A Survey on ECG Analysis

Amar Pratap Singh (211EE212)
Bharat Singh (211EE218)
M Shreyas Shenoy (211EE236)
Manish Shakya (211EE239)
Rohit Merotha (211EE248)
Sujal Ojha (211EE259)
Department of Electrical and Electronics Engineering
National Institute of Technology, Karnataka - 575025

***Abstract*—An electrocardiogram (ECG) signal is associated with the electrical activity of the heart [6]. The analysis of Electrocardiogram (ECG) signals plays a vital role in clinical cardiology and has evolved significantly over the years. ECG Analysis was a groundbreaking development in 20th century.It is crucial technology in the field of medical sciences, today a detailed study of the process and methodology of the ECG process are still being discussed and optimized.**

I. Introduction

ECG analysis (depending on the type of test) can have several steps, such as preprocessing, feature extraction, feature selection, feature transformation, and classification.It is important that each step is followed for a proper analysis. Also, the basic measurements of the ECG signal and the corresponding constitution play an important role in the analysis. This work provides a comprehensive review of the literature on ECG analysis, especially from the last decade, reviewed based on all the key aspects mentioned above.
The objectives of this project include:

- To learn and understand the procedure involved in analysis of ECG signals.
- To use existing algorithms and develop a method to analyse the ECG signals from MIT-BIH database [1].
- To implement in MATLAB and get the desired results.

II. Literature Survey

The primary challenge in diagnosing heart diseases using ECGs is the variability of ECG signals among individuals. Different patients with the same disease can exhibit distinct ECG morphologies. Additionally, there are instances where two different diseases may manifest similar characteristics in an ECG signal. These issues complicate the process of diagnosing heart diseases.

To detect abnormalities in heartbeats, it is essential to analyze the electrical signals of each heartbeat. Consequently, the analysis of long-term ECG records, particularly in situations such as bedside monitoring or wearable online health care monitoring, can be arduous and time-consuming for individuals. Furthermore, fatigue can lead to personal errors during ECG analysis, and interpreting these signals requires a deep understanding. As a result, computer-assisted methods that offer automatic ECG analysis are employed to address these challenges.

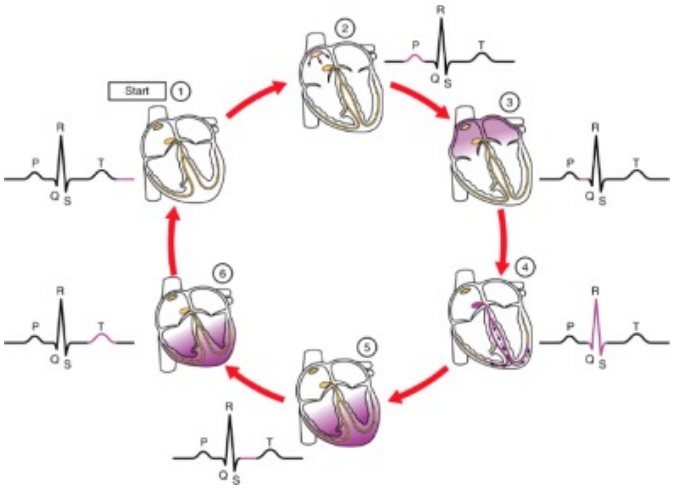

Fig. 1. A diagram of cardiac cycle with the associated ECG signal [5]

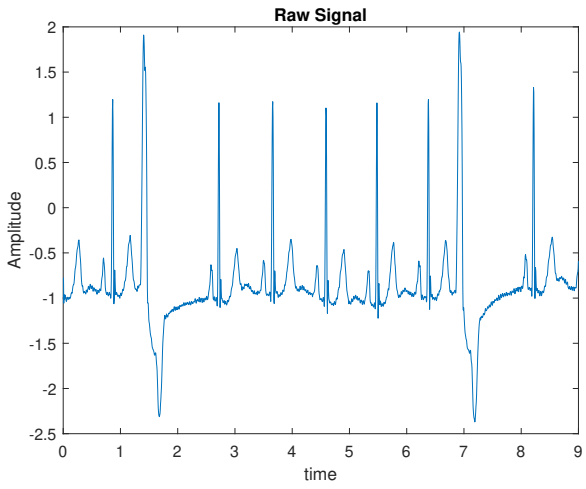

Fig. 2. Typical ECG Signal

A. History

The origins of ECG can be traced back to the early 20th century when Einthoven introduced the concept of lead-based recordings, laying the foundation for modern ECG analysis. Over several years, ECG technology has advanced.

B. Components of ECG Signal

ECG signal consists of components such as: P-wave, QRS complex, and T-wave, each representing different parts of the cardiac cycle. These components are necessary for diagnosis of arrhythmias, myocardial infarctions, and other heart conditions. It can also be used for Biometrics and other applications as well.

C. Challenges and Difficulty

The major challenge in analysis of ECG signals is removing the unnecessary components. ECG recordings often suffer from various forms of interference and unwanted signals. During the initial data preprocessing stage, the objectives include mitigating these interferences and artifacts, identifying essential fiducial points (P, Q, R, S, T) and standardizing signals to facilitate comparisons between different patients. Various types of noise can be summarized and categorized as follows [3], [4] :

- **Electrode connectivity issues:** Noise arising from poor contact between the electrode and the skin, effectively disconnecting the measurement system from the patient.
- **Power line interference:** This interference occurs at a frequency of 50 or 60 Hz, with a bandwidth of less than 1 Hz.
- **Baseline drift:** A low-frequency noise ranging from 0.15 Hz to 0.3 Hz. This noise is a result of the patient's breathing and causes a gradual shift in the ECG baseline.
- **Artifacts due to electrode movement:** Disturbances resulting from changes in electrode-skin impedance caused by electrode motion.
- **Electrosurgical interference:** Disruption produced by other medical equipment in the patient care setting, typically occurring at frequencies between 100 kHz and 1 MHz.
- **Muscle activity interference (Electromyography noise):** Noise originating from the contraction of muscles other than the heart.

In the classification of cardiovascular diseases, the presence of different levels of noise can lead to inaccurate assessments by medical professionals and a decrease in diagnostic accuracy. So it necessary to remove these noises.

D. Modern ECG analysis methods

In 1985, Jiapu Pan and Willis J. Tompkins published a paper titled "A Real-Time QRS Detection Algorithm" in the IEEE Transactions on Biomedical Engineering. This algorithm helped identify the critical feature points of the ECG waveform. The algorithm involves the use of bandpass filtering to enhance the QRS complex, differentiation to enhance the slope of the R-wave, and adaptive thresholding to detect the QRS complex. The Pan-Tompkins algorithm has become a standard method for QRS detection and is widely implemented in ECG analysis both in software and medical devices due to its effectiveness and efficiency.

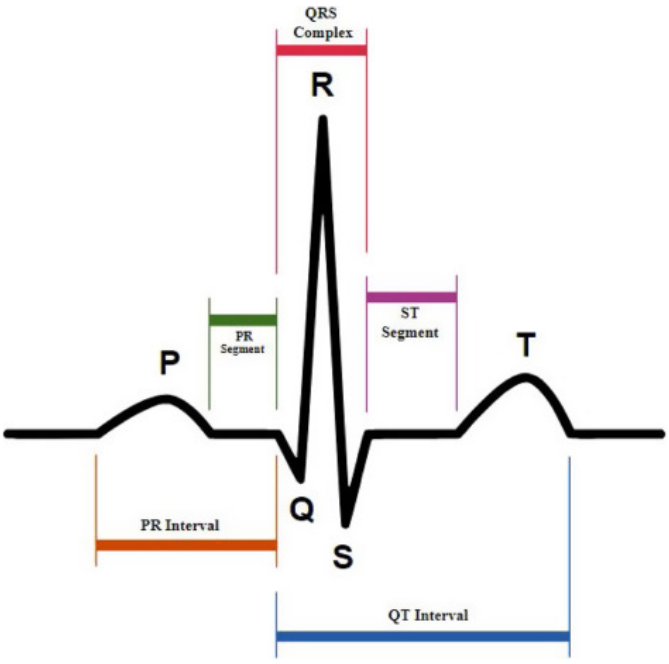

Fig. 3. Components of the ECG Signal

III. Methodology

A. *Overview*

For our project, we have decided to use the Pan-Tompkin's algorithm for Pre-processing, Feature detection and extraction. Further we have also calculated heart beat rate and classified several arrhythmia diseases based on the features extracted such as: QRS width, RR interval and average RR interval. The ECG signal data is taken from the MIT-BIH arrhythmia database [1] and sampled at a rate of 200 samples/sec. Next the raw signal which has noise from muscles, baseline wander, powerline interference etc. needs to be filtered using an appropriate bandpass filter. In our case we are using an Integer coefficient based bandpass filter that only allows QRS components of the signal and removes all other noises, which can help reduce computation and processing requirements. The filter is a cascaded low and high pass filter that only allows 5 to 11 Hz frequencies, since most of the QRS spectrum lies here. Next it is passed through a differentiator and squaring to get the slope of the QRS complex and to reduce the chances of false positive. Next it passes through a moving window integrator that clearly shows the slope and width of the QRS complex while reducing errors and false positives caused by T wave [3].

The entire Methodology may be described as per the below flowchart which includes: pre-processing, thresholding, feature extraction.

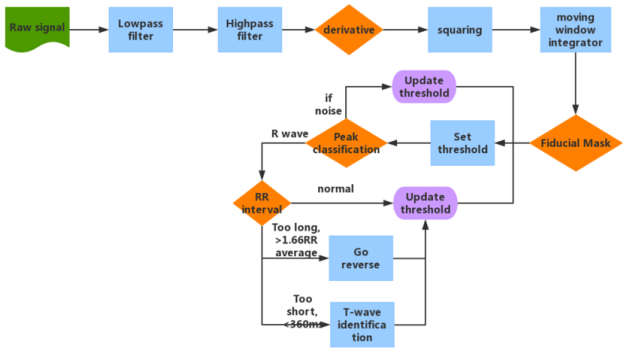

Fig. 4. FlowChart

B. *Pre-Processing*

BAND PASS FILTER (Filtering) : A band-pass filter is a cascade of low and high-pass filters. It isolates signal energy in the 5-15 Hz range, which is mainly QRS energy. The low-pass filter effectively suppresses high-frequency interference, such as 60Hz power line interference and T-wave interference, while allowing low-frequency components. Meanwhile, the high-pass filter effectively removes low-frequency noise such as Baseline Wander while allowing for higher-frequency components such as power line interference [3].

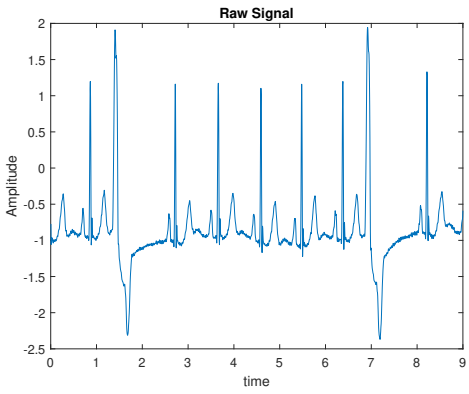

Fig. 5. Original ECG Signal

Low Pass Filter

Second order Low Pass Filter. [3]

$$H(Z) = \frac{(1 - Z^{-6})^2}{(1 - Z^{-1})^2} \tag{1}$$

Difference equation,

$$y(nT) = 2y(nT - T) - y(nT - 2T) + x(nT) \\ -2x(nT - 6T) + x(nT - 12T) \tag{2}$$

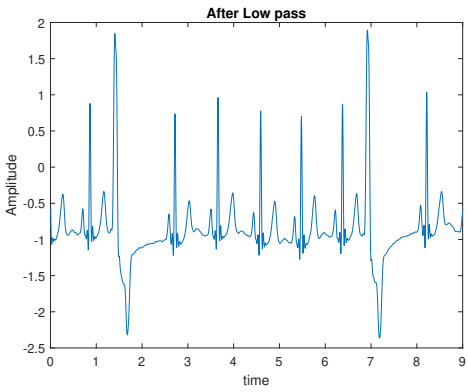

Fig. 6. Signal After Low Pass

Here 11 Hz is the cutoff frequency.

High Pass Filter

Transfer function of the High Pass Filter. [3]

$$H(Z) = \frac{-1 + 32Z^{-16} + Z^{-32}}{1 + Z^{-1}} \tag{3}$$

Difference equation,

$$y(nT) = 32y(nT - 16T) - [y(nT - T) + x(nT) \\ -x(nT - 32T) \tag{4}$$

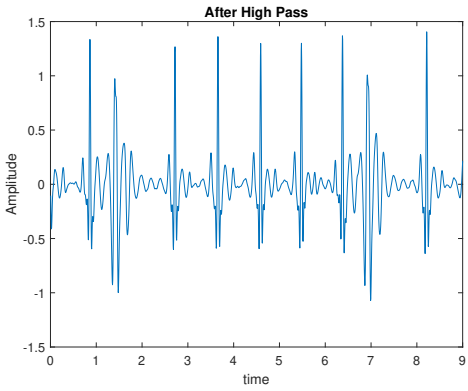

Fig. 7. Signal After High Pass

Here 5 Hz is the low cutoff frequency.

Derivative :

Differentiation is a process that is performed after the ECG data is passed through a bandpass filter. The signal is differentiated to provide information about the slope of the QRS complex. One of the main purposes of differentiation is to identify R-waves (peaks) and S-waves (valleys) in the QRS complex [3].

$$H(Z) = \left(\frac{1}{8T}\right)(-Z^{-2} - 2Z^{--1} + 2Z^{1} + Z^{2}) \tag{5}$$

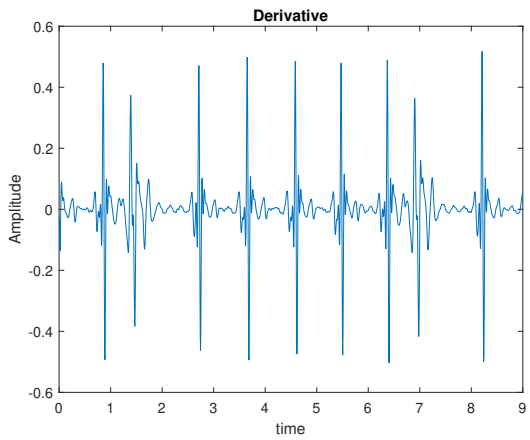

Fig. 8. Signal After Differentiation

Difference equation,

$$y(nT) = \left(\frac{1}{8T}\right)[-x(nT-2T) - x(nT-T) \quad (6)$$
$$+2x(nT+T) + x(nT+2T)]$$

These points are important for heart rate measurement and arrhythmia detection. It enhances the specific characteristics of the ECG signal, making it easier to identify important landmarks and subtle changes in the waveform. This process will attenuate some high-frequency noise components, which increases the accuracy of further analysises.

SQUARING :

After the differentiation process, the signal was amplified so that all signal components were positive values. Quadrature of the ECG signal amplifies the R-waves in the QRS complex while reducing the amplitude of other parts of the signal, such as P-waves and T-waves. This makes it easier to identify the peaks of the QRS complex in the presence of noise and interference [3].

$$y(nT) = [x(nT)]^2 \quad (7)$$

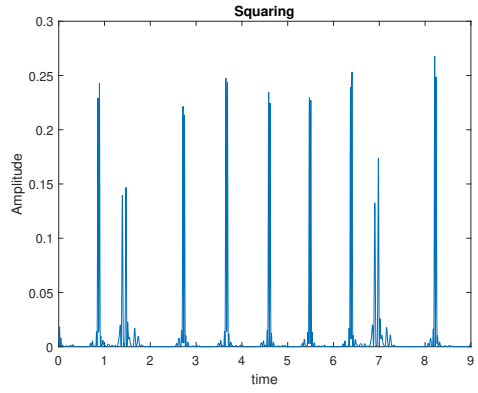

Fig. 9. Signal After Squaring

Signal quadrature attenuates high-frequency components, which helps isolate the QRS complex and suppress low-frequency components such as baseline deviation and noise. The higher amplitudes of the signal related to the QRS complex were further amplified after this quadratic function.

Moving window integration :

Moving window integration is done after squaring the signal and is a very important process. It helps reduce noise and high-frequency components in the signal. By performing a moving window integration, the signal is smoothed and the impact of noise is reduced, making it easier to distinguish the QRS complex from other waveform and noise components. Difference equation, [3]

$$y(nT) = \left(\frac{1}{N}\right)[x(nT-(N-1)T) - \quad (8)$$
$$x(nT-(N-2)T) + ... + x(nT)] \quad (9)$$

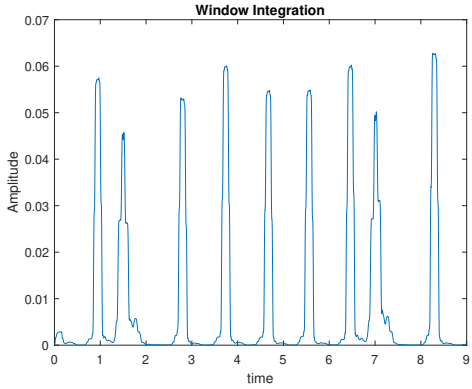

Fig. 10. Signal After Moving Window Integration

Through this process, the peaks of the R-wave are enhanced so that they are more identifiable. This step is crucial for accurate peak detection, which is a key part of the Pan-Tompkins algorithm.

C. *Feature Extraction*

The QRS complex represents the ascent of the integration waveform, with its rising edge's duration being equivalent to the QRS complex's width. A fiducial marker for determining the temporal location of the QRS complex can be established based on the preferred waveform characteristic. This marker is identified either as the peak slope of the rising edge or as the peak of the R wave. For the initial signal analysis, two distinct threshold levels are applied to each of two sets. If no QRS complex is detected within a specific time interval, a lower threshold is employed, necessitating a search method to track the QRS complex's time [3]. Initially, a set of thresholds applied to the integration waveform is calculated using the following formulae:

$$SPKI = 0.125 * PEAKI + 0.875 * SPKI \quad (10)$$
$$(if PEAKI is the signal peak) \quad (11)$$

$$SPKI = 0.125 * PEAKI + 0.875 * SPKI \quad (12)$$

$$(if PEAKI is the noise peak) \quad (13)$$

$$NPKI = 0.125 * PEAKI + 0.875 * NPKI] \quad (14)$$

$$(if PEAKI is the peak volume) \quad (15)$$

$$ABANIl = NPKI + 0.25 * (SPKI - NPKI) \quad (16)$$

$$thresholdI2 = 0.5 * thresholdIl \quad (17)$$

All variables pertain to the integrated waveform: PEAKI represents the peak total, SPKI represents the peak signal level, NPKI stands for the peak volume index, [3] THRESHOLD Il denotes the first threshold used, and THRESHOLD I2 is the second threshold employed. The peak of the SPKI signal indicates the intended QRS complex, whereas NPKI volume level is unrelated to the QRS, such as the Twave. The calculation of SPKI and NPKI is primarily based on their previous values. A new peak is classified as either noise or signal peak, requiring it to surpass THRESHOLD I1 if it is a signal peak or THRESHOLD 12 if searching for the QRS is necessary. For the QRS complex, SPKI is set to 0.25 * PEAKI + 0.75 * SPKI.

To identify a QRS complex, the peak must be recognized in both the integrated and band-pass filtered waveforms.

Adjusting Average RR Interval and Rate Limit: Two RR intervals are averaged: one is the average of the last eight consecutive beats, while the other is the average of the last eight beats with RR intervals [3]. Maintaining these two averages enables adaptation to a fast or irregular heartbeat. The first average is computed as follows:

$$RRAVERAGE1 = 0.125 * (RRn - 7 + ... + RRn) \quad (18)$$

The second average is based on selected beats:

$$RRAVERAGE2 = 0.125 * (RR'7 + RR'6 + ... + RR') \quad (19)$$

The RR interval limits are defined as follows:

$$RRlowlimit = 92 \quad (20)$$

$$RRupperlimit = 92 \quad (21)$$

$$RRMISSEDLIMIT = 166 \quad (22)$$

If a complex QRS is not detected within the RR MISSED LIMIT interval, the highest peak of the QRS between the defined limits is calculated.

T wave detection: If the RI interval is less than 360 ms (with a 200 ms delay), a decision is made to determine whether the current QRS complex is correctly detected or if there is indeed a T wave, based on the peak slope during this waveform being less than half of the previous QRS wave [3].

D. Disease Classification

Efforts to analyze ECG signals are crucial for diagnosing heart diseases. The main objective is early detection, which can save lives and prevent organ damage. Most studies focus on arrhythmias, which can be life-threatening, with two main categories: life-threatening (e.g., ventricular fibrillation and tachycardia) and non-life-threatening arrhythmias. Various classification algorithms and dimensionality reduction methods are used in these studies. Researchers also explore specific ECG arrhythmias like atrial fibrillation.

In addition, research extends to other medical conditions, including sleep apnea detection, Parkinson's disease severity classification, ischemia detection, heart rate turbulence, sleep bruxism, myocardial infarction detection, myocardial scar identification, hypertrophic cardiomyopathy detection, inferior myocardial infarction detection, and coronary artery disease classification. These efforts aim to improve the understanding and diagnosis of various medical conditions beyond arrhythmias.

After getting the QRS width, RR intervals and the average RR interval of the signal we can then classify the signals based on the following criteria [2]:

Category	Criteria
Bradycardia	RR_t>1.5s or AR_t>1.2s
Tachycardia	AR_t<0.6s
Premature ventricular contractions	RR_{t-1}<0.875AR_{t-2} QRS width>0.12s $RR_{t-1}+RR_t$=2AR_{t-2}
Atrial premature beats	RR_{t-1}<0.875AR_{t-2} QRS width normal $RR_{t-1}+RR_t$<2AR_{t-2}

RR_t : RR interval
AR_t : Average RR interval. $AR_t = (RR_t + RR_{t-1} + \cdots + RR_{t-7})/8$

Fig. 11. Disease Classification Criteria

IV. Results

The Pan-Tompkins algorithm is a computationally efficient algorithm which uses integer coefficients based filters and can be run on microcontrollers. Using this, we were able to successfully implement in MATLAB and extract the QRS complex features. Further, we also classified the signals, diagnosed the signals based on the arrythmia category of diseases.

When the results were compared with the MIT-BIH database diagnosis results, the results were always matching.

Further, we also got the heartbeats from the signals and plotted the ECG Signals. The results are obtained as follows:

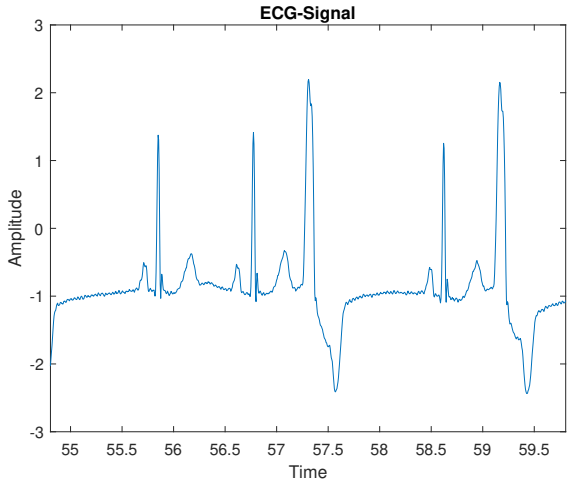

Fig. 12. ECG signal used for diagnosis

```
Average heart rate: 65.9/min
There seems to be Premature ventricular contractions, 2 records in all have been found
 at time=12.5s 29.0s >>
```

Fig. 13. Heart rate and diagnosis

V. Conclusion

In conclusion, our survey ECG analysis project using the PAN-Tompkins algorithm for the detection of arrhythmias has been a significant and valuable effort. We have learned and achieved quite a lot in this project.

First, we successfully applied the PAN-Tompkins algorithm to process ECG signals. This algorithm helps us accurately detect irregular heartbeats and arrhythmias, which is essential for diagnosing and treating heart problems. We also discussed the various phases of ECG signal analysis, including preprocessing, feature extraction, classification, and more. These steps are necessary to understand and interpret ECG data.

In addition, we looked at various databases and application areas related to ECG analysis. Having access to good databases and knowing where ECG analysis is used can be of great benefit to further research in this area.

In the future, we recommend that researchers continue to explore and improve ECG analysis techniques, especially for arrhythmia detection. Sharing databases and findings with the research community can help advance our understanding of heart disease and improve patient care.

In summary, our project using the PAN-Tompkins algorithm for ECG analysis has contributed to a broader effort to better understand and diagnose cardiac problems. We hope our work will help improve the lives of those affected by arrhythmias and other heart diseases.

VI. Individual Contribution

Rohit, Manish - Understanding and implementation of filters, differentiator, square and integrators.

Sujal, Bharath - Understanding Pan tompkins algorithm, thresholding, feature extraction

Shreyas, Amar - Disease classification and heart beat calculation.

At the end of this we all sat together and discussed and shared our understanding of the topics clarified doubts and finally implemented in MATLAB.

VII. Github Repo

The codes can be found in the Github Repository here : Github Repo Link

References

[1] MIT-BIH arrhythmia database
[2] Selcan Kaplan Berkayaa, Alper Kursat Uysal a, Efnan Sora Gunal b, Semih Erginc, Serkan Gunal a, M. Bilginer Gulmezoglu, A survey on ECG analysis
[3] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. 32 (1985) 230–236.
[4] F.A. Elhaj, N. Salim, A.R. Harris, T.T. Swee, T. Ahmed, Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals, Comput. Methods Programs Biomed. 127 (2016) 52–63.
[5] Mario Merone a, Paolo Soda a, Mario Sansone b, Carlo Sansone b, ECG databases for biometric systems: A systematic review .
[6] A. Alberdi, A. Aztiria, A. Basarab, Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review, J. Biomed. Inform. 59 (2016) 49–75.