

Comparison of Four Methods for Premature Ventricular Contraction and Normal Beat Clustering

G Bortolan¹, I Jekova², I Christov²

¹Institute of Biomedical Engineering ISIB - CNR, Padova, Italy

²Centre of Biomedical Engineering – Bulgarian Academy of Science, Sofia, Bulgaria

Abstract

The learning capacity and the classification ability for normal beats and premature ventricular contractions clustering by four classification methods were compared: neural networks (NN), K-th nearest neighbour rule (Knn), discriminant analysis (DA) and fuzzy logic (FL). Twenty-six morphology feature parameters, which include information of amplitude, area, specific interval durations and measurement of the QRS vector in a VCG plane, were defined. One global and two local learning sets were used. The local classifiers achieved better accuracies because of their good adaptability to the patients, while the capacity of the global classifier to process new records without additional learning was expectedly balanced by lower accuracies. NN assure the best results (high and balanced indices for specificity and sensitivity) using one of the local learning set, while the Knn provides the best results with the other local learning set. Using the global learning set DA and the FL methods perform better than the NN and Knn.

1. Introduction

Detection and classification of different types of heartbeats in the electrocardiogram (ECG) is of major importance in the diagnosis of cardiac dysfunctions. Some arrhythmias appear infrequently, and in order to capture them the clinicians use Holter devices. The use of specific algorithms for automatic analysis of ECG recordings may facilitate the analysis of the very long Holter ECG recordings.

Several algorithms for the discrimination between normal beats (N) and premature ventricular contractions (PVC) have been proposed in literature, some of them using heart beat morphology parameters [1-6] or frequency-based parameters [7,8].

In addition numerous classification methods have been studied, and they include: adaptive signal processing for on-line estimation of non-stationary signals that present a recurrent behaviour [9-13], linear discriminants [4,5], neural networks [14,15,3,8], fuzzy adaptive resonance

theory mapping [16], operation on vectors in the multidimensional space [6] and self-organized maps [17].

A particular aspect of the learning strategy is studied, paying attention to the organization of the classifiers' training set, and considering two main strategies: *local* learning set and *global* learning set [18,4,6]. In the first case the learning set is customized to the tested patient, while in the latter it is built from a large ECG database. The local learning set requires a cardiologist to annotate a set of normal (N) and premature ventricular contraction (PVC) heart beats of the patient under consideration. On the other hand, the capacity of the global learning set to classify new records without additional training is balanced by a lower accuracy, since the morphology of N and PVC differ not only from patient to patient, but also from lead to lead of a same individual.

In the present work, the local and global classifiers were investigated, considering 26 morphology heart beat parameters for the classification of normal beats and premature ventricular contractions in the electrocardiogram. For this purpose, the learning capacity and the classification ability of four classification methods were compared: neural networks (NN), K-th nearest neighbour rule (Knn), discriminant analysis (DA) and fuzzy logic (FL).

2. Methods and material

2.1. ECG database

All 48 ECG recordings from the Massachusetts Institute of Technology - Beth Israel Hospital (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 [19]. 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 Normals.

Since we focused only on the PVC classification, we followed the American Heart Association (AHA) records equivalent annotation [19], 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 Normal group [3,5,6]. In addition, fusion premature ventricular contractions, ventricular flutter waves, ventricular escape beats, blocked APB, 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.

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. It involves suppression of power-line interference, high-frequency electromiographic noise and low-frequency drift, described in details in [3,5].

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 (*AvI*), 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 QRS slope velocity (*SI*) between the beat onset, and the first peak and the slope velocity (*S2*) between the first and the second peaks were obtained. From the VCG signal, the following parameters were considered: the maximum amplitude of the vectorcardiogram vector in the plane formed by the two MIT-BIH leads (*VCGamp*) 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 Normal and PVC beats.

2.4. Classification methods

2.4.1. Neural networks (NN)

Multiple feed-forward neural networks were used for the classification of the QRS complexes. The back-propagation learning phase with an approximation of a second-order algorithm, based on Levenberg-Marquardt method, was applied for improving the speed of convergence. One input layer, one hidden layer and one output layer with one output node were used. The assessed heart beat was labeled as belonging to one of the two classes: N or PVC, which corresponds to the highest output of the neural network.

2.4.2. K-th nearest neighbour rule (Knn)

This classification rule operates on vectors in a multidimensional space [20,21]. In our case the heart beats are presented by 26-dimensional vectors:

$$\mathbf{x} = \{Width, Pp_1, Pn_1, Ar_1, ArP_1, ArN_1, Av_1, No_1, Ima_1, Imi_1, SI_1, S2_1, Pp_2, Pn_2, Ar_2, ArP_2, ArN_2, Av_2, No_2, Ima_2, Imi_2, SI_2, S2_2, VCGam, VCGsin, VCGcos\},$$

where index 1 and 2 indicate the two ECG lead from which the parameter is obtained.

The reference sets are formed by vectors of both N and PVC classes. Each vector was previously labelled as belonging to one of the two classes by experts' classification (in this case – the database signals annotations). A new vector \mathbf{x} (with unknown classification) is classified on the basis of the nearest vector from the used reference set. The distance between the vector \mathbf{x} and each of the clusters \mathbf{z}^j is computed as Euclidean distance to the mean vector of the corresponding cluster:

$$d_j = \sqrt{\sum_{i=1}^n (x_{i-norm} - z_i^j)^2} \quad (1)$$

where j is the cluster index, i is the parameter index, and n is number of the parameters used. Vector \mathbf{x} is classified to the class of the cluster \mathbf{z}^j at which d_j has a minimum.

All parameters used in this equation must be previously scaled with respect to the standard deviation, because their different magnitude or variances may cause unequal weight in the calculation of d_j with inappropriate effects:

$$x_{i-norm} = \frac{x_i}{\sqrt{\frac{1}{Nr} \sum_{j=1}^{Nr} (x_i^j - \bar{x}_i)^2}} \quad (2)$$

Here Nr is the number of all reference vectors and \bar{x}_i is the mean value of the i -th parameter.

2.4.3. Discriminant analysis (DA)

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

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

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

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 as corresponding to one of two classes: N or PVC, depending which of the values of F' or F'' was higher.

2.4.4. Fuzzy logic (FL)

A further method for the classification of QRS complexes was used with a partition of the feature space with a fuzzy clustering and a consequent fuzzy classification [22-24]. In this way a fuzzy inference system whose membership function parameters are tuned using a fuzzy clustering with a least squares type method was developed. The tested QRS complexes were classified as the type which corresponds to the maximal membership function.

2.5. Learning datasets

The particular aspect of the organization of the training strategies is considered, and one global and two local learning sets were investigated. The global learning set (G) consists of 260 vectors – 130 for N and 130 for PVC, taken almost equally from any of the patients in the database. In order to ensure unbiased classification accuracy we used the leave one out (LOO) rule. The first local learning set (L1) includes 76 vectors - 38 PVCs collected from all the patients in the database, and the first 38 heart beats of the tested patient (which are usually Ns). The second local learning set (L2) consists of the first 30% of the heartbeats of the tested patient with additional artificial PVCs, if necessary.

3. Results

In order to compare the classification ability of the four tested methods and the learning capacity of the three

learning datasets, we process all the QRS complexes of the 48 ECG recordings, described in section 2.1 and classified as N or PVC. The specificity, representing the accuracy of the classification of Normal QRS complexes, and the sensitivity, representing the PVC clustering for any of the tested classification methods with the three learning datasets are given in Table 1.

Table 1. Specificity (Sp) and Sensitivity (Se) of neural networks (NN), K-th nearest neighbour rule (Knn), discriminant analysis (DA) and fuzzy logic (FL) with the three learning datasets - global (G) and local (L1 and L2).

Classifier	Learning dataset					
	L2		L1		G	
	Sp [%]	Se [%]	Sp [%]	Se [%]	Sp [%]	Se [%]
<i>NN</i>	98.3	95.8	94.5	97.7	82.1	80.7
<i>Knn</i>	98.7	91.3	96.7	96.9	75.4	80.9
<i>DA</i>	94.4	97.0	95.4	94.8	88.5	81.7
<i>FL</i>	98.4	92.8	93.4	78.7	81.8	85.8

4. Discussion and conclusions

The local classifiers (L1 and L2) achieved better accuracies because of their good adaptability to the patients. On the other hand, the easier operation with the global classifier, which has the capacity to process new records without additional learning, was expectedly balanced by lower accuracies.

The performance of the four heartbeat classification methods is dependent on the use of local or global learning sets. With global learning set, the DA and the FL methods perform better than the other two classifiers. On the other hand, NN assure the best results (high and balanced Sp and Se) using the L2 learning set, while the Knn provides the best results with the L1 learning set.

The results, obtained in this study are helpful for selection of the proper type of learning set for each of the tested classifiers. Using all 48 ECG recordings from the MIT – BIH database we believe that we had provided reliable statistical outcome and adequate comparison of the assessed methods and learning datasets.

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] 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.
- [2] 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.
- [3] Christov I, Bortolan G. Neural networks for pattern recognition parameters ranking in premature ventricular contraction classification. *Physiological Measurements* 2004;25:1281-1290.
- [4] Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans Biomedical Engineering* 2004; 51:1196-1206.
- [5] Jekova I, Bortolan G, Christov I. Pattern Recognition and Optimal Parameter Selection in Premature Ventricular Contraction Classification. *Computers in Cardiology* 2004; 31:357-360.
- [6] Christov I, Jekova I, Bortolan G. Premature ventricular contraction classification by the K-th nearest neighbours rule. *Physiological Measurements* 2005;26:123-130.
- [7] Senhadji L, Carrault G, Bellanger JJ, Passariello G. Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Eng Med Bio Mag* 1995;14:167-173.
- [8] Herrero G G, Gotchev A, Christov I, Egiazarian K. Heartbeat classification using independent component analysis and matching pursuits. *IEEE, Int. Conf. Acoustics, Speech and Signal Processing, ICASSP'2005, Philadelphia, Pennsylvania, USA* 2005;4:725-728.
- [9] Ferrara ER, Widrow B. The time-sequenced adaptive filter. *IEEE Trans Acoustics, Speech, and Signal Processing* 1981;28:519-523.
- [10] Widrow B, Stearns SD. *Adaptive signal processing*. New Jersey: Prentice-Hall, 1985.
- [11] 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.
- [12] Laguna P, Jane R, Caminal P. Adaptive feature extraction for QRS classification and ectopic beat detection. *IEEE Computer Society Press* 1992; 613-616.
- [13] 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.
- [14] Al-Nashash H. Cardiac arrhythmia classification using neural networks. *Technol. Health Care* 2000;8:363-372.
- [15] Maglaveras N, Stamkopoulos T, Diamantaras K, Pappas C, Strintzis M. ECG pattern recognition and classification using non-linear transformations and neural networks: a review. *IJ Medical Informatics* 1998;52:191-208.
- [16] Ham FM, Han S. Classification of cardiac arrhythmias using Fuzzy ARTMAP. *IEEE Trans Biomedical Engineering* 1996;43:425-430.
- [17] Lagerholm M, Peterson G, Braccini G, Edenbrandt L, Sörnmo L. Clustering ECG complex using Hermite Functions and self-organizing maps. *IEEE Trans Biomedical Engineering* 2000;47:838-848.
- [18] Hu YH, Palreddy S, Tompkins WJ. A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Trans Biomedical Engineering* 1997;44:891-900.
- [19] Mark R, Moody G. *MIT-BIH Arrhythmia data base directory*. Cambridge: Massachusetts Institute of Technology, 1988.
- [20] Fukunaga K. *Introduction to statistical pattern recognition*. New York: Academic Press, 1972.
- [21] Kissiov VT, Hadjitodorov ST. A fuzzy version of the K-NN method. *Fuzzy Sets and Systems* 1992;49:323-329.
- [22] Pedrycz W, Bortolan G, Degani R. Classification of Electrocardiographic signals: a fuzzy pattern matching approach. *Artificial Intelligence in Medicine*, 1991;3:31-46.
- [23] Chiu S. Fuzzy Model Identification Based on Cluster Estimation. *J Intelligent & Fuzzy Systems* 1994;2:267-268.
- [24] Jang JSR. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Trans Systems Man and Cybernetics* 1993;23: 665-685.

Address for correspondence
Giovanni Bortolan
Institute of Biomedical Engineering ISIB - CNR
Corso Stai Uniti, 4
35127 Padova, Italy
E-mail: bortolan@isib.cnr.it