

ECG Signal De-Noising and Feature Extraction using Discrete Wavelet Transform

Raaed Faleh Hassan^{#1}, Sally Abdulmunem Shaker^{#2}

[#]Department of Medical Instrument Engineering Techniques, Electrical Engineering Technical College, Middle Technical University
Iraq.

Abstract

Electrocardiogram (ECG) provides an important information about cardiovascular performance, it is a noninvasive technique utilized as a main diagnostic appliance for cardiovascular diseases. The clear ECG signal supply valuable information about the electrophysiology of the heart diseases and ischemic changes that may happen. This research aims to extract the common features of the ECG signals based on Discrete Wavelet Transform(DWT). As the ECG signal suffers from two types of interferences: power line interference and baseline wander interference, therefore DWT has used firstly for de-noising the ECG signal then extracting its features. Daubechies 4 (db4) for R wave detection is employed, because R wave has high amplitude so it is easily detected and used as a reference point to detect the other waves by creating a window for each wave and search in minima and maxima amplitude, then extract other feature R-R, P-R interval, QRS width, heart rate and ST Deviation. The analysis of ECG signal is executed in MATLAB environment. This work is tested with signals imported from physionet and use field MIT-BIH Arrhythmia database and PTB database. The collected results appear a sensitivity of 99.6% and a positive predictivity of 100% for QRS complex wave detection, the P peaks detection show sensitivity of 97.9% and a positive predictivity of 98.1% and T peaks detection show sensitivity of 97.8% and a positive predictivity of 98.6%.

Keyword:

ECG signal, MATLAB, DWT, QRS, Daubechies, db4.

I. INTRODUCTION

The electrocardiogram is the recording of electrical activity of the heart and its effective method to a diagnosis of heart disease. ECG signal is acquired using electrodes which are placed on the skin of the chest and connect to the machine. These electrodes determine the voltage on the body surface produced by the electrical activity of the heart.

The ECG signal includes P-QRS-T waves as appear in figure(1), each wave reflect a certain electrical activity of the heart, P wave represents atrial depolarization, QRS complex represents ventricular

depolarization and T wave represents ventricular repolarization [1]. The bandwidth of ECG signal between (0.5 -100) Hz.

Commonly, the ECG analysis can be performed through 4 steps: Acquiring step, in which the ECG signal is acquired as a result of leads voltage differences. According to its dynamic range which is between (0.5-100) Hz, the ECG signal suffers from the baseline wonder noise and the power line interference, this shows the need for a step to remove these noises namely de-noising step. The third step in the ECG analysis system is the feature extraction of the clean ECG signal, while the fourth step is the diagnosis of the ECG signal based on its features [2]. There are many methods used for ECG signal analysis, the earlier methods are based on time domain approach but this approach cannot be extracted all feature of ECG signal. Recently, the frequency domain approach has been employed for analyzing ECGsignal [3].

Firstly, Fast Fourier Transform (FFT) is used but its failed to provides knowledge related the accurate time location of the frequency component [4].

To overcome this drawback, STFT (Short Term Fourier Transform) is employed, but the prime obstacle of this STFT is that its time-frequency accuracy is poor [5]. According to its reliability in the signal processing field, wavelet transform became available for analyzing ECG signal.

The wavelet transform is configured based on a set of analyzing and scaling wavelets which decompose the ECG signal into a sequence of coefficients. These coefficients produce by (WT) reflect the ECG components in a specified time duration and frequency band [5]. The majority feature extracted of ECG is originated from duration and amplitude of each wave. Normal ECG features for the healthy person is summed up in table 1 [6].

A significant deviation from these normal values in table 1 confirm the existence of abnormality [5], [7]. The remaining sections of this paper are as follows, Section II gives an overview of Wavelet Transform; methodology of this work contains preprocessing and feature extraction of ECG signal in Section III; Section IV contains simulation results and finally the conclusion in section V

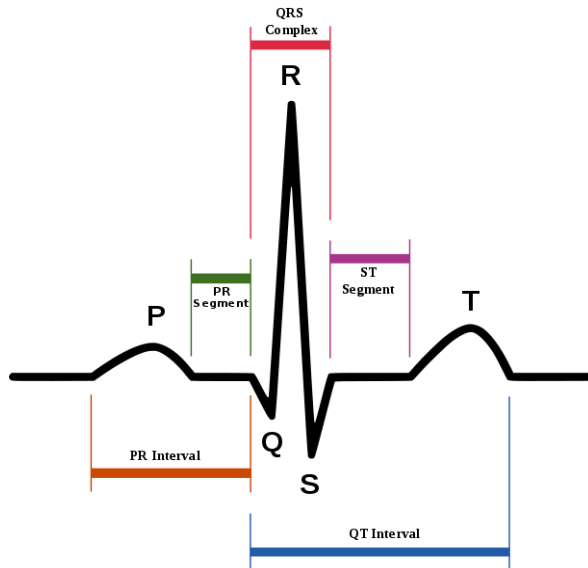


Figure 1: Typical ECG Signal

Wave	Amplitude
P wave	0.25 mv
R wave	1.6 mv
Q wave	25% R wave
T wave	0.1 to 0.5 mv
Interval	Duration
P-R interval	0.12 to 0.2 sec
QRS interval	0.09 sec
P wave interval	0.11 sec
S-T interval	0.05 to 0.15 sec
Q-T interval	0.35 to 0.44 sec
R-R interval	0.6 to 1.0 sec

Table 1 Normal ECG signal features

II. WAVELET TRANSFORMER

In general, wavelets are consciously constructed to have particular properties that make them eligible to be used for signal processing. Using a well-known technique called convolution, wavelets can be integrating with a known segment of a scrambled signal to extract features of the unknown segment. The wavelet can be moving to different positions (translation operation) on the signal and it can be prolonged or compressed (scaling operation) [2]. Essentially, the wavelet transformation is acquired by convolving the input signal $x(n)$ with the scaling function $g(n)$ to obtain approximation coefficients where $g(n)$ is the impulse response of Low pass filter and $x(n)$ convolved with wavelet function $h(n)$, the impulse response of a high pass filter to produce detailed coefficients [5].

The wavelet transforms consist of two processes decomposition and reconstruction, in the decomposition process the signal pass through number of a filter (low pass filter and high pass filter) and produce a number of the coefficient (approximation and details) each coefficient represents the definite range of frequency. The frequency content is then down-sampling in each level to prevents the redundancy.

The reconstruction process occurs as a reverse of the decomposition process, in each level is followed by up-sampling until synthesis the original signal figure 2. Discrete wavelets are not continuously scalable and translatable but can only be scaled and translated into discrete steps [8]. The basic function of the orthonormal wavelet is given by

$$\Psi_{m,n}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (1)$$

where the integers m and n control the wavelet dilation and translation respectively. Use binary discretization set $a = 2^{-m}$ and $b = 2^{-m}n$. The dyadic grid wavelet can be written as

$$\Psi_{m,n}(t) = 2^{\frac{m}{2}} \Psi(2^m t - n) \quad (2)$$

Now the discrete wavelet transform (DWT) can be written as follows:

$$T_{m,n} = \int_{-\infty}^{+\infty} x(t) \Psi_{m,n}(t) dt \quad (3)$$

Where $T_{m,n}$ is defined as the wavelet coefficient at scale(m) and location(n). The selection of the mother wavelet and number of levels for decomposition process are crucial for the analysis of ECG signals. The mother wavelet which is applied for decomposition was the Daubechies wavelet, it has been more complex but it gets the detail that is lost by Haar wavelet algorithm. Daubechies (db4) has properties similar to the properties of ECG signal, it identical in shape to QRS complex and their energy spectrum are focused on the low frequency [2], [5], [8]- [13].

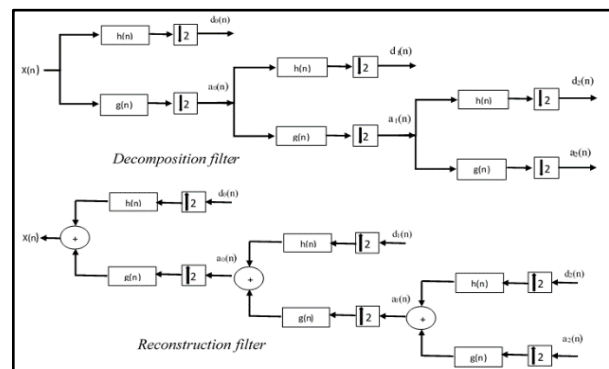


Figure (2) DWT based decomposition and reconstruction process.

III. RESEARCH METHOD

The aim of this paper is to extract the features of the ECG signal. The acquiring step is replaced by importing noisy ECG signal from the physiobank database. The remaining two steps; ECG de-noising or preprocessing and ECG feature extraction are shown in fig. 3.

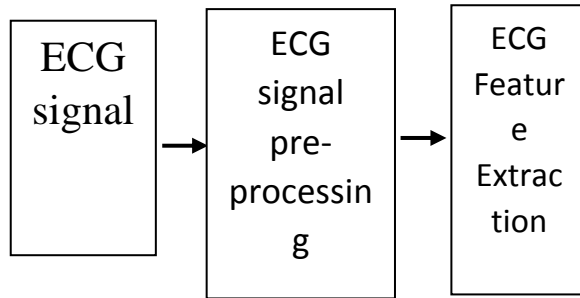


Figure (3) the general block diagram.

A. ECG Signal Database

ECG Signal collected from PhysioBank. The MIT-BIH arrhythmia database and PTB database are used for the verification of the suggested work. The recordings were digitized at (360 and 1000) samples per second.

Each ECG waveform contains three files: header file, a binary file, and a binary annotation file. The header file contains an information about a number of sampling, sample frequency, number of ECG leads, age, gender, and diagnosis, a binary file contains ECG signal that download in mat form and annotation file contain beat annotations, describe each beat signal [14].

B. ECG Signal Pre-Processing

The first step in ECG signal analysis is to make ECG signal free from the noise or de-noising ECG signal. The de-noising process contains removing of high frequency, baseline wander and eliminate the effect of power line interface using multi-resolution wavelet transform [15]. Baseline wander noise represents low-frequency range usually below 0.7 Hz. When a decomposing the signal with db4 into corresponding Approximation and Detail coefficients up to level 10, approximation coefficient A10 contain this range of frequency, to reconstruct ECG signal free form baseline noise subtract A10 from original ECG signal. This is in the case of sampling frequency 1000 Hz. In the case of sampling frequency of 360 Hz subtracted A9 from the original signal. Then using the average filter to removeglitches

to increase the performance of peak detection by using instruction in Matlab "smooth". The various bands generated by the decomposition process are as illustrated in the tables (2) and table (3). To eliminate the effect of power line interface, notch filter with cutoff frequency =50 Hz is employed [15], [16].

The high frequency represents by detail coefficient. De-noising ECG signal by wavelet consists of three steps: in first decompose ECG signal by discrete wavelet transform, instruction in Matlab (wavedec). Then calculate detail coefficient by instruction (detcoef) and then apply a threshold to all detail coefficient for each level. In the final step, ECG signal is reconstitute based on the original approximation coefficients of level N and the adjusted detail coefficients of levels from 1 to N [10], [8], [11]- [13], [15]- [17].

Table (2): components frequency range when sampling frequency 1000 Hz.

Level	Approximation a(n)	Detail d(n)
1	0 - 500 Hz	500 - 1000 Hz
2	0 - 250 Hz	250 - 500 Hz
3	0 - 125 Hz	125 - 250 Hz
4	0 - 62.5 Hz	62.5 - 125 Hz
5	0 - 31.25 Hz	31.25 - 62.5 Hz
6	0 - 15.625 Hz	15.625 - 31.25 Hz
7	0 - 7.8125 Hz	7.8125 - 15.625 Hz
8	0 - 3.906 Hz	3.906 - 7.8125 Hz
9	0 - 1.9 Hz	1.9 - 3.906 Hz
10	0 - 0.976 Hz	0.976 - 1.9 Hz

Table (3): components frequency range when sampling frequency 360 Hz.

Level	Approximation a(n)	Detail d(n)
1	0 - 180 Hz	180 - 360 Hz
2	0 - 90 Hz	90 - 180 Hz
3	0 - 45 Hz	45 - 90 Hz
4	0 - 22.5 Hz	22.5 - 45 Hz
5	0 - 11.25 Hz	11.25 - 22.5 Hz
6	0 - 5.625 Hz	5.625 - 11.25 Hz
7	0 - 2.8125 Hz	2.8125 - 5.625 Hz
8	0 - 1.40625 Hz	1.40625 - 2.8125 Hz
9	0 - 0.703 Hz	0.703 - 1.40625 Hz

A generalized algorithm for baseline wander removal and the de-noising process is presented in figure (4).

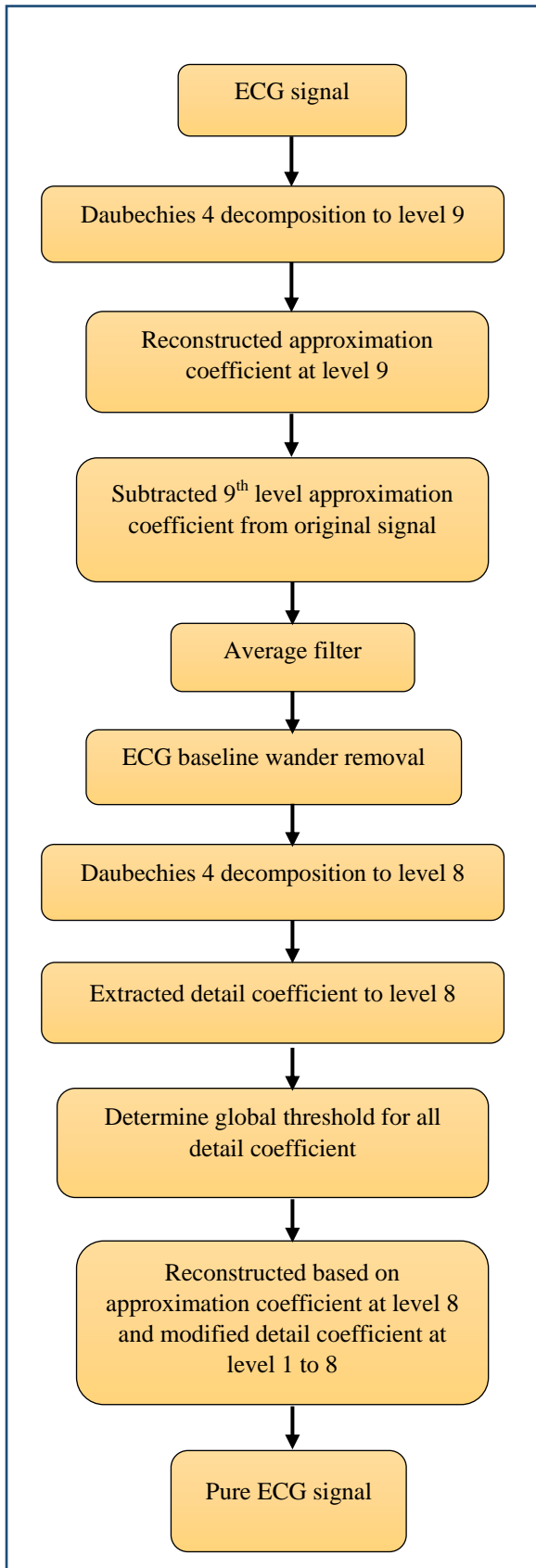


Figure (4): algorithm flow chart for baseline wander removal and de-noising from high frequency.

C. ECG Signal Feature Extraction

After the de-noising process, the next step is to extract the ECG signal features. In this paper, extract 12 features have been considered these are: (P, Q, R, S, T) peaks and locations, R-R interval, QRS interval, P-R interval, P wave duration, T wave duration, ST deviation and heart rate. A generalized algorithm for feature extraction process is presented in figure (5). First, pure ECG signal is decomposed using discrete wavelet transform, Daubechies (db4) into 4 level approximation coefficient. In each level ECG signal decomposed by down-sampling 2, the signal in the first level has half number of original signal samples, 2nd level has half number of the samples at the 1st level and so on. As the decomposition progress, the signal becomes smooth more and more. The second level approximated coefficient is selected, it contains a quarter number of samples in the original signal, use threshold is set to 60% of the max amplitude in the signal [5], [6], [15], [18]. Any peaks are higher than the threshold are R peaks, R peak in down-sampling signal is detected. R peak has higher amplitude is first detected and R peak is considered as a reference peak to detect other peaks. To detect R peak in original signal multiply 2nd level approximation coefficient by 4 and a window of 20 samples of width is applied to inspect the original signal after each R [17]. Each peak is higher than the selected threshold are R peaks. To determine R-R interval by discrete wavelet transform after detecting R peaks, in this work, the detail coefficient D3, D4, D5 and the approximated coefficient A5 are preserved and all other coefficients are discarded and to get R wave more Pronounced, the signal received from (D3, D4, D5, and A5) are squared and apply the threshold. The lower from the threshold are discarded to remove pseudo peaks [9], [10], [16], [19]. At last the interval between R-R interval are calculated. After calculate R-R interval can calculate heart rate by the equation: Square no. = $(R-R \text{ interval} / 0.2)$, 0.2 is a duration of big square in second, Heart Rate in minute = $300 / \text{Square no.}$ The P, Q, S and T waves are detected with a reference to R peaks location by creating a window and search with this window to detect other peaks. Detect P peaks by creating a window of R location-100 to R location-50, in this samples P peaks are located in maximum amplitude, Detect Q Peaks in a window of R location-100 to R location-10 but search for the minimum amplitude. Detect T peaks create a window of R location+25 to R location+100 and search for maximum amplitude and to detect S peaks use the window of R location+25 to R location+50 but search for minimum amplitude. Now all peaks and waves are detected [11], [12]. The next step has to found start point of every wave and the end point of every wave, this procedure was done by the onset and offset of every peak, in onset create a window from peak to the left of peak and Using if condition nested inside for loop for every type of wave (P, R, T), If condition

checks for a sample that proceed the sample which is bigger than zero and follows a sample which is less than zero. In offset, the same procedure was done but a window from peaks to the right of peak to find the end point of waves. Can find the duration of every wave if all onset and offset of every peak have, by subtracting offset from the onset [11]. Finally, calculate ST deviation by subtracted ST segment from PQ segment, and calculate if difference more than 1 mm this indicate an abnormality. Or determine J point (first point after S wave end) and determine reference point (first point before start Q wave) then

calculate the difference between j point and the reference point in each cycle If more 1 mm (0.1 mv) this mean abnormality [20].

IV. RESULTS AND ANALYSIS

Discrete wavelet transform (db4) analysis most ECG signal, the signal import from the website. The algorithm is examined with 100 ECG signal obtained from MIT-BIH Arrhythmia database and PTB database with duration of 10 seconds.

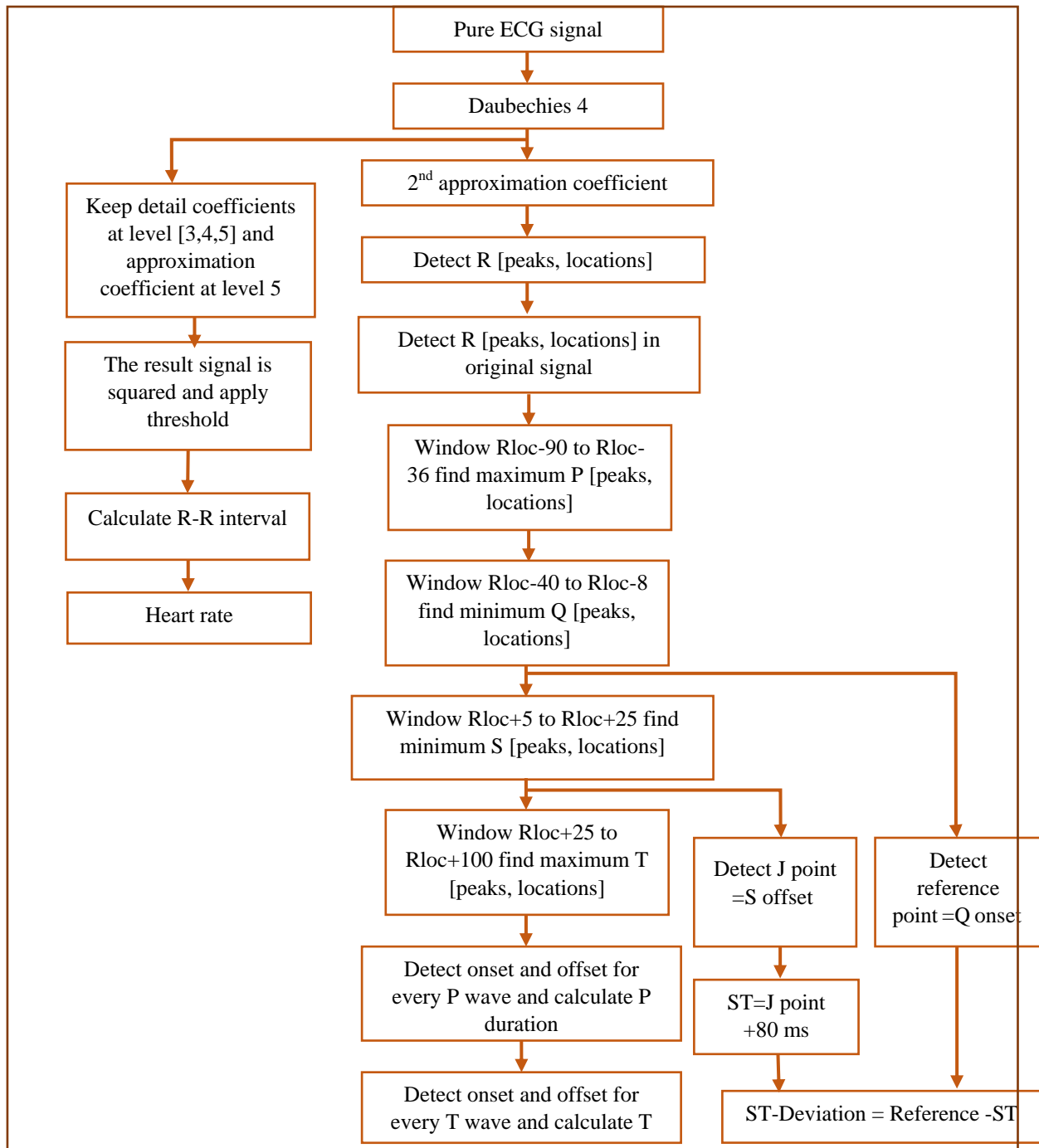


Figure (5): algorithm Flowchart for feature extraction.

Using Daubechies 4 to de-noising and extracted the most feature of ECG signal. Simulation results divided into 2 parts, de-noising part and feature extraction part.

In the following figures 6(a, b, c) show the noisy original signal and pure signal. By an accurate visual checking of the simulation results, determined that the algorithm exhibit a positive response particularly for the portion of baseline wander removal and high-frequency noise removal.

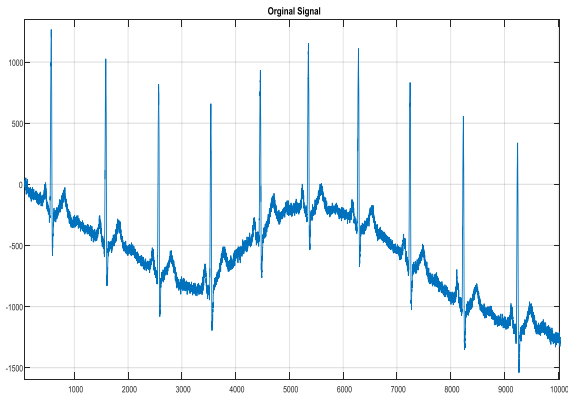


Figure (6a): Original ECG Signal

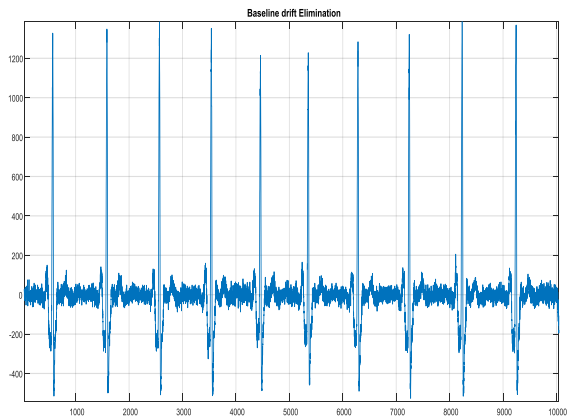


Fig. (6b): ECG signal after completely removing baseline noise.

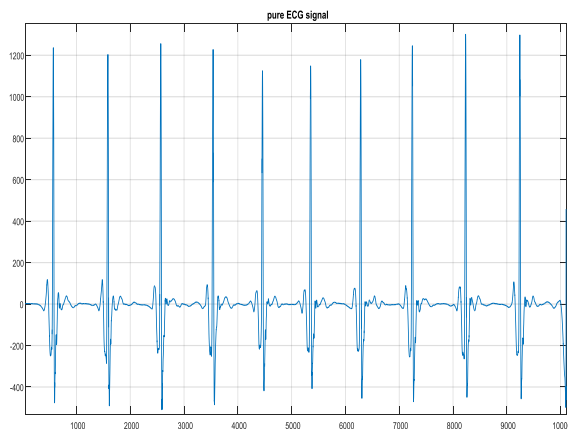


Fig. (6c): ECG signal after removing high frequency noise.

In next part show simulation results for detection peaks and feature extraction. In order to appraisal the performance of the suggested algorithm in the detection of peaks, two statistical measurements were employed. These are the Sensitivity (Se), is used to assess the capability of the algorithm to discover true beats and is defined by equation

$$Se(\%) = \frac{TP}{TP+FN} \%$$

and Positive Predictivity (PP) is used to assess the capability of the algorithm to distinguish between true and false beats and is defined by equation

$$PP(\%) = \frac{TP}{TP+FP} \%$$

where TP is the correctly detected beats (True Positive), FP is falsely detected beats (False Positive) and FN is the undetected beats (False Negative).

The proposed algorithm reaches a good degree of Se =99.6% and PP =100% for QRS complex wave detection, P peaks detection Se=97.9% and PP=98.1% and for T peaks detection Se=97.8% and PP=98.6%.

In the following figures 7(a, b, c, d, e, f) show the simulation result for detection peaks.

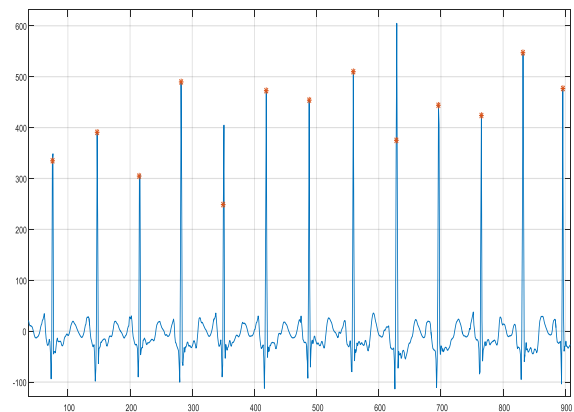


Figure (7a): detected R waves in level 2 approximation coefficient.

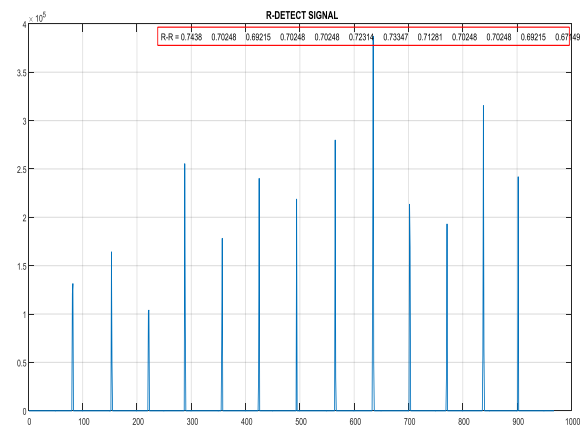


Figure (7b): detect R-R interval duration.

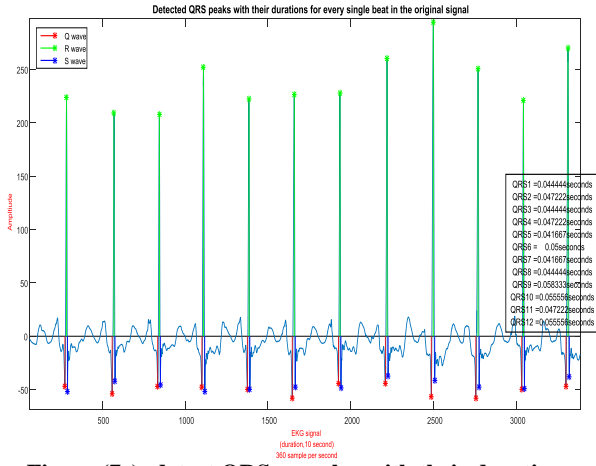


Figure (7c): detect QRS complex with their duration.

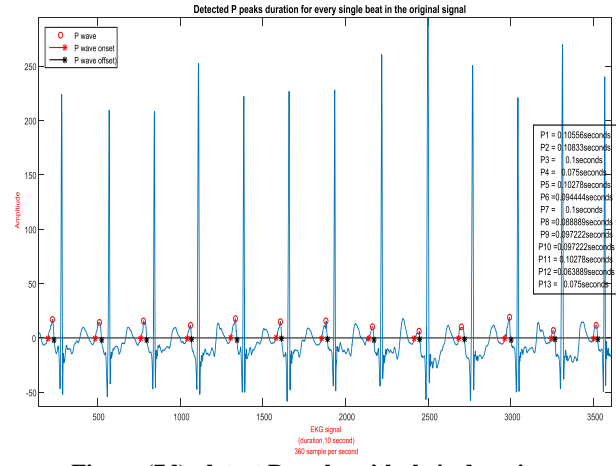


Figure (7d): detect P peaks with their duration.

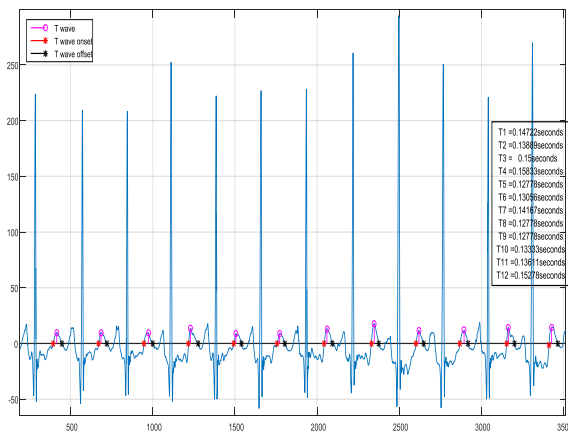


Figure (7e): detect T peaks with their duration.

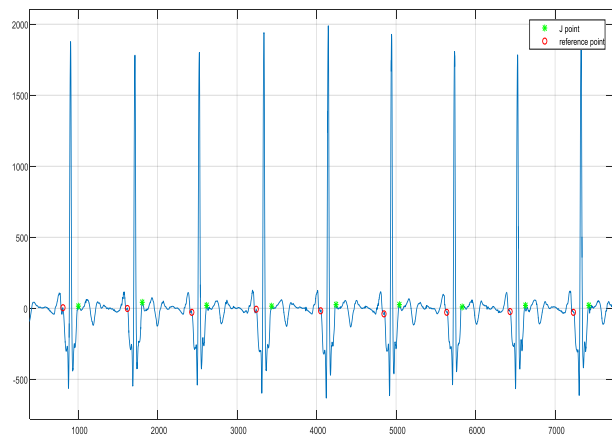


Figure (7f): j point and reference point

Table 4 shows different heights and intervals which are extracted by the suggested algorithm. All heights are measured in millivolts (mV) and all periods are measured in seconds (s).

Table 4: Extracted Features from 10 MIT-BIH and PTB Database records

Record	P mv	R mv	Q mv	S mv	T mv	Heart rate	Mean R-R sec	R-R1 sec	R-R2 sec	PR sec	QRS sec	ST deviation
100m	0.066	1.1951	-0.245	-0.229	0.06	76	0.759	0.75	0.76	0.18	0.048	0.1913
111m	-0.001	0.844	-0.0699	-0.321	0.27	68	0.86	0.84	0.83	0.27	0.115	0.1738
108m	0.146	0.71	-0.0821	-0.474	0.136	58	---	0.61	1.41	0.20	0.096	0.101
201m	-0.031	0.98	-0.0753	-0.104	0.07	106	---	0.55	0.95	---	0.154	0.11
109m	0.12	1.608	-0.289	-0.642	0.167	79	0.709	0.71	0.72	0.27	0.22	0.16
106m	0.15	1.88	-0.143	-0.8	0.3	57	---	1.13	0.7	0.14	0.2	0.17
234m	-0.009	1.422	-0.095	-0.133	-0.004	145	0.44	0.42	0.44	0.27	0.11	0.365
203m	0.39	1.88	-0.777	-0.655	0.99	155	0.4	0.38	0.33	0.12	0.23	---
s0021arem	0.0711	0.74	-0.17	-0.144	-0.007	82	0.74	---	---	0.21	0.078	1.79
s0390lrem	0.027	2.74	-0.649	-0.78	0.14	70	0.88	---	---	0.2	0.082	2.02

V. CONCLUSION

In this paper, we analysis ECG signals obtained from PhysioBank. We have investigated the ability of the wavelet analysis to process the biomedical ECG signal. The DB4 wavelet is employed for the R peaks detection, R-R interval, the baseline wander cancellation and eliminate high-frequency noise. All other primary peaks (P, Q, S, T) are extracted with respect to the location of R peaksthrough creating windows proportional to their normal intervals. Then determine QRS interval, PR interval and ST deviation.

REFERENCES

- [1] A.A.M.A.A.B.Zahia Zidelmala, "QRS detection based on wavelet coefficients," Elsevier, 2011.
- [2] M. K. EL Mimouni El Hassan *, "An FPGA-Based Implementation of a Pre-Processing," IEEE, no. 978-1-4799-4647-1/14/\$31.00, 2014.
- [3] A.K. C. P. N. Neophytou, "ECG analysis in the Time-Frequency domain," in IEEE 12th International Conference, Larnaca, Cyprus, 2013.
- [4] S. K. K. Himanshu Gothwal, "Cardiac arrhythmias detection in an ECG beat signal using fast Fourier transform and artificial neural network," J. Biomedical Science and Engineering, 2010.
- [5] G. Kavya, "Abnormality Diagnosis in ECG signal using Daubechies Wavelet," in 5th International Conference on Biomedical Engineering and Informatics, 2012 IEEE.
- [6] M. A. D. Muhidin A. Mohamed, "An Approach for ECG Feature Extraction using Daubechies 4 (DB4) Wavelet," International Journal of Computer Applications, 2014.
- [7] R. P. JAAKKO MAMIVUO, "Principles and Applications," in Bio electromagnetism, 1995.
- [8] U. A. C. Sambhu D.1, "Automatic Classification of ECG Signals Features Extracted Using Wavelet Transform and Support Vector Machines," International Journal of Advanced Research in Electrical, vol. ISSN (Online): 2278 – 8875, 2013.
- [9] S.-Y. J. K. W. a. K. N. S. M. Abed Al RaoofBsoul, "Detection of P, QRS, and T Components of ECG Using Wavelet Transformation," IEEE, 2009.
- [10] J.Seena V, "A REVIEW ON FEATURE EXTRACTION AND DENOISING OF ECG SIGNAL USING WAVELET TRANSFORM," in 2nd International Conference on Devices, Circuits, and Systems (ICDCS), 2014.
- [11] GZSCCET, "Automatic Delineation of ECG Characteristics Points using the Window Search & Multi-resolution Wavelet Transform approach," in Proc. of Int. Conf. on Emerging Trends in Engineering and Technology, ACEEE, 2013.
- [12] K. P. I. A. B. K. K. *. Remya R Sa, "Classification of Myocardial Infarction Using Multi-Resolution Wavelet Analysis of ECG," in International Conference on Emerging Trends in Engineering, Science and Technology, Elsevier, 2015.
- [13] Y.-S. N. a. J.-Y. W. Kun-Chih (Jimmy) Chen*, "Electrocardiogram Diagnosis using Wavelet-based Artificial Neural Network," in IEEE 5th Global Conference on Consumer Electronics, 2016.
- [14] "https://www.physionet.org/cgi-bin/atm/ATM," [Online].
- [15] A.T. Hari Mohan Rai, "ECG Signal Classification using Wavelet Transform and Back Propagation Neural Network," in 5th International Conference on Computer and Device for Communication (CODEC), IEEE, 2012.
- [16] J. S. Vanisree K, "Automatic Detection of ECG R-R Interval using Discrete Wavelet Transformation," International Journal on Computer Science and Engineering (IJCSE), 4 April 2011.
- [17] A. B. Rachid HADDADI, "Discrete Wavelet Transform Based Algorithm for Recognition of QRS Complexes," World of Computer Science and Information Technology Journal (WCSIT), vol. 4, no. ISSN: 2221-0741, 2014.
- [18] M. I. Mahesh A Nair, "ECG Feature Extraction using Time-Frequency Analysis," Springer Science Business Media B.V., 2010.
- [19] A.A. M. D. A. M. E. J. H. B. S. Z. Mahmoodabadi, "ECG Feature Extraction Based on Multiresolution Wavelet Transform," in Engineering in Medicine and Biology 27th Annual Conference IEEE, Shanghai, China, 2005.
- [20] [20] S. G. A.-K. A. R. T. *, "Real-Time Detection of Myocardial Infarction by Evaluation of ST-Segment in Digital ECG," Journal of Medical Imaging and Health Informatics, Vol. 1, 1–6, 2011.