

Heart Rate Feature Extraction based on Neurokit2 with Python

Le H. Nguyen, Tu H. Diep, Anh Q. Phan, Anh Q. Le, Quoc H. Le, and Arjon Turnip

Abstract—Cardiovascular disease (CVD) is caused by disorders of the heart and blood vessels. Cardiovascular disease includes coronary artery disease (myocardial infarction), cerebrovascular accident (stroke), hypertension (high blood pressure), peripheral artery disease, rheumatic heart disease, and congenital heart disease. heart failure. However, an estimated 80% of strokes are preventable, based on diet, exercise, and "listening" to your body's cues before a stroke has occurred. Up to now, heart disease is still a potential risk affecting the health and life of patients. We analyzed algorithms to filter ECG signals and gain feature extraction in order to process experiment data. The methods of Neurokit2 were proposed to analyze the sample signal and acquire details of feature extraction. The results show the numeric difference in 3 states: Relax, Walk and Run.

Index Terms— Cardiovascular disease (CVD), feature extraction, Neurokit2, Electrocardiograph (ECG).

I. INTRODUCTION

HEART rate monitoring is a crucial aspect of maintaining heart health. The maximum and minimum heart rate (number of beats per minute) are different between different age groups, the monitoring system, therefore, should be compatible enough to handle this scenario. The heart is one of the most important organs in the human body. It acts as a pump for circulating oxygen and blood throughout the body, thus keeping the functionality of the body intact. A heartbeat is produced due to the contraction of the heart and is defined as a two-part pumping action that takes about a second. When blood collects in upper chambers (the right and left atria), the SA (Sino Atrial) node sends out an electrical signal that causes the atria to contract. This contraction then pushes the blood through tricuspid and the mitral valves into the resting lower chambers (the right and left ventricles) [1][2]. This two-part pumping phase called diastole. The next phase begins when the ventricles are full of blood. The electrical signals generated from the SA

node reach the ventricle causing them to contract. This phase of the pumping system is called systole. To prevent the back flow of blood, the tricuspid and mitral valves are closed and the pulmonary and aortic valves are pushed open. When the blood moves into the pulmonary artery and aorta, the ventricles relax and the pulmonary and aortic valves close. Due to the lower pressure in the ventricles that cause the tricuspid and mitral valves to open, and another cycle start again.

Nowadays, many health problems are related to the heart. Especially, heart diseases are one of the most important causes of death among men and women. It has been shown that many heart problems can be identified through heart rate. Therefore, heart rate monitoring is crucial in the study of heart performance and thereby maintaining heart health [3]. In this paper, algorithms for heart rate feature extraction are proposed. The electrocardiogram (ECG) signals are measured by using the AD8232 ECG sensor.

An electrocardiogram records the electrical signal from your heart to check for different heart conditions. Electrodes are placed on your chest to record your heart's electrical signals, which cause your heart to beat. The signals are shown as waves on an attached computer monitor or printer. There are three main components to an ECG: the P wave, which represents the depolarization of the atria; the QRS complex, which represents the depolarization of the ventricles; and the T wave, which represents the repolarization of the ventricles. The main features of ECG signal are PSQRT peak, ST segment, QT segment, PR interval, and QT interval [5]. Digital filters are more accurate and precise than analog filters. Digital filters are Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) [8][9].

In this paper, FIR filter with Kaiser Window is used to remove the artifacts from ECG with less modification in the waveform and Neurokit2 [16] is then utilized to process feature extractions.

II. SIGNAL PROCESSING METHOD

A. Pre-filtering – Feature Extraction

The ECG signals are encountered by mainly four types of noises: baseline wander, powerline interference, EMG noise, and electrode motion artifacts. In order to remove these noises, digital filters are widely used because they are more accurate and precise than analog filters [8][9]. There are two fundamental types of digital filters: Finite Impulse Response (FIR) and Infinite Impulse Response (IIR). The FIR filters have the following advantages over IIR filters: 1- FIR filters do not have feedback, therefore they are always stable for any type of

Le Hoa Nguyen, Faculty of Advanced Science and Technology, The University of Danang - University of Science and Technology, Danang, Vietnam (nlhoa@dut.udn.vn)

Tu Hieu Diep, Faculty of Advanced Science and Technology, The University of Danang - University of Science and Technology, Danang, Vietnam (123170055@sv1.dut.udn.vn)

Anh Quan Phan, Faculty of Advanced Science and Technology, The University of Danang - University of Science and Technology, Danang, Vietnam, (123170067@sv1.dut.udn.vn)

Anh Quan Le, Faculty of Advanced Science and Technology, The University of Danang - University of Science and Technology, Danang, Vietnam (123170066@sv1.dut.udn.vn)

Quoc Huy Le, Faculty of Advanced Science and Technology, The University of Danang - University of Science and Technology, Danang, Vietnam, (lqhuy@dut.udn.vn)

Arjon Turnip, Padjadjaran University, Bandung, Indonesia, (turnip@unpad.ac.id*)

input signal; 2- FIR filters can be easily designed to have a linear phase; 3- FIR filters can be customized easily due to its arbitrary frequency response. Due to these advantages, FIR filters are widely deployed in biomedical signal enhancement applications.

Baseline wander is a low-frequency artifact in ECG that is caused by improper electrodes (poor contact), body movement, and respiration. The occurrence of baseline wander may interfere with signal analysis, rendering the clinical interpretation inaccurate and misleading. The frequency content of baseline wander is usually less than 1 Hz, but it may contain higher frequencies during strenuous exercise [10].

The powerline interference represents a common noise source in the ECG and other physiologic signals recorded from the body surface. Depending on the country region, such noise is characterized by a 50 or 60 Hz sinusoidal interference, possibly accompanied by harmonics [11].

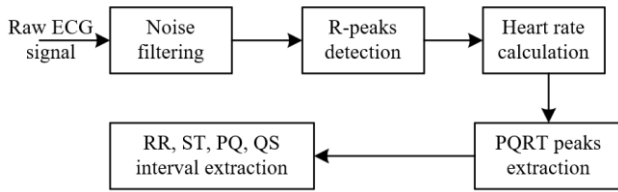


Fig. 1. Diagram of ECG signal processing

B. Using FIR Low Pass Filter to Denoising

A low-pass filter is a filter that allows signals with low frequencies (lower than cutoff frequency) to pass through and attenuates signals with high frequencies (higher than the cutoff frequency). The exact frequency response of the filter depends on the filter design. The filter is sometimes called a high-cut filter, or treble-cut filter in audio applications.

Equations use to build a FIR low pass Filter in direct form realization [12-13].

$$H_z(\omega) = \begin{cases} 1 & |\omega| < \omega_c \\ 0 & |\omega| > \omega_c \end{cases} \quad (1)$$

where ω_c is the cutoff frequency.

We want to design low pass filter with a cutoff frequency of ω_c the desired frequency response will be:

$$H(z) = \sum_{k=0}^M b_k z^k \quad (2)$$

To find the equivalent time-domain representation, we calculate the inverse discrete-time Fourier transform.

$$H_d[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(\omega) e^{j\omega n} d\omega \quad (3)$$

Substitute Eq. (2) into Eq. (3), we obtain the impulse response of an ideal lowpass filter.

$$H_d[n] = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega = \frac{\sin(n\omega_c)}{n\pi} \quad (4)$$

C. Proposed Algorithm for ECG signal After Filtering R Peak Features Extraction

NeuroKit2 is an open-source, community-driven, and user-centered Python package providing easy access to advanced neurophysiological signal processing routines. These processing routines include high-level functions that enable data processing in a few lines of code using validated pipelines, which we illustrate in two examples covering the most typical scenarios, such as an event-related paradigm and an interval-

related analysis. The package also includes tools for specific processing steps such as rate extraction and filtering methods, offering a trade-off between high-level convenience and fine-tuned control. Its goal is to improve transparency and reproducibility in neurophysiological research, as well as foster exploration and innovation. Its design philosophy is centered on user experience and accessibility to both novice and advanced users [16].

III. EXPERIMENT

In this study, ECG data were recorded on 10 people in 3 different states: resting, walking, and running. The AD8232 ECG sensor was attached to the Raspberry Pi device with a battery.

All samples are recorded in both normal and abnormal conditions with 3 of these above states. All the recorded data will save on the SD card and we will use them to test with our model code in order to find out the different ECG feature extraction between each person.



Fig. 2. Students at Universitas Padjadjaran was taking samples in 3 different states

IV. RESULTS AND DISCUSSIONS

In order to choose a suitable algorithm to detect R-peaks, we calculated errors like: out of bounds for the axis with different indexes as the results are shown in Fig. 3. We also calculated the execution time and accuracy between the neurokit library [16] and 9 popular algorithms were published by the authors Pan Tompkins [17], Hamilton [18], Martinez [19], Christov [20], Gamboa [21], Elgendi [22], Engzeemod [23], Kalidas [24], and Rodrigues [25]. The obtained results are depicted in Fig. 4 and Fig. 5. From Fig. 3, it seems that gamboa2008 and martinez2003 are particularly prone to errors, especially in the case of a noisy ECG signal. Aside from that, the other algorithms are quite resistant and bug-free. Fig. 4 shows that gamboa2008 and neurokit are the fastest methods, followed by martinez2003, kalidas2017, rodrigues2020, and hamilton2002. The other methods are then substantially slower. As shown in Fig. 5, the accuracy is computed as the absolute distance from the original “true” R-peaks location. As such, the closer to zero,

the better the accuracy, therefore it seems that neurokit, kalidas2017 and martinez2003 are the most accurate algorithms to detect R-peaks.

From the calculation results, we found that the neurokit library has the best results. Therefore, in this article, we will continue to develop the neurokit library to process ECG signals.

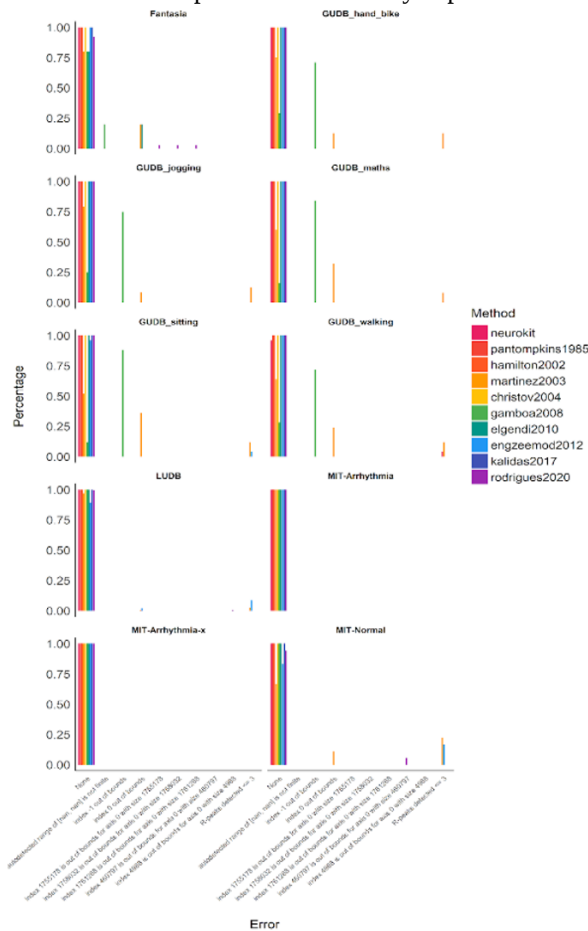


Fig. 3. Errors between 10 algorithms [16]

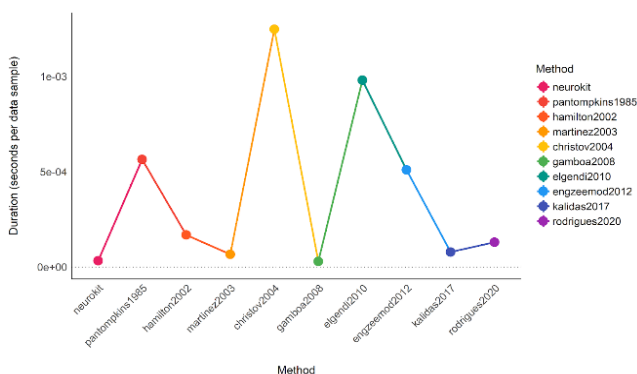


Fig. 4. Execution time between 10 algorithms [16]

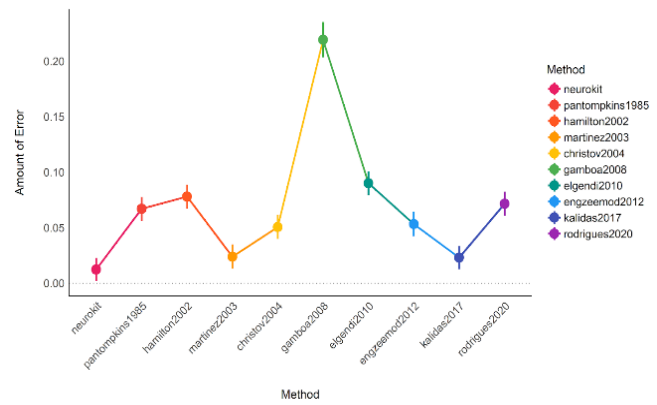


Fig. 5. Accuracy between 10 algorithms [16]

To filter raw ECG signals for feature extraction, we used a low pass FIR filter as mentioned in Section II, and the obtained result of object 1 in resting state is shown in Fig. 6. It is clear from Fig. 6 that, before filtering ECG signal is interfered by baseline wander and powerline interference. After filtering, the noise had been removed and the signal becomes smoothy pointing out details of the ECG beat signal.

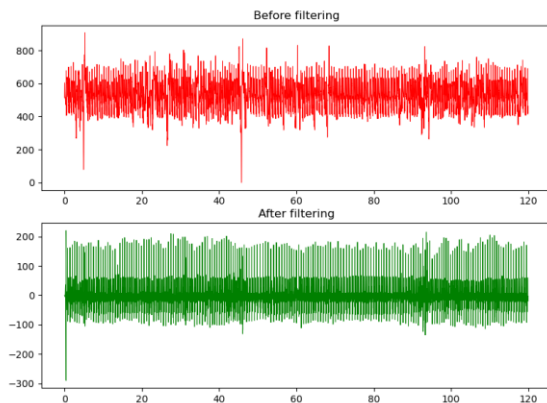


Fig. 6. ECG signal before and after filtering

After filtering ECG signals, we continued to detect R peaks based on the Neurokit library and calculate the average heart rate. The obtained result is shown in Fig. 7. In this figure, it shows that the Neurokit library [16] detected almost R-peak in filtering ECG signals.

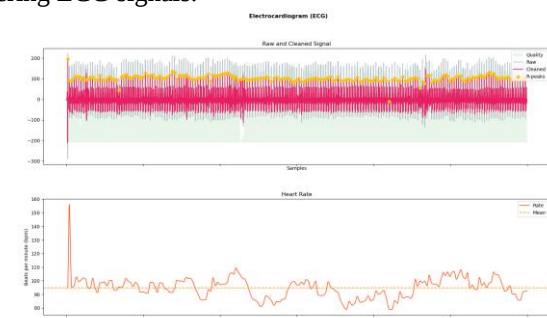


Fig. 7. Detect R peak and Heart Rate after filtering (Object1)

After detecting R peaks and average heart rate, we continued to detect the other peaks of ECG signals (P, Q, S, T) as shown in Fig. 8. The details of the P, Q, S, and T peaks were very precise. The left of Fig. 8 shows the full detection of the signal

and the right of Fig. 8 zooms into the first 3 beats of detection.

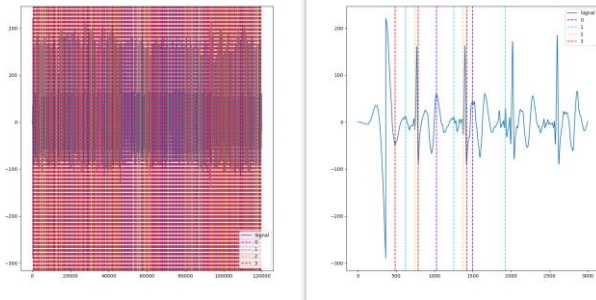


Fig. 8. Detect P Q S T peak of ECG signal after filtering

After we detected all peaks of the ECG signal, we continued to detect R-R and S-T. P-Q. Q-S interval in 3 different states (relaxing, walking, running). The obtained result is shown in Fig. 9. It shows that in a relaxing state the R-R, S-T, P-Q, and Q-S have the highest amplitude. Meanwhile, the running state has the lowest amplitude.

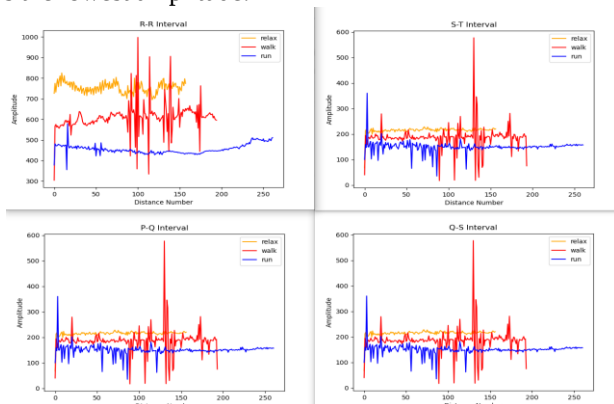


Fig. 9. Detect R-R, S-T, P-Q, and Q-S interval of ECG signal after filtering in 3 different states

We continued to run the rest of the real data sample in a relaxing state. The result is obtained in Table 1. In this table the distance of the average R-R interval from Table 1 is between 558.89 ms and 804.43 ms and the standard deviation is between 26.05 and 169.25. The S-T interval is between 159.76 and 216.38 and the standard deviation is between 12.83 and 96.36. The P-Q interval is between 76.38 ms and 207.46 ms and the standard deviation is between 33.15 and 149.78. The Q-S interval is between 37.68 ms and 102.80 ms and the standard deviation is between 3.39 and 31.43.

TABLE I
ECG FEATURES EXTRACTION FOR RELAXING SAMPLES

Relaxing								
RR			ST		PQ		QS	
	Avg Distance (ms)	SD	Avg Distance (ms)	SD	Avg Distance (ms)	SD	Avg Distance (ms)	SD
1	753.87	26.05	214.47	12.83	76.38	41.47	51.57	18.27
2	855.33	57.33	218.97	30.18	123.56	80.44	37.68	6.01

3	723.75	82.46	203.82	10.06	207.46	116.09	45.71	3.62
4	670.70	44.72	189.83	21.93	107.37	79.27	102.80	24.94
5	633.87	49.33	207.96	67.72	150.36	67.66	50.51	18.73
6	634.89	25.91	186.41	34.12	90.50	33.15	50.33	30.13
7	711.40	169.25	194.73	96.36	121.22	116.29	82.62	31.43
8	804.84	144.48	159.76	28.35	159.72	149.78	98.11	18.51
9	558.89	52.07	164.85	70.70	86.02	83.65	79.81	30.64
10	804.43	53.56	216.38	16.42	136.87	61.12	40.79	3.39
av								
g	715.20	70.52	195.72	38.87	125.95	82.89	63.99	18.57

We continue to run the rest of the real data sample in a walking state. The result obtained is shown in Table 2. In the walking state, the distance of the average R-R interval from table 4.2 is between 500.27 and 737.83 and the standard deviation is between 3.39 and 106.76. The S-T interval is between 121.35 and 224.81 and the standard deviation is between 3.39 and 106.76. The P-Q interval is between 80.95 and 157.04 and the standard deviation is between 31.91 and 119.81. The Q-S interval is between 41.88 and 99.55. The standard deviation is between 10.56 and 46.30.

TABLE II
ECG FEATURES EXTRACTION FOR WALKING SAMPLES

Walking								
RR			ST		PQ		QS	
	Avg Distance (ms)	SD	Avg Distance(m s)	SD	Avg Distance (ms)	SD	Avg Distance (ms)	SD
1	608.70	72.42	185.64	48.17	96.78	101.18	58.59	25.14
2	599.21	65.56	162.41	52.31	125.27	81.91	48.24	25.81
3	624.81	106.76	142.06	58.33	157.04	111.60	61.14	24.69
4	580.45	25.65	177.25	36.70	99.46	49.13	99.55	28.30
5	500.27	19.93	124.22	64.92	140.22	77.82	68.91	32.12
6	505.43	21.21	121.35	64.54	80.95	31.91	43.38	20.89
7	737.83	14.62	224.81	23.06	132.08	119.81	71.60	46.30
8	622.85	40.45	132.05	13.24	98.80	42.03	83.72	15.52
9	454.37	43.53	150.71	73.38	101.50	72.08	71.72	31.02
10	589.24	40.51	169.82	20.37	155.47	65.60	41.88	10.56
Avg	582.32	45.06	159.03	45.50	118.76	75.31	64.87	26.04

We continue to run the rest of the real data sample in a running state. The result is obtained in Table 3. In the running state, the distance of the average R-R interval from table 3 is

between 455.30 ms and 621.98 ms and the standard deviation is between 14.80 and 354.64. The S-T interval is between 103.41 ms and 174.14 ms and the standard deviation is between 17.62 and 66.63. The P-Q interval is between 80.95 ms and 166.23 ms and the standard deviation is between 27.04 and 192.27. The Q-S interval is between 39.83 ms and 97.94 ms and the standard deviation is between 8.73 and 30.37.

TABLE III
ECG FEATURES EXTRACTION FOR RUNNING SAMPLES

Running								
RR		ST		PQ		QS		
Avg Distance (ms)	SD	Avg Distance (ms)	SD	Avg Distance (ms)	SD	Avg Distance (ms)	SD	
1	455.30	23.10	149.83	20.83	80.95	68.18	48.52	19.03
2	508.72	40.25	150.36	35.11	109.11	54.11	39.83	13.67
3	481.69	58.90	129.81	36.18	122.18	64.99	48.96	15.97
4	557.75	26.46	174.14	17.62	104.30	44.00	97.94	8.73
5	470.12	14.80	142.75	66.63	136.37	69.02	57.31	30.37
6	621.98	354.64	109.46	43.07	82.18	39.66	47.58	21.54
7	581.13	210.71	156.66	65.10	166.23	192.27	59.99	31.46
8	470.02	41.83	103.41	29.76	90.14	27.04	76.86	23.99
9	482.57	236.01	116.70	62.16	109.32	78.76	65.31	28.87
Avg								
g	514.36	111.86	137.01	41.83	111.20	70.89	60.26	21.51

V. CONCLUSION

In this study, we proposed to use FIR low pass filter to denoise raw ECG signals and utilized Neurokit 2 algorithms for feature extraction. The experiment was performed on 10 people in 3 different states (resting, walking, and running). The average values of all subjects show the difference between the amplitude of peaks, intervals, and standard deviation very clearly. This experiment helps us to propose that Neurokit2 algorithms work more effectively than the others. Average numerical results of the resting state demonstrated that the RR, ST, PQ, and QS of the relaxing state are 715.20ms, 195.72ms, 125.95ms, and 63.99ms, respectively. The corresponding SD numbers are 70.52, 38.87, 82.89, and 18.57 respectively. For the walking state, the average values of RR, ST, PQ, and QS are 582.32ms, 159.03ms, 118.76ms, and 64.87ms, respectively. The corresponding SD numbers are 45.06, 45.50, 75.31, and 26.04, respectively. During the running state, the average values of RR, ST, PQ, and QS are 514.36ms, 137.01ms, 111.20ms, and 60.26ms, respectively. The corresponding SD numbers are 111.86ms, 41.83ms, 60.26ms, and 21.51ms, respectively.

ACKNOWLEDGMENT

This research is mainly funded by the “Program Penelitian Kolaborasi Indonesia” research program run by the Indonesian Ministry of Education, Culture, Research, and Technology and supported by Universitas Padjadjaran, Indonesia and The University of Danang - University of Science and Technology.

REFERENCES

- [1] F. R. Pebriansyah, P. Turnip, N. S. Syafei, A. Trisanto, and A. Turnip, “Design of Arrhythmia Early Detection Interface Using Laravel Framework”. In *2021 International Conference on Artificial Intelligence and Mechatronics Systems*, IEEE, pp. 1-6, 2021.
- [2] C. N. Ginting, I. N. E. Lister, M. Silaen, E. Girsang, Y. Laila, M. Tumip, and A. Turnip, “Fetal Heart Detection Based Wide Area Network Technology With Wireless Sensor Transmission,” In *Journal of Physics: Conference Series*, IOP Publishing, vol. 1230, no. 1, pp. 012037, 2019.
- [3] D. Balakrishnan, T. Dhiliphan Rajkumart, and S. Dhanasekaran, “An Intelligent And Secured Heart Rate Monitoring System Using IOT.”
- [4] <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [5] S. Kumar, “Efficient Heart Rate Monitoring,” Applications Engineer, Cypress Semiconductor Corp.
- [6] K. S. Rathore, G. Amritha, G. Kanimozhi, “IoT based system for Heart Rate Monitoring and Heart Attack Detection,” TV Sethuraman,
- [7] A Survey on Tracing Heart Attacks by Pulse Monitoring in IoT, article: C. Santhanakrishnan et al 2019 J. Phys.
- [8] A. K. Wadhvani, and M. Yadav, “Filtration of ECG signal By Using Various Filter” *International Journal of Modern Engineering Research (IJMER)*, vol.1, Issue2, pp. 658-661 ISSN: 2249-6645.
- [9] A. S. Khaing, and Z. M. Naing, “Quantitative Investigation of Digital Filters in Electrocardiogram with Simulated Noises,” *International Journal of Information and Electronics Engineering*, vol. 1, no. 3, pp. 210-216, 2011.
- [10] L. Sömmo, and P. Laguna, in *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, 2005.
- [11] J. P. do Vale Madeiro, Priscila Rocha Ferreira Rodrigues, in *Developments and Applications for ECG Signal Processing*, 2019.
- [12] FIR Filter Design by Windowing: Concepts and the Rectangular Window - Technical Articles (allaboutcircuits.com)
- [13] W. Caesarendra, M. Ariyanto, K. A. Pambudi, M. F. Amri, and A. Turnip, “Classification of EEG signals for eye focuses using artificial neural network,” *Internetworking Indones J*, vol. 9 no. 1, pp. 15-20, 2011.
- [14] Defi, R. Irma, I. Shelly, C. Septiana, T. Arjon, and N. Dessy, “Healthcare Workers’ Point of View on Medical Robotics During COVID-19 Pandemic—A Scoping Review.” *International Journal of General Medicine*, vol. 15, 2022.
- [15] M. S. Chavan, R. A. Agrawala, M. D. Uplane, “Digital Elliptic Filter Application for Noise Reduction in ECG Signal” *4th Wseas International Conference on Electronics, Control & Signal Processing* Miami Florida USA, vol. 17, no.19, pp. 58-63, 2005.
- [16] NeuroKit2: A Python toolbox for neurophysiological signal processing <https://neurokit2.readthedocs.io/en/latest/>
- [17] J. Pan, and W. J. A. Tompkins, “Real-time QRS detection algorithm,” *IEEE transactions on biomedical engineering*, vol. 3, pp. 230-236, 1985.
- [18] P. Hamilton, “Open source ECG analysis,” In *Computers in cardiology*, pp. 101-104, IEEE, 2002.
- [19] W. Zong, T. Heldt, G. B. Moody, and R. G. Mark, “An open-source algorithm to detect onset of arterial blood pressure pulses”. In *Computers in Cardiology*, pp. 259-262, IEEE, 2003.
- [20] Ivaylo Christov Bulgarian Academy of Sciences, “Real time electrocardiogram QRS detection using combined adaptive threshold, 2004.
- [21] H. Gamboa, “Multi-modal behavioral biometrics based on hci and electrophysiology,” PhD Thesis Universidade, 2008.
- [22] W. A. H. Engelse, and C. Zeelenberg, “A single scan algorithm for QRS detection and feature extraction *IEEE Comput Cardiol*”. Long Beach: IEEE Computer Society, 1979.
- [23] A. Lourenço, H. Silva, P. Leite, R. Lourenço, and A. L. Fred, “Real Time Electrocardiogram Segmentation for Finger based ECG Biometrics”. In *Biosignals*, pp. 49-54, 2012.

- [24] M. Nabian, Y. Yin, J. Wormwood, K. S. Quigley, L. F. Barrett, and S. Ostadabbas, "An Open-Source Feature Extraction Tool for the Analysis of Peripheral Physiological Data," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1-11, 2018. doi:10.1109/jtehm.2018.2878000.
- [25] Rodrigues, Tiago and Samoutphonh, Sirisack and Plácido da Silva, Hugo and Fred, Ana. "A Low-Complexity R-peak Detection Algorithm with Adaptive Thresholding for Wearable Devices", 2021.