Automatic Classification of Heartbeats using Neural Network Classifier based on a Bayesian Framework

Conference Paper · August 2006

DOI: 10.1109/IEMBS.2006.4398330

2 authors:

George Karraz
Damascus University
18 PUBLICATIONS **40** CITATIONS
SEE PROFILE

Giovanni Magenes
University of Pavia
265 PUBLICATIONS **5,223** CITATIONS
SEE PROFILE

Automatic Classification of Heartbeats using Neural Network Classifier based on a Bayesian Framework

G.Karraz, G. Magenes

Department of Informatics and Systems, Università degli Studi, Pavia, Italy.

Abstract— **This paper presents a method of automatic processing the electrocardiogram (ECG) signal for the classification of heart beats. Data were obtained from 48 records of the MIT-BIH arrhythmia database[1] (only one lead). Five types of arrhythmic beats were classified using our method, Premature Ventricular Conduction beat (PVC), Atrial Premature Conduction beat (APC), Right Bundle Branch Block beat (RBBB), Left Bundle Branch Block beat(LBBB), and Paced Rhythm Beat (PRB), in addition to the Normal Beat (NB). A learning dataset for the neural network was obtained from a five records set (124, 214, 111, 100, and 107) which were manually classified using MIT-BIH Arrhythmia Database Directory and documentation, taking advantage of the professional experience of a cardiologist. Feature set was based on ECG morphology and time intervals. Our system resulted in a minimal sensitivity of 86% and minimal specificity of 90%.**

Keywords—**Automatic ECG analysis, Ectopic beats, Neural networks**

I. Introduction

Each year, arrhythmia disease claims approximately 400.000 lives, and costs up to $ 30 billion in combined health care spending in France, Germany, Italy, Spain, and in the United Kingdom [2].

Heart arrhythmia results from any disturbance in the rate, regularity, site of origin, or conduction of the cardiac electric impulse.

Arrhythmias can be divided into two groups. The first one is a life threatening and require immediate therapy, like as the ventricle fibrillation. Detection of these arrhythmias is well researched, and successful detectors have been designed. The second one which is investigated in this study, includes arrhythmias that are not imminently life-threatening, but may require therapy to prevent future problems.

In this work we consider five type of arrhythmic beats:

1-Premature Ventricular Conduction (PVC) is an ectopic cardiac pacemaker located in the ventricle. PVCs are characterized by the premature occurrence of bizarre shaped QRS complexes, usually greater than 120 ms. These complexes are not preceded by the P wave, and the T wave is usually large and opposite in direction to the major deflection of the QRS.

2- Atrial Premature Conduction (APC) is an ectopic cardiac pacemaker located in the atrias; APCs are characterized by the premature occurrence of QRS.

3 & 4- *Blocks of the cardiac conduction system*: they cause prolonged PR interval more than 0.2sec. In this sense cardiac blocks were classified in three levels. The first one is caused by intra-atrial, AV node, or intra-ventricle slow conductions. The second level shows either the Wenckebach form (increasing PR interval until not connected P wave), or the Mobitz one (fixed PR interval and non connected P wave). The third level causes prolonged PR interval, but P wave and QRS complexes are dissociated.

Our classificator takes as a case of study the slow conduction in the left and right ventricles, that caused by a left or right bundle branch blocks (LBBB, or RBBB respectively). These types of arrhythmia show a prolonged PR interval and a wide abnormal QRS complex.

5 -Paced Beats: they are a sequence of 3 or more junctional escapes. There may be an AV dissociation, or the atria may be captured retrogradely by the junctional pacemaker.

The main goal of this work is to realize a robust classifier able to identify all the above cited types of ectopic beats by using neural networks on a Bayesian framework.

II. Methodology

1- Pre –Processing: It is intended for filtering and reducing the influence of noise superimposed on the ECG. First, a pass-band (0.05-40Hz) FIR filter was used to remove the continues shift and draft baseline. Its coefficients were designed with Sptool [4]. Second, a notch filter was used to reduce the 60Hz powerline interference. Finally the signal was resampled to obtain a sampling frequency of 1000 s/s.

2- QRS detection algorithm: In order to enhance the desired signal features, the QRS detection represents the first step to be executed. Many algorithms were proposed in the literature [5], [6] and guarantee a good performance level. In this paper we developed a QRS detection algorithm based on a non linear operator which exploits the curve-length concept [7]. Consider $X = [x_1, x_2, \ldots . x_n]$ a vector representing the ECG signal, since the output of this non-linear operator is:

$$U_i = U_{i-1} - (X_{i-w-2} - X_{i-w})^2 + (X_{i+w-2} - X_{i+w})^2 \quad (1)$$

$i \in (1 \ldots n)$, and w is the length of the processing window, that corresponds to the QRS width. Figure 1 shows the application of this filter on a MIT-BIH database recording after the pre-processing was performed .

As we see, there are two main type of components: the first one corresponds to the QRS complexes, and the second one corresponds to the T and P waves, that disturb QRS detection process.

In order to eliminate it and to threshold the QRS complexes, we propose a new threshold method based on the fuzzy clustering c-mean algorithm [8].

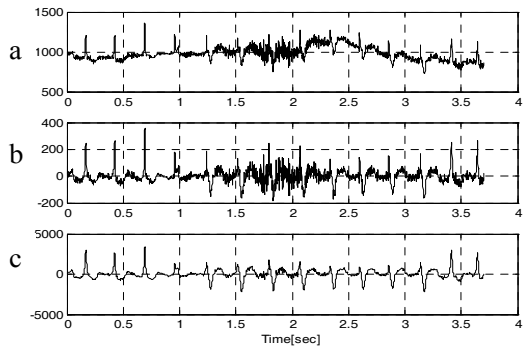

Figure1. Processing steps from the upper to the lower tracings; signal, Pass-band filter, Notch filter)

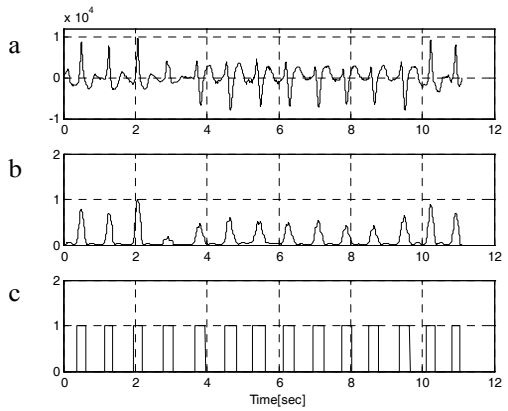

Figure 2. QRS-detection (a. signal, b. Non-linear operator (normalized output between 0-1), and c. thresholding result)

Let the output of the nonlinear operator be $U=[u_1,.....u_n]$, we want to partition it into J clusters, each represented by a centre *cent* $_j$.

Denote C the partition matrix made by rows C_j, where $C_j=[c_{j,1}\ .....c_{j,n}]$ and each element represents the membership degree of the data vector U in the Jth cluster ($j=1,..J$) and ($i=1,..n$). The fuzzy clustering algorithm searches for the partition matrix and cluster centres such that the objective function E is minimized

$$E=\sum_{j=1}^{J} C_j^m d^2(U,cent_j) \qquad (2)$$

subject to

$$\sum_{j=1}^{J} C_j = 1 \qquad (3)$$

the parameter m controls the fuzziness of the clusters (usually $m=2$).The function $d_j=d(U,cent_j)$ measures the distance between the data vector U and the centre j of the jth cluster.

$$d^2(U,cent_j)=(U-cent_j)^T M_j(U-cent_j) \qquad (4)$$

where M_j is a positive definite value adjusted to the actual shapes of the jth cluster. M_j is defined as:

$$M_j=\sqrt[n]{F_j}F^{-1} \qquad (5)$$

where F_j is the cluster covariance

$$F_j = C_j^m (U-cent_j)(U-cent_j)^T \qquad (6)$$

Given a data vector U, we choose the number of clusters J (in our case 2 or 3), the weighting coefficient m, and the termination tolerance ξ. The partition matrix C is initialized randomly in such a way, that condition (3) is satisfied. Then the algorithm iterates through the following steps:

1- Determine the cluster prototypes $cent_j$ *(j=1,...J)*

$$cent_j = \frac{\sum C_j^m U}{\sum C_j^m} \qquad (7)$$

2- Calculate F_j according to equation (6).

3- Compute the distance d_j^2 according to (4).

4- Update the partition matrix $C_j = \frac{1}{(d_j)^{2/(m-1)}}$ (8)

if $d_j(i)$=0 then take $C_j(i)$=1.

5- iterate until $\left\|C-C^{-1}\right\| \le \xi$.

6- Take the C_j vector that corresponds to the max centre of clusters, and threshold it by choosing the threshold value, according to the standard deviation value (STD) of the signal vector:

If $C_j(i) > std * max(C_j) \Rightarrow C_j(i)=1 \Rightarrow$ ***QRS complex.***

else if $C_j(i) \le std * max(C_j) \Rightarrow C_j(i)=0 \Rightarrow$ ***P&T waves.***

Three fiducial point were determined, the QRS-on, the QRS-off, and the R-peak or amplitude.

Figure (2.c) illustrates the performance of this method on the signal reported in Figure 1.

A control procedure was added to the QRS detection process for:

1. controlling the periods that correspond to the distance between the QRSs detected. In the case of a too long distance with respect to the preceding one (based on signal STD), it means that there are missed QRSs. So the algorithm searches them using another adaptive threshold value, less

than the prevoius one. Then the procedure repeats this step, until all the missed QRSs are detected.
If the STD is negligible, this means that there is a signal pause, and the procedure stops to search the QRSs.
2. correcting the width of all detected QRSs, in order to ignore the very small ones with respect to the mean of the others. A false positive QRS might be caused by the high T-wave amplitude.

3-T&P waves' detection: In order to detect the T-waves, the previous algorithm was used within the time window between the QRS complexes. Three fiducial points were determined at every beat, the T-wave on the T-wave off, and T-wave amplitude. After the T-wave was been detected, the P-waves were searched within the distance between the present T-wave and the successive QRS, using the same algorithm. Other three fiducial points were determined, P wave on, P-wave off, and P-wave amplitude.

4-Feature extraction: From the two former steps ten indicative parameters were extracted representing the features of each beat: 1) P-amplitude, 2) P-wide, 3)R-amplitude, 4) Q-amplitude, 5) S-amplitude, 6) QRS-wide, 7) T-amplitude, 8) T-wide, 9) PR-period, and 10) RR-period.
They were used to build the input vector of our neural network.

5-Bayesian Artificial Neural Network (BANN): Bayesian neural networks can be found in related literature [9], [10], and [11]. Our Bayesian ANN classifier's configuration is shown schematically in figure 3, where *i=1,2,...10* is the index of input nodes, *j=1,2,...J* is the index of hidden nodes, and *k=1,2,..K* is the class label of the output nodes. Assume we have the training set *D*, consisting of *N* input-output pairs:

$$D = \left[\left(x^n , y^n \right) | \, n = 1, 2, .. N \right] \qquad (9)$$

where x is an input vector consisting of I elements, every input vector corresponds to a heart beat, and y is the corresponding class label consisting of K classes. The goal is to use an ANN to model the input-output relation $(y = k \mid x)$. In our case we have 6 classes $k \in (k1, k2, k3, k4, k5, k6)$, where the first five classes correspond to the five types of arrhythmia. Any one of them takes a values (-1) for the normal, and (1) for the arrhythmic heart beat. The sixth class corresponds to the normal case, and takes a value (1) for the normal beat and (-1) in the case of arrhythmic one. To realize a logistic regression model based on a Bayesian method, we estimated the class probability for the given input by: $P(Y=k \mid X)$.
Assume that the outputs for the summation operation, and the sigmoid activation function, in the hidden and output neurons, denoted by *Sj, Sk* respectively can be written as follows:

$$S_j = tanh\left(\sum_i \omega_{ij} X_i + \omega_{j0} \right) \qquad (10)$$

$$S_k = \left(\sum_k \omega_{kj} S_j + \omega_{k0} \right) \qquad (11)$$

where *tanh* is the tangent hyperbolic function. ω_{ij} denotes the weight matrix in the input layer and ω_{kj} the weight matrix in the output layer. To ensure that the outputs can be interpreted as probabilities, logistic regression is used to model the risk (or probability) of occurrence of arrhythmia. Let $P(Y=k \mid X)$ be the probability of the event $Y=1$, given the input X. This is modelled by function of a network output y by:

$$P(Y=k|X) = \frac{1}{1+exp(-S_k)} \qquad (12)$$

The logistic regression model is simply a non linear transformation of the linear regression. The 'logistic' distribution is a *S-shaped* distribution function which is similar to the standard-normal distribution (which results in a probit regression model) but easier to work within most applications. The logistic distribution constrains the estimated probabilities to lay between 0 and 1.
The final part of the system is a multi-layer perceptron neural network, trained using the BNN algorithm. The network is optimized using a log-likelihood cost function, given by :

$$C(\omega) = -\frac{1}{k} \sum_k \sum_i Y_i(k) \ln\left[(Y=k|X) \right] \qquad (13)$$

Where $\omega=(\omega_{ij}, \omega_{jk})$ is the vector of network weights. To minimize the cost function between the actual and desired outputs of the network, the BNN algorithm passes the information from the output neuron backwards to all hidden units to form error terms which are used to update the weights of the multi-layer network.

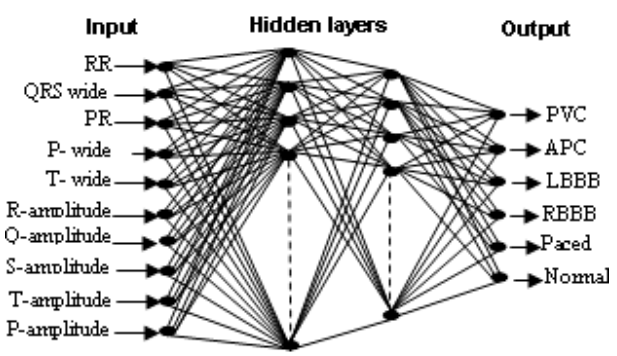

Figure3. Neural network structure

III. Results

In order to determine the performance of our classifier in recognizing arrhythmias, we used an evaluation method based on the calculation of system sensitivity and specificity, which determine the classifier ability to recognize the abnormal and normal cardiac rhythm respectively. In addition we used ROC curve to graphically represent the performance of our classifier.

Let NP be the number of positive (abnormal), and NN the number of negative (normal) cases. Let again TP be the true positive, TN the true negative, FN the false negative, and FP the false positive cases as they are classified from our system. The sensitivity and specificity are denoted and calculated respectively as follows:

$$S_n = \frac{Tp}{NP + NN} \qquad S_p = \frac{TN}{NN + Np} \qquad (14)$$

To calculate the percentage error, the normal and pathological cases must be equally distributed. To this purpose we use two weights ω_p for abnormal cases and ω_n or the normal one, so $\omega_n NN = \omega_p NP$

TABLE I
SYSTEM SENSITIVITY AND SPECIFICITY

Sensistivity (Average)	Specificity (Average)	Error (Average)
96.26%	95.86%	0.24%

Let $\omega_n = 1 \Rightarrow \omega_p = \frac{NN}{NP}$, the percentage error is:

$$P(e) = \frac{FP + \omega_p FN}{NN + \omega_p NP} \qquad (15)$$

We illustrate in Table I the resulting average of sensitivity, specificity, and error, applied to MIT database records, and in Figure 4 the results of ROC curve.

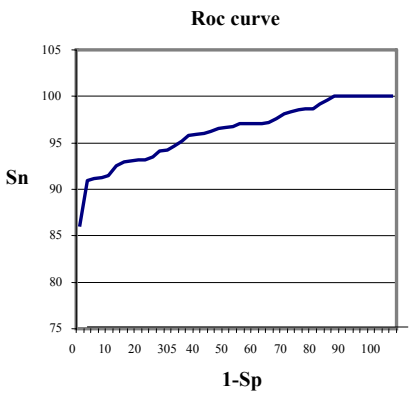

Figure 4 – ROC curve of the classifier

IV. Discussion and Conclusions

In this paper, an automatic classification for arrythmia detection is presented, and its performance is evaluated by measuring the system sensitivity, specificity and error after the all MIT-BIH positive and negative cases have been normalized. The system errors are caused by two main reasons: 1-The pre-processing and filtering processes in some cases, are not enough to remove all type of noise affecting the the ECG signal, and sometimes they cause a signal distortion, so the QRS detection and the feature extraction procedure don't perfom as well as they are supposed to do; 2- Other types of arrhythmia are rarely present in some records; thus we haven't sufficient cases to be inserted in the dataset and they resulted as unrecognized patterns from our system.

The future work will focus on developing our ECG filtering method to reduce the first cause of system error, and we will try to find more data containing different cases of arrhythmia to overcome the second one. We hope this system can be further developed, in order to realize a low-cost, high performance, simple to use and portable equipment for ECG signal monitoring.

REFERENCES

[1] MIT-BIH Arrhythmia Database Directory (Records), Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center). Available: http://www.physionet.com.

[2] European Markets for Arrhythmia Management Products, 278 Pages Report # RP-481420, February 2001.

[3] F. G. Yanowitz, M.D., ECG Learning Center. University of Utah School of Medicine Medical Director, ECG Department LDS Hospital Salt Lake City, Utah.. Available http://medstat.med.utah.edu/kw/ecg/..

[4] G. Clifford, Matlab software, http://www.gnu.org/

[5] S.E. Dobbs, N.M. Schmitt, H.S. Ozemek "QRS detection by template matching using real-time correlation on a microcomputer" *Journal of Clinical Engineering*, 9: 197-212, 1984.

[6] Hamilton, P.S. and Tompkins, W.J. "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database" Trans.Biomed.Eng. BME-33: 1157-65, 1986

[7] C. Marchesi, M. Paoletti "Model based characterisation for long-term personal monitoring". The Internet Journal of Medical Technology. Volume 1, Number 2, 2004.

[8] S. Osowski, and Tran Hoai Linh, " Fuzzy clustering neural network for classification of ECG beats", International Joint Conference on Neural Networks, (IJCNN 2000)-Volume 5, 2000.

[9] D.E. RRumelhart, G. E. Hinton, R. J. Williams, Learning internal representation by error propagation. In: Rumelhart, D. E., McClelland, J. L, and the PDP research group (Eds), Parallel distributed processing, MIT press, Cambridge, MA 318-362, 1996..

[10] D. Gao, Y. Kinouchi, K. Ito, X. Zhao, Neural Networks For event Extraction from Time Series: A Back Propagation Algorithm Approach. Future Generation Computer Systems (FGC), Volume 21, No 7.,1096-1105, 2005.

[11] D. Gao, Y. Kinouchi, K. Ito, X. Zhao, Time Series Identifying and Modeling with Neural Networks. In: Proc. IEEE Int. Joint Conference on Neural Networks, Portland, USA.; 2454-59, 2003.

[12] D. Gao, M. Madden, M. Schukat, D. Chambers, G. Lyons. "Arrhythmia Identification from ECG Signals with a Neural Network Classifier Based on a Bayesian Framework", the Twenty-fourth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, December 2004.