Efficient ECG multi-level wavelet classification through neural network dimensionality reduction

Rodrigo Varejão Andreão[1], Bernadette Dorizzi[1],
Paulo César Cortez[2] and João César Moura Mota[2]
[1]INT, dépt EPH, 9 rue Charles Fourier 91011Evry France
[2]UFC, Deti, CP 6005 Campus do Pici 60455-760 Fortaleza Brasil
[1]{Rodrigo.Andreao, Bernadette.Dorizzi}@int-evry.fr
[2]{cortez, mota}@deti.ufc.br

Abstract - In this article, we explore the use of a unique type of wavelets for ECG beat detection and classification. Once the different beats are segmented, classification is performed using at the input of a neural network different wavelets scales. This improves the noise resistance and allows a better representation of the different morphologies. The results, evaluated on the MIT/BIH database, are excellent (97,69% on the Normal and PVC classes) thanks to the use of a regularization technique.

1. INTRODUCTION

ECG signal processing has been the subject of an intense research in the past years due to its strategic place in the detection of several cardiac pathologies. In this article, we propose the joint use of wavelets and neural networks to segment heartbeats in the ECG records of the patients and to classify these beats according to predefined pathologies. This approach has already been proposed by other authors [12,1,3,11].

Indeed, ECG signal is always affected by noise, baseline draft and artifacts. Hence, it is difficult to objectively measure the time intervals and subsequently use them in the classification of disease. Wavelet transform provides a powerful tool that allows decomposing the signal into components that appear at different scales. Thus, even in the presence of noise, the important information can be extracted from the signal. On the other hand, artificial neural networks (ANN) have been shown to be very efficient classification tools, and it is rather tempting to combine these two efficient and complementary techniques.

However, our originality relies on the following points.

First, we use the same wavelet transform to perform beat detection (according to Cuiwei Li approach [8]) and to compute the features that will be used at the input of the network. We choose as mother wavelet, the Mexican hat. Most authors use at the input of the network features that have been extracted from the segmented

This work was supported by the CAPES/COFECUB project n° 266/99-I.

0-7803-7616-1/02/$17.00 (c) 2002 IEEE

beats, like widths of the QRS complex by example. We have chosen to present at the network input, a downsampling of the wavelet filtered signal at some scales.
Moreover, we will use several filtering scales which will allow a better extraction of the informant parts of the signal, especially in presence of noise. Thus, the ANN will have an important number of inputs, which has been proven to lead to bad generalization performance when the training database has an insufficient size. In order to overcome this difficulty, we use a rather efficient regularization scheme [9], which permits an important reduction of the number of independent parameters. This way, only the input features that are necessary to achieve a good classification are kept in the network.
Finally, we performed a lot of experiences on the well-known MIT/BIH database, using a large amount of data, aiming at showing the quality of our approach. We trained a system with records of several patients and we tested its robustness to new patients who did not participate to the system training. Concerning the number of classes, different situations are tested, namely the case of two or more diseases.

In section 2, we describe the database at hand (MIT/BIH) as well as the division we performed on it. In section 3, we recall the QRS detection procedure, adapted from the one of Cuiwei Li [8]. The wavelet family we have considered is described in this section. In section 4, we report the regularization technique which has been applied to the neural network, and finally section 5 presents our experiments.

2. THE ECG DATABASE

This work uses a two-channel ECG database called MIT-BIH that contains records of many patients with heart troubles or anomalies [6]. This database has been extensively used in the literature so far [3,7,8,12]. It presents many examples of heart anomalies such as ventricular premature beat (PVC), atrial premature beat, fusion beats, left bundle branch block, right bundle branch block and others. The data are sampled at 360 Hz. Only channel 1 was used in the experiments presented in this work. Each record has its respective annotation file, which was made by cardiologists. The annotations indicate the position of the QRS complex, and also the class of the beat.
In most of the works which have been performed using this database, tests performances are calculated on only very few data. On the contrary, in our work we use an important part of the entire database which was divided in two groups corresponding to different patients. Group 1 contains 13 records while group 2 contains 12 records. The record labels have been chosen in order that the number of examples of each anomaly is well balanced between group 1 and 2.
Data of group 1 will be used in order to train and test the recognition system while data of group 2 will only be used in order to test the robustness of the system to new patients (whose data had not been learned so far).

3. QRS DETECTION

QRS detection consists in finding the point R of the heartbeat (Figure 1) which is in general the point where the heartbeat has the highest amplitude. This detection allows also the evaluation of the heart beat rate by measuring the distance between two adjacent QRS. Besides, QRS detection is fundamental for the classification step because each beat is later analyzed separately.

A lot of methods have been proposed in order to perform this task. We have followed the one developed by Cuiwei Li [8] because of its good performance. It uses multi resolution analysis by wavelet transform. We will just recall how this method works and present the results obtained on our database.

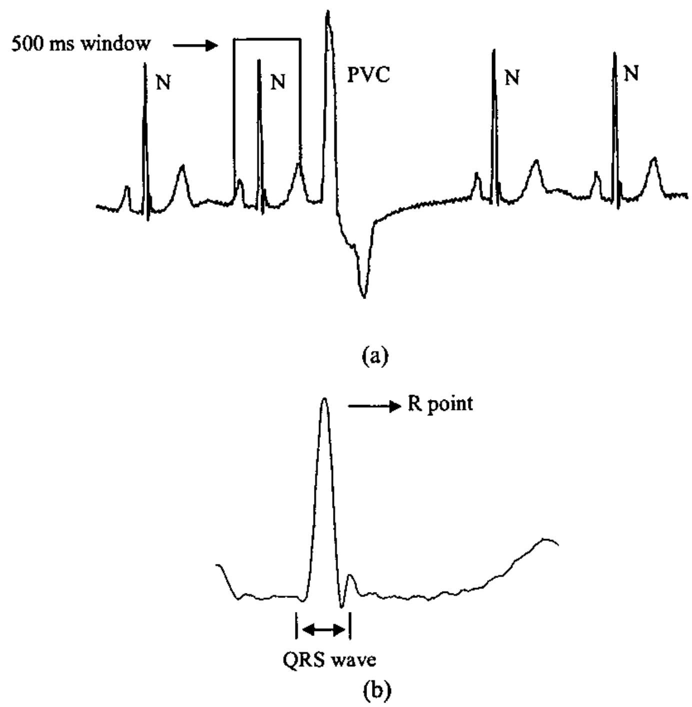

Figure 1. One ECG signal. (a) Four beats with labels N (normal) and PVC (b) Detected beat extracted by a 500 ms window.

Wavelet Transform

The wavelet transform represents the signal in a scale-time space, where each scale is the result of a pass-band filtering. The frequency bands depend on the scale and also on the type of the chosen wavelets. The type of the wavelets is specified by the choice of a mother wavelet, and in the case of Cuiwei Li, the quadratic spline (first derivative of the gaussian function) was chosen. However, in our work, we preferred the Mexican Hat (second derivative of the gaussian function) as the mother wavelet:

$$\psi(t)=\frac{1}{\sqrt{2\pi}}\left(1-t^2\right)e^{\left(-\frac{t^2}{2}\right)} \tag{1}$$

where t is the time and $\psi(t)$ is the mother wavelet with zero mean.

We do indeed believe that the shape of this function is more adapted to the shape of the QRS signal than other wavelets that can be found in the literature.

From the mother wavelet, the class of wavelets is then [10]

$$\overline{\psi}_s(t)=\frac{1}{\sqrt{s}}\psi^*\left(\frac{-t}{s}\right) \tag{2}$$

where $\overline{\psi}_s(t)$ is a wavelet in the scale s and ψ^* represents the wavelet complex conjugate. In this work, the scale s is called dyadic (it is a power of 2).

Thus, the wavelet transform is represented as

$$Wf(t,s)=f*\overline{\psi}_s(t) \tag{3}$$

where f is the signal.

For more details on wavelet transforms, we refer to [10].

Detection method

The signal is analyzed over consecutive blocks of 500 samples. In the detection procedure, the wavelet transform of the ECG signal is calculated over four scales. The choice of the scales depends on the mother wavelet adopted. For this work, scales 2^2, 2^3, 2^4 and 2^5 were selected. The main idea of the method is to use the complementary information obtained from each wavelet scale in other to improve the detection performance. As we have already mentioned, the noise influence over the ECG signal may cause many detection mistakes. Thus, a right wavelet scale with a good signal-noise ratio makes possible the detection of the ECG important regions. Then, using an iterative process, these regions are analyzed on other wavelet scales which allow better resolutions and more precise detection of the point R.

In order to take into account the variation of the signal amplitude, an adaptive threshold has to be calculated.

The algorithm description and some details are available in [8].

Evaluation of the QRS detection was made over 25 records. The automatically detected beats were compared to the beat labels, which are at disposal in this database. Following the AAMI recommendation [4], we used a beat-by-beat matching within an interval of 150 ms between the beat labels and those detected automatically. Two types of errors can thus occur: when there is no beat detection in the interval of 150 ms around the labeled R point, a false negative (FN) detection is noted, and if the detected beat is outside the interval, a false positive detection is computed (FP). Table 1 shows the detection results on the whole database (group 1 + group 2).

Most of the errors are due to the rapid decrease of the signal amplitude (this is particularly true with record #203), which is always a problem for the threshold adaptation.
The method is robust to the noise that may appear on the ECG signal. This is a good reason for using the same wavelet transform in order to generate the parameters for the classification phase. This way, we minimize the computation complexity as segmentation of the signal and feature extraction will use the same process.

ECG record name	Total (beats)	FP (beats)	FN (beats)	Errors (beats)
100	2273	0	0	0
101	1867	2	1	3
102	2186	0	1	1
103	2084	0	0	0
104	2228	3	1	4
105	2571	37	4	41
106	2020	1	7	8
107	2136	0	0	0
118	2278	0	0	0
119	1984	5	3	8
200	2596	4	4	8
201	1941	0	22	22
202	2134	0	2	2
203	2907	17	95	112
205	2653	0	3	3
207	1856	7	5	12
208	2926	2	27	29
209	3005	1	0	1
210	2636	2	15	17
212	2748	0	0	0
213	3250	0	1	1
214	2257	2	5	7
215	3362	0	1	1
217	2207	1	0	1
219	2154	0	0	0
Totals	60259	85	197	282 (0,47%)

Table 1. QRS detection results

4. CLASSIFICATION OF ECG BEATS

The classification task consists in associating a label to each heartbeat case. Neural networks have proved their efficiency for performing this kind of tasks [7,1,3,11]. Among the different neural network architectures available, in this work, we have used the multilayer perceptrons (MLP).

Preprocessing

As was explained in the last section, the classification task uses the same features as those used for QRS detection. However, to reduce the amount of data, we took a window of 500 ms (180 samples) around the detected QRS complex and performed a downsampling over each scale, followed by a Hanning window multiplication to ensure the end points are zero [2]. Only three scales were selected: 2^3, 2^4 and 2^5. Scale 2^2 was removed because it suffers from the influence of noise. The downsampling, respecting the frequency components of each scale, was done at a ratio of 1/3. As a result, each scale corresponds to 60 features.

Network training

Three database subsets were constructed from group 1: training subset, validation subset and a test subset. All these subsets contain examples of each patient. On the other hand, group 2 database is not considered during training. It is used to evaluate the classifier performance robustness to other patients than those considered during the training phase of the network.

A neural network with one hidden layer was trained. After some empirical trials, the hidden layer size has been fixed to 10 neurons. The number of neurons in the input layer depends on the number of features that one wants to consider. One of the interests of considering neural networks as classifiers is their ability to fuse a great number of features. In this work, we will consider at the input of the network different reconstructions of the signal at different scales. However, when the training database is not large enough, an oversized network considerably degrades the network performance. It is thus necessary to perform some features reduction strategies.

Some other works use wavelet features as inputs in a neural network, but they commonly proceed a features reduction before the network training phase [1,3,11]. This requires a priori computation of the features that should be useful. Instead, in this work we will use a training procedure, called Bayesian regularization [9], which prunes the network weights (and hence the features) during the learning process. Thus, the features, which do not improve the classification accuracy, are eliminated automatically, while those that help this process are kept.

Bayesian Regularization

Regularization is employed, for example, to solve the neural network over-fitting to the noise on the training data caused by the excess of neural networks free parameters. Thus, reducing the size of the network weights, a smooth mapping can be achieved and hence significant improvements in network generalization may result. To proceed the regularization, an additional term is put in the error function

$$\tilde{E} = \beta E_D + \alpha E_W \tag{4}$$

where E_D is the sum of the squared errors (these errors represent the difference between the network response and the target), E_W is the sum of squares of the networks weights, and α and β are the objective function parameters. In [9], MacKay proposes a solution for the computation of α and β parameters through a Bayesian framework. Furthermore, a measure of the network weights reduction at the end of the procedure is provided. The effective number of weights is represented by the parameter γ_E and the total number of weights by the parameter γ_T.

The optimal regularization technique requires the computation of the Hessian matrix and then, as introduced by Foresee [5], its Gauss-Newton approximation is computed by the Levenberg-Marquard algorithm.

5. EXPERIMENTATION AND RESULTS

Several experiments have been developed to evaluate the classification performance. The total and effective number of network weights is also reported. First, only two classes of beats were analyzed: normal (N) and ventricular premature beat (PVC). Second, all the beats, excluding the N and PVC ones, were combined in a third class, which corresponds to 9 different labels in MIT/BIH database. This procedure is important to test the rejection capability of the N/PVC classifier. Finally, some classification errors are reported and the noise robustness is discussed. The number of examples of each class is shown in Table 2.

Class	Group 1 Subsets			Group 2
	Training	Validation	Test	
N beats	1963	400	1192	3343
PVC beats	1295	224	650	2533
Other	2847	492	1441	8966

Table 2. Number of examples for each beat class

Two Classes Results

The classifier performance is evaluated in terms of good beat classification on group 1 and group 2 separately. We have also tested different scales at the input of the network (namely 2^2, 2^3, 2^4 and 2^5). The number of output neurons was fixed at two, each neuron corresponding to a class.

Features	γ_T	γ_E	Group 1 (test subset) (%)			Group 2 (%)		
			N	PVC	Total	N	PVC	Total
Scale 2^3	632	217	96.47	98.00	97.01	99.43	94.75	97.41
Scale 2^4	632	156	97.57	99.38	98.21	97.46	94.55	96.20
Scales 2^3 and 2^4	1232	238	96.98	98.15	97.39	99.55	95.23	97.69

Table 3. Classification results for 2 beat classes

Three Classes Results

This time, we consider all the pathologies but our aim is not to build a classifier able to classify them all, as we do not have at our disposal enough examples of each class. We will, instead, test the ability of a network trained in order to discriminate between N and PVC to reject the examples belonging to other classes. For this, we build a neural network with two outputs, and we train it on examples labeled N, PVC and Other, with the following target output: N→(1,0), PVC→(0,1) and Other→(0,0). During the test phase, a threshold has to be defined in order to reject the examples whose output will be near (0,0) (note that we have also built a network with three output neurons which has shown poorer results than this one).

Features	γ_T	γ_E	Group 1 (test subset) (%)				Group 2 (%)			
			N	PVC	Other	Total	N	PVC	Other	Total
Scale 2^3	632	432	94.30	94.62	95.63	94.94	76.85	83.10	33.07	53.89
Scale 2^4	632	261	95.05	91.38	96.25	94.85	83.52	80.38	56.77	66.82
Scale 2^5	632	163	93.04	93.08	95.49	94.12	54.44	57.60	73.44	66.46
Scales 2^3 and 2^4	1232	462	94.21	96.31	96.32	95.55	70.95	86.66	60.93	67.58
Scales 2^4 and 2^5	1232	311	94.38	93.85	97.09	95.46	64.19	62.50	66.98	65.58
Scales 2^3, 2^4 and 2^5	1832	440	95.47	95.08	95.07	95.22	61.59	74.85	86.25	78.75

Table 4. Classification results for 3 beat classes

Results analysis

From Table 3, we observe the regularization benefits to network generalization. Indeed the classification results of group 2 are still better than the classification results of the group 1 test subset.

The improvement in classification performance due to the combination of several scale features is clearly observed in Table 4. It can be noted that the best performance rate of the different classes is not obtained using, at the input of the network, the same scale: class PVC – scale feature 2^3, class N – scale feature 2^4 and class Other – scale features 2^5. Hence, as the database is composed of a great

variety of morphologies, using several features scales allow a better modelization of the different morphologies.
We verify from Table 4 that the inclusion of the class Other caused an important performance reduction in the classification results of group 2. This happened because many examples of classes N and PVC of group 2 have been confused with the class Other. Anyhow, the errors between classes N and PVC remained almost unchanged in relation with the 2 classes results. Looking at the examples of class Other in group 2, we remark that a lot of them are similar to N or PVC examples, which seems to be less frequent in group 1. This may explain why the performance of N and PVC classes decrease significantly in group 2.
For the neural network trained with 3 classes, we evaluated the classification results over the QRS wave that were falsely detected (Table 1). As a result, 41.7 % of the 85 false detection examples were classified as belonging to the class Other. Therefore, the classifier could help to reject bad detected beats.
It is important to remark that the same anomaly may correspond to different morphologies. This is commonly observed between different patients, but this can also be true in the same patient record, which makes the automatic classification task a hard one.

6. CONCLUSION

In this article, we have presented a technique for ECG beat detection and classification using the same mother wavelet, namely the Mexican Hat. We use several wavelet scales at the input of the neural network. This coding has the interest of being very resistant to noise, while offering a good representation of the different morphologies. Despite the great complexity of the resulting network, good performances are obtained thanks to the use of a regularization technique. Our test protocol involves two groups of patients. Group 1 is used to train and evaluate the network, while group 2 allows the test of the system on data corresponding to patients, different from those used during the training phase. Results on a large subset of the MIT/BIH database show very good performance for a 2 classes discrimination test (N and PVC), on both groups 1 and 2. Finally, other pathologies than N and PVC, which are presents in the MIT/BIH database, can also be fairly rejected. Comparisons with others systems using the same database are difficult to make as in general in these works the test set is further less important than ours and the experimental conditions are different (nature and number of classes). One can only say that our results are in the norm of those commonly encountered in the literature.
These preliminary results suggest some further work.
We think of regrouping into the same class, morphologies that look the same but do not correspond to the same label. This way, the global classification task will be made easier.
Moreover, the noise robustness of this approach will be tested on more degraded signals (acquired with more simple sensors).

REFERENCES

[1] B.G. Celler, "Low computational cost classifiers for ECG diagnosis using neural networks," **Proceedings of the 20th International Conference of the IEEE Engineering in Medicine and Biology Society**, vol. 20, no. 3, pp. 1337-1340, 1998.

[2] P.de Chazal et al, "Classification of the electrocardiogram using selected wavelet coefficients and linear discriminants," **Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing**, vol. 6, pp. 3590-3593, 2000.

[3] Z. Dokur et al, "ECG waveform classification using the neural network and wavelet transform," **Proceedings of the First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology,** pp. 273, October 1999.

[4] Association for the Advancement of Medical Instrumentation (AAMI), "Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms," **Tech. Rep. AAMI ECAR-1987,** AAMI, Arlington, VA, 1987.

[5] F.D. Foresee and M.T. Hagan, "Gauss-Newton approximation to bayesian learning," **IEEE International Conference on Neural Networks,** vol. 3, pp. 1930-1935, 1997.

[6] A.L. Goldberger et al, "PhysioBank, PhysioToolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals," **Circulation** 101(23):e215-e220.

[7] Y.H. Hu et al, "A patient-adaptable ECG beat classifier using a mixture of experts approach," **IEEE Transactions on Biomedical Engineering,** vol. 44, no. 9, pp. 891-900, September 1997.

[8] C. Li et al, "Detection of ECG characteristic points using wavelet transforms," **IEEE Transactions on Biomedical Engineering,** vol. 42, no. 1, pp. 21-28, January 1995.

[9] D.J. MacKay, "Bayesian Interpolation," **Neural Computation**, vol. 4, pp. 415-447, 1992.

[10] S. Mallat, **A Wavelet Tour of Signal Processing**, Academic Press, 1998.

[11] S. Wu et al, "A novel method for beat-to-beat detection of ventricular late potentials," **IEEE Transactions on Biomedical Engineering,** vol. 48, no. 8, pp. 931-935, August 2001.

[12] M.Y. Yang et al, "ECG events detection and classification using wavelet and neural networks," **Proceedings of the 19th International Conference of the IEEE Engineering in Medicine and Biology Society**, pp. 280-281, 1997.