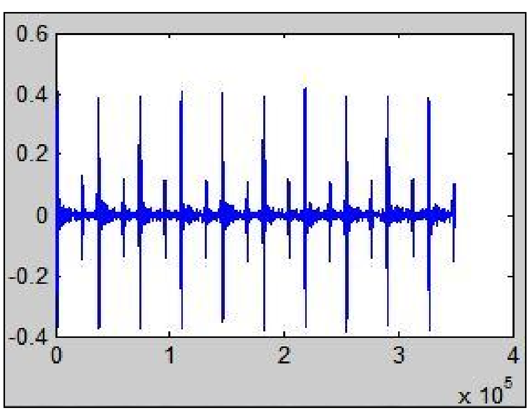

Figure 2: raw phonocardiogram signal

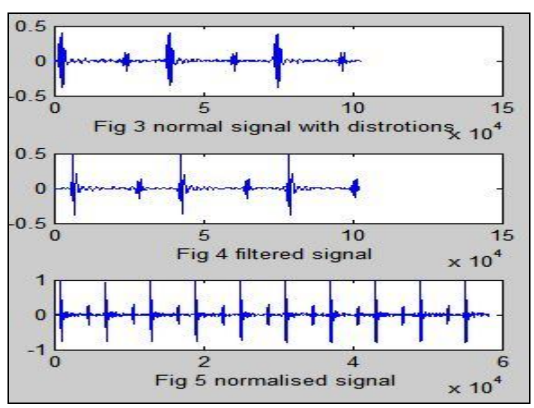

Figure 3, 4, 5

Heart sound signal is a one dimensional signal. The major components in the heart sound signal is S1 (lub) and S2 (dub) also the third and fourth heart sound i.e. S3 and S4 may be heard. If there was any pathology then murmur sound will be heard significantly.. The presence of extra heart sound either in the systole and diastole is not a major problem, but identifying them at an early stage could a serve a lot. The heart sound was converted to .*wav*format so it can be given as input to the Matlab. There may be some external noise embedded along with heart sound. Usually the heart sound is low frequency component and noise is high frequency component. A low pass filter of cut off frequency 159 Hz was used. The filtered signal was then down sampled by a factor of 6 using the decimate function in Matlab so that the details and approximations can result in frequency bands which contain the maximum power of S1 and S2.In the Artifact category there are a wide range of different sounds, including feedback squeals and echoes, speech, music and noise. There are usually no discernable heart sounds, and thus little or no temporal periodicity. This category is the most different from the others. It is important to be able to distinguish this category from the other three categories, so that someone gathering the data can be instructed to try again. We found the signal to noise ratio of all the heart signals those with minimum values were considered as Artifact.

$$x_{norm} = x/|\max(x)|$$

The signals were then normalized to the absolute maximum because the signals may be obtained using different devices, in order to bring all the signals in the common range of -1 to +1 the signals are normalised. Where x is the actual signal

2.2. Peak Finding Process

2.2.1. Todetect the peaks of S1 and S2 sound lobes using a threshold value

We need to identify the location of S1, S2 and also calculate the time period of systole and diastole. A peak detection algorithm was implemented to identify the peaks of S1 and S2

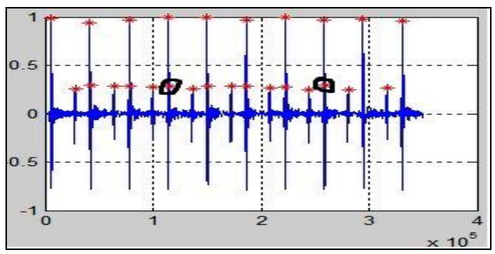

Figure 6: finding the peaks of the signal

A point is said to be peak in the signal if that instant value is greater than the previous three instant values and the following three instant values which can be implemented by a simple *for* loop. Now we will get several peaks in the signal so we need to set a threshold value to distinguish S1 and S2, usually the amplitude of S2 is greater than that S1 .

2.2.2. Rejecting Extra Peaks

There may also be extra peaks adjacent to the determined peaks. These extra peaks can be rejected, when a peak is determined within 80 ms of the previous peak, the lower amplitude of the two peaks can be rejected.

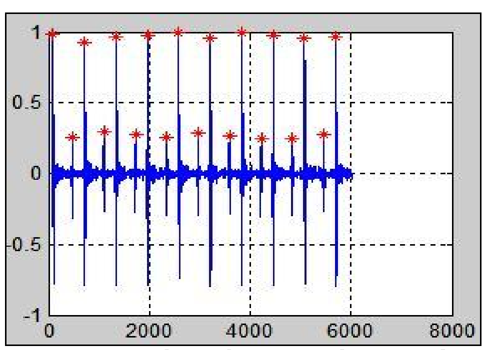

Figure 7: plot with extra peaks rejected

The longest distance between the two sounds is considered as the diastolic period, the sound at right side was assigned as S1 and that of left side was S2. The distance between S1-S2 was calculated for every segment compared. But these intervals vary from file to file. So this cannot be taken as a feature , we then figured several time domain and frequency domain features.

3. Segmentation of Sound Lobes

To segment the sound lobes S1 and S2 from the entire signal, initial we need to find the zero crossing points in the signal.

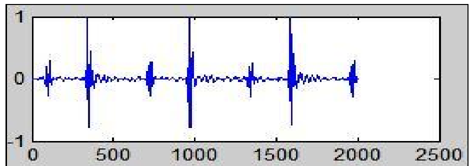
Fig 8 pre-processed pcg signal

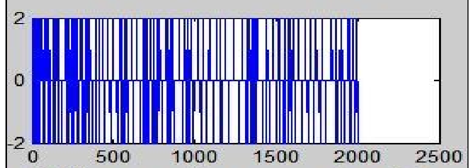
Fig 9 corresponding zero crossing points

There will be positive going and negative going zero crossing points.

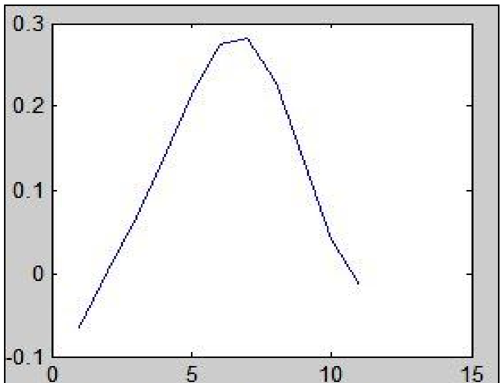
Fig 10 segmented sound lobe S1

For an instant peak location we need to find the nearest positive going and negative going zero crossing point, say for example a1 and a2. Nowwhen we plot from a1 to a2 of the signal we will get the segmented sound lobe.

4. Feature Extraction

4.1. Time Domain Features

Several time domain and frequency domain features were extracted. A ratio is calculated between the mean of the segmented sound lobe and the pre-processed signal. Likewise standard deviation and variance can also be calculated. Mean is the average of numbers, a calculated “central” value of a set of numbers.standard deviation is the “mean of mean”. varianceis the measure of spread between numbers in a data set .the variance measures how far each number in the set is from the mean. Then entropy and total harmonic distortion is determined for then signal.

Entropy(H) = - ∑ p(x) log p(x)

Entropy refers to the relative degree of randomness. The higher the entropy, the more frequent are signaling errors. Entropyis directly proportional to the noise and bandwidth of the signal.

$$\mathrm{THD} + \mathrm{N} = \frac{\sum_{n=2}^{\infty} \text{harmonic powers} + \text{noise power}}{\text{fundamental power}}$$

Total harmonic distortionis an amplifier or pre-amplifier specification that compares the output signal of the amplifier with the input signal and measures the level differences in harmonic frequencies between the two.

4.2. Frequency Domain Features

Frequency domain features is calculated for the signal by taking discrete fourier transform by means of fast fourier transform algorithm. In frequency domain, for particular frequency range alone entropy and power is calculated.

Power = ∑ (x.^2) / length(x)

Power is defined as the amount of signal energy consumed per unit time.

Z=(length(indx_up)+length(indx_down))/length(x)

Zero crossing rate(z)is the ratio of sum of positive going and negative going zero crossing point to the length of the signal.Where x is the pre-processed signal.indx_up is the positive going zero crossing points while indx_down is negative going zerocrossing points with respect to any instant peak location.

5. Classification Using K-Means Clustering Algorithm

K-means is an algorithm to classify or to group the objects based on attributes/features Into K number of group. Where K is a positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data. The process used in this method is very simple. Initially K value (i.e.) the number of clusters into which we need to classify the data is mentioned. Then we assume the centroid or the centre of these clusters. The initial centroids can be any random objects.

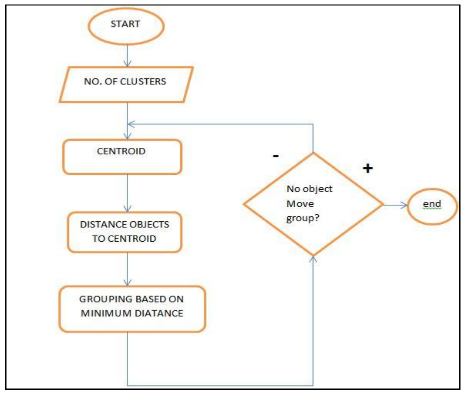

Figure 11: Flowchart for K-means clustering algorithm

The following steps take place as a backend process. The centroid coordinate is determined, then the distance from each object from the centroid is calculated. Then grouping is done based on the minimum distance (by the closest centroid).

6. Experimental Results

Both pathological and non-pathological sounds were obtained from online databases such as http://www.peterjbentley.com/heartchallenge/http://www.med.umich.edu/lrc/psb/heartsounds/ also heart sounds were collected from clinical databases. Many PCG signals were rejected as they were corrupted by background noise, weak heart sounds and respiratory sounds.To assess the algorithm performance, sensitivity (SE) and specificity (SP) are considered.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specitifity = \frac{TP}{TP + FP}$$

To calculate these parameters 3 variables are used, true positives (TP), false positives (FP), false negatives (FN). A TP is considered when a sound is correctly identified, a FP is considered when a sound is incorrectly detected, and a FN is considered when a sound is not detected. The calculation of SP uses a slightly different expression from the regular one, which included the variable true negatives (TN), correct classification of no event.

A GUI was developed to process the PCG signals and classify the signal by means of k-means clustering algorithm.

Signal	FMS1	FMS2	FSDS1	FSDS2	THD	ZCR	ENTROPY	POWER
1.	0.85	2.771	7.683	4.459	-11.41	0.104	3.2575	0.0285
2.	2.48	2.906	2.176	8.449	-11.80	0.108	3.2674	0.0271
3.	1.32	2.884	2.134	8.318	-11.19	0.114	3.1940	0.0257
4.	0.96	3.281	2.390	10.76	-13.15	0.135	3.1481	0.0208

Table 6.1: features obtained for normal signal

SIGNAL	FMS1	FMS2	FSDS1	FSDS2	THD	ZCR	ENTROPY	POWER
1.	4.1149	2.588	2.014	6.698	-11.44	0.108	3.1770	0.0263
2.	4.0339	2.052	2.912	4.211	-11.30	0.137	3.2949	0.0272
3.	5.3604	2.050	2.944	4.205	-11.99	0.148	3.1987	0.0210
4.	4.5094	2.539	2.045	6.449	-11.13	0.139	3.2980	0.0273

Table 6.2: features obtained for murmur signal

SIGNAL	FMS1	FMS2	FSDS1	FSDS2	THD	ZCR	ENTROPY	POWER
1.	-0.1723	-0.069	4.056	3.448	-193.18	0.34	1.2313	76.2854
2.	-0.3705	-0.240	4.986	1.868	-188.66	0.331	2.9092	2.3132
3.	-0.4375	-2.072	3.676	1.939	-185.22	0.33	2.9767	2.9540

Table 6.3: features obtained for extraheart sound signal

signal	Classified pattern
Normal-1	1
Normal-2	2
Normal-3	1
Normal-4	1
Normal-5	1
Murmur-1	2
Murmur-2	2
Murmur-3	2
Murmur-4	1
Murmur-5	2
Murmur-6	2
Murmur-7	2
Extra heart sound-1	3
Extra heart sound-2	3
Extra heart sound-3	3

Table 6.4: Result of k-means clustering algorithm

7. Conclusion

The classification of different Heart sounds are carried out and an overall efficiency of 86.67 % and sensitivity of 92.857 % from a total of 52 PCG signals that were obtained from clinical database.

The analysis of heart sounds in new born babies can be very useful for deciding to release or send them to echocardiogram. Hence, it is very important to devise an effective method for analysing heart sound defects. This paper introduced a new method for heart sound segmentation, and feature extraction which is applicable even in presence of murmurs. Heart murmurs are diagnosed using several features, including Shannon energy, ZCR, THD and classifying them using k-means clustering. The front end software tool created for this work using MATLAB GUI.

Further work is under way to improve feature extraction and classification, so that the efficiency of this work can be improved.

8. References

1. A book on PHONOCARDIOGRAPHY SIGNAL PROCESSING by Abbas k. Abbas, RashaBassamisbn, Lectures on Biomedical Engineering, 2009, Vol. 4
2. Faizan Javed, P A Venkatachalam and Ahmad Fadzil M H "A Signal Processing Module For The Analysis Of Heart Sounds And Heart Murmurs". Journal of Physics: Conference Series, 2006 *J. Phys.: Conf. Ser.* 34 1098
3. Deng, Y. and Bentley, P. J. (2012) A Robust Heart Sound Segmentation and Classification Algorithm using Wavelet Decomposition and Spectrogram. Extended Abstract in the First PASCAL Heart Challenge Workshop, held after AISTATS 2012, La Palma. March 25 2012
4. Gomes, E.F., Bentley, P.J., Coimbra, M., Pereira, E., Deng, Y. (2013) Classifying Heart Sounds: approaches and results for the PASCAL Challenge. In Proc. 6th International Conference on Health Informatics, HealthInf 2013, Barcelona, Spain, Feb 2013.
5. A REMOTE MONITORING PLATFORM FOR AUSCULTATION by Tiago Filipe TeixeriaSapata in Faculty of sciences and technology, University of Coimbra(2010)
6. Ahlstrom C, Hult P, Rask P, Karlsson JE, Nylander E, Dahlström U, Ask P. Feature extraction for systolic heart murmur classification" Ann Biomed Eng. 2006 Nov;34(11):1666-77. Epub 2006 Oct 4.
7. Olmez, T. and Z. Dokur "Classification Of Heart Sounds Using An Artificial Neural Network" Pattern Recogn. Lett.24:617–629, 2003