Development of an algorithm for heartbeats detection and classification in Holter records based on temporal and morphological features

A García 1, H Romano 1, E Laciar1 and R Correa1

1 Gabinete de Tecnología Médica, Facultad de Ingeniería, Universidad Nacional de San Juan, San Juan, Argentina

E-mail: alfredorg86@gmail.com, marauder.7@hotmail.com, laciar@gateme.unsj.edu.ar, rcorrea@gateme.unsj.edu.ar

Abstract. In this work a detection and classification algorithm for heartbeats analysis in Holter records was developed. First, a QRS complexes detector was implemented and their temporal and morphological characteristics were extracted. A vector was built with these features; this vector is the input of the classification module, based on discriminant analysis. The beats were classified in three groups: Premature Ventricular Contraction beat (PVC), Atrial Premature Contraction beat (APC) and Normal Beat (NB). These beat categories represent the most important groups of commercial Holter systems. The developed algorithms were evaluated in 76 ECG records of two validated open-access databases "arrhythmias MIT BIH database" and "MIT BIH supraventricular arrhythmias database”. A total of 166343 beats were detected and analyzed, where the QRS detection algorithm provides a sensitivity of 99.69 % and a positive predictive value of 99.84 %. The classification stage gives sensitivities of 97.17% for NB, 97.67% for PCV and 92.78% for APC.

1. Introduction

Cardiovascular diseases (CVDs) cause the death of around 17 million people per year, representing almost a third of all deaths globally. The most commonly cardiovascular diseases are heart disease and stroke. It is estimated that in 2015 will die about 20 million people by these diseases and is expected to remain the leading cause of mortality [1].

Holter ECG system is a useful ambulatory diagnostic tool, which has evolved considerably not only in the technological aspect, but in the clinical field as well. It allows a non-invasive diagnosis of heart-rate abnormalities and other cardiovascular diseases for a long period of time (24 – 48 hours). Besides, it is desirable to record for a long period of time, in order to eliminate wrong interpretations from normal spontaneous variabilities.

Through this study it is possible to analyze heart rate variability, detect extra-systoles, tachyarrhythmias and bradyarrhythmias, identify ST disorders or perform more complex analysis like the estimation of sudden death risk using QT interval analysis. Other utilities are the evaluation of the antiarrhythmic drugs efficacy, the performance of pacemakers and defibrillators re-synchronizers, as well as the study of ventricular late potentials [3].

The detection of Premature Ventricular Contraction beats (PVC) and Atrial Premature Contraction beats (APC) is very important for the identification of two kinds of arrhythmias. The APCs are due to auricular arrhythmias, which begin in an atrium or in the AV node [4]. On the
contrary, the PVCs are related to ventricular arrhythmias, which begin in a ventricle, being potentially lethal because they can produce ventricular fibrillation and/or sudden death [5]. PVCs have an aberrant morphology, allowing an easy identification. The APCs have a similar morphology as NB, being their detection very difficult. Both PVCs and APCs generate an early contraction of the heart, which produces an increase of instantaneous heart rate.

Diverse techniques were used in different works to detect and classify heartbeats in ECG and/or Holter records, such as classification based on ECG morphology and heartbeat interval features [6] frequency domain analysis [7] and neural networks [8, 9]. The main objective of this work consists of detecting, featuring and classifying the heartbeats in Holter records in the following categories: PVC, APC and Normal Beats (NB), using a modified version of Pan and Tompkins algorithm for the heartbeats detection and the discriminant analysis technique for their classification.

2. Materials

Real Holter records from open-access database such as “MIT BIH arrhythmia database” and “MIT BIH supraventricular arrhythmia database” were used in this study. They are available free of charge in the website Physionet [10, 11] and they are widely used for the validation of commercial Holter analyzers.

The “MIT BIH arrhythmia database” has 48 records of 30 minutes duration. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. The subjects were 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years. This database has different morphologies of PVCs and NBs. The “MIT BIH supraventricular arrhythmia database” has 78 records of 30 minutes of length. They are sampled at 128 Hz with an amplitude resolution of 11 bits. This database has mainly APCs and NBs.

In most records of both databases, channel one is a modified limb lead II, and channel two is usually a modified lead V1 (occasionally V2 or V5, and in one instance V4). In the development of this work channel one was only used.

3. Methods

The algorithm has 3 stages: the first consist on the QRS complexes detection in the Holter record; this should be done to locate each heartbeat within the recording. In the next stage, the most important characteristics of QRS complexes were extracted. Finally, each heartbeat was classified using discriminant analysis.

Figure 1 shows the block diagram summarizing the different stages of the proposed algorithm. For the development of algorithms and the graphical interface the program MatLab® was used.

![Figure 1](image.png)

**Figure 1.** Block diagram of the proposed algorithm for heartbeats detection and classification in Holter records.

3.1. QRS complex detection

The QRS detector implemented is based on the detection method developed by Pan and Tompkins [12] and consists of five stages. Figure 2 shows the block diagram of the implemented QRS detector, showing in each case the result of each stage for an ECG example.

The first stage is a band pass filter, type Butterworth, order 12, with lower cutoff frequency of 5Hz and higher cutoff frequency of 15Hz to reduce the P and T waves, leaving only the QRS complex. After filtering, the signal is differentiated to provide the QRS-complex slope information;
a five-point derivative filter was used. In the third stage, the signal is squared point by point to eliminate negative values, as well as mitigate the points whose value is less than unity. In the fourth stage, a moving average filter of 100 ms was applied to smooth the signal and obtain well-defined maximum. Finally, a robust thresholding technique was proposed to locate the QRS complex in the record. It used an adaptative detection based on the RR intervals.

In order to decrease incorrect detections, the algorithm eliminates QRS marks that are less than 200 ms of an earlier mark (decreasing incorrect detections). If an excessively long RR interval is sensed, the algorithm repeats the analysis in the area with a lower threshold.

![Figure 2. Block diagram of the QRS detector. (Modified from [5], [12]).](image)

3.2. Feature Extraction

After detecting the position of each QRS complex in the Holter recording, a number of features were extracted from each heartbeat, which allowed their subsequent classification. This stage is very relevant because the parameters defined here are fundamental for the correct identification of the beating and subsequent clustering.

After several tests and empirical studies, 12 features were selected based on different temporal and morphological parameters, which are very similar to the used for the specialist, and are described below.
3.2.1. Temporal features

Eight temporal features were selected. One of them is the RR interval, defined as the time delay between two QRS peaks. It was found the interval between the current and previous beat, as well as the one between the current and subsequent beat, which are called $RR1$ and $RR2$ respectively. Another interval used is the distance between the previous beat and its predecessor, called $RR0$ (see Figure 3).

![Figure 3](image-url)  
**Figure 3.** Identification of RR intervals in the Holter record. In red, it is illustrated the current beat. (record # 119 of “MIT BIH arrhythmia database”)

After that, adimensional time-independent features were looked for, which are adapted to signals coming from different patients. The relations between RR1 and RR2 were defined, as well as the one between RR1 and RR0, called $Ratio1$ and $Ratio2$ respectively [13]. $Ratio3$ was obtained as the relationship between RR1 and RRM, being RRM the average of all RR intervals corresponding to each record. The $MRATIO$ was calculated as the mean value of Ratio1, Ratio2 and Ratio3. Finally, the robust median ($MRR$) was computed by applying to a series of RR intervals a median filter, taking a scan window of 10 intervals.

3.2.2. Morphological features

In order to extract 4 characteristics, the QRS complex of each beat was firstly extracted, using the previous 70 ms before the QRS mark up to the 140 ms after the mark. So, a window of 210 ms wide was generated, which contains all QRS complex morphology, without including the P and T waves. Then, each QRS complex was normalized to obtain an independent feature of the range.

\[
QRSN_i(n) = \frac{QRS_i(n) - \max(QRS_i(n))}{\max(QRS_i(n)) - \min(QRS_i(n))} \quad (1)
\]

\[i = 1, 2, \ldots, N \text{ beats}\]

where:

- $QRS_i(n)$: is the original $i$-th QRS complex detected in the record.
- $QRSN_i(n)$: is the normalized $i$-th QRS complex detected in the record.
Figure 4 shows an example of the normalized QRS complex for: (a) NB, (b) PVC, and (c) APC beats.

Using the normalized QRS complexes, the maximum of cross-correlation function between each detected beat and the following beat was calculated, as well as the maximum of cross-correlation between the current beat and the previous beat detected, called respectively $\text{Corr}_1$ and $\text{Corr}_2$ [13].

Another feature was the maximum of cross-correlation between a template of normal beat, with each QRS complex detected, called $C_{xy}$, was computed. For each Holter record, the template was calculated as the averaged beat of a sequence of 20 normal sinus beats.

Finally, a feature was defined as the QRS duration (in ms) when $\text{QRS}_i(n) = 0.5$ in the normalized QRS complex, as shown in Figure 5.

Figure 5. Feature A: QRS duration at 50% of the amplitude of the template ($\text{QRS}_i(n)=0.5$). (record # 119 of “MIT BIH arrhythmia database”)
3.3. Heartbeat classification

Figure 6 shows the block diagram of the proposed heartbeats classifier.

![Block diagram of the proposed heartbeats classifier.](image)

The method used is the discriminant analysis method, which is a multivariate statistical technique whose aim is to examine if there are significant differences between groups. Since there are three possible categories for each beat, the classification model has 2 linear discriminant functions. To create that model, 5800 heartbeats were selected from validated open access databases "MIT BIH arrhythmia database" and "MIT BIH supraventricular arrhythmia database", choosing 2000 NBs, 2000 PVCs and 1800 APCs. With these 5800 selected heartbeats, two groups were created:

- A training group, which contains 70% of the heartbeats (4060), the first 1400 corresponding to a NBs, the second 1400 to PVCs and the last 1260 to a APCs.
- A validation group, which contains 30% of the heartbeats (1740), the first 600 corresponding to a NBs, the second 600 to PVCs and the last 540 to APCs.

The heartbeats for both groups were randomly selected. Once materialized this stage, it was proceeded to select the best combination of the 12 features described above, that best discriminated different types, concluding that the best combination was the one that contained the 12 features used in this work. Figure 7 shows a segment of a record, where the different categories of heartbeats detected by the classifier are highlighted with different colours.

![Classifier applied to a real Holter record (record # 872 of “MIT BIH supraventricular arrhythmia database”)](image)

Figure 7. Classifier applied to a real Holter record (record # 872 of “MIT BIH supraventricular arrhythmia database”)
3.4. Post – Processing
After detecting and classifying PCV, a search of episodes of ventricular bigeminy and trigeminy was made, as well as an identification of ectopic foci. Ventricular bigeminy is the occurrence of a PVC every other beat, where as trigeminy is every third beat. Figure 8 shows an episode of bigeminy and Figure 9 an episode of trigeminy. In Figure 10 are represented PVCs from different foci.

![Figure 8](image1.png)  
**Figure 8.** Episode of ventricular bigeminy detected, PVCs in red. (record # 106 of “MIT BIH arrhythmia database”)

![Figure 9](image2.png)  
**Figure 9.** Episode of ventricular trigeminy detected, PVCs in red. (record # 106 of “MIT BIH arrhythmia database”)

![Figure 10](image3.png)  
**Figure 10.** Different PVCs morphologies from two ventricular foci. (record # 106 of “MIT BIH arrhythmia database”)

4. Results

4.1. Heartbeat Detector.
The QRS detection algorithm was validated with the signals of the databases presented above, where 76 records were used. These records contain 166343 heartbeats from different patients,
acquired under different conditions (noise, sampling frequency, channel used). The marks obtained by the detector were compared with annotations made by professionals and according to them, the following parameters were calculated.

- **True Positives (TP):** the number of positive detections that correspond to the annotations of the specialist.
- **False Positives (FP):** the number of detections that do not correspond with the annotations of the specialist.
- **False Negative (FN):** the number of heartbeats that were annotated by the specialist that were not detected.
- **Sensitivity (Sen):** the percentage of heartbeats correctly identified by the algorithm.
  \[
  Sen(\%) = \frac{TP}{TP + FN} \times 100
  \]
- **Predictive value (PPV):** the detection rate given by the algorithm corresponding to marks made by the specialist.
  \[
  PPV(\%) = \frac{TP}{TP + FP} \times 100
  \]
- **Classification error rate (Err):** the percentage of false detections over the total number of detected heartbeats.
  \[
  Err(\%) = \frac{FP + FN}{TP} \times 100
  \]

Table 1 shows the results obtained by the QRS detector.

|                  | Heartbeats | TP      | FN   | FP   | Sen (%) | PPV (%) | Err (%) |
|------------------|------------|---------|------|------|---------|---------|---------|
| MIT BIH arrhythmia database | 85944      | 85642   | 302  | 83   | 99.66   | 99.89   | 0.45    |
| MIT BIH supraventricular arrhythmia database | 80399      | 80214   | 185  | 143  | 99.71   | 99.80   | 0.49    |
| Both databases   | 166343     | 165856  | 487  | 226  | 99.70   | 99.86   | 0.43    |

In comparison with other works [14] the QRS detector has similar values, but for the validation a 50% more of heartbeats were used, because another database was utilized.

4.2. Heartbeat classification

For the classification stage, the 70/30 method validation was used, the 70% of the heartbeats corresponding to the computation of the discriminating functions and the 30% for their validation. The random selection process of the groups of heartbeats was repeated 100 times. Table 2 shows the average values of sensitivity, specificity, positive predictive value (PPV), TP (true positives), TN (true negatives), FP (false positives) and FN (false negatives) for each of the categories of heartbeats.
Table 2. Results of the heartbeat classification for the validation stage.

|       | NB       | PVC       | APC       |
|-------|----------|-----------|-----------|
| TP    | 583      | 586       | 501       |
| TN    | 1119     | 1119      | 1172      |
| FP    | 21       | 21        | 28        |
| FN    | 17       | 14        | 39        |
| Sensitivity | 97.17%   | 97.67%    | 92.78%    |
| Specificity | 98.16%   | 98.16%    | 97.67%    |
| PPV   | 97.01%   | 97.98%    | 96.15%    |

Figure 11 shows the scatter diagram obtained, showing that PVCs are notably separated of the NB. On the contrary, the APCs overlap the other two groups.

5. Discussion and conclusions.

In this paper, it was proposed an algorithm for heartbeat detection and classification, which can be used in ambulatory or Holter ECG records.

The results achieved with the QRS detector were similar in comparison to other algorithms, giving a sensitivity of 99.70%, a specificity of 99.86% and an error of 0.43% in contrast with the work of Martinez et al [14] which obtains a sensitivity of 99.66%, PPV of 99.56%. The algorithm was adapted to different kinds of signals with different environmental conditions.

The heart beat classification algorithm grouped them in three categories and was compared with the annotations of the specialists, being very efficient in the identification of PVCs (with a 97.67% of sensitivity), since they have very different morphological characteristics from the rest. This performance was similar than others algorithms, such as the proposed for Couceiro et al [9] and better than the proposed for de Chazal et al [6] in the classification of heartbeats, they obtains sensitivities of 96.35% and 94.4% for the PVCs classification respectively. However, APCs were more difficult to separate, obtaining a sensitivity of 92.78%, because they are morphologically identical to NB, and they also have similar time features as the PVCs. Anyway, better results were
achieved compared with other papers, Karraz et al obtains a sensitivity of 86% for the APCs classification [15].

In conclusion the classifier proposed in this paper can be applied to real Holter records and can analyze automatically a record beat to beat. It takes approximately 10 seconds to analyze each of the records. This allows an automated analysis without a physician who classifies the different beats. It is also a useful tool for the specialist to make appropriate diagnosis.

In the future, this work will be integrated in a Holter analyzer system, which will have the most important functions of a commercial Holter analyzer.

References
[1] Hoffmann M, Heinke S, Lang K, Leder U, Hoffmann G, Heinke M, Dannberg G, Kiinhert H and Muller S 1996. Comparative studies of heart rate variability (HRV) using Holter - ECG - systems and high solution ECG - systems in patients with acute myocardial infarction and dilatative cardiomyopathy and healthy volunteers. 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam. pp. 1256-57.
[2] Rojas Malgarejo E. Monitoría electrocardiográfica ambulatoria de 24 horas (Holter). Historia, indicaciones y elaboración de un informe.
[3] Gamiz Palma JL 1983. Electrocardiografía de Holter. Bases prácticas y aplicaciones clínicas. Ediciones NORMA.
[4] Harrison 2006 Principios de Medicina Interna 16a edición http://es.wikipedia.org/wiki/Taquicardia_supraventricular. 21/12/2010 18:58 hs.
[5] Laciar E. and Valentinuzzi M. 2010. Chapter 4. Ventricular Fibrillation Detection, in Cardiac Fibrillation-Defibrillation. Clinical and Engineering Aspects by Max E Valentinuzzi, World Scientific Publishing Co, pp. 115-138, 2010.
[6] de Chazal P, Reilly R 2006. A patient-adapting heartbeats classifier using ECG morphology and heartbeat interval features. IEEE transactions on biomedical engineering, vol. BME-53, NO. 12.
[7] Lin C, Du Y, Chen Y and Chen T 2006. Multiple ECG beats recognition in the frequency domain using grey relational analysis. EMBS annual international conference, proceedings of the 28th IEEE.
[8] Palaniappan R, Krishnan S. 2004. Detection of ectopic heartbeats using ECG and blood pressure. International conference on signal processing & communications.
[9] Couceiro R, Carvalho J, Henrique J and Antunes M. 2008. On the detection of premature contractions. 30th annual international IEEE EMBS conference.
[10] MIT-BIH 1990, “MIT-BIH supraventricular arrhythmia database directory”, Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center.
[11] MIT-BIH 1999, “MIT-BIH arrhythmia database directory”, Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center.
[12] Pan J and Tompkins W 1985. A real time QRS Detection Algorithm. IEEE transactions on biomedical engineering, vol. BME-32, NO. 3.
[13] Gharavari A, Dehghan F, Teshnelab M and Mohgaddam H. 2008. Comparison of neural network, ANFIS, and SVM classifiers for PVC arrhythmia detection. Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kumming, pp. 12-15.
[14] Martinez JP, Almeida R, Olmos S, Laguna P. 2004. A Wavelet-Based ECG Delineator: evaluation on Standard Databases. IEEE IEEE transactions on biomedical engineering, VOL. 51, NO. 4.
[15] Karraz G, Magenes G 2006. Automatic Classification of Heartbeats using Neural Network Classifier based on a Bayesian Framework. Proceedings of the 28th IEEE EMBS Annual International Conference (New York City, USA).