Pattern Recognition and Optimal Parameter Selection in Premature Ventricular Contraction Classification

I Jekova[1], G Bortolan[2], I Christov[1]

[1]Centre of Biomedical Engineering – Bulgarian Academy of Science, Sofia, Bulgaria
[2]Institute of Biomedical Engineering ISIB - CNR, Padova, Italy

Abstract

Analyses of electrocardiographic pattern recognition parameters for premature ventricular contraction (PVC) and Normal (N) beat classification are presented. Twenty-six parameters are defined: 11x2 for the two ECG leads, 3 for vectorcardiogram (VCG) and width of the complex. Some of them include: amplitudes of maximal positive and negative peaks, area of the absolute values, area of positive and negative values, number of samples with 70% higher amplitude then that of the highest peak, amplitude and angle of the QRS vector in a VCG plane. They are measured for all N and PVC heart beats in the MIT-BIH arrhythmia database. The classification ability of each parameter is tested using discriminant analysis. Considering both leads 7 parameters with highest discriminant power for N and PVC are extracted and a specificity of 96.6% and a sensitivity of 90.5% are obtained. Taking into account relatively all parameters a specificity of 97.3% and a sensitivity of 93.3% are achieved.

1. Introduction

ECG signal analysis is the most common way to study and diagnose cardiac dysfunctions. The normal signal is characterised by recurrent or periodic waveforms with each beat. Beat-to-beat detection and classification of the QRS complexes allows to follow the heart rhythm evolution and to detect arrhythmias such as premature ventricular contractions (PVC). Detection and classification of ventricular beats changes is of considerable importance in real-time critical care or operating room patient monitoring. In these applications it is important to develop signal-processing techniques that allow real-time feature extraction for classification of QRS and other ventricular beat patterns.

Adaptive signal processing has been used for on-line estimation of non-stationary signals that present a recurrent behaviour [1-4].

The QRS and PVC detection algorithm of Millet et al. [5] included a set of characteristic parameters of the QRS waves, including different area measures and specific interval durations. They have obtained statistical indices for sensitivity (Se) and specificity (Sp) of 94.6% and 98% respectively, from exploration of a set of seven MIT-BIH records.

Christov and Bortolan [6] derived 26 pattern recognition parameters for heart beat classification. Employing neural networks they ranked the huge quantity of parameters for the PVC and N clustering with the idea that this ranking will further be used in real-time clustering implementation.

Laguna et al. [7] have presented an adaptive Hermite model for on-line beat-to-beat estimation of QRS features. They have considered the width of the complex as most relevant for ectopics detection.

Lagerholm et al. [8] have used Hermite functions and self-organizing maps for ECG complex clustering, decomposing each QRS complex into Hermite basis functions.

Moraes et al. [9] have extracted four QRS complex features presenting best results: width, total sum of the areas under the positive and negative curves, total sum of the absolute values of sample variation and total peak-to-peak amplitude. They used the Mahalanobis distance as classification criterion. Forty-four records have been used from the MIT-BIH database for testing and the results have been 90.74% sensitivity and 96.55% positive predictivity.

Ham and Han [10] have used fuzzy adaptive resonance theory mapping for classification of cardiac arrhythmias. The authors have reported 97% sensitivity and 99% specificity from a selected set of six MIT-BIH records.

Al-Nashash [11] and Maglaveras et al., [12] have used neural networks for PVC classification. The first author has reported 98.1% sensitivity and 94.7% predictivity from 14 MIT-BIH database records.

We developed a quantitative analysis of the pattern recognition parameters for PVC classification. 23 ECG and 3 VCG parameters were included. The classification ability of each of them was assessed by discrimanant analysis and optimal parameter grouping was proposed for real time processing improvement. No selection based

on the quality of the signal was performed, thus the analysis was carried out even in the presence of artifact or noise in the ECG signal such as power-line interference, baseline wander, electromyogram noise, poor electrical contact, electrode loose, etc.

2. Methods and material

2.1. ECG database

All 48 ECG recordings from the MIT-BIH arrhythmia database were used. Each recording has a duration of 30 min and includes two leads – the modified limb lead II and one of the modified leads V1, V2, V4 or V5 [13]. The sampling frequency is 360 Hz and the resolution is 200 samples per mV. Two cardiologists have annotated all beats in the database. Approximately 70% of the beats have been annotated as N. Since we focused only on the PVC classification, we followed the AHA records equivalent annotation, including some of the abnormal beats (left bundle branch block, right bundle branch block, aberrantly conducted beat, nodal premature beat, atrial premature beat, nodal or atrial premature beat, nodal escape beat, left or right bundle branch block, atrial ectopic beat and nodal ectopic beat) in the N group [13]. In addition, fusion premature ventricular contractions, ventricular flutter waves, ventricular escape beats, blocked, atrial premature beats, paced beats, missed beats and questionable beats were excluded from the study. No selection based on the quality of the signal was performed, thus the analysis was applied even in the presence of artifact or noise in the ECG signal.

For the measurement of the VCG parameters it is important to know that forty-two of the MIT-BIH arrhythmia database recordings are of leads II and V1. The remaining are of leads II and V5 (100, 114, 123), V2 and V5 (102, 104), II and V4 (124).

2.2. Preprocessing

The preprocessing of the ECG signal was consistent to subsequent real-time application of the PVC/N clustering of the beat complexes, involving:

- moving averaging of samples in one period of the power-line interference; this filter is meant to eliminate the powerline interference by its frequency response having a first zero at the interference frequency.
- moving averaging of samples in 30 ms time-interval; this low-pass filter with a first zero at about 35 Hz suppresses the electromyogram noise.
- high-pass recursive filter for drift suppression [14, 15]; the phase characteristic of this filter is constant. The small distortion thus introduced by the absence of backward filtration does not impede the PVC/N classification.

The high-pass recursive filter is given by the formula:

$$Y_n = C_1(X_n - X_{n-1}) + C_2 Y_{n-1} \quad (1)$$

where Y_n is the filtered samples sequence, X_n is the samples sequence of the original signal and n is the consecutive number of samples. The constants C_1 and C_2 are calculated by the formulae:

$$C_1 = \frac{1}{1+\tan(Fc\pi T)} \qquad C_2 = \frac{1-\tan(Fc\pi T)}{1+\tan(Fc\pi T)} \quad (2)$$

where T is the sampling period and Fc=2.2 Hz is the chosen cut-off frequency.

2.3. Pattern recognition parameters

Several heart beat parameters for pattern recognition were derived for each complex annotated as N or PVC in the MIT-BIH arrhythmia database. First, examining the two ECG leads, the onset and the offset of the complex were identified and the width (Width) was computed. Then, from each ECG lead the following 11 parameters were derived: maximal positive peak (Pp), maximal negative peak (Pn), the area as sum of absolute values in the beat interval width (Ar), the area of positive values in the same width (ArP), the area of negative values in the width (ArN), the sum of absolute velocity values in the width (Av1), the number of samples with 70% higher amplitude than that of the highest peak (No), the time interval duration from the onset to the maximal positive peak (Ima), the time interval duration from the onset (Imi) to the maximal negative peak. Considering the time sequence obtained by the beat onset, the maximal positive peak and the maximal negative peak, the ECG slope velocity between 1st and 2nd point (S1) and between 2nd and 3rd point (S2) were computed, where S1 is the onset of the beat, S2 is its maximum point and S3 is its offset.

From the VCG signal, the following parameters were considered: the maximum amplitude of the vectorcardiogram vector (VCGamp) in the plane formed by the two leads and the angle of the maximal amplitude vector (VCGang). For more relevant information in the classification task, the VCGang parameter was split in the sine and cosine components (VCGsin, VCGcos). Therefore, 23 ECG and 3 VCG parameters were considered for the classification of N and PVC beats.

2.4. Discriminant analysis

Using discriminant analysis to differentiate between N and PVC beats, two linear discriminant functions of the n-dimensional vector x were calculated – equations 3 and 4.

$$F'(x) = \sum_{i=1}^{n} w_i' x_i + a' \quad (3)$$

$$F''(x) = \sum_{i=1}^{n} w_i'' x_i + a'' \quad (4)$$

Here w_i', w_i'' and a', a'' are the corresponding discriminant coefficients and constants. Equation 3 relates to the possibility the heart beat described by vector x to be N, and on the opposite equation 4 gives the possibility to be PVC. These two discriminant functions were computed for the assessed heart beat and it was labeled to one of two classes: N or PVC, depending which of the *F'* or *F"* value was higher. The classification ability of each parameter was tested, and several optimal parameter' groups were detected. After that stepwise discriminant analysis was applied on the whole parameter set. At the beginning, the weight of each one of the twenty-six parameters was estimated (the ability of each of them to separate the two classes), by comparing with a predefined constant, called Fisher constant (F). For F>4 the corresponding parameter was included in the two discriminant functions (3) and (4). For F≤4 the parameter was not included. The best discriminating parameter was first included. After that its combinations with the remaining parameters were analyzed. The best combination, which satisfies the *F* criteria, was included at the final version of the discriminant functions. Iteratively, the best combination was combined with each of the remaining *(n-1)* parameters, until the inclusion of a new parameter did not improve the classification.

3. Results

The specificity (Sp) representing the accuracy of the classification of Normal QRS complexes, and the sensitivity (Se) representing the PVC clustering for any of the parameters are given in Table 1, where the mean values on 48 MIT-BIH records are reported. The indices are derived for each individual ECG lead, as well as for both leads. Considering both leads, the first top 6-ranked parameters for N clustering are: Pn, Width, ArN, VCGsin, Ar and Ima, while for the PVC clustering are: Width, Ar, Pn, ArN, Ima and Pp. Considering the above mentioned 7 (5x2 + 2) parameters a specificity of 96.6% and a sensitivity of 90.5% were obtained.

Applying stepwise discriminant analysis we exclude 4 (4x2) parameters (ArN2, Ima2, Pn2 and ArP1), which do not improve the classification of N and PVC beats. Taking into account the remaining parameters, the obtained discriminant function provided Sp=97.3% and Se=93.3%.

We made an experiment in order to measure the clustering performance when working with only one ECG channel. The results are shown in Table 2. Only those parameters, which could be derived separately for the leads were taken into account, i.e. VCG parameters and the QRS width were excluded.

There is a decrease of the results for both N and PVC clustering when using parameters derived from only one channel. It is more expressed in the 2nd Channel, where in some patients from the MIT-BIH arrhythmia database the ECG has a very low information value (Figure 1) due to the fact that this lead axis happens to be almost perpendicular to the polarization vector.

Table 1. Specificity (Sp) and Sensitivity (Se) of each parameter for Lead1, Lead2 and Lead1+2

Parameter	Lead1		Lead2		Lead1 + 2	
	Sp %	Se %	Sp %	Se %	Sp %	Se %
Pp	47.7	60.5	73.1	64.9	75.6	70.2
Pn	94.2	74.3	63.4	50.8	94.1	74.2
Ar	90.4	73.8	73.6	58.9	90.5	74.5
ArP	88.2	67.8	73.4	53.7	81.5	65.0
ArN	91.0	72.6	70.8	60.6	91.5	73.5
Av	69.1	57.4	69.7	53.0	68.1	56.5
No	50.4	56.7	82.9	60.3	77.4	55.8
Ima	96.1	62.3	71.0	56.3	89.8	72.9
Imi	80.9	54.2	80.3	49.4	77.8	62.0
S1	89.8	56.3	48.8	48.9	89.7	55.9
S2	77.7	59.6	46.2	46.6	77.3	59.4
Width					92.5	76.9
VCGamp					62.8	56.0
VCGsin					91.0	47.9
VCGcos					50.5	50.5

Table 2. Clustering of Lead 1, Lead 2 and Lead1+2.

Lead	Sp %	Se %
Lead 1	97.1	89.5
Lead 2	86.7	78.3
Lead 1+2	97.2	91.7

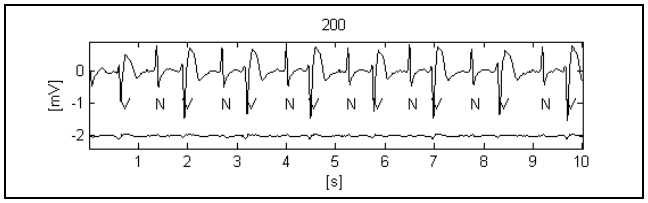

Figure 1. Very low amplitude in the 2nd Channel (the second trace) of record 200 from the MIT-BIH arrhythmia database.

All values for Se and Sp were obtained by the program package SPSS. The leave-one-out discrimination used guarantees significant results, because the assessed signal is removed from the data set.

The discriminant functions, which provided the best results for Sp (97.3%) and Se (93.3%) are given below:

$$FN = 19.3Pn_1 + 5.2Av_1 + 0.65Width - 0.14Ar_1 + 0.2Ima_1 + 5.8VCGamp + 3.7Pp_2 - 28.6S2_2 + 76.5S1_1 - 3.8VCGsin + 0.3No_2 - 0.02ArP_2 - 0.24Imi_1 - 21.7Pp_1 + 0.06Imi_2 - Av_2 - 0.2Ar_2 - 0.3ArN_1 + 0.3No_1 + 1.3VCGcos - 227.7S1_2 - 1.9S2_1 - 27.6 \quad (5)$$

$$FPVC = 10.1Pn_1 + 5Av_1 + 0.8Width + 0.03Ar_1 + 0.3Ima_1 + 1.7VCGamp + 11.4Pp_2 - 3S2_2 + 122.2S1_1 - 1.9VCGsin + 0.4No_2 - 0.16ArP_2 - 0.2Imi_1 - 29.2Pp_1 + 0.04Imi_2 - 1.6Av_2 - 0.13Ar_2 - 0.36ArN_1 + 0.35No_1 + 0.06VCGcos - 193.7S1_2 - 9.4S2_1 - 42.8 \quad (6)$$

4. Discussion and conclusions

Analyses of electrocardiographic pattern recognition parameters for PVC and N beat classification have been presented. Twenty-six parameters were defined and measured form the two ECG leads, and the VCG signal. The presence of two ECG leads improves the classification accuracy of the single ones, and the evident higher accuracy in Lead 1 is due in general by a better signal-to-noise ratio. The results show that there are different groups of parameters, which provide different classification results. Yet it is evidently that increasing the parameter number leads to improving the classification ability of the method. The specificity of 97.3% and sensitivity of 93.3%, obtained with 22 (18x2 + 4) ECG parameters is comparable with the reported values of the referred authors.

The parameters ranking and grouping is helpful and practical for premature ventricular and normal contraction pattern recognition. It orientates which parameter, or combination group should be used for the PVC/N clustering, and what result should be expected.

Acknowledgements

This study was supported by a joint project between the Italian National Research Council (ISIB-CNR) and the Bulgarian Academy of Sciences (Center of Biomedical Engineering-BAS).

References

[1] Widrow B, Stearns SD. Adaptive signal processing. In: Englewood Cliffs, editors. Place Published: Prentive-Hall (New Jersey), 1985.

[2] Ferrara ER, Widrow B. The time-sequenced adaptive filter. IEEE Trans. 1981; CAS-28: 519-523.

[3] Thakor NV, Yi-Sheng Z. Application of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. IEEE Trans. Biomedical Engineering 1991; 38: 785-794.

[4] Laguna P, Jane R, Caminal P. Adaptive feature extraction for QRS classification and ectopic beat detection. IEEE Computer Society Press 1992; 613-616.

[5] Millet J, Perez M, Joseph G, Mocholi A, Chorro J. Previous identification of QRS Onset and Offset is not essential for classifying QRS complex in a single lead. Computers in Cardiology 1997; 24: 299-302.

[6] Christov I, Bortolan G. Ranking of pattern recognition parameters for premature ventricular contractions classification by neural networks, Physiological measurements 2004; 25: 1281-1290.

[7] Laguna P, Jane R, Olmos S, Thakor NV, Rix H, Caminal P. Adaptive estimation of QRS complex wave features of ECG signal by the Hermite model. Medical & Biological Engineering & Computing 1996; 34: 58-68.

[8] Lagerholm M, Peterson G, Braccini G, Edenbrandt L, Sörnmo L. Clustering ECG complex using Hermite Functions and self-organizing maps. IEEE Transaction on Biomedical Engineering 2000; 47, No 7: 838-848.

[9] Moreas JCTB, Seixas MO, Vilani FN, Costa EV. A real time QRS complex classification method using Mahalanobis distance. Computers in Cardiology 2002; 29: 201-204.

[10] Ham FM, Han S. Classification of cardiac arrhythmias using Fuzzy ARTMAP. IEEE Transaction on Biomedical Engineering 1996; 43, No 4: 425-430.

[11] Al-Nashash H. Cardiac arrhythmia classification using neural networks. Technol. Health Care 2000; 8: 363-372.

[12] Maglaveras N, Stamkopoulos T, Diamantaras K, Pappas C, Strintzis M. ECG pattern recognition and classification using non-linear transformations and neural networks: a review. Int. J. Med. Inf. 1998; 52: 191-208.

[13] Mark R, Moody G. MIT-BIH Arrhythmia data base directory. Cambridge: Massachusetts Institute of Technology 1988

[14] Daskalov IK, Dotsinsky IA, Christov II. Developments in ECG acquisition, preprocessing, parameter measurement and recording. IEEE Engineering in Medicine & Biology 1998; 17: 50-58.

[15] Christov II. Bidirectional high-pass recursive filter for drift suppression. Electrotechnics & Electronics E+E, ISSN 0861-47172004; 1-2: 43-51.

Address for correspondence

Irena Jekova.
Centre of Biomedical Engineering
Bulgarian Academy of Science
acad. G. Bonchev st. bl.105, Sofia 1113, Bulgaria
E-mail irena@clbme.bas.bgaddress.