ECG SIGNAL CLASSIFICATION USING DISCRETE WAVELET TRANSFORM AND PAN TOMPKINS ALGORITHM

Yan Naung Soe[1], Khet Khet Khaing Oo[2]

Faculty of Computer Systems and Technologies, University of Computer Studies, Myitkyina and 1011, Myanmar

Abstract

Electrocardiogram (ECG) is one of the most widely used techniques for heart beat to diseases. The system presents the methods to analyze electrocardiogram (ECG) signal, detect the QRS complex, P and T wave and notch detection and extract features according to the different data.The input ECG signal is often contaminated by noise. In order to extract useful information from the noisy ECG signals, the raw ECG signal has to be diagnosing. The affect ECG signal analysis to baseline wandering is significant and can strongly. The inputs noisy ECG signal has been eliminated with the Discrete Wavelet Transform method for P and T wave detection. Then, the extracted features from the ECG signals achieve using Pan Tompkins Algorithm and QRS complex Location. The system classifies the heart beats types based on Rule Based Algorithm. Data are obtained from the records of the MIT-BIH database. The implementation of the approach is accomplished using Matlab programme software and the Evaluation results for the system quality is measured the performance accuracy.

Keyword: ECG classification; Normal; Left Bundle Branch Block; Right Bundle Branch Block; Premature Atrial Contraction; Premature Ventricular Contraction; paced beat algorithm.

1.INTRODUCTION

Some metabolic problems an important diagnostic tool in the diagnosis of as well as by using ECG. To learn an ECG correctly, one has to be thorough with the basic knowledge of electromechanical system of the human's heart. Frequently associated with signals of electrical, chemical or acoustic origin to the function of human body.

The electrical activity of the heart is records for ECG signal. ECG signals can provide deices diagnosis clinicians with valuable information about the patient health condition. The Electrocardiogram (ECG) is an important and commonly used diagnostic in cardiac disease. A typical ECG signal of a normal heartbeat (or cardiac cycle) consists of a P wave, a QRS complex and a T wave. P-wave is depolarization of atria. QRS-complex is depolarization of ventricles. T-wave is repolarization of ventricles. QRS complex detection is the most important task in ECG analysis. Its detection is the first step of all kinds of feature extraction. QRS detector must be able to define a lot of number different QRS morphologies. Most of the energy of the QRS complex exits between 3 Hz and 40 Hz. P and T waveform can be give with some main information about physiological conditions of patient distress from heart disease.

The system presents the methods to analyze electrocardiogram (ECG) signal, detect the QRS complex, and extract the morphological features and temporal features according to the different data. The extracted features from the ECG signals will achieve using Discrete Wavelet Transform and Pan Tompkins Algorithm. The system classifies the heart beats types on the extracted features using Rule-Based Algorithm.

The experimental procedure is as follows:

1) ECG signal acquisition
2) Normalization & denoising
3) Features extraction
4) Classification of heart beat types

In this research, signals to test the algorithm are obtained from the MIT/BIH database. The first stage is to remove unwanted noises from the raw ECG signal. Since the QRS complex is generally the most distinct of the various ECG waveform features, it is the feature which can most easily be identified with an algorithm.

2.THEORETICAL BACKGROUND

Bioelectrical signal generated to the human body during cardiac cycle from Electrocardiogram (ECG) is a graphical record. Diagnosing of heart problems to the duration slope and amplitude between characteristic points of ECG signal are important. ECG tracings have a very predictable direction, duration and amplitude by under normal conditions.

In ECG signal processing, it is very important to detect very accurately heartbeats, because it is the base for further analysis. The energy of heartbeats is located in the QRS complex. Thus ECG analysis an accurate QRS detector is essential. The signal varies along the time and different types of noise can be present in QRS detection. The accurate signal analysis and diagnosis to the process of automated analysis, the noises present in ECG signal are needed to be considered and eliminated.

2.1. Noise Artifacts in ECG

Electrocardiographic (ECG) signals may be destroy by various kinds of noise. Typical examples are:

1. ower line interference
2. Electrode contact noise
3. Motion artifacts
4. Muscle contraction (Electromyography, EMG)

2.2. Time Frequency Analysis

Time-frequency representations are used to characterize signals whose energy distribution varies in time and frequency. These representations figure the one-dimensional time-domain signal into a two-dimensional function of time and frequency. Time–frequency analysis studies a two-dimensional signal function whose domain is the two-dimensional real plane, obtained from the signal via a time–frequency transform. A time-frequency representation indicates the change of spectral energy and musical score describes the variation of musical pitch over time. The Fourier transition is similar to the Discrete Wavelet Transform (DWT) in that it is a decomposition of a signal in terms of a basis set of functions. In Discrete Wavelet Transform (DWT) the basis set consists of sines and cosines and the expansion has a single parameter. In wavelet transform has two generated from a single "mother" wavelet using dilation and offsets corresponding to the two parameters. [1]

2.2.1. Discrete Wavelet Transform (DWT)

Exhibits zero redundancy to dyadic grid and orthonormal wavelet basis functions .DWT but is determined only on a discretized grid of a scales and b locations from the transform integral remains continuous. In use, the input signal is treated as an initial wavelet approximation to the constant signal used to a multiresolution algorithm, the wavelet and inverse transform can be computed and without loss of signal information. The parameters a and b is to use a logarithmic discretization of the scale and link this, to the size of to link b to an it move each location b in which are proportional to the scale. The wavelet has the form

$$\varphi_{m,n}(t)=\frac{1}{\sqrt{a_0 m}}\varphi\left(\frac{t-nb_0{a_0}^2}{{a_0}^m}\right)$$

Where the whole number m and n control the wavelet dilation and translation respectively; a_0 is a defined stable dilation step parameter set at a value only great than 1, and b_0 is the location parameter which must be greater than zero. A common Selection for discrete wavelet parameters a_0 and b_0 are 2 and 1 respectively. The dyadic grid wavelet can be written compactly, as

$$\varphi_{m,n}(t) = 2^{-\frac{m}{2}}\varphi(2^{-m}t-n)$$

Discrete dyadic grid wavelets are typically selected to be orthonormal, i.e. they are both orthogonal to each further and are normalized to have unit energy. This is expressed as

$$\int_{-\infty}^{\infty}\varphi_{m,n}(t)\varphi_{m',n'}(t)dt = \begin{cases}1 & if\ m=m'and\ n=n' \\ 0 & otherwise\end{cases}$$

2.3. ECG Waves

Each heart cycle consists of the same depolarization/ repolarization phase from the atria to the ventricles. The heart electrical action of the screen projection likely is therefore a pseudo-periodic signal it means that the cardiac cycle constant according to heart rate. In each cycle, the locations of different waves on the ECG are arbitrarily marked by the letters P, Q, R, S, T and U. [11] it is necessary to noted that the baseline voltage of the electrocardiogram is known as the isoelectric line. Typically the isoelectric line is measured as the T wave and preceding the next P wave.

2.4. Pan Tompkins Algorithm

The algorithm automatically adjusts the thresholds and parameters periodically to adapt to changes in QRS morphology and heart rate. In summary, it consists of the following:

2.4.1. Band Pass Filter

Cut back the impact of muscle noise, 60 Hz interference, baseline wander, and T-wave interference for the band-pass filter. The desirable pass band of the band-pass filter to maximize the QRS energy is approximately 5-15 Hz. The purpose of low-pass filter is to reduce electromyography noise which can occur due to electrode probe.

2.4.2. Derivative

After band-pass filtering, it is needed to know the rate of change of ECG signal amplitude. So, the signal is differentiated to provide the QRS complex slope information.

2.4.3. Squaring

The derivative of the signal contains positive slope and negative. To emphasize amplitude change (positive or negative), the signal is squared point by point. All data points positive and amplifies by using squaring makes to the output of the derivative emphasizing the higher frequencies.

2.4.4. Moving window integration

The slope of the R wave can be use moving -window integration is to obtain waveform feature information. It is called moving average filter. Normal QRS complex width is about 150 ms, so, window width 54 is chosen because of the sample rate is 360 samples per second.

2.4.5. Thresholds adjustment

The thresholds on moving average filtered signal are applied to decide the location of R-peak approximately. These two slopes are compared and threshold to identify noise-peak or R-peak. The slope of noise-peak is less than R-peak.

4.EXPERIMENTAL RESULTS

The classification of beat types are compared with MIT/BIH annotation records which represents the actual time intervals from patient. While comparing with annotations, the performance of Rule Based algorithm can be defined with the accuracy. Accuracy = $\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \times 100\%$ where TP is the number of true classification of the beat types and TN is the number of true classification of absent beat type. FP is the number of false classification of absent of beat types and FN is the number of false classification of beat types.

No	Data Name	TP	TN	FP	FN	Accuracy
1	100	14	0	0	0	100.00%
2	101	11	0	0	1	91.67%
3	102	13	0	0	0	100.00%
4	103	12	0	0	0	100.00%
5	104	14	0	0	0	100.00%
6	105	15	0	0	0	100.00%
7	106	11	0	0	0	100.00%
8	107	11	0	2	0	84.62%
9	108	12	0	0	0	100.00%
10	109	17	0	6	0	73.91%
11	111	13	0	0	0	100.00%
12	112	16	0	0	0	100.00%
13	113	10	0	0	0	100.00%
14	114	10	0	0	0	100.00%
15	115	11	0	0	0	100.00%
16	116	14	0	0	0	100.00%
17	117	10	0	0	0	100.00%

18	118	13	0	0	0	100.00%
19	119	11	0	0	1	91.67%
20	121	11	0	0	0	100.00%
21	122	17	0	0	0	100.00%
22	123	8	0	1	0	88.89%
23	124	9	0	0	0	100.00%
24	200	16	0	4	0	80.00%
25	201	14	0	1	0	93.33%
26	202	10	0	0	0	100.00%
27	203	21	0	1	0	95.45%
28	205	6	0	0	5	54.55%
29	207	16	0	0	0	100.00%
30	208	16	0	0	0	100.00%
31	209	17	0	0	0	100.00%
32	210	16	0	0	2	88.89%
33	212	15	0	1	0	93.75%
34	213	19	0	1	0	95.00%
35	214	13	0	0	0	100.00%
36	215	20	0	0	0	100.00%
37	217	7	0	5	0	58.33%
38	219	14	0	0	0	100.00%
39	220	13	0	0	0	100.00%
40	221	13	0	0	0	100.00%
41	222	13	0	0	1	92.86%
42	223	14	0	0	0	100.00%
43	228	13	0	0	0	100.00%
44	230	15	0	0	0	100.00%
45	231	11	0	0	0	100.00%
46	232	9	0	0	0	100.00%
47	233	18	0	0	0	100.00%
48	234	17	0	0	0	100.00%
Average		639	0	22	10	**95.48%**

Table 1 Ten Seconds Accuracy for ECG Data Annotation

Table 1 describes the result of testing in ten seconds for all data. Numbers of true classification of heart beat types is 639, numbers of true classification of absent of beat types is 0, numbers of false classification of absent beat types is 22 and number of false classification of the beat types is 10. Then the average accuracy for 10 seconds data annotation is 95.48%.

No	Data Name	TP	TN	FP	FN	Accuracy
1	100	75	0	0	0	100.00%
2	101	72	0	0	0	100.00%
3	102	74	0	0	0	100.00%
4	103	71	0	0	0	100.00%
5	104	67	0	2	0	97.10%
6	105	76	0	3	4	91.57%
7	106	68	0	0	0	100.00%
8	107	69	0	3	0	95.83%
9	108	60	0	0	0	100.00%
10	109	88	0	8	2	89.80%
11	111	70	0	1	0	98.59%
12	112	87	0	0	0	100.00%
13	113	56	0	1	1	96.55%
14	114	55	0	0	0	100.00%
15	115	64	0	0	0	100.00%
16	116	78	0	0	0	100.00%
17	117	51	0	0	0	100.00%
18	118	72	0	2	0	97.30%
19	119	51	0	0	15	77.27%
20	121	60	0	1	0	98.36%
21	122	89	0	0	0	100.00%
22	123	46	0	2	2	92.00%
23	124	50	0	0	0	100.00%
24	200	80	0	3	7	88.89%
25	201	86	0	2	2	95.56%
26	202	54	0	1	0	98.18%
27	203	103	0	2	0	98.10%
28	205	40	0	1	5	86.96%
29	207	86	0	1	7	91.49%
30	208	86	0	1	7	91.49%
31	209	95	0	0	0	100.00%
32	210	91	0	0	1	98.91%
33	212	90	0	2	0	97.83%
34	213	111	0	1	0	99.11%
35	214	77	0	0	0	100.00%
36	215	112	0	1	2	97.39%

37	217	61	0	9	2	84.72%
38	219	70	0	0	5	93.33%
39	220	74	0	0	0	100.00%
40	221	78	0	1	0	98.73%
41	222	67	0	0	9	88.16%
42	223	78	0	1	1	97.50%
43	228	63	0	6	2	88.73%
44	230	80	0	0	0	100.00%
45	231	63	0	0	1	98.44%
46	232	57	0	0	1	98.28%
47	233	105	0	0	0	100.00%
48	234	94	0	0	0	100.00%
Average		3550	0	55	76	**96.38%**

Table 2 One Minute (60 seconds) Accuracy for ECG data Annotation

Table 2 describes the result of testing in one minute (60 seconds) for all data. Numbers of true classification of heart beat types is 3550, numbers of true classification of absent of beat types is 0, numbers of false classification of absent beat types is 55 and number of false classification of the beat types is 76. Then the average accuracy for 10 seconds data annotation is 96.38%.

No	Data Name	TP	TN	FP	FN	Accuracy
1	100	369	0	0	3	99.19%
2	101	339	0	2	1	99.12%
3	102	329	0	14	24	89.65%
4	103	356	0	0	0	100.00%
5	104	317	0	6	0	98.14%
6	105	400	0	3	14	95.92%
7	106	312	0	17	3	93.98%
8	107	348	0	6	2	97.75%
9	108	281	0	2	5	97.57%
10	109	426	0	14	3	96.16%
11	111	343	0	36	0	90.50%
12	112	426	0	1	2	99.30%
13	113	268	0	7	14	92.73%
14	114	257	0	10	6	94.14%
15	115	315	0	1	1	99.37%
16	116	385	0	0	11	97.22%
17	117	252	0	0	0	100.00%
18	118	343	0	16	3	94.75%
19	119	292	0	0	35	89.30%
20	121	291	0	9	4	95.72%
21	122	414	0	8	1	97.87%
22	123	233	0	5	12	93.20%
23	124	238	0	4	3	97.14%
24	200	393	0	23	31	87.92%
25	201	410	0	14	20	92.34%
26	202	260	0	6	0	97.74%
27	203	466	0	10	18	94.33%
28	205	230	0	22	24	83.33%
29	207	406	0	9	28	91.65%
30	208	406	0	9	28	91.65%
31	209	480	0	4	3	98.56%
32	210	416	0	8	20	93.69%
33	212	437	0	28	0	93.98%
34	213	410	0	16	33	89.32%
35	214	372	0	3	7	97.38%
36	215	552	0	7	8	97.35%
37	217	299	0	33	1	89.79%
38	219	368	0	0	13	96.59%
39	220	355	0	0	0	100.00%
40	221	400	0	2	6	98.04%
41	222	346	0	2	20	94.02%
42	223	374	0	2	20	94.44%
43	228	322	0	14	14	92.00%
44	230	392	0	3	3	98.49%
45	231	268	0	3	23	91.16%
46	232	291	0	0	5	98.31%
47	233	505	0	1	10	97.87%
48	234	445	0	0	19	95.91%
Average		17137	0	380	501	**95.10%**

Table 3 Five Minutes (300 seconds) Accuracy for ECG data Annotation

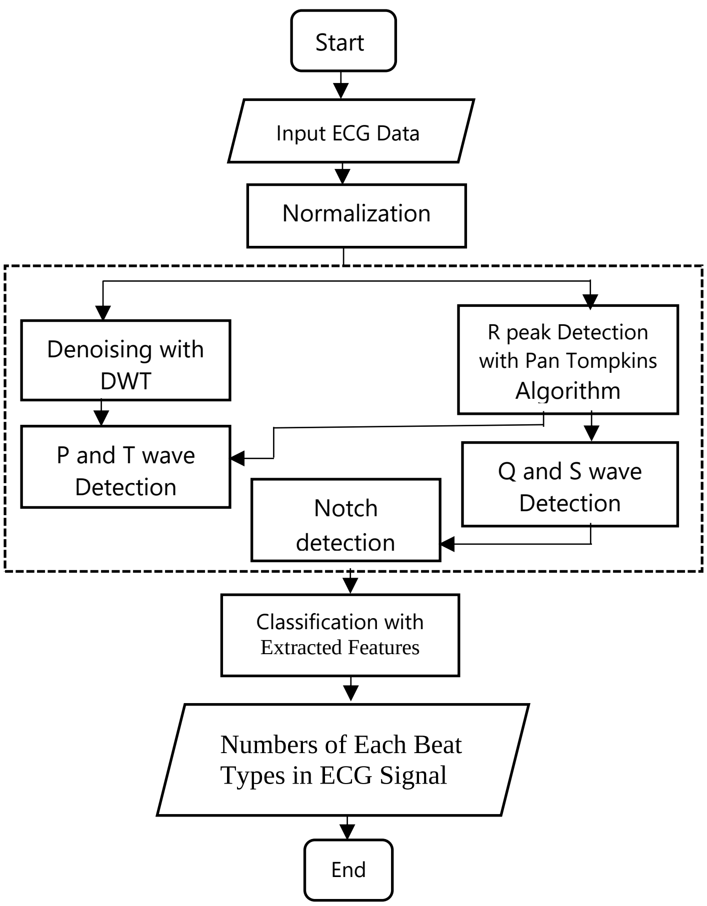
Table 3 describes the result of testing in five minutes (300 seconds) for all data. Numbers of true classification of heart beat types is 17137, numbers of true classification of absent of beat types is 0, numbers of false classification of absent beat types is 380 and number of false classification of the beat types is 501. Then the average accuracy for five minutes data annotation is 95.10%.

5.SYSTEM IMPLEMENTATION

The patient health condition given to analysis of ECG signals can provide clinicians with valuable information. A typical ECG signal of a normal heartbeat (or cardiac cycle) consists of a P wave, a QRS complex and a T wave. Figure 1 describes the flowing steps:

- ECG data Acquisition and signal normalization
- Normalization & denoising
- Feature extraction
  - R-wave detection
  - Waves detection factor are Q,S,P and T
  - Notch Detection
  - Heart beat types of classification by using extracted features.

  -



6.CONCLUSION

The system has been implemented for ECG signal classification using Discrete Wavelet Transform and Pan Tompkins Algorithm. Discrete wavelet transforms and Pan Tompkins algorithm is used to extract the important features such as R peaks, P, Q, S, T wave and notch. The system classifies the heartbeat types by using rule based algorithm. The system accurately classifies and differentiated normal and abnormal heartbeats such as left bundle branch block (LBBB), right bundle branch block (RBBB), premature atrial contraction (PAC), premature ventricular contraction (PVC) and paced beat (PB) with adequate level of accuracy. In future, it can also try different classification methods for better result for categorization of various kinds of abnormalities. ECG signals are obtained to open source MIT-BIH cardiac arrhythmia database. All of the operation steps are completed in MATLAB for computerized interpretation of heartbeat types of classification of the ECG signal. Cloud computing can be used to make this system available in the world.

REFERENCES

[1] Paul S Addison , "Wavelet transforms and the ECG: a review", Institute of physics publishing 8 August 2005

[2] Mallat S G, "A theory for multiresolution signal decomposition", the wavelet representation IEEE Trans. Pattern Anal. Mach. Intell. , September 1989.

[3] Lee S-H, Zahouani H, Caterini R and Mathia T G "Morphological characterisation of engineered surfaces by wavelet transform" Int. J. Mach. Tools Manuf , July 1998..

[4] Mallat S G, "A Wavelet Tour of Signal Processing" (San Diego, CA: Academic), June1998.

A. ddison P S, "The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science Engineering, Medicine and Finance" (Bristol: Institute of Physics Publishing), April 2002.

[5] S.Mukhopadhyay, S.Biswas, A.B. Roy and N. Dey, "Wavelet Based QRS complex Detection", Vol. 2, Issue 3, May-Jun 2012, pp.2361-2365

[6] O.Brien, Dean, "Investigation of peak detection methodologies for ECG signals", Doctoral dissertation, Charles Darwin University, 2014