Evaluation of Lead I ECG Features Discriminant Power for Cardiac Diseases Identification

Renato Pereira[1]
rfmfp@icloud.com

Pedro Rodrigues[1]
prodrigues@porto.ucp.pt

Bruno C. Bispo[2]
bruno.bispo@ufsc.br

[1] Universidade Católica Portuguesa
CBQF - Centro de Biotecnologia e Química Fina –
Laboratório Associado
Escola Superior de Biotecnologia, Porto, Portugal.

[2] Department of Electrical and Electronic Engineering
Federal University of Santa Catarina, Florianópolis, Brazil

Abstract

This work proposes to analyze the capacity of several ECG features of Lead I 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 66.7% and 96.9% using feature selection algorithm and SVM classifiers.

1 Introduction

The electrocardiogram (ECG) is the recording of the rhythmic alterations of the heart electrical activity and represents the cardiac cycle [2]. 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. Any minor change in the normal pattern of an ECG signal can be interpreted as malfunction of the heart [9]. 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. During the last years, several works have proposed methods to detect ECG features (morphological or not) and then to diagnose cardiac pathologies. 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 tried 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 from Lead I to discriminate 28 pairs of study groups, combining 7 pathological groups and 1 control group.

2 Methods

2.1 ECG Database

This work used the PTB Diagnostic ECG Database available in [4, 5]. The database contains 549 ECG records from 268 subjects, including healthy subjects (52 subjects) and patients diagnosed with 7 pathologies, namely Cardiomyopathy(18 subjects), bundle branch block (15 subjects), valvular heart disease (6 subjects), myocarditis (4 subjects), MI (148 subjects), myocardial hypertrophy (7 subjects) and Dysrhythmia (14 subjects). Each ECG record contains all the 12-lead system signals with a sampling rate of 1000 Hz, but only Lead I signals were used on this work due to its simplicity in acquisition. The ECG signals had their DC levels removed and their amplitudes were normalized between -1 and 1.

2.2 Peaks detection

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

- R peak: A Discrete Wavelet Transform (DWT) translation analysis, using the wavelet symlet 4, is applied in order to calculate the cross-correlation between the signal and the DWT. The R peaks are the maximum of each correlation over the channel. To find all R peaks in the ECG a 70% signal amplitude threshold of the DWT's

maximum was used to make sure that just the most prominent correlation peak in each heart cycle is detected.

- S peak: The S peak was identified as being the first negative minimum wave after the R peak [8].
- Q peak: Contrary to S peak, the Q peak was identified as being the negative minimum wave right before a R peak [8].
- T peak: For this peak detection all the other previously identified waves are used as references. The signal between the Q and S waves was set to zero, as well as all the negative signal amplitudes. The remaining signal samples were amplified by fifth rooting the signal. Finally to find T peak, the ECG signal was set to zero in each cardiac cycle (CC) before the first CC1 S peak is detected and after CC2 S peak is detected. After it the last 45% of the signal between CC1 and CC2 was also set to 0 and T wave is estimated as the maximum of the remaining signal (Figure 1).
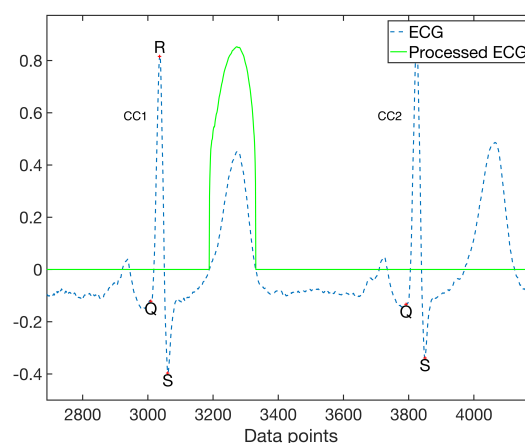

Figure 1: T peak detection.

- P peak: For the P wave detection, the signal was amplified square rooting it. Thus, all the signal between the Q peak and T peak was set to -1. The signal between this peaks was not set to zero because there is a chance of the P peak having a negative value, in some cases. After it, the 55% of the signal between CC1 and CC2 was also set to -1. The P peak is estimated the maximum of the resulting signal.

2.3 ECG Features Analyzed

After detecting the peaks of the P, Q, R, S and T waves, through Wavelet Transform correlation, which are hereinafter called P, Q, R, S and T peaks, of each cardiac cycle, several features of each Lead I ECG signal are calculated in order to analyze their discrimination capacities. The analyzed features are: energy between Q and S peaks, power between Q and S peaks, energy of Wavelet Transform (WT) between Q and S peaks, detail (det) and approximation (app) levels using symlet 4 and symlet 8 waves, time between peaks, peaks amplitude (Amp.) and amplitude difference (a Amp. diff.) between peaks. The features are summarized in the Table 1.

2.4 Classification

Leave-one-out and 5-fold cross validations were used for training and testing the Support Vector Machines (SVM) models with Gaussian, linear and polynomial kernel functions plus a cost of 10. The classification were performed for 28 pairs of the study groups between 7 pathological groups

Table 1: List of analyzed features and their indexes.

Index	Feature	Index	Feature	Index	Feature
1	Energy Q-S	14	Time P-R	27	Amp. T
2	Power Q-S	15	Time P-S	28	Amp. diff. P-Q
3	WT 'sym4' app. level 4 energy	16	Time P-T	29	Amp. diff. P-R
4	WT 'sym4' det. level 1 energy	17	Time Q-R	30	Amp. diff. P-S
5	WT 'sym4' det. level 2 energy	18	Time Q-S	31	Amp. diff. P-T
6	WT 'sym4' det. level 3 energy	19	Time Q-T	32	Amp. diff. Q-R
7	WT 'sym4' det. level 4 energy	20	Time R-S	33	Amp. diff. Q-S
8	WT 'sym8' app. level 4 energy	21	Time R-T	34	Amp. diff. Q-T
9	WT 'sym8' det. level 1 energy	22	Time S-T	35	Amp. diff. R-S
10	WT 'sym8' det. level 2 energy	23	Amp. P	36	Amp. diff. R-T
11	WT 'sym8' det. level 3 energy	24	Amp. Q	37	Amp. diff. S-T
12	WT 'sym8' det. level 4 energy	25	Amp. R		
13	Time P-Q	26	Amp. S		

and 1 control group. Finally, a Genetic algorithms with entropy criterion were applied for feature selection from matrices with 37 features per X subjects (X is the number of the patients involved in each classification pair). The 37 features resulted from the mean features values extracted from each ECG cycle per subject (I lead), as described in Table 1.

3 Results and Discussion

As previously said, in subsection 2.4, the classifications were performed for 28 pairs of the study groups combining 7 pathological groups and 1 control group. The results shown in Figure 2 are the best accuracies achieved from the six trained/tested SVM with Leave-one-out and 5-fold cross validations using gaussian, linear and polynomial functions as kernels and a cost of 10. As can be observed from Figure 2, the maximum accuracy classification was achieved for the pairs Valvular heart disease vs. MI and Myocarditis vs. MI with 96.9% and the Myocarditis vs. Myocardial hypertrophy provided the lowest accuracy classification (66.7%). 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 91% for the exception of Bundle branch block and MI where the reached accuracies were 85.9% and 77.8%, respectively. For the pairs Healthy controls vs. Cardiomyopathy and Healthy controls vs. Dysrhythmia the classifiers achieved an accuracy of 93.7% and 92.3%. The distinguish accuracies between Healthy controls against MI, Dysrhythmia and Cardiomyopathy are slightly under the results of those in the state-of-art (Teble 2) 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.

Table 2: List of works found in the literature.

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

Therefore, using only Lead I analysis, the SVM classifier was able to achieve an accuracy between 66.7% and 96.9% 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.

4 Conclusion

Autonomous and accurate discrimination of cardiac pathologies through ECG is an important tool to assist in the diagnosis of heart diseases. The works found in the literature have the common goal of discriminating between 2 study groups, one pathological and one control, even when ECG records from patients diagnosed with several pathologies are available in the databases. This work have 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 SVM classifier was able to achieve an accuracy between 66.7% and 96.9% for the 28 pairs of study groups. These results become even more relevant considering that only 3 of these
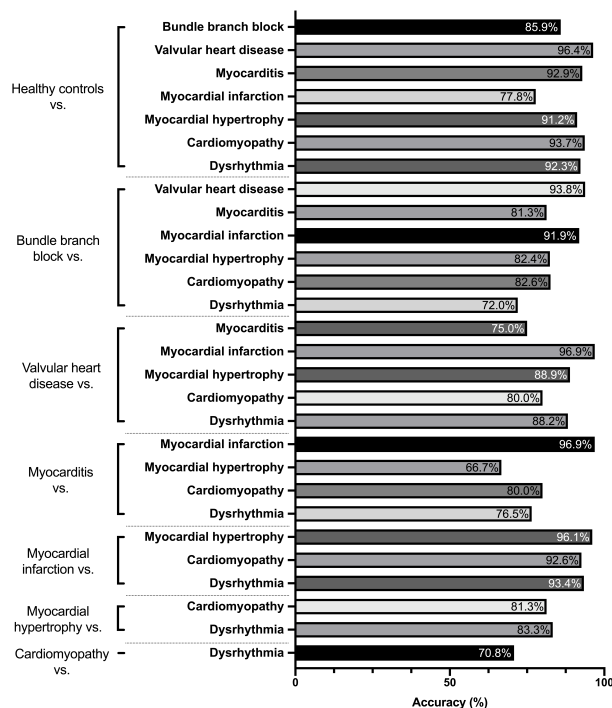

Figure 2: Pathologies discrimination accuracies.

pairs are commonly analyzed in the literature: MI, Dysrhythmia and Cardiomyopathy. This study also proves Lead I has good capacity for heart pathologies discrimination, however the low number of records for some pathologies should be taken in consideration.

References

[1] S. G. Al-Kindi, F. Ali, A. Farghaly, M. Nathani, and R. Tafreshi. Towards 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, Feburary 2011. doi: 10.1109/MECBME.2011.5752162.

[2] John S. Barlow. *The electroencephalogram: its patterns and origins*. MIT Press, 1993.

[3] Rabiya Begum and Ramesh Manza. Detection of cardiomyopathy using support vector machine and artificial neural network. *International Journal of Computer Applications*, 133(14):29–34, January 2016. doi: 10.5120/ijca2016908178.

[4] R Bousseljot, D Kreiseler, and A Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1):317–318, 1995.

[5] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[6] Yongqin Li, Joe Bisera, Max Harry Weil, and Wanchun 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.

[7] 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.

[8] CA Steinberg, S Abraham, and CA Caceres. Pattern recognition in the clinical electrocardiogram. *IRE Transactions on Bio-Medical Electronics*, 9(1):23–30, 1962.

[9] Galen S. Wagner and David G. Strauss. *Marriott's Practical Electrocardiography*. Lippincott Williams&Wilki, 2013. ISBN 1451146256.