

Wavelet Feature Extraction for ECG Beat Classification

Sani Saminu, Nalan Özkurt and Ibrahim Abdullahi Karaye

Department of Electrical and Electronics Engineering

Yasar University

İzmir, Turkey

{sani.saminu, ibrahim.abdullahi }@stu.yasar.edu.tr, nalan.ozkurt@yasar.edu.tr

Abstract—Electrocardiography (ECG) signal is a bioelectrical signal which depicts the cardiac activity of the heart. It is a technique used primarily as a diagnostic tool for various cardiac diseases. ECG provides necessary information on the electrophysiology and changes that may occur in the heart. Due to the increase in mortality rate associated with cardiac diseases worldwide despite recent technological advancement, early detection of these diseases is of paramount importance. This paper has proposed a robust ECG feature extraction technique suitable for mobile devices by extracting only 200 samples between R-R intervals as equivalent R-T interval using Pan Tompkins algorithm at preprocessing stage. The discrete wavelet transform (DWT) of R-T interval samples are calculated and the statistical parameters of wavelet coefficients such as mean, median, standard deviation, maximum, minimum, energy and entropy are used as a time-frequency domain feature. The proposed hybrid technique has been tested by classifying three ECG beats as normal, right bundle branch block (Rbbb) and paced beat using the signals from Massachusetts Institute of Technology Beth Israel Hospital (MIT-BIH) arrhythmia database and processed using Matlab 2013 environment. Classification has been performed using neural network backpropagation algorithm because of its simplicity. While equivalent R-T interval features gives average accuracy of 98.22%, the proposed hybrid method gives a promising result with average accuracy of 99.84% with reduced classifier computational complexity.

Keywords— ECG, DWT, Mobile devices, ECG Feature extraction, Pan Tompkins

I. INTRODUCTION

Human body consists of different organs that are interconnected together for proper and efficient body function. Heart is one of the most critical organs in the human body because it supplies blood to different part of the body organs; therefore a need in the development of methods and systems for monitoring its functionality is very high. One of the most powerful diagnostic tools in medical application that is commonly used for the assessment of the functionality of the heart is Electrocardiography (ECG). The ECG is a real-time non-invasive and conventional method for interpretation of the electrical activity of the heart [1].

Generally, healthcare is one of the emerging areas of research in this century both in hospital and health care community, there are considerable commercial interests in the wireless and automatic classification of the ECG signals. Because cardiovascular diseases (CDV) remains as the dominant causes of death all over the world with an estimated of 17.3 million people died from CDV in 2008 which account to 30% of all global death and 23.6 million people will die from these diseases by the year 2030 based on the prediction and statistics from World Health Organization (WHO) [2]. Also, according to a recently published (2014) report by Heart failure Working Group of the Turkish Society of Cardiology (TDK), there are 15 million heart-failure patients in Europe and 6 million in the United States (US), in Turkey there are 1 million patients suffering from heart failure. With another 2 million who are at serious risk of this disease and those figures will increase about two fold within 10 years [3]. It is very important to detect and diagnose as early as possible and accurately these cardiac arrhythmias since they usually cause sudden cardiac death. It is tedious and time consuming to used visual inspection in ECG analysis even for an expert cardiologist. Therefore, the usage of computer software to automatically detect the ECG beats and diagnose the ECG classes as well as simple and low cost acquisition system is cost effective and significantly improves diagnostic accuracy and patient healing outcomes [4].

Currently, concepts such as mobile healthcare and tele-medicine that require the use of low power and portable battery operated system incorporated with wireless ECG sensors are continuously growing and are becoming integral part of our daily life for continuous assessment of cardiac functionality and ECG beat classification. However, extracting significant and useful features from ECG signal characteristics is a very crucial for successful implementation of these devices since they need to be fast, simple and computationally efficient.

There are several studies proposed for the analysis of the ECG beats. Gradient-based algorithm and time domain morphology was presented in [5]. Also, in [6] statistical method of comparison between relative magnitudes of ECG samples and their time domain slope has been described.

Another classifier based on ECG morphological features was reported in [7,8]. Wavelet transform finds application in ECG beats detection and feature extraction as reported in [9,10]. Also, Mahesh used wavelet and Pan-Tompkins algorithm to extract time-frequency features for ECG beat detection system [11]. In [12] they presented classification of normal and abnormal signal using R-R interval features of ECG waveform. In [13], the principal component of 4th-levels DWT with db4 mother wavelet is used to classify normal and arrhythmic beats with accuracy of 95.60%.

In this work, a system for ECG feature extraction and classification suitable for mobile devices is presented. The system uses a proposed hybrid technique using features from time and time-frequency domain. The classification of ECG beats has been implemented by using artificial neural networks. After a brief introduction of the ECG wave and arrhythmias considered, the wavelets will be summarized at Section 3. Then, the feature extraction and arrhythmia classification for the proposed method will be explained in Section 4. Finally, the results of the experiments will be presented and discussed.

II. ELECTROCARDIOGRAPHY AND ARRHYTHMIAS

ECG signal is a bioelectrical signal which depicts the cardiac activity of the heart and it is a technique used primarily as a diagnostic tool for various cardiac diseases because of its simplicity. By attaching electrodes at different outer surface of the human skin, electrical cardiac signals can be recorded by an external device. These currents cause the contractions and relaxations of heart by stimulating cardiac muscle [1] and travel as electrical signals through the electrodes to the ECG device, which records them as characteristic waves. Different waves and fiducial points of ECG reflect the activity of different parts of the heart which generate the respective flow of electrical currents. Figure 1 below shows a schematic representation of a normal ECG and its various waves.

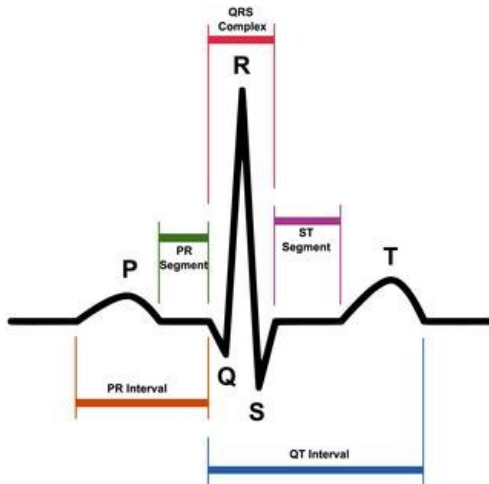


Fig. 1. Normal ECG wave [2]

The most important features include the information lying in the P, Q, R, S, and T waves of the ECG signal, ECG beats should be classified based on these features in order to detect different types of cardiovascular diseases. The length of a normal QRS wave is between 80 to 120ms [13].

Two different arrhythmias which are not critical in terms of emergent care but important to detect for future cardiac problems are considered in this study: right bundle branch block (Rbbb) and paced beats. When one bundle branch is blocked: Electrical impulse will travel through intact branch and stimulate ventricle supplied by that branch. Ventricle affected by blocked or defective bundle branch is activated indirectly. There is a delay caused by this alternate route and QRS complex will represent widening beyond usual time interval of 0.12 sec. Classified as either complete [QRS measures 0.12 sec or greater] or incomplete blocks [QRS measures between 0.10 and 0.11 second]. The paced beat is the artificial beat form from the device called pacemaker. A pacemaker is a treatment for dangerously slow heart beats. Slow heart beats can be the result of metabolic abnormalities or occur as a result of blocked arteries to the heart's conduction system. These conditions can often be treated and a normal heart beat will resume. Slow heart beats can also be a side effect of certain medications in which case discontinuation of the medicine or a reduction in dose may correct the problem. It can be characterized in ECG by a large peak after QRS complex [13].

III. WAVELET ANALYSIS

The continuous wavelet transform (CWT) has been developed as a method to obtain simultaneous, high resolution time and frequency information about a signal. The CWT unlike Short Time Fourier Transform (STFT) uses a variable sized window region. Since the wavelet may be dilated or compressed; different features of the signal are extracted. While a narrow wavelet extracts high frequency components, a stretched wavelet picks up the lower frequency components of the signal [14].

The CWT is computed by correlating the signal $s(t)$ with families of time-frequency atoms $\Psi(t)$, it produce a set of coefficients $C(a,b)$ given by :

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \Psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

where b is the time location(translation parameter), a is called scale factor and it is inversely proportional to the frequency ($a>0$), $*$ denotes a complex conjugate and $\Psi(t)$ is the analysing wavelet (mother wavelet).

The Discrete Wavelet Transform (DWT) is a time-scale representation of the digital signal and is obtained using digital filtering techniques. It is found to yield a fast computation of wavelet transform, easy to implement and adopts dyadic scales and translations in order to reduce the

amount of computation time, which results in better efficiency of calculation. DWT can be obtained by

$$C_{mn} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \Psi_{mn}(t) dt \quad (2)$$

where the dyadic scaled and translated wavelet is defined as

$$\Psi_{mn}(t) = 2^{-m/2} \Psi(2^{-m}t - n) \quad (3)$$

The most common wavelets providing the orthogonality properties are Daubechies, Symlets, Coiflets and Discrete Meyer in order to provide reconstruction using the fast algorithms [14].

The successive low-pass and high-pass filters calculating three levels of DWT is shown in Fig.2.

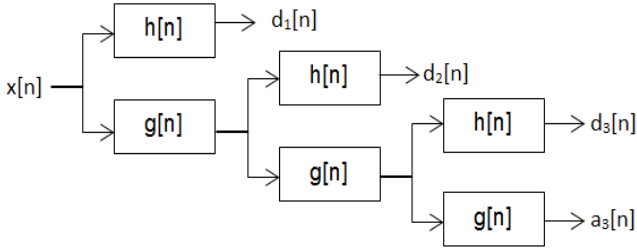


Figure 2. Three level Wavelet decomposition tree

Each stage consists of two digital filters and two downsamplers by 2 to produce the digitized signal. The low pass filter is denoted by $g[n]$ while the high pass filter is denoted by $h[n]$. At each level, the high pass filter produces detail information; $d[n]$, while the low pass filter associated with scaling function produces coarse approximations, $a[n]$. The filtering and decimation process is continued until the desired level is reached. The maximum number of levels depends on the length of the signal. Only the last level of approximation is save among all levels of details, which provides sufficient data. The filter coefficients can be obtained from mother wavelet and scaling functions [14].

IV. DATA ACQUISITION AND FEATURE EXTRACTION

The flowchart of the overall working principle of the automatic beat classification system is given in Figure 3. After the acquisition of the data, the preprocessing steps are applied to remove noise and artifacts, then R waves are detected to localize QRS complexes using Pan-Tompkins algorithm.

For each R-R interval 200 samples are obtained and the features are extracted by DWT and statistical techniques. Finally, each ECG beats are classified by artificial neural network due to its simplicity and the performance is analyzed. In this section the details of the procedure will be explained.

A. ECG Data Acquisition

In this study, the source of the ECG data used for training and testing is MIT-BIH Arrhythmia database from Physionet website [15]. The database contains 48 recordings of both routinely clinical waveforms and some complex arrhythmias sampled at 360Hz of 30 min durations selected from 24 hr recording with two channels obtained from 47 patients [16]. Only one channel of 1 min long for each record is used in this work.

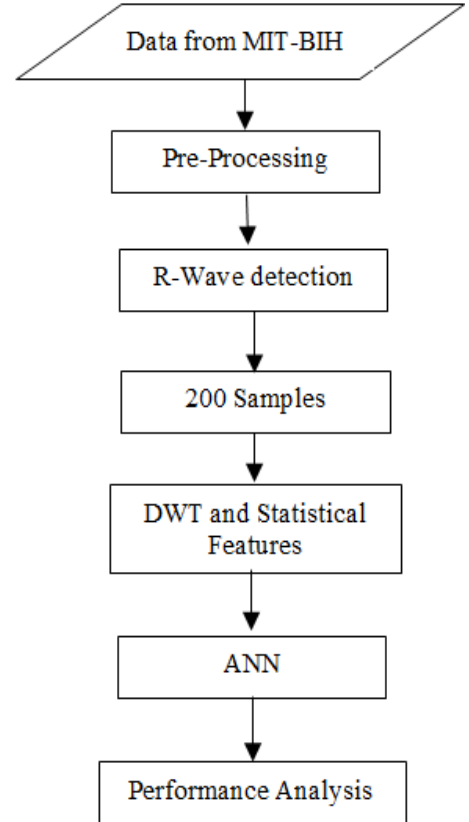


Fig. 3. Automatic ECG Beat Classification System Development Flow Chart

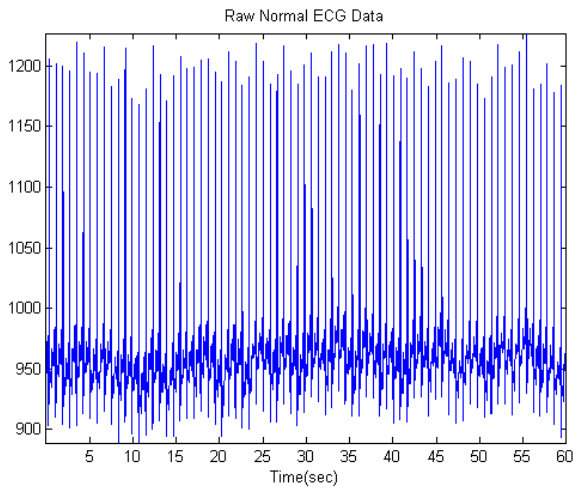
B. Preprocessing

Preprocessing step involves removal of noise from sources such as electrode contact noise, baseline drift, muscle contraction, power line interference and motion artifacts. Also QRS detection [17] was carried out in this stage. A well known and acceptable Pan Tompkins algorithm is employed as a real time QRS detection algorithm based on the analysis of slope, amplitude and width of QRS complexes [18]. The steps of preprocessing are given as

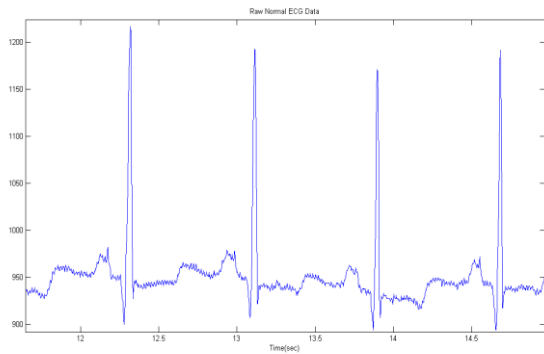
- Removing DC component

- Removing high frequency noise (low pass filter)
- Removing low frequency noise (high pass filter)
- Removing power line interference (comb filter)
- Derivative operation
- Squaring operation
- Integrator
- Thresholding
- Search procedure for R-peaks

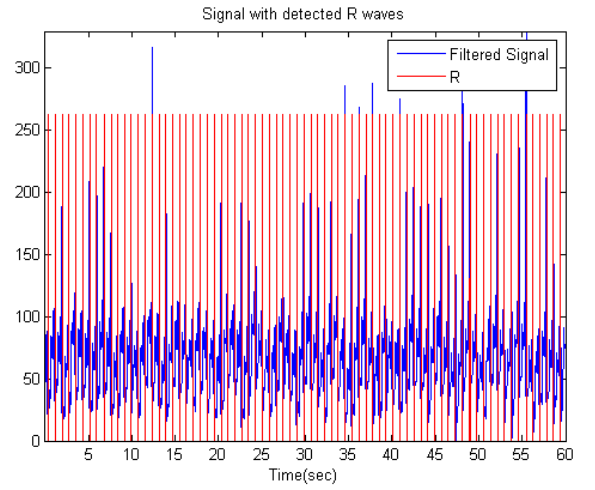
One sample of the original normal ECG recording and preprocessed ECG wave with detected R waves are shown in Fig. 4. The red lines in Fig.4.c point the detected R waves. In the feature calculation R-R intervals are taken into account.



(a)



(b)



(c)

Fig. 4.(a) Original Normal ECG for 1 minute signal, (b) zoomed in for a few beats, (c) preprocessed ECG signal

C. Feature extraction

ECG signal consists of many parameters and data points which characterize its behavior, extracting significant and smaller number of parameters without sacrificing accuracy of classifier is particularly important in ECG beat detection and classification using mobile devices. To achieve this, feature extraction in this work are in three stages.

- 1) **Equivalent R-T interval features:** Only 200 samples from detected R-peaks have been extracted from R-R interval which corresponds to R-T interval. The feature vector is constructed depending on the number of R-peaks in each ECG record which is between 60 and 90 per record.
- 2) **DWT decomposition:** In this stage, discrete wavelet decomposition is applied to feature vector extracted from R-T interval above. Different wavelet families are considered to find the best and suitable wavelet. Another important point is to select the wavelet decomposition level. The level is chosen to cover the frequency range of the normal and abnormal ECG signals.
- 3) **Statistical features:** In this stage, seven standard statistical parameters are used over the set of wavelet coefficients in order to reduce the feature vector dimension and to increase robustness of ECG features. The mean, median, maximum, minimum, standard deviation, energy and entropy are the features that represent the time-frequency distribution of the ECG signals.

III. RESULTS AND DISCUSSION

In this paper, ECG signal from MIT-BIH database were used for training and testing. 40 ECG records were

downloaded as .mat files and distributed randomly for classification of normal, paced and rbbb beats. 1602 ECG beats are fed to artificial neural network and approximately 70% is used for training and the remaining is used for validation and testing. The classification performance is considered in terms of sensitivity, specificity, positive predictive value, negative predictive value, accuracy and recognition rate. The used measures are defined as

$$Sensitivity = \frac{TP}{TP + FN} * 100\% \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} * 100\% \quad (5)$$

$$Positive Predictive Value = \frac{TP}{TP + FP} * 100\% \quad (6)$$

$$Negative Predictive Value = \frac{TN}{TN + FN} * 100\% \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} * 100\% \quad (8)$$

$$Recognition Rate = \frac{TP}{TP + FP + FN + TN} * 100\%$$

Where TP is True Positive (correctly identified), FP is False Positive (incorrectly identified), FN is False Negative (incorrectly rejected), TN is True Negative (correctly rejected) [19].

A. Performance Analysis of Equivalent R-T Interval Features

Several numbers of hidden layers are experimented in terms of recognition rate and the best result has been obtained with 15 hidden neurons. This network was used in our subsequent experiments.

Table 1 below shows the performance of equivalent R-T interval features extracted after QRS detection using Pan Tompkins algorithm.

TABLE I
PERFORMANCE MEASURES FOR R-T INTERVAL SAMPLES

ECG Class Beat	Sens. (%)	Spec. (%)	Pos. Pred. (%)	Neg. Pred. (%)	Accuracy (%)
Normal	100	98.91	97.85	100	99.27
Rbbb	97.01	97.49	94.89	98.55	97.34
Paced	95.07	99.63	99.26	97.47	98.06
Average	97.36	98.68	97.33	98.67	98.22

B. Performance Analysis of Hybrid Features

After decomposing R-T equivalent features using DWT, statistical parameters of the DWT coefficients have been calculated to further reduce the feature dimension to suit our

target. While the raw feature vector contains 200 samples for each beat, after processing the size is reduce to only 77 representing features of hybrid technique. There is an improvement in the proposed hybrid system of almost 100% accuracy as shown in Table 2 below. Also, the performance measures sensitivity, specificity, positive predictive value and negative predictive value show the success of the classification.

TABLE III
THE PERFORMANCE OF HYBRID FEATURES

ECG Class Beat	Sens. (%)	Spec. (%)	Pos. Pred. (%)	Neg. Pred. (%)	Accuracy (%)
Normal	100	100	100	100	100
Rbbb	99.25	100	100	99.64	99.76
Paced	100	99.63	99.30	100	99.76
Average	99.75	99.88	99.77	99.88	99.84

C. Comparison on Wavelet Families for Hybrid Features

As another experiment, different wavelet families are compared in terms of recognition accuracy. Table 3 gives the results of this comparison. According to the table, Daubhechies 4 and 10, Bior 3.7 and Coiflets 5 perform better than the other wavelets.

TABLE III
THE PERFORMANCE COMPARISON OF WAVELET FAMILIES

Wavelet Name	Correct recognized patterns	Recognition rates (%)
Db2	347	84.02
Db4	411	99.52
Db7	409	99.03
Db10	411	99.52
Bior1.5	351	84.99
Bior2.6	392	94.92
Bior3.7	412	99.76
Bior6.8	408	98.79
Coif2	409	99.03
Coif5	411	99.52
Sym5	410	99.27
Sym8	410	99.27

V. CONCLUSION

This paper addresses one of the challenging issues in health care services especially for developing countries by proposing a novel feature extraction technique suitable for ECG portable devices. The method used ECG signals data from MIT-BIH database and different filters were designed to reduce unwanted signal like baseline wander and power line interference. QRS complex was detected using well known and acceptable Pan Tompkins algorithm, R-R intervals features without applying any operation or transform are used as benchmark. DWT was used to decompose R-R intervals

and provide a time-frequency representation of the signal. The statistical parameters of DWT coefficients were calculated and used as hybrid feature for training and testing using neural network classifier. Based on the result obtained, RT equivalent feature and DWT with statistical feature gives 98.22% and 99.84% respectively. Thus, the proposed feature gives higher accuracy in classifying three types of ECG beats as normal, rbbb and paced beat. When different wavelet families are compared in terms of classification performance, Daubhechies 4 and 10, Bior 3.7 and Coiflets 5 performs better than the other wavelets. The more the mother wavelet resembles the ECG waveforms, the beats are represented better.

The design and implementation of a portable ECG recording and arrhythmia detection system which uses the proposed feature extraction algorithms is the concern of the future studies.

REFERENCES

- [1] A. C. Guyton, J. E. Hall, Textbook of Medical Physiology (11th ed.). Philadelphia: Elsevier Saunder. 2006.
- [2] R. Murugavel, Heart-Rate and EKG Monitor Using the MSP430FG439, Application Report SLAA280A, pp.1-12. 2005.
- [3] C. Yuksel, "Heart Failure Working Group of the Turkish Society of Cardiology (TDK)" <http://www.todayszaman.com/news>, Date accessed: May 10, 2014.
- [4] R.A. Bruce, S.R. Yarnall, Computer-aided diagnosis of cardiovascular disorders, J.ChronicDis.19, pp473-484. 1966.
- [5] E.B. Mazomenos, T. Chen, A. Acharyya, A. Bhattacharya, J. Rosengarten, K. Maharatna, A timedomain morphology and gradient based algorithm for ECG feature extraction. In: Proceedings of IEEE international conference on industrial technology (ICIT), pp. 117-122, 2012.
- [6] H.K. Chatterjee, R. Gupta and M. Mitra, A statistical approach for determination of time plane features from digitized ECG. Comput Biol Med. 41(5):278-84, 2011.
- [7] P. Chazal, M. O'Dwyer, R.B. Reilly, Automatic classification of heart beats using ECG morphology and heart beat interval features. IEEE Tran Biomed Eng. 51(7):1196-206, 2004.
- [8] P. Chazal, R.B. Reilly, A patient- adapting heart beat classifier using ECG morphology and heartbeat interval features. IEEE Tran Biomed Eng. 53(12):2535-43, 2006.
- [9] S.C. Saxena, V. Kumar, S.T. Hamde, Feature extraction from ECG signals using wavelet transforms for disease diagnosis. Int J Syst Sci. 33(13):1073-85, 2003.
- [10] J.P. Martinez, R. Almeida, S. Olmos, Rocha, A.P., Laguna, P., A wavelet-based ECG delineation: evaluation of standard databases. IEEE Tran Biomed Eng. 51(4):570-81, 2004.
- [11] A.N. Mahesh, ECG feature extraction using Time-Frequency analysis, Innovation in Computing Sciences and Software Engineering, 2014.
- [12] C. Marlar, S.K. Aung, Implementation of ECG beat classification, International journal of Societal Applications of Computer Science, 2014.
- [13] R.J. Martis, U.R.K. Acharya, M. Mandana, A.K. Ray, C. Chakraborty, Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, Expert Syst. Appl.39, 2012, pp11792-11800.
- [14] L. Williams and Wilkins, ECG Interpretation made incredibly easy. 5th edition, Kluwer Wolters, 2011.
- [15] P.S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, CRC Press, 2002.
- [16] Physionet, MIT-BIH, ECG database, <http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm>, 2014.
- [17] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, Peng, C.K., PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000.
- [18] Y. Ozbey, B. Karlik, A New Approach for Arrhythmias Classification, Proc. Of Annual International Conference of IEEE of Medicine and Biology Society. IEEE Transactions on Biomedical Engineering, 32-230-236, 1985.
- [19] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering, v.32,pp.230-236, 1985.
- [20] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, July 2011.