Towards the development of a new wavelet for ECG classification

Baby Paul, K.T. Shanavaz, P. Mythili
babypaul@cusat.ac.in, shanavazkt@cusat.ac.in, mythili@cusat.ac.in
Division of Electronics, School of Engineering
Cochin University of Science and Technology
Kochi, Kerala
India

Abstract— **In this paper an attempt has been made to determine the number of Premature Ventricular Contraction (PVC) cycles accurately from a given Electrocardiogram (ECG) using a wavelet constructed from multiple Gaussian functions. It is difficult to assess the ECGs of patients who are continuously monitored over a long period of time. Hence the proposed method of classification will be helpful to doctors to determine the severity of PVC in a patient. Principal Component Analysis (PCA) and a simple classifier have been used in addition to the specially developed wavelet transform. The proposed wavelet has been designed using multiple Gaussian functions which when summed up looks similar to that of a normal ECG. The number of Gaussians used depends on the number of peaks present in a normal ECG. The developed wavelet satisfied all the properties of a traditional continuous wavelet. The new wavelet was optimized using genetic algorithm (GA). ECG records from Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) database have been used for validation. Out of the 8694 ECG cycles used for evaluation, the classification algorithm responded with an accuracy of 97.77%. In order to compare the performance of the new wavelet, classification was also performed using the standard wavelets like morlet, meyer, bior3.9, db5, db3, sym3 and haar. The new wavelet outperforms the rest.**

***Keywords*- Electrocardiogram, Eigenvalue, Genetic Algorithm, K-means, new wavelet, Principal Component Analysis , Premature Ventricular Contraction .**

I. INTRODUCTION

Bio-electrical signals express the electrical functionality of the different organs in the human body. The electrocardiogram is one of the important signal among all bio-electric signals. The ECG reflects the performance and properties of the human heart and conveys important hidden information about the structure and its working. Extracting this information or feature from ECG has been found to be very helpful in explaining and identifying various pathological conditions. Although this work can be done by analyzing ECG visually on paper or on a screen, the complexity and duration of the ECG signals make the manual analysis a very time consuming process. Therefore ECG signal processing has become an important tool in extracting clinically important information from ECG signals.

This work is limited to focus on the classification of an arrhythmia called PVC. In a normal heart beat the ventricles contract after the atria has been sufficiently filled with blood so that they can pump a maximum amount of blood to the lungs and other parts of the body. In PVC the ventricles contract first, which means that the circulation of blood is insufficient [1]. PVC is caused by an ectopic cardiac pacemaker located in the ventricle. This is characterized by unusually shaped QRS complex usually wider than 120 ms in lead II configuration. These QRS complexes will not be preceded with P-wave and the T wave. The clinical significance of PVC depends on their frequency, complexity and hemodynamic response.

Previous literature shows that a large number of works has been done in the classification of premature ventricular contraction. Al Nashash [2] explained ECG arrhythmia classification using PCA and reported a classification sensitivity and positive predictivity of 98.1% and 94.7%, respectively. Nur Asyiqin et al [3] in their work have tried to find an optimal Discrete Wavelet Transform (DWT) that would accurately classify PVC and normal beats using Probabilistic Neural Network (PNN). It was observed that “haar”, “db3” and “sym3” wavelets produced sufficiently accurate results with 400 beats. V. Mahesh et al [4] presented a diagnostic system for classification of cardiac arrhythmia from ECG data, using Logistic Model Tree (LMT) classifier. The results obtained indicate prediction accuracy of 98% using 1281 beats from MIT-BIH database [5]. Awadhesh Pachauri et al [6] described a wavelet and energy based technique for the detection of ventricular premature arrhythmic beats in ECG. Using four records from MIT-BIH data base [5] an overall accuracy of 86.48% was obtained. P. Ghorbanian et al. [7] developed an algorithm using Continuous Wavelet Transform (CWT), PCA and multi-layered perceptron neural network (MNLPNN) to detect and classify six classes of ECG. Using 600 beats from different records of MIT BIH database [5] they obtained 99.5% sensitivity, 99.66% positive predictive accuracy and 99.17% total accuracy. S Almagro et al. [8] designed a new mother wavelet for abdominal electrocardiogram (AECG) analysis. The algorithm is evaluated by AECG data from the Database for the identification of systems shows low MSE <<1.2%, RMS <<0.016 μ volt, and excellent visual similarity between the original and the reconstructed AECG.

From the above literature it is seen that even though the percentage accuracy obtained in the work done by Ref. [3, 4, 7] is high, the number of ECG cycles used for classification is less. Ref. [6] uses 4 records from MIT-BIH [5] database but the percentage of accuracy obtained was just 86.48%. On the other hand this work which uses five full records from MIT BIH

978-1-4673-0449-8/12/$31.00 ©2012 IEEE

database [5] (8694 beats) gives a better classification accuracy. In this work normal ECG and PVC beats are classified using a new wavelet. The classification is based on CWT, PCA and K-means classifier. The rest of the paper is structured as follows. Section II describes the proposed work and its methodology. Section III provides results and discussion and section IV concludes the paper.

II. PROPOSED WORK

A. Data source and content

The database selected for analysis is the MIT-BIH Arrhythmia Database [5]. It contains a large number of long-term Holter recordings that were obtained by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. 48 records selected from this are available in the website named Physionet [5]. Each of these is nearly 30 minutes long. Among these records 106, 114, 116, 119 and 217 were selected for our analysis. These records contain a total of 7416 normal beats and 1278 premature ventricular contractions.

B. Methodology

Fig. 1 shows the block diagram of the proposed ECG classification and GA based wavelet evolution system. In the preprocessing stage each ECG record was selected and preprocessing steps were applied to remove the baseline wander using two stages of a moving average filter [9]. In baseline wander, the isoelectric line change position, possibly due to various reasons like cable movement during reading, patient movement, dirty lead wires or electrodes, loose electrodes. Moving average filter is optimal for reducing random noise while retaining a sharp steep response. Pan-Tompkins algorithm [10] was used for detecting the R-peak in the ECG cycle. This is a real time QRS detection algorithm based on analysis of the slope, amplitude and width of the QRS complexes. The algorithm includes a series of filters and methods that perform low pass, high pass, derivative, squaring, integration, adaptive thresholding and search procedures. Once the sample number corresponding to the R wave has been identified, the ECG cycle is extracted by selecting sufficient samples to the left and right of the R-wave. In the next stage a new custom made mother wavelet was constructed in such a way that its shape was almost similar to that of a normal ECG. The amplitude and duration of the PQRST complex (5 peaks) of the normal ECG were taken in to consideration in the design of the mother wavelet.

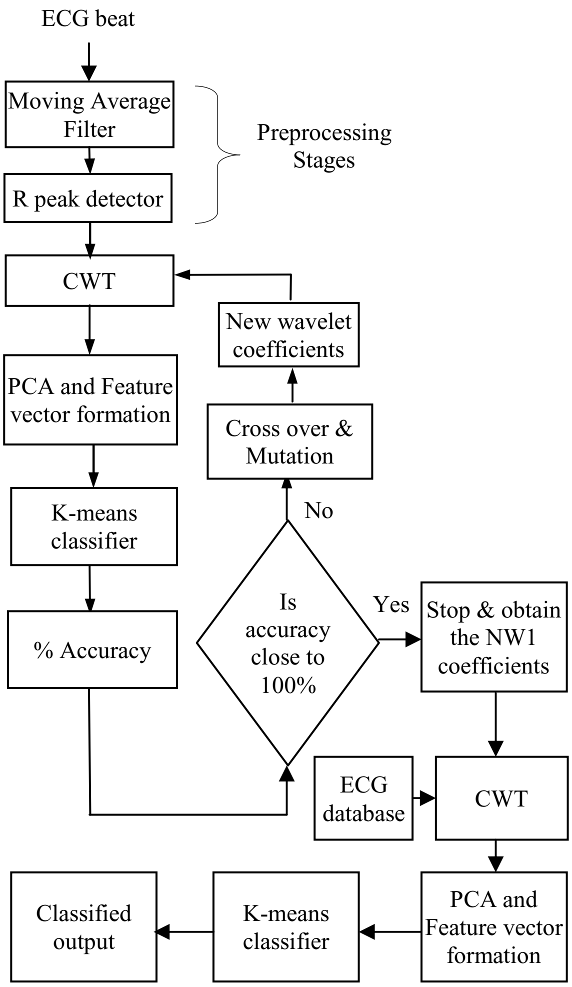


Figure 1. Block diagram of the GA based wavelet evolution and classification system

Since a Gaussian is symmetric around its mean, gains its maximum value at the mean and goes very fast to zero, similar to an ECG peak, it has been used to model the peaks. In one dimension the Gaussian function is the probability density function of the normal distribution, and is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}, \qquad (1)$$

where $\mu = \text{mean}, \sigma^2 = \text{variance}$ and x is a real number.

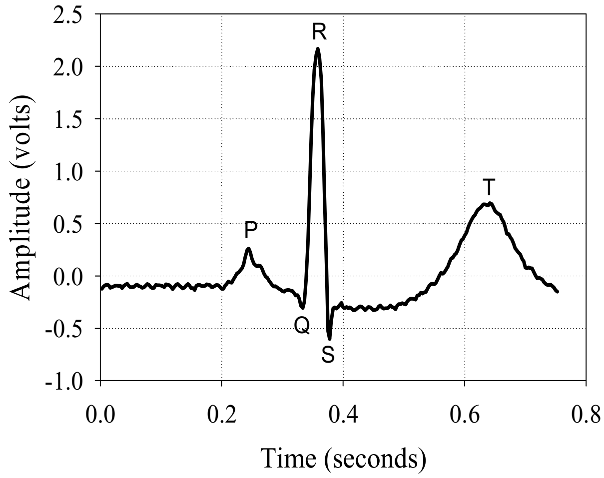
The unoptimized custom made mother wavelet was constructed by adding five different gaussians corresponding to the five peaks of the ECG. The equation corresponding to the custom made wavelet is given by

$$f(x) = P + Q + R + S + T. \qquad (2)$$

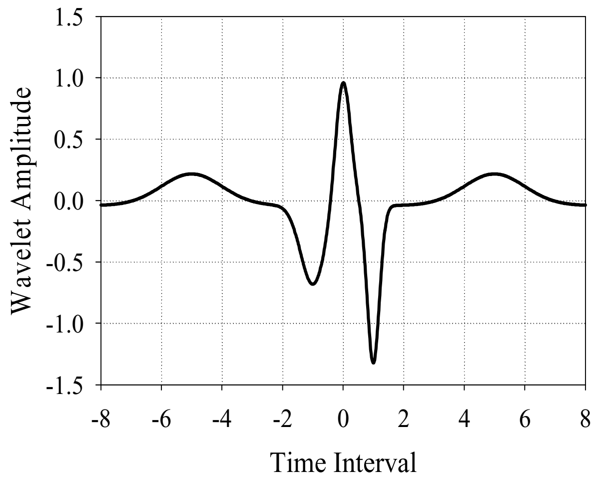
where $f(x)$ = custom made wavelet.

$$\left.\begin{aligned} P &= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x+5)^2 / 2\sigma_1^2}, \\ Q &= \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x+1)^2 / 2\sigma_2^2}, \\ R &= \frac{1}{\sigma_3\sqrt{2\pi}} e^{-(x)^2 / 2\sigma_3^2}, \\ S &= \frac{1}{\sigma_4\sqrt{2\pi}} e^{-(x-1)^2 / 2\sigma_4^2}, \\ T &= \frac{1}{\sigma_5\sqrt{2\pi}} e^{-(x-5)^2 / 2\sigma_5^2}. \end{aligned}\right\} \qquad (3)$$

Here $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ and σ_5 are the parameters to be optimized using GA to obtain the optimum mother wavelet for better ECG classification. In other words the optimum custom made mother wavelet directly depends on the variances of the various Gaussians. To begin with a wavelet which is similar to the ECG waveform is constructed with $\sigma_1 = 1$, $\sigma_2 = -0.4$, $\sigma_3 = 0.25$, $\sigma_4 = -0.2$ and $\sigma_5 = 1$. The properties including the admissibility condition was tested for the new wavelet [11,12]. The plot of the normal ECG and the constructed mother wavelet which looks similar to a normal ECG is shown in Fig. 2(a) & (b).



(a)



(b)

Figure 2. Plot of (a) Normal ECG (b) Unoptimized new wavelet

A new wavelet was then evolved from this custom made wavelet using GA to produce better classification of ECG cycles. This was achieved by choosing the candidate solutions for the mother wavelet to be closer to the one constructed using the variance values mentioned in the previous paragraph. Since GA combines survival of the fittest among chromosomes with a structured and yet randomized information exchange [13] it has the ability to identify the optimum mother wavelet. Reproduction, Cross over and Mutation are the basic operators in GA which helps in the convergence of the solution. For evolving the new wavelet using GA, the accuracy of classification of the ECG cycles was selected as the fitness function. The number of generations was chosen as 100 and population size was selected as 25. The initial range of the σ's was set in the range of -5 to +5. The initial population of GA was created by the randomly mutated copies of 'σ' values corresponding to normal ECG. Roulette wheel selection technique has been used for selecting the solutions to apply cross over operators. Elite count was set as 2 and cross over fraction is 0.8. The scale and shrink parameters which decide the standard deviation of the Gaussian distribution used in the Gaussian mutation function was set as '1'. Two patient records from the same database [5] were selected for training. The optimum evolved wavelet which gave maximum classification accuracy is shown in Fig. 3.

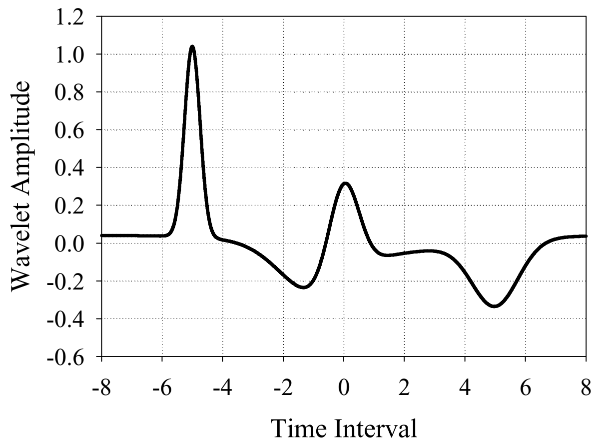


Figure 3. Plot of NW1

Continuous wavelet transform was performed using the new evolved wavelet. The wavelet coefficient matrix obtained contains the details corresponding to the different waves in the ECG cycle. This output has to be given to a classifier. The size of this data is considerably large. It is desirable to have lesser number of inputs for a classifier, which can be achieved using Principal Component Analysis (PCA). The central idea of PCA is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the dataset. The coefficient matrix is transformed by PCA to a new set of variables, the Principal Components, which are uncorrelated, and that are ordered so that the first few retain most of the variation present in all of the original data [14]. After PCA, a feature vector matrix is constructed.

This feature vector matrix to be given as input to the classifier was constructed by taking the highest Eigen value which corresponds to the first principal component [15]. The values are then normalized and input to the classifier. The K-means algorithm is used to classify ECG data into normal and PVC cycles. In this algorithm the given feature vector data set is partitioned into two clusters such that all data in a given feature vector subset are closest to the same cluster centroid. First the algorithm randomly selects two values of the feature vector to represent the initial clusters. Then it uses an iterative algorithm that minimizes the distance from each value to its cluster centroid, over all clusters. The algorithm then computes the new centroid by taking the mean of all the values of the feature vector belonging to the same class. The operation continues until there are no changes in the values of the centroids and hence the class assignment for a particular cluster. Since the Eigen values for normal and PVC beats are well spaced the k-means clustering performs well.

III. RESULTS AND DISCUSSION

Accuracy is a statistical measure of how well a classifier correctly identifies or excludes a condition. To evaluate the efficiency of the setup, the accuracy of the classification was calculated using the equation [16] where

$$Total\ accuracy\ (TA) = \frac{TP+TN}{(TP+FP+FN+TN)}\ x\ 100\ \%, \qquad (4)$$

TP is the number of true positives, TN is the number of true negatives, FP is the number of false positive and FN is the number of false negatives. The actual number of normal and PVC cycles were obtained from the annotation files supplied with each record. The comparison classification accuracies of the different ECG records using NW1 and some of the commonly used wavelets are given in Table 1. From the table it can be seen that the wavelet NW1 evolved using GA outperformed other wavelets with an average classification accuracy of 97.77%. Fig. 4 shows one beat of PVC together with the wavelet scalogram obtained using NW1 and sym3.

TABLE 1. COMPARISON OF CLASSIFICATION ACCURACY

Wavelet	Total accuracy (%)					
	119 [a]	***217*** [a]	***116*** [a]	***106*** [a]	***114*** [a]	***Average***
NW1	100	96.02	99.58	94.02	99.25	97.77
meyer	100	95.27	99.63	90.90	99.13	96.98
morlet	100	94.77	99.63	91.89	99.13	97.08
haar	100	92.79	99.54	88.63	98.66	95.92
sym3	100	96.02	99.83	93.33	98.60	97.56
bior3.9	100	95.77	99.63	90.65	98.38	96.88
db3	100	93.53	99.54	90.01	98.82	96.38
db5	100	96.26	99.58	91.10	98.70	97.13

[a]Record

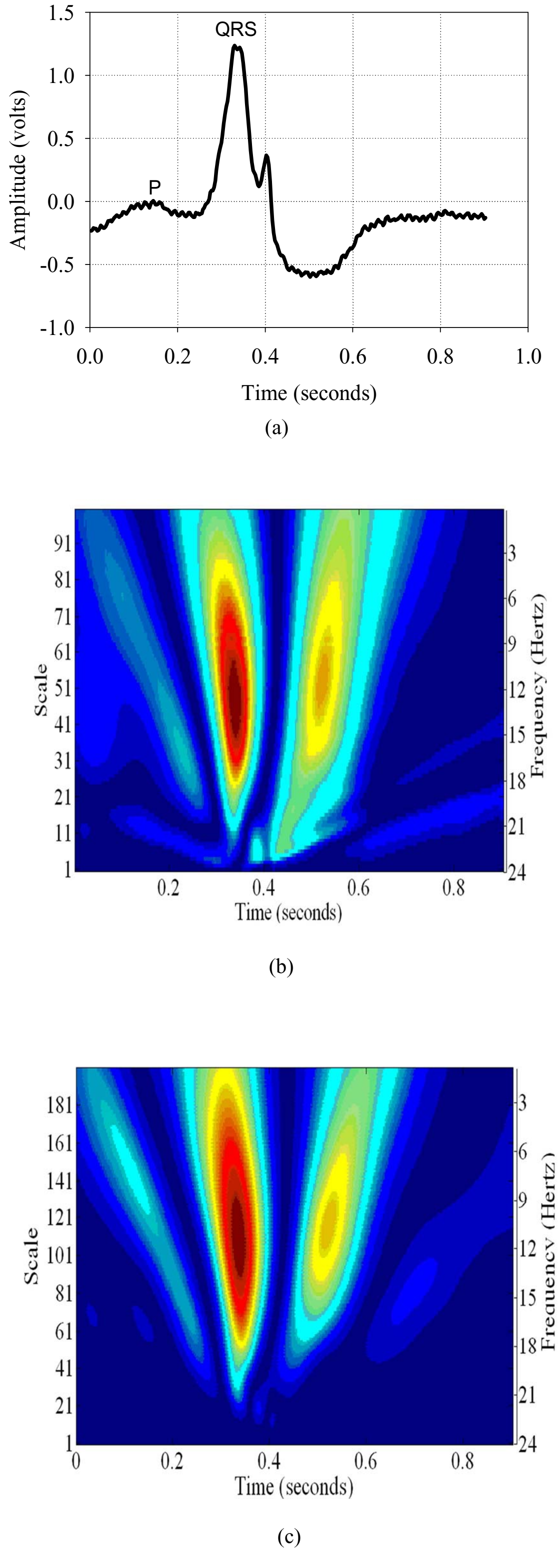


Figure 4. Plot and Wavelet scalogram of ECG. (a) Plot of one cycle of PVC. (b) Wavelet scalogram of (a) using NW1. (c) Wavelet scalogram of (a) using sym3.

Wavelet scalograms communicate the time frequency localization property along with the magnitude of the CWT. The points corresponding to P wave and QRS complex are clearly visible in the scalogram corresponding to NW1, especially a high frequency spike present in the falling edge of the R wave. These figures highlight the ability of the new wavelet to separate out signal components in the ECG.

IV. CONCLUSION

A new mother wavelet for classification of ECG has been proposed. The wavelet was able to successfully differentiate normal and PVC beats. The classification was performed using records from MIT-BIH arrhythmia database [5]. It is seen that the new evolved wavelet gave a classification accuracy equally and even more competent with that of the traditional wavelets. The average accuracy obtained was 97.77% with the K-means classifier. One of the drawbacks of K-means classifier is that the classification accuracy decreases when the sample in any of the class is less. Since the main aim in this work is to have a comparative study with the other wavelets, this classifier is adequate. The advantage of this method is that it uses only one feature for classification.

Inclusion of more heart abnormalities and enhancement of the evolved wavelet for better classification are being currently investigated.

REFERENCES

[1] Stanfield C, Germann W, Principles of Human Physiology, 3rd edition, Pearson International Edition, 2008.

[2] Al-Nashash, H, “Cardiac arrhythmia classification using neural networks,” Technol Health Care 2000, 8: 363-372.

[3] Nur Asyiqin, Amir Hamzah and RosH Besart, “Comparison of Different Mother Wavelets in PVC detection using PNN,” 10th International Conference on Information Science, Signal Processing and their Applications , ISSPA 2010.

[4] V. Mahesh, A. Kandaswamy, C. Vimal, B. Sathish, “ECG arrhythmia classification based on logistic model tree,” Journal of Biomedical Science and Engineering, Vol.2, No.6, pp.405-411, October 2009..

[5] www.physionet.org.

[6] Awadhesh Pachauri and Manabendra Bhuyan, “Wavelet and Energy Based Approach for PVC detection,” International Conference on Emerging Trends in Electronic and Photonic Devices & Systems, Institute of Technology, Banaras Hindu University, Varanasi, 2009.

[7] P. Ghorbanian, A. Jalali, A. Ghaffariand C. Nataraj, “An improved procedure for detection of heart arrhythmias with novel pre-processing techniques,” Expert Systems, The Journal of Knowledge Engineering, June 2011.

[8] S Almagro, MM Elena, MJ Bastiaans, JM Quero, “A New Mother Wavelet for Fetal Electrocardiography,” Computers in Cardiology, 33:157−160, ISSN 0276−6547, 2006.

[9] Steven W. Smith, Digital Signal Processing: A Practical Guide for Engineers and Scientists, Newnes, Elsevier Science, Burlington, MA 01803, 2003.

[10] R.M. Rangayyan, Biomedical Signal Analysis, IEEE Press and Wiley, New York, 2002, p.187.

[11] Soman K.P, Insight into Wavelets- From theory to practice, 2nd edition, Prentice Hall of India, 2006, p.43.

[12] Jaideva C. Goswami, Fundamentals of wavelets, Theory, algorithms and applications, John Wiley and sons, Inc., Hoboken, New Jersey, 2011, p.67.

[13] David E. Goldberg, Genetic Algorithms in search, optimization & Machine learning, Pearson Education, India, 2009.

[14] I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer-Verlag, New York Berlin Heidelberg, 2002, p.21.

[15] Granino Arthur Korn, Theresa M. Korn, Mathematical Handbook for Scientists and Engineers, Courier Dover Publications, New York, USA, 2000.

[16] John Robert Taylor, An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, University Science Books, Sausalito, CA 94965, 1999.