

## Heart Disease Detection Using ECG Lead I and Multiple Pattern Recognition Classifiers

Renato M. Pereira<sup>1</sup>, Bruno C. Bispo<sup>2</sup> and Pedro M. Rodrigues<sup>1,\*</sup>

<sup>1</sup>Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Rua Diogo Botelho 1327, 4169-005, Porto, Portugal, (\*) corresponding author's email: prodrigues@porto.ucp.pt

<sup>2</sup>Department of Electrical and Electronic Engineering, Federal University of Santa Catarina, 88040-370, Florianópolis, SC, Brazil.

Received 18 March 2020; Accepted 04 April 2020

**Abstract:** ECG is an important tool to assist in heart diseases diagnosis. The works found in the literature have the common goal of discriminating between binary study groups, one pathological and one control, even when ECG records from patients diagnosed with several pathologies are available in the databases. This work proposes a method to detect ECG morphological features and to analyze the capacity of this ECG features to discriminate 28 pairs of study groups, combining 7 pathological groups and 1 control group, presented in the PTB Diagnostic ECG Database. For each pair, it was achieved an accuracy between 77.4% and 100%, with an average of 94%, using several pattern recognition classifiers.

**Key Word:** Heart diseases; ECG features; Pattern recognition; PTB Diagnostic ECG Database; Classifiers.

### I. INTRODUCTION

The electrocardiogram (ECG) is the recording of the rhythmic alterations of the heart electrical activity and represents the cardiac cycle [1]. A typical ECG is usually recorded by means of a 12-lead system (i, ii, iii, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6). The amplitude and direction of the current flow in the heart are detected by the electrodes, resulting in different ECG signals according to the leads axis. An ECG signal of a healthy subject is cyclically formed by a P wave, a QRS complex and a T wave [1], which represent the atria depolarization, the ventricular depolarization and the ventricular repolarization, respectively [2]. Other important time intervals and signal segments are also described in Figure no 1. The time between the beginnings of the P wave and the QRS complex is the PQ interval, which is often called PR interval because the Q wave is usually very small [3]. During the PR interval, which lasts approximately 0.16 seconds, the auricle contracts and begins to relax [4]. The QT interval extends from the beginning of the QRS complex to the end of the T wave, lasting approximately 0.36 seconds, and represents the approximate duration required for the ventricles to contract and relax [5].

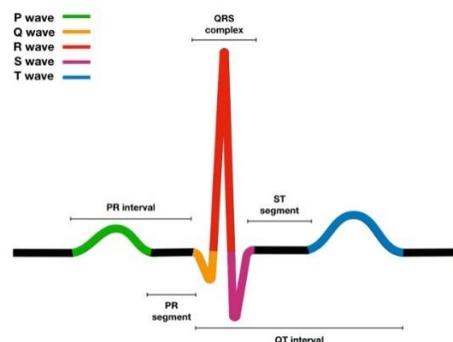


Figure no 1: ECG wave and its morphological features.

Any minor change in the normal pattern of an ECG signal can be interpreted as malfunction of the heart [6,7]. Thus, autonomous and accurate discrimination of cardiac pathologies through ECG is an important tool to assist in the diagnosis of these diseases, especially considering that the detection of cardiac disorder is an exhausting task for cardiologists [8]. During the last years, several works have proposed methods to detect ECG features (morphological

or not) and then to diagnose cardiac pathologies. These works can be classified according to: pathologies diagnosed, ECG database used, number of ECG leads used, ECG analysis method and classification method. Considering only those that use public databases, some works found in the literature and their results are summarized in the Table no 1. The great majority of these methods aimed to discriminate Myocardial Infarction (MI) from healthy controls, which can be explained by the greater number of ECG records from patients diagnosed with this disease in the public databases, mainly in the PTB Diagnostic ECG Database. Few works aimed to discriminate Dysrhythmia or Cardiomyopathy from healthy controls. However, all works have the common goal of discriminating between 2 study groups, one pathological group and one control group, even when ECG records from patients diagnosed with several pathologies are available in the databases. This work proposes to analyze the capacity of several ECG parameters to discriminate 28 pairs of study groups, combining 7 pathological groups and 1 control group, presented in the PTB Diagnostic ECG Database.

**Table no 1: List of works found in the literature.**

References	Pathologies	Number of Leads	Method and Classification	Accuracies
[9]	MI	12 leads	ST segment elevation and threshold classification	92.5%
[10]	MI	3 leads	Q peak depth and ST segment elevation. Classification by a simple adaptive threshold	90.56%
[11]	Dysrhythmia	12 leads	Template construction from CWT features using a morphological consistency classifier	93.0%
[12]	Cardiomyopathy	12 leads	PR, RR, QT and QRS intervals analysis. Classification through BPNN	95.2%

## II. MATERIAL AND METHODS

### ECG Database

This work used the PTB Diagnostic ECG Database available in [13, 14]. The database contains 549 ECG records from 268 subjects, including healthy subjects. Table no 2 summarizes de database. Each ECG record contains all the 12-lead system signals with a sampling rate of 1000 Hz, but this work used only the Lead I signals.

**Table no 2: Diagnostic classes and number of records found in PTB database.**

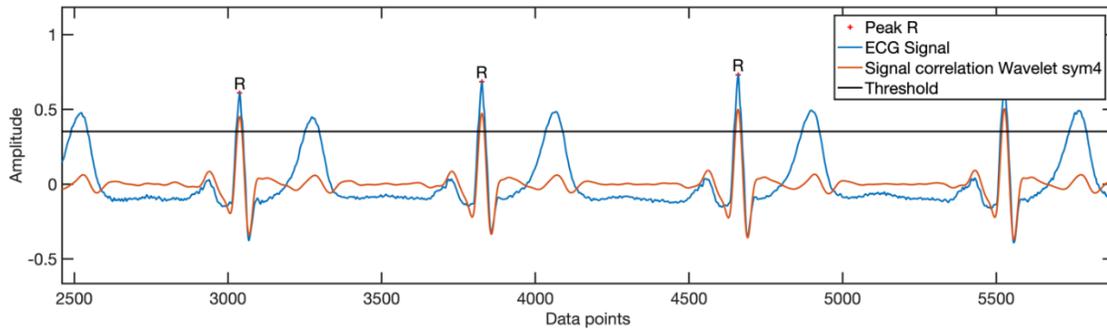
Pathologies	Number of patients
MI	148
Cardiomyopathy	18
Bundle Branch Block	15
Dysrhythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy controls	52

### Peak detection

The method detects the R, S, Q, P and T peaks of the ECG signal, in that order, as follows:

**Peak R:** A Wavelet Transform (WT) translation analysis, using the wavelet 'symlet 4'(orange in Figure no 2), is applied in order to calculate the cross-correlation between the signal and the WT. The R peaks are the maxima of each correlation over the channel. To find all the R peaks in the ECG a 70% signal amplitude threshold (black line in Figure

no 2) of the WT's maximum was used to make sure that just the most prominent correlation peak in each heart cycle is detected.

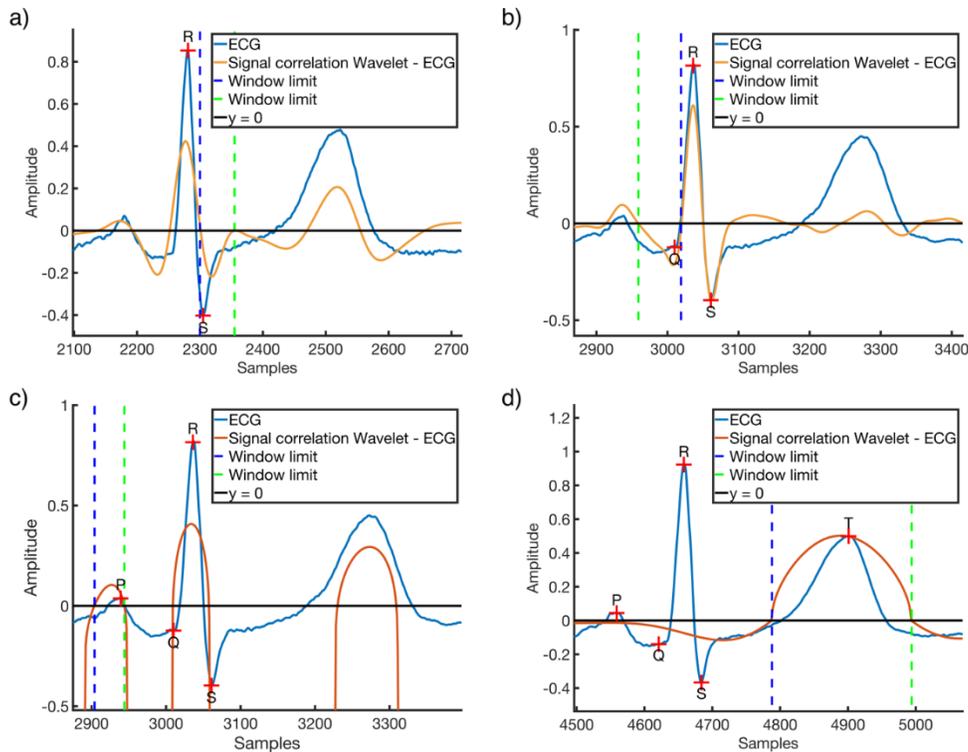


**Figure no 2: ECG R peak detection.**

**S peak:** The S peak was identified as being the first negative minimum wave after the first zero of WT ‘symlet 4’ translation after the R peak (Figure no 3 a)) [15].

**Q peak:** Contrary to S peak, the Q peak was identified as being the negative minimum wave right before the first zero before of WT ‘symlet 4’ translation before the R peak (Figure no 3b)) [15].

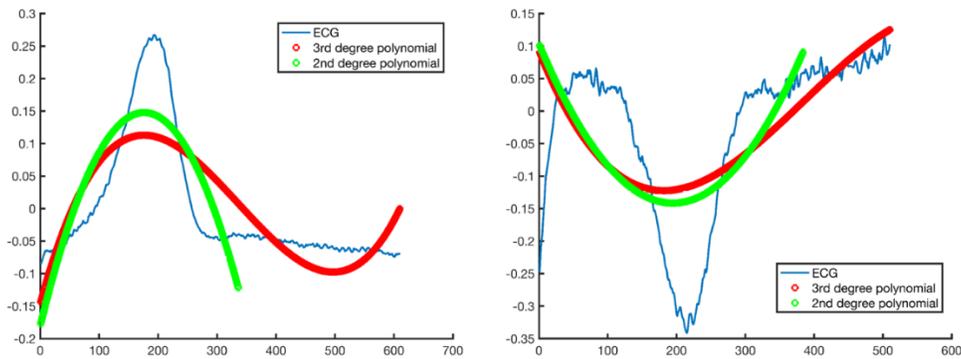
**P peak:** WT translation was performed using WT ‘symlet 4’. The WT was then amplified fifth rooting it. The first three zeros of the WT wave before each R peak were calculated. The last two zeros were used as windows to compute the maximum of the ECG signal, corresponding to the P peak (Figure 3 c)).



**Figure no 3: ECG a) S peak, b) Q peak, c) P peak and d) T peak detection.**

**T peak:** For the T wave analysis the ECG signal was set to zero before and after two consecutive R peaks, knowing T waves are comprehended in this interval. To verify if the T wave was in an inverting or non-inverting state a

3<sup>rd</sup> degree polynomial approximation was applied to each SP interval. Computing the inflection points and applying again a 2<sup>nd</sup> degree polynomial approximation from the beginning of the SP interval to the first inflection point, the coefficient signal dictates if positive the T wave is inverted and if negative, non-inverted (Figure no 4). If non-inverted the WT maintains the same and if inverted the WT was inverted. The T peaks are the maxima of each correlation over the channel. To find the T peak in each RR window of the ECG a 99% signal amplitude threshold was used to make sure that just the most prominent correlation peak is detected. If the maximum of WT being offset from T peak, the first zero, before and after the WT maximum, were used as windows to compute the maximum of the ECG signal, corresponding to the T peak (Figure no 3 d)).



**Figure no 4: T wave state (inverted/non-inverted) detection.**

After all peak's identification, it is easy to find the segments that represent the QRS complex, PR interval, PR segment, QT interval and ST segment were followed identified.

**ECG Features Analyzed**

Several features of each Lead I ECG signal are calculated in order to analyze their discrimination capacities. The features are summarized in the Table no 3.

**Table no 3: List of analyzed features and their respective index.**

Index	Features	Index	Features	Index	Features
1	Energy QRS complex	38	WT 'sym8' det. level 3 energy QT interval	75	Time P-T
2	Power QRS complex	39	WT 'sym8' det. level 2 energy QT interval	76	Time Q-R
3	Entropy QRS complex	40	WT 'sym8' det. level 1 energy QT interval	77	Time Q-S
4	Shannon Entropy QRS complex	41	Duration P wave	78	Time Q-T
5	Log Energy Entropy QRS complex	42	Energy P wave	79	Time R-S
6	WT 'sym4' det. level 4 energy QRS complex	43	Power P wave	80	Time R-T
7	WT 'sym4' det. level 3 energy QRS complex	44	Entropy P wave	81	Time S-T
8	WT 'sym4' det. level 2 energy QRS complex	45	Shannon Entropy P wave	82	Amplitude P peak
9	WT 'sym4' det. level 1 energy QRS complex	46	Log Energy Entropy P wave	83	Amplitude Q peak
10	WT 'sym8' det. level 4 energy QRS complex	47	WT 'sym4' det. level 4 energy P wave	84	Amplitude R peak
11	WT 'sym8' det. level 3 energy QRS complex	48	WT 'sym4' det. level 3 energy P wave	85	Amplitude S peak

12	WT 'sym8' det. level 2 energy QRS complex	49	WT 'sym4' det. level 2 energy P wave	86	Amplitude T peak
13	WT 'sym8' det. level 1 energy QRS complex	50	WT 'sym4' det. level 1 energy P wave	87	Amplitude difference P-Q
14	Energy PR interval	51	WT 'sym8' det. level 4 energy P wave	88	Amplitude difference P-R
15	Power PR interval	52	WT 'sym8' det. level 3 energy P wave	89	Amplitude difference P-S
16	Entropy PR interval	53	WT 'sym8' det. level 2 energy P wave	90	Amplitude difference P-T
17	Shannon Entropy PR interval	54	WT 'sym8' det. level 1 energy P wave	91	Amplitude difference Q-R
18	Log Energy Entropy PR interval	55	Duration T wave	92	Amplitude difference Q-S
19	WT 'sym4' det. level 4 energy PR interval	56	Energy T wave	93	Amplitude difference Q-T
20	WT 'sym4' det. level 3 energy PR interval	57	Power T wave	94	Amplitude difference R-S
21	WT 'sym4' det. level 2 energy PR interval	58	Entropy T wave	95	Amplitude difference R-T
22	WT 'sym4' det. level 1 energy PR interval	59	Shannon Entropy T wave	96	Amplitude difference S-T
23	WT 'sym8' det. level 4 energy PR interval	60	Log Energy Entropy T wave	98	Energy ECG
24	WT 'sym8' det. level 3 energy PR interval	61	WT 'sym4' det. level 4 energy T wave	99	Power ECG
25	WT 'sym8' det. level 2 energy PR interval	62	WT 'sym4' det. level 3 energy T wave	100	Entropy ECG
26	WT 'sym8' det. level 1 energy PR interval	63	WT 'sym4' det. level 2 energy T wave	101	Shannon Entropy ECG
27	Energy QT interval	64	WT 'sym4' det. level 1 energy T wave	102	Log Energy Entropy ECG
28	Power QT interval	65	WT 'sym8' det. level 4 energy T wave	103	WT 'sym4' det. level 4 energy ECG
29	Entropy QT interval	66	WT 'sym8' det. level 3 energy T wave	104	WT 'sym4' det. level 3 energy ECG
30	Shannon Entropy QT interval	67	WT 'sym8' det. level 2 energy T wave	105	WT 'sym4' det. level 2 energy ECG
31	Log Energy Entropy QRS complex	68	WT 'sym8' det. level 1 energy T wave	106	WT 'sym4' det. level 1 energy ECG
32	WT 'sym4' det. level 4 energy QT interval	69	Duration PR segment	107	WT 'sym8' det. level 4 energy ECG
34	WT 'sym4' det. level 3 energy QT interval	70	Duration ST segment	108	WT 'sym8' det. level 3 energy ECG
35	WT 'sym4' det. level 2 energy QT interval	71	Time P-Q	109	WT 'sym8' det. level 2 energy ECG

36	WT 'sym4' det. level 1 energy QT interval	73	Time P-R	110	WT 'sym8' det. level 1 energy ECG
37	WT 'sym8' det. level 4 energy QT interval	74	Time P-S		

**Classification**

A sequential feature selection algorithm with deviance of the fit (generalization of the residual sum of squares) criterion were applied for feature selection from matrices with 110 features per X subjects (X is the number of the patients involved in each classification pair). The 110 features resulted from the mean features values extracted from each ECG cycle per subject (I lead), as described in Table no 3. A 5-fold cross validation was used for training and testing several different machine learning classifiers presented in Table no 4. The classifications were performed for 28 pairs of the study groups between 7 pathological groups and 1 control group.

**Table no 4: List of used classifiers.**

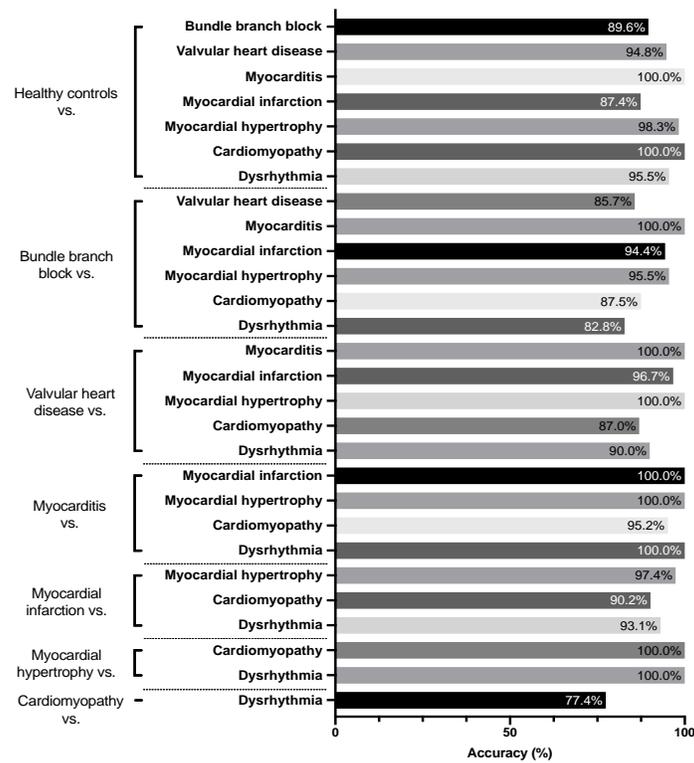
<b>Decision Trees</b>	<b>Support Vector Machines</b>	<b>Nearest Neighbor Classifiers</b>	<b>Ensemble Classifiers</b>	<b>Discriminant Analysis</b>	<b>Logistic Regression Classifiers</b>
Medium Tree	Linear SVM	Fine KNN	Boosted Trees	Linear	Logistic
Coarse Tree	Quadratic SVM	Medium KNN	Bagged Trees	Discriminant	Regression
	Cubic SVM	Coarse KNN	Subspace	Quadratic	
	Fine Gaussian SVM	Cosine KNN	Discriminant	Discriminant	
	Medium Gaussian SVM	Cubic KNN	Subspace KNN		
	Coarse Gaussian SVM	Weighted KNN	RUSBoosted Trees		

**III.RESULTS**

As previously said, the classifications were performed for 28 pairs of the study groups combining 7 pathological groups and 1 control group. The results shown in Figure are the best accuracies achieved from the trained/tested classifiers.

As can be observed from Figure no 5, the maximum accuracy classification was achieved for the pairs Healthy controls vs. Myocarditis and Cardiomyopathy; Bundle branch block vs. Myocarditis; Valvular heart disease vs. Myocarditis and Myocardial hypertrophy; Myocarditis vs. MI, Myocardial hypertrophy and Dysrhythmia; Myocardial hypertrophy vs. Cardiomyopathy and Dysrhythmia; with an outstanding 100% precision rate. Moreover, it can be noticed that the best features were capable of discriminating Healthy controls from any other heart disease with an accuracy higher than 95% for the exception of Bundle branch block and MI where the reached accuracies were 89.6% and 87.4%, respectively. For the pairs Healthy controls vs. Dysrhythmia the classifiers achieved an accuracy of 95.5%. The distinguish accuracy between Healthy controls against MI is slightly under the results of those in the state-of-art (Table no 1) and can be explained by the fact that the methods available in the literature use multiple leads and, as previously mentioned, not just one lead as this work.

Using only Lead I analysis, the classifiers were able to achieve an accuracy between 77.4% and 100%, with an average of 94%, for the 28 pairs of study groups showing Lead I has good capacity for heart pathologies discrimination, however, the low number of records for some pathologies should be taken in consideration.



**Figure no 5: Pathologies discrimination accuracies.**

#### IV. CONCLUSION

This work has analyzed the capacity of several ECG features to discriminate 28 pairs of study groups, combining 7 pathological groups and 1 control group, presented in the PTB Diagnostic ECG Database. Using only Lead I, the classifiers were able to achieve an accuracy between 77.4% and 100%, with an average of 94%, for the 28 pairs of study groups. These results become even more relevant considering that only 3 of these pairs are commonly analyzed in the literature: MI, Dysrhythmia and Cardiomyopathy. This study also proves that Lead I has good capacity for heart pathologies discrimination, however the low number of records for some pathologies should be taken inconsideration.

#### ACKNOWLEDGMENTS

This work was supported by National Funds from FCT - Fundação para a Ciência e a Tecnologia through project UIDB/50016/2020.

#### REFERENCES

- [1]. J. S. Barlow. The electroencephalogram: its patterns and origins. MIT Press, 1993.
- [2]. M. B. Conover, Understanding electrocardiography: physiological and interpretive concepts, 3rd ed. Mosby, 1980.
- [3]. R. R. Seeley, P. Tate, and T. D. Stephens, Anatomia & Fisiologia. McGraw-Hill Higher Education, 2003
- [4]. B.-U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," IEEE Engineering in Medicine and Biology Magazine, vol. 21, no. 1, pp. 42–57, January 2002.
- [5]. S. Sharma, J. A. Drezner, A. Baggish, M. Papadakis, M. G. Wilson, J. M. Prutkin, A. L. Gerche, M. J. Ackerman, M. Borjesson, J. C. Salerno, I. M. Asif, D. S. Owens, E. H. Chung, M. S. Emery, V. F. Froelicher, H. Heidbuchel, C. Adamuz, C. A. Asplund, G. Cohen, K. G. Harmon, J. C. Marek, S. Molossi, J. Niebauer, H. F. Pelto, M. V. Perez, N. R. Riding, T. Saarel, C. M. Schmied, D. M. Shipon, R. Stein, V. L. Vetter, A. Pelliccia, and D. Corrado, "International recommendations for electrocardiographic interpretation in athletes," Journal of the American College of Cardiology, vol. 69, no. 8, p. 1057–1075, February 2017.

- [6]. D. Sopic, A. Aminifar, A. Aminifar, and D. Atienza, "Real-time event driven classification technique for early detection and prevention of myocardial infarction on wearable systems," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 5, pp. 982–992, oct 2018.
- [7]. G. S. Wagner and D. G. Strauss, *Marriott's Practical Electrocardiography*. Lippincott Williams & Wilki, 2013.
- [8]. E. S. Jayachandran, P. Joseph, and R. Acharya, "Analysis of myocardial infarction using discrete wavelet transform," *Journal of Medical Systems*, vol. 34, no. 6, pp. 985–992, 2009.
- [9]. S. G. AlKindi, F. Ali, A. Farghaly, M. Nathani, and R. Tafreshi. To-wards real-time detection of myocardial infarction by digital analysis of electrocardiograms. In *Proceedings of 1st Middle East Conference on Biomedical Engineering*, pages 454–457, Sharjah, United Arab Emirates, February 2011. doi: 10.1109/MECBME.2011.5752162.
- [10]. R. S. Remya, K. P. Indiradevi, and K. K. Anish Babu. Classification of myocardial infarction using multi resolution wavelet analysis of ECG. *Procedia Technology*, 24:949–956, July 2016. doi: 10.1016/j.protcy.2016.05.195.
- [11]. Y. Li, J. Bisera, M. H. Weil, and W. Tang. An algorithm used for ventricular fibrillation detection without interrupting chest compression. *IEEE Transactions on Biomedical Engineering*, 59 (1):78–86, January 2012. doi: 10.1109/TBME.2011.2118755.
- [12]. R. Begum and R. Manza. Detection of cardiomyopathy using support vector machine and artificial neural network. *International Journal of Computer Applications*, 133(14):29–34, January2016. doi: 10.5120/ijca2016908178.
- [13]. R. Boussejot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das internet," *Biomedizinische Technik/Biomedical Engineering*, pp. 317–318, July 2009.
- [14]. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol.101, no. 23, June 2000.
- [15]. C. A. Steinberg, S. Abraham, and C. A. Caceres. Pattern recognition in the clinical electrocardiogram. *IRE Transactions on Bio-Medical Electronics*, 9(1):23–30, 1962.

Renato M. Pereira, et al. "Heart Disease Detection Using ECG Lead I and Multiple Pattern Recognition Classifiers." *IOSR Journal of Engineering (IOSRJEN)*, 10(4), 2020, pp. 01-08.