# ECG DISEASE DETECTION & FEATURE EXTRACTION BY WAVELET TRANSFORMATION

**[1]TASBIHA TABASSUM ANTARA, [2]SAIF IMAM, [3]ZARIN IRTIZA**

[1,2]Postgraduate Student, BRAC University, Bangladesh [3]ICT Lecturer, Scholars School & College, Bangladesh
E-mail: [1]tasbihaantora@gmail.com, [2]supersaiif@gmail.com, [3]zarinirtiza@gmail.com

**Abstract—** ECG is the most common as well as basic test to run on patients to check any kind of anomalies in the heart. In the ECG result 10 to 20 minutes long continuous data of a patient's heart is down sampled and printed as a 1D graph. We have developed a program which will take the continuous dataset from the ECG machine and analyses the data and extracts various features of the ECG wave. At first we decompose the data using Wavelet decomposition. Then the data is reconstructed in 4 levels which removes the noise from the signal. In the same time we detect major components of the ECG wave which is P wave, QRS complex and T wave. Then we calculate ST deviation, heart rate and extract other features such as location and amplitude of each wave in order to detect anomalies. Finally our output provides the heart status (healthy, if any disease found, if any major or minor risk) in a language that the patient can understand and also some detailed wave properties in medical term for the doctors.

**Index Terms—** Dataset of ECG machine, Wavelet decomposition, Removing noise, Detecting anomalies etc

## I. INTRODUCTION

Most of the stuffs that perform tests are not doctors. Thus they cannot detect any problems instantly. They forward the averaged printed result to the doctors and then the doctors provide medical explanation and recommendation which wastes a lot of time for a patient. Our program can detect diseases based on the properties of the ECG wave. So the patient gets to know the heart status as soon as the ECG is done without visiting a doctor which saves a lot of time for the patient and he/she can still visit the doctor for expert opinions and our program output will also provide wave properties which are helpful for the doctors. Our program may be integrated with the ECG machine in the diagnostic centre to provide accurate output, heart status (if any disease detected, if any future risks, heart rate) and wave properties.

## II. FEATURES OF THE PROGRAM

### A. Input Flexibility

Our Input which is continuous dataset of ECG/EKG test result, is taken from MIT-BIH Database (physionet.org) which is available is various formats (*.mat, *.csv, *.dat, etc.). At this moment we prefer using the *.mat file because *.csv and *.dat files require additional algorithm to remove the header. Our program can also be modified to take live feed from ECG/EKG machine.

### B. Informative Medical Database

Our Program contains useful medical terms and data that the doctors use and need to treat a patient. All the waves in the ECG/EKG result are described in terms of shape and amplitude. Medical information corresponding with wave properties are included. The normal and abnormal range of each wave, segment and duration is displayed in our output taken from medical books.

### C. Disease Detection

Our program detects diseases based on the wave properties and amplitude and duration of certain wave(s). Some properties give hints or probability for a certain diseases but those are not 100% certain and require other medical test results. For this kind of probable disease our program output is given by Risk or Probability.

### D. Statistical Information

Our program also calculates each wave amplitude, mean and deviation. They are compared with normal range to provide medical information and probable cause.

## III. ECG/EKG ANALYSIS

### A. Medical Uses of ECG

Electrocardiography (ECG or EKG*) is the process of recording the electrical activity of the heart over a period of time using electrodes placed on a patient's body. These electrodes detect the tiny electrical changes on the skin that arise from the heart muscle depolarizing during each heartbeat. **

Reasons for performing electrocardiography include:

1. Suspected heart attack
2. Suspected pulmonary embolism
3. A third heart sound, fourth heart sound, a cardiac murmur or other findings to suggest structural heart disease
4. Cardiac arrhythmia / dysrhythmias
5. Fainting or collapse
6. Seizures
7. Monitoring the effects of a heart medication
8. Assessing severity of electrolyte abnormalities, such as hyperkalemia

### B. TYPES OF ECG WAVE

The P wave represents atrial depolarization. In a normal ECG, the P-wave precedes the QRS complex. It looks like a small bump upwards from the baseline. The amplitude is normally 0.05 to 0.25mV. Normal duration is 0.06-0.11 seconds. The shape of a P-wave is usually smooth and rounded.

The QRS complex indicates ventricular depolarization. Depolarization triggers contraction of the ventricles. Because of the larger tissue mass, the QRS complex is larger than the P wave. While the prototypical QRS complex consists of three wave components, one or two of these components may be missing.QRS interval is measured from the end of the PR interval to the end of the S wave.

The T wave indicates the depolarization of the ventricles. It is a slightly asymmetrical waveform that follows (after a pause), the QRS complex. Take note of T waves have a downward (negative) deflection or of T waves with tall, pointed peaks. The U-wave is a small upright, rounded bump. When observed, it follows the T-wave.

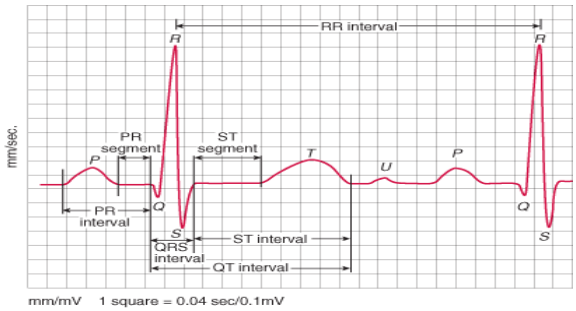



**Fig. 1: All types of wave in ECG**

## IV. SYSTEM IMPLEMENTATION

### A. ECG Dataset as an input

ECG Dataset as an input: We have collected a lot of continuous ECG dataset of patients which is a *.mat file from MIT-BIH Database (physionet.org). It is a matrix data. We have collected the real time continuous data of the patient for high accuracy disease detection. We have used mat lab R2016a for our program.

Append 100 zeros before and after the signal to remove the possibility of window crossing the signal boundaries while looking for peak locations.

Plotting the matrix using the algorithm gives us the following diagram.

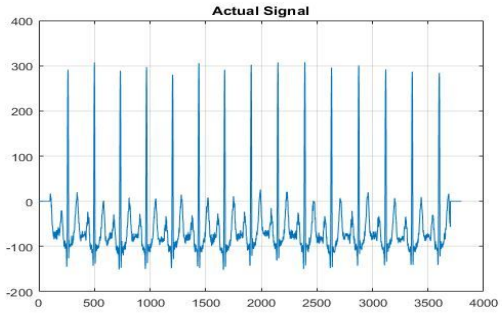



**Fig. 2: Original ECG Signal**

### B. Wavelet Decomposition

Perform wavelet decomposition. The process of wavelet decomposition down samples the signal which essentially means taking the samples at a much lower frequency than the original signal. Therefore details are reduced and QRS complex is preserved.

```
[c,l]=wavedec(s,4,'db4');
```

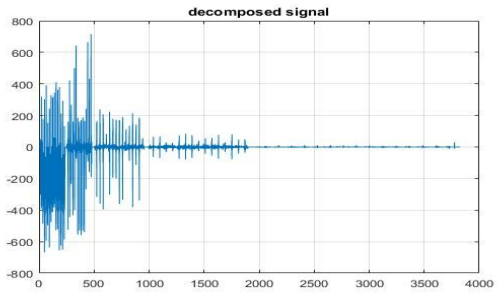



**Fig. 3: Decomposed Signal**

### C. Reconstruct Signal

Extract the Coefficients after the transform.

```
ca1=appcoef(c,l,'db4',1);
ca2=appcoef(c,l,'db4',2);
ca3=appcoef(c,l,'db4',3);
ca4=appcoef(c,l,'db4',4);
```

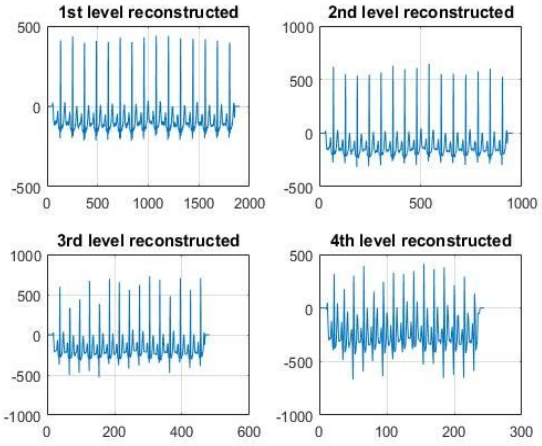



**Fig. 4: Reconstructed Signal**

While plotting the coefficients we observed that the frequency bands are separated and ca1, ca2, ca3 and ca4 are cleaner signal. But they will have less number of samples than the actual signal due to down sampling. We can see that first signal resembles to the actual signal but has exactly one forth number of samples because the signal was decomposed in 4 levels. 2nd level has exactly half number of samples that of 1st level, 3rd level has exactly half number of samples than the 2nd level. Because the number of samples is reduced, such signals are also called down-sampled signal.

It is clear that 2nd level decomposed data is noise free. Therefore we consider this signal as ideal ECG signal from which QRS must be detected. But the first R is located in 3rd level decomposition signal at approximately 40th sample whereas the same is

located in the original signal at 260th location. Therefore once R peak is detected in 3rd level reconstructed signal, it must be cross validated in the actual signal.

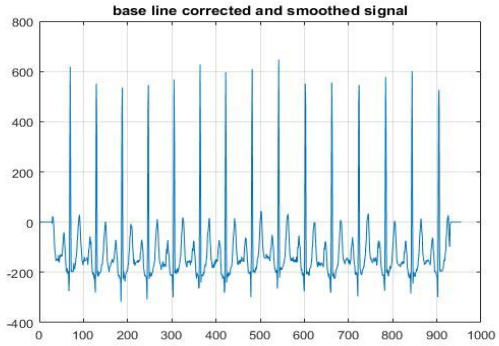



**Fig. 5: Base line Corrected Signal**

### D. Detecting R-peak

First we find the values which are greater than 60% of the max value of the actual signal. Invariably these are R peaks. As the decomposed signals are noise free signals, First R peak needs to be detected in the Noise free signal. But the ultimate goal is to detect the Peak in the original Signal. The sample values in Original Signal will be different than the decomposed signal. So our strategy here will be to first detect the R peaks in the down sampled signal and then cross verify those point the actual signal.

Let y1 be the decomposed signal.

```
m1=max(y1)*.60;
P=find(y1>=m1);
```

So, P is now set of points which satisfies the above criteria. If we observe the signal very closely, R-Peak is not a single impulse peak; therefore there are chances of multiple points in the same peak satisfying the criteria. One thing to remember is in 500Hz sampled signal number to R-Location will be found below 350 samples. In 4th Level decomposition order this value is around 20. So first we will remove the R locations that are too close.

Variable P2 represents the position of R-Peaks in the down sampled signal.

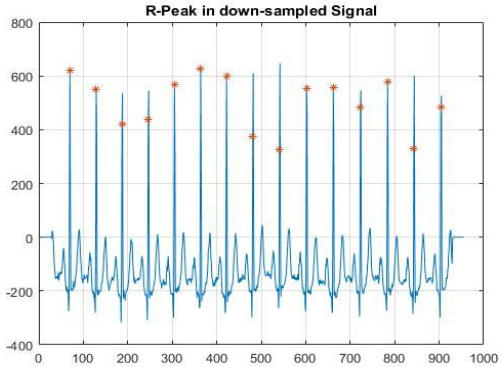



**Fig. 6: R-peak detected in Down-sampled Signal**

### E. Detecting R-peak in the Original Signal

Search for the position of all the location in signal y1 which are greater than this value m1. They are R locations. We know that a R Location in Rt is at least 1/4th of the actual R location of the same point. Hence we will first map the detected positions to original signal by first multiplying with 4.

```
P3=P2*4;
%Multiply the current location with 4 to get the actual scale.
```

R location in down sampled signal will never be on the original signal at a scale of 4. Down sampling process always devide the signal positions. Hence we need to search for the maximum value in the Original Signal in a window of +-20 samples from the reference R point obtained as P3.

Now Ramp and Rloc represents the R peak amplitude and location at the original scale.

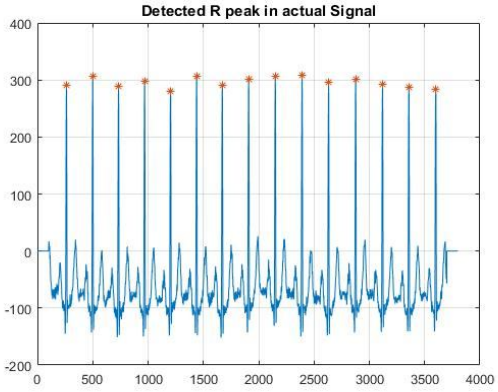



**Fig. 7: R-peak in the Original Signal**

### F. Detecting other with Reference to R-peaks

From R-Peak Traverse Forth and Back and Search for Minima and Maxima, these are P, Q, T, S peaks respectively. So loop in R location and search for the other peaks.

Firstly, if we observe the waveform, it will be very clear that from R location if we select a window of Rloc-100 to Rloc-50 and find the maximum, than that maximum is P peak.

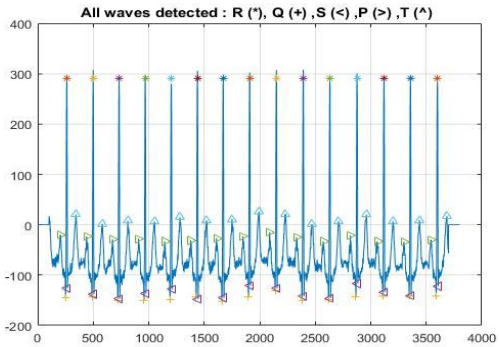



**Fig. 8: All waves detected in Original Signal**

## V. FEATURE EXTRACTION & DISEASE DETECTION

### A. ST Deviation

ST Deviation calculation is necessary to detect Myocardial Infarction (MI) which causes Cardiac arrhythmia in common language Heart Attack.

Normal Range: ST Deviation/100 < 1
Heart Attack: ST Deviation/100 > 1

**B. PR Interval**

Measured from beginning of P to beginning of QRS in the frontal plane

Normal Range: 0.12 - 0.20s

Short PR: < 0.12s

Syndromes:

- WPW (Wolff-Parkinson-White) Syndrome: An accessory pathway (called the "Kent" bundle) connects the right atrium to the right ventricle or the left atrium to the left ventricle, and this permits early activation of the ventricles (delta wave).
- LGL (Lown-Ganong-Levine): An AV nodal bypass track into the His bundle exists, and this permits early activation of the ventricles without a delta-wave because the ventricular activation sequence is normal.
- AV Junctional Rhythms with retrograde atrial activation (inverted P waves in II, III, aVF): Retrograde P waves may occur before the QRS complex (usually with a short PR interval), in the QRS complex or after the QRS complex.
- Ectopic atrial rhythms originating near the AV node (the PR interval is short because atrial activation originates close to the AV node; the P wave morphology is different from the sinus P)

Prolonged PR :> 0.20s

Syndromes:

- First degree AV block (PR interval usually constant)
- Intra-atrial conduction delay (uncommon)
- Slowed conduction in AV node (most common site)
- Slowed conduction in His bundle (rare)
- Slowed conduction in bundle branch(when contralateral
- bundle is blocked)
- Second degree AV block (PR interval may be normal or prolonged; some P waves do not conduct)
- Type I (Wenckebach): Increasing PR until nonconducted P wave occurs
- Type II (Mobitz): Fixed PR intervals plus nonconducted P waves
- AV dissociation: Some PR's may appear prolonged, but the P waves and QRS complexes are dissociated.

**C. QRS Duration**

Duration of QRS complex in frontal plane

- Normal: 0.06 - 0.10s

Prolonged QRS Duration:

QRS duration > 0.10s:

- QRS duration 0.10 - 0.12s
- Incomplete right or left bundle branch block
- Nonspecific intra ventricular conduction delay (IVCD)
- Some cases of left anterior or posterior fascicular block

QRS duration ≥ 0.12s:

- Complete RBBB or LBBB
- Nonspecific IVCD
- Ectopic rhythms originating in the ventricles (e.g., ventricular tachycardia, pacemaker rhythm)

**D. QT Interval**

Measured from beginning of QRS to end of T wave in the frontal plane

a. Normal: heart rate dependent (corrected QT = $QT_C$ =measured QT, sq-root RR in seconds; upper limit for $QT_C$ = 0.44 sec)
b. Long QT Syndrome: "LQTS" (based on upper limits for heart rate; $QT_C \geq 0.47$ sec for males and ≥ 0.48 sec in females is diagnostic for hereditary LQTS in absence of other causes of increased QT)

This abnormality may have important clinical implications since it usually indicates a state of increased vulnerability to malignant ventricular arrhythmias, syncope, and sudden death. The prototype arrhythmia of the Long QT Interval Syndromes (LQTS) is Torsade-de-pointes, a polymorphic ventricular tachycardia characterized by varying QRS morphology and amplitude around the isoelectric baseline. Causes of LQTS include the following:

- Drugs (many antiarrhythmics, tricyclics, phenothiazines, and others)
- Electrolyte abnormalities ($\downarrow K^+$, $\downarrow Ca^{++}$, $\downarrow Mg^{++}$)
- NS disease (especially subarrachnoid hemorrhage, stroke, trauma)
- Hereditary LQTS (e.g., Romano-Ward Syndrome)
- Coronary Heart Disease (some post-MI patients)

**E. No Disease/ Healthy Heart**

For a healthy heart the primary condition is:

- ST Deviation/100 < 1

And for a perfect heart condition which is very rare includes these conditions:

- PR Interval: 0.12 - 0.20s
- QRS Duration: 0.06 - 0.10s
- Heart Rate: 60 - 90 bpm

## VI. RESULT VALIDATION & ACCURACY CHECK

### A. Case Study 1

We have collected a lot of ECG data from MIT-BIH Arrhythmia Database, the largest Bio-Engineering Research Organization. This Arrhythmia Database contains ECG data along with the disease analysis from doctors and they have also developed their own method for detecting wave properties. Then we take only the ECG data from the database and cross-verify with our program output with MIT-BIH report analysis.

Although MIT-BIH does not provide PR interval and QRS duration, our program successfully detects the disease Myocardial Infarction (Heart Attack) validated by MIT-BIH database patient description.

### B. Case Study 2

'ECG Q R S wave online detector' Program is found from Matlabworks.com. The program detects only Q, R, S wave and thus heart rate. We run this program and also our program on the same input to cross-validate our result.

As per Medical Books, Q and S wave is negative. The comparison shows all our Q and S amplitudes are negative and consistent. On the other hand, 'ECG Q R S wave online detector' detects some false positive points which are not accurate. In our Program R-peak is also consistent and matches with actual signal. Both the programs detect irregular Heart Rate but our program offers addition feature such as PR Interval and QRS duration. Thus we can conclude that our program is more accurate.

## ACKNOWLEDGMENT

We feel honored in expressing our heartfelt indebtedness and gratitude to our respected supervisor Dr. Md. Haider Ali. He guided us with continuous encouragement, technical suggestion and valuable instructions throughout the thesis work.

## REFERENCES

[1] Uddin, J., Van, D. N., & Kim, J. (2015). Accelerating 2D Fault Diagnosis of an Induction Motor using a Graphics Processing Unit. International Journal of Multimedia and Ubiquitous Engineering IJMUE, 10(1), 341-352. doi:10.14257/ijmue.2015.10.1.32

[2] R Wave Detection in the ECG. (n.d.). Retrieved April 15, 2016, from https://www.mathworks.com/examples/wavelet/mw/wavelet-ex77408607-r-wave-detection-in-the-ecg

[3] PhysioNet. (n.d.). Retrieved April 15, 2016, from https://www.physionet.org/

[4] Normal ECG. (n.d.). Retrieved April 15, 2016, from https://meds.queensu.ca/central/assets/modules/ECG/normal_ecg.html

[5] Nixon, M. S., &Aguado, A. S. (2002). Feature extraction and image processing. Oxford: Newnes.

[6] Medscape. (n.d.). Retrieved April 15, 2016, from http://www.medscape.com/viewarticle/545741_1

[7] MATLAB Speaks SDR. (n.d.). Retrieved April 15, 2016, from http://www.mathworks.com/

[8] Lesson 4. (n.d.). Retrieved April 15, 2016, from http://www.ieee-uffc.org/ultrasonics/software/matlab/lecture4/lecture4_6.htm

[9] Grasshopper Network Discussion Forum. (n.d.). Retrieved April 15, 2016, from http://www.grasshoppernetwork.com/

[10] EKG Nomenclature. (n.d.). Retrieved April 15, 2016, from http://www.practicalclinicalskills.com/ekg-lesson.aspx?CourseCaseOrder=2

[11] ECG – A Pictorial Primer. (n.d.). Retrieved April 15, 2016, from http://www.medicine-on-line.com/html/ecg/e0001en_files/05.htm

[12] ECG processing — R-peaks detection. (n.d.). Retrieved April 15, 2016, from http://www.librow.com/cases/case-2

[13] ECG analysis using wavelet transform: Application to myocardial ischemia detection. (n.d.). Retrieved April 15, 2016, from http://www.sciencedirect.com/science/article/pii/S1297956203000032

[14] ECG Basics. (n.d.). Retrieved April 15, 2016, from http://lifeinthefastlane.com/ecg-library/basics/

[15] ECG. (n.d.). Retrieved April 15, 2016, from https://en.wikipedia.org/wiki/ECG

[16] Detection of ECG characteristic points using Multiresolution Wavelet Analysis based Selective Coefficient Method. (n.d.). Retrieved April 15, 2016, from http://www.sciencedirect.com/science/article/pii/S0263224109002139

[17] CodeProject - Forum Discussion. (n.d.). Retrieved April 15, 2016, from http://www.codeproject.com/

[18] Characteristics of the Normal ECG. (n.d.). Retrieved April 15, 2016, from http://ecg.utah.edu/lesson/3

[19] BioMedical Engineering OnLine. (n.d.). Retrieved April 15, 2016, from https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-10-23

[20] Annadurai, S., &Shanmugalakshmi, R. (2010). Fundamentals of digital image processing. Noida: Pearson.

[21] Algorithm Links. (n.d.). Retrieved April 15, 2016, from http://www.mit.edu/~gari/algos.html

[22] Acharya, T., & Ray, A. K. (2005). Image processing: Principles and applications. Hoboken, NJ: John Wiley.

[23] Citation Machine automatically generates citations in MLA, APA, Chicago, Turabian, and thousands more! (n.d.). Retrieved April 15, 2016, from http://www.citationmachine.net/

★★★