Heart Diseases Identification System Using Fuzzy Cluster Algorithm

Naveena Joshy
PG student
Department of Electronics and Instrumentation Engineering
Karunya University, Coimbatore

D. Pamela
Assistant Professor
Department of Electronics and Instrumentation Engineering
Karunya University, Coimbatore

Abstract: **In order to diagnose many heart diseases, physicians and health professionals performs electrocardiogram analysis. The development of a portable heart diagnosis system able to process the signal in real time is important to monitor high risk patients and to assist physicians to make decisions. This paper presents the development of heart diseases identification system offers reduction on electrocardiogram signal – ECG signal processing. The ECG signals are applied into system that performs an initial filtering, fuzzy cluster algorithm and correlation. The result shows that the developed system can perform the detection of heart diseases on reduced set of data than the original signal, while maintaining the same effectiveness. The simulation and coding of heart diseases identification system is done in MATLAB.**

Keywords: - Electrocardiogram – ECG signal, Heart diseases, Fuzzy clustering

I. INTRODUCTION

Diagnose silent heart diseases and the risk for a heart attack in someone in apparent good health and without symptoms such as chest discomfort or breathlessness is a very common dilemma. Unfortunately, among cardiologist it is an area of great confusion. The analysis of an electrocardiogram is performed in real time by physicians and other health professionals in order to identify many diseases. The analysis of electrocardiogram signal consists of extracting information from the peak and time intervals of the signal waveform.

Developing of an algorithm for separating P wave, QRS complex and T wave from PQRST waveform is a very complicated problem due to the time varying morphology of the signal subject to the physiological conditions and presence of noise. The algorithms used earlier adapt a range of different approaches to yield a procedure leading to the denoising and identification of the waves under consideration. Some algorithms are based on signal processing technique that uses computer algorithms for the extraction of the electrocardiogram characteristics, thus allowing preliminary diagnosis of a cardiopathy. Some others introduce a cardiac arrhythmia classification system using fuzzy classifiers that uses algorithms based on artificial intelligence and a knowledge based to classify arrhythmias. Some systems are using neural network for ECG signal classification and was implemented in hardware.

Electrocardiography is an interpretation of the electrical activities of the heart, as detected and recorded by using electrodes attached to the surface of the body and device external to the body. The development of an embedded system to extract main characteristics of ECG signal faces challenges such as time varying signals and physiological variables, including patient gender and age and the noises generated by the external world is also taken into account. Filtering is necessary to use filtering, so that the clustering can be conducted later. By using correlation technique, the heart diseases are identified.

II. METHOD

For test and evaluation, the system uses the European electrocardiogram ST-T Database (EDB) as a reference. The data bank consists of 90 ECG records. Each record consists of two hours long signals which are sampled at a rate of 250 samples per second with 12 bit resolution in 20mV intervals. To find the full cycle, the signal is scanned using ECG pattern detection algorithm and analyzed using clustering processes. The Fig 1 shows the full cycle of ECG signal starts with a P segment and ends with a T segment and consists of 213 samples per ECG cycle for all signals.

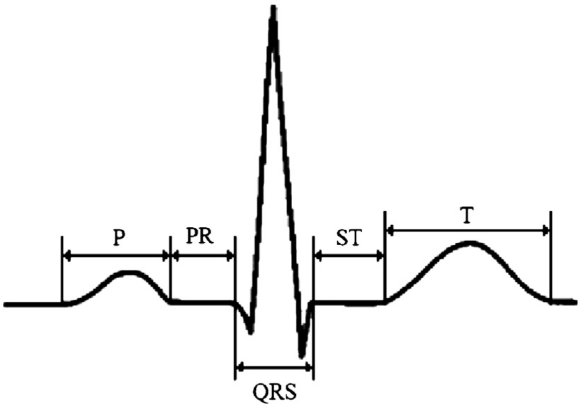

Fig 1: Typical electrocardiogram waveform with segment indication

The basic ECG signal processing algorithm is comprised of three modules:

1) Filtering process,
2) Fuzzy cluster algorithm and
3) Correlation with previously diagnosed data bank.

III. FILTERING

Electrocardiogram signal of different patients having angina, myocardial infraction and normal signal are acquired. Fig 2 shows the ECG signal acquired from the data bank. Signal acquisition must be noise free, for the diagnostic quality of ECG recordings. The disturbances from other biological and environmental source interfere with the signal during the acquisition. The noise generated by other electrical equipments in the vicinity and by the patient's muscle movement may cause wrong decision of diseases. Due to the alteration of the signal, the noise can cause errors on the process under analysis. The filtering is the most relevant component in an ECG signal processing.

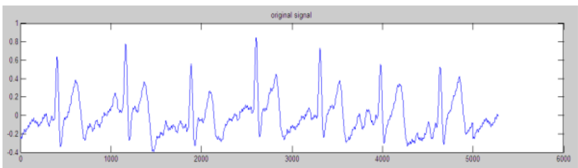

Fig 2: ECG signal with noise and DC level

Filtering is used for noise reduction and also to identify and to classify the signal. The preprocessing of ECG signal consists of baseline wandering, denoising and filtering. In order to mask the key features of the electrocardiogram signal baseline wandering is used and is desirable to remove the noise for proper analysis and display of the ECG signal. Filtering eliminates noise and DC level of the ECG signal. The average obtained from the signal samples represents the DC level of the signal. The DC level is eliminated by subtracting the average from each sample. The noise at the frequencies higher than the ECG is eliminated by using a low pass filter.

The preprocessing performs wavelet denoising, baseline wandering and low pass Butterworth filtering. The Butterworth filter offers a flat pass band and a null rejoinder outside it. Only slope is changed as the filter order is altered is the advantage of Butterworth filter. The Butterworth filter was implemented by the basic equation is given by [1]:

$$|T(j\omega)| = \frac{1}{\sqrt{(1+e^2(\frac{w}{w_p})^{2N})}}$$

Threshold wavelet denoising [2] technique is used to remove the DC level of the ECG signal. In wavelet denoising technique, thresholding is used for noise removal. Wavelet thresholding of ECG signal is done by thresholding the wavelet coefficients to eradicate their noisy part. The preprocessing unit containing filter and denoising eliminates the DC component and noise on the ECG, results in a clean ECG, which is ready for further processing.

IV. ALGORITHM

A membership function indicates the membership degree of a particular element gazing at an event. The range of the membership function is stuck between 0 and 1, thus pointing towards the level influence of an element regarding an event. In order to trim down the number of samples and thus to minimize the amount of input signal, it is used a fuzzy clustering algorithm. The test conducted shows that Gaussian membership function is superior to the triangular, sigma and sine membership functions, in view of the fact that it provided the smallest error compared to other functions. Fig 3 shows the Gaussian membership fuzzy function that is given as [1]:

$$\mu_{Ai}(x) = \exp\left(-\frac{(c_i - x)^2}{2\sigma_i^2}\right)$$

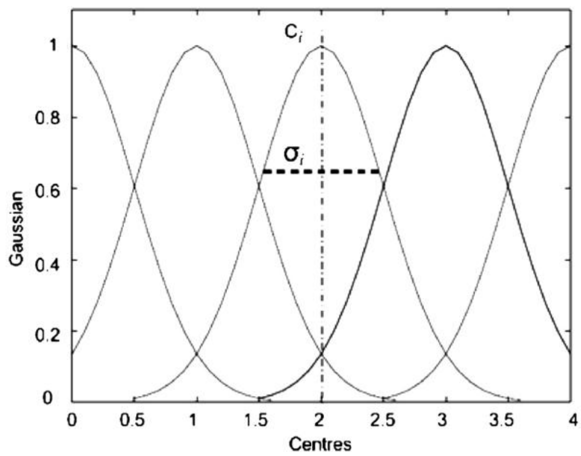

Fig 3 : Gaussian membership function

Where c_i and σ_i are the center width of i^{th} fuzzy set, respectively.

A. Gustafson- Kessel fuzzy cluster algorithm:

Gustafson and Kessel (Gustafson and Kessel, 1979) expanded the standard fuzzy c means algorithm by make use of an adaptive distance norm, in order to perceive the clusters of diverse geometrical shapes in one statistics set. Each cluster is the owner of their own norm-inducing matrix [6] ***Ai*,** which acquiesces the following inner-product norm:

$$D_{ikAi}^2 = (Z_k - V_i)^T A_i (Z_k - V_i)$$

In the *c* – means functional, the matrices ***Ai*** are applied as the optimization variables. These optimization variables are thus allowing each cluster to acclimatize the distance norm to the local topological structure of the data. The GK algorithm is defined by (Gustafson and Kessel, 1979) [6]:

$$J(Z; U, V\{A_i\}) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m D_{ikA}^2$$

The objective function cannot be straightly reduced with respect to ***Ai***, because it is linear in ***Ai***. ***Ai*** should be constricted in some way in order to obtain a practicable solution. The normal way of constrict is accomplished by contrain the determinant of ***Ai***:

$$|A_i| = \rho_i\ , \qquad \rho_i > 0,\ \forall_i$$

Allowing the matrix ***Ai*** to show a discrepancy with its determinant which is fixed corresponds to optimizing the cluster's shapes while its volume remains invariable. The expression for the matrix ***Ai*** [6] is obtained by using the language multiplier methods, and is given as follows:

$$A_i = [\rho_i \det(F_i)]^{\frac{1}{n}} F_i^{-1}$$

Where F*i* is the fuzzy covariance matrix [1] of the *i*th cluster is given by:

$$F_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m (Z_k - V_i)(Z_k - V_i)^T}{\sum_{k=1}^{N} (\mu_{ik})^m}$$

In distance norm, the membership degree in U is exploited to weight the covariance. The GK algorithm is computationally more engrossed, since the inverse and the determinant of the cluster covariance matrix must be computed in each iteration. The Gustafson-Kessel fuzzy cluster algorithm reduces the number of samples to be process the ECG signal than the original signal.

V. CORRELATION

By using correlation technique, the ECG signal is compared with other previously sampled signals which are already stored in the data base [1]. The comparison presents the resemblance factor between the electrocardiogram signal and the previously identified data bank signals stored in the system memory. The evaluation offering the largest comparison factor is distinguished as the possible patient's disease diagnosis.

It is permissible to reach three conclusions from the analysis of the correlation between two phenomena [1]. A correlation factor close to -1 (negative correlation) point towards a strong inverse correlation between the two phenomena. If the correlation factor close to 0, either positive or negative, specifies that there is no correlation between the phenomena. A correlation factor close to 1 showing that there exist a strong correlation between the data under analysis.

It is appraised the correlation between the data generated by the fuzzy clustering algorithm and the previously fuzzy clustered data bank. The system will recognize the diagnosis as the signal from the data base that receives the highest correlation with the assessed signal.

VI. RESULTS AND DISCUSSIONS:

This section presents the results and discussion of heart diseases identification system. The method discussed in this paper presents a simple and consistent method for detecting the heart diseases. As previously argued, the results were tested and evaluated using the European electrocardiogram ST-T Database (EDB) as a reference. The signal acquired from Physionet is filtered to eliminate noise and the DC component to obtain a clean ECG signal. Fig 4 shows the filtered signal and fig 5 shows the denoised signal.

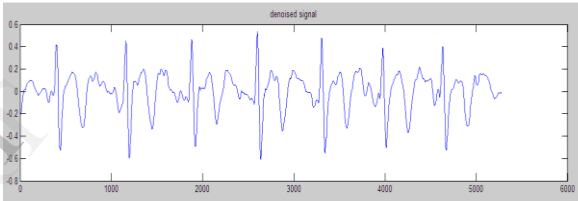

Fig 4: Filtered ECG signal

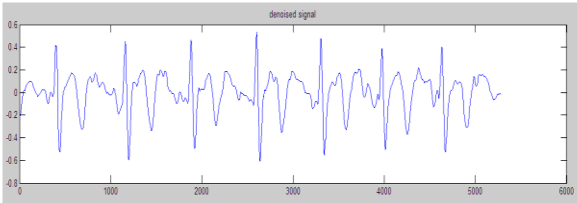

Fig 5: Denoised signal

It is applied the Gustafson – Kessel fuzzy cluster algorithm to the preprocessed signal shown in fig 5. It reduces the number of samples which is to be correlated as shown in fig 6. It is observed that this much of samples are enough to draw waveform that maintains the main characteristics of electrocardiogram signal.

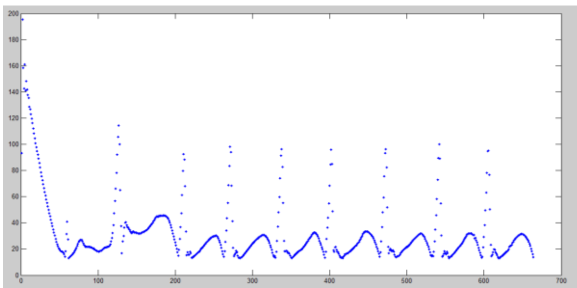

Fig 6: Reduced samples after fuzzy clustering algorithm

VII. CONCLUSION AND FUTURE WORK:

The novel method in this paper provides a simple and reliable method for heart diseases detection. As previously pointed, the results were obtained by using the Physionet EDB databank for test and evaluation. A database of previously known ECG diagnoses was created in order to implement the proposed system. All data were preprocessed and only the clusters containing the main features of the signal were stored in system memory. This method is more efficient because it reduces the number of operations by reducing number of samples to be processed which results in reduced processing time required to be processed.

The future work includes simulation and implementation of heart diseases identification embedded system. It is also possible to include more database corresponds to many other diseases. This system can be validated in a Xilinx Spartan FPGA.

REFERENCES

[1] Helton Hugo de Carvallo Juniora, Robson Luiz Morenob, Tales Cleber Pimentab, Paulo C. Crepaldib, Evaldo Cintrab "A heart diseases recognition embedded system with fuzzy cluster algorithm" *ELSEVIER,(Jan 2013).*

[2] Mrs. V. R. Lele, Prof. K. S. Holkar, " Removal of baseline wander from ECG signal" *Special Issue of International journal of Electronics, Communication and Soft Computing Science and Engineering, ISSN (2013).*

[3] A. Baraldi, P. Blonda, "A Survey of fuzzy clustering algorithms for pattern recognition ", *IEEE Transaction on system, Man and Cybernetics, Part B: Cybernetics 29 (December(6)),(2009).*

[4] Naser Safdarian, Keivan Mghooli and Nader Jafarnia Dabanloo,"Classification of cardiac arrhythmias with TSK Fuzzy System using Genetic Algorithm", I*nternation Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 5, No. 2, (June, 2012).*

[5] F. A. Afsar M.U Akram M. Arif J. Khurshid,"A pruned fuzzy k-nearest neighbour classifier with application to electrocardiogram based cardiac arrhythmia recognition", *IEEE international multi topic conference, December 23-24, (2008).*

[6] Hui WANG Chengjun Huang, Linpeng Yao, Yong Qian, Xiuchen Jiang, "Application of Gustafson- Kessel clustering algorithm in the pattern reconition for GIS", *PRZEGLAND ELEKTROTECHNICZNY (electrical review), ISSN 033-2097, (2011).*

[7] Magne setnes, " Supervised Fuzzy Clustering for Rule Extraction" *IEEE Trasactions on Fuzzy Systems, Vol. 8, No. 4,(August 2008).*

[8] Mrs. B. Anuradha, V.C. Veera Reddy, "Cardiac arrhythmia classification using fuzzy classifiers" *Journal of theoretical and applied information technology (2009).*

[9] Andrea Baraldi and Palma Blonda, Member of IEEE "A Survey of Fuzzy Clustering Algorithm for Pattern Recognition- Part II" *IEEE transactions on system, man, and cybernetics- Part b: cybernetics, Vol. 29, No. 6, (December 2009).*