An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance

Mariano Llamedo* and Juan Pablo Martínez

***Abstract*—In this paper, we present a patient-adaptable algorithm for ECG heartbeat classification, based on a previously developed automatic classifier and a clustering algorithm. Both classifier and clustering algorithms include features from the RR interval series and morphology descriptors calculated from the wavelet transform. Integrating the decisions of both classifiers, the presented algorithm can work either automatically or with several degrees of assistance. The algorithm was comprehensively evaluated in several ECG databases for comparison purposes. Even in the fully automatic mode, the algorithm slightly improved the performance figures of the original automatic classifier; just with less than two manually annotated heartbeats (MAHB) per recording, the algorithm obtained a mean improvement for all databases of 6.9% in accuracy A, of 6.5% in global sensitivity S and of 8.9% in global positive predictive value P^+. An assistance of just 12 MAHB per recording resulted in a mean improvement of 13.1% in A, of 13.9% in S, and of 36.1% in P^+. For the assisted mode, the algorithm outperformed other state-of-the-art classifiers with less expert annotation effort. The results presented in this paper represent an improvement in the field of automatic and patient-adaptable heartbeats classification, concluding that the performance of an automatic classifier can be improved with an efficient handling of the expert assistance.**

***Index Terms*—Clustering, heartbeat classification, linear classifier, patient adaptable.**

I. Introduction

The World Health Organization places cardiovascular diseases as the first single cause of death globally in the present, and forecasts the same ranking up to 2030 [1]. The heart function can be analyzed by means of the electrocardiographic signal (ECG), which is a noninvasive, inexpensive, and well-established technique. The computerized analysis of the ECG is nowadays a well-established practice, and many improvements were achieved to aid cardiologists in the task of analyzing long-term ECG recordings. One of the analysis performed is the classification of heartbeats, for the subsequent study of arrhythmias. Arrhythmias are understood as any disturbance in the rate, regularity, site of origin, or conduction of the electrical impulses through the heart [2]. The classification of thousands of heartbeats is a challenging task even for experienced cardiologists, and moreover, if the analysis is focused in subtle or infrequent arrhythmias. Note that subtle arrhythmias are not always imminently dangerous, but may represent a long-term threat without a proper treatment. Therefore, the aid in the analysis provided by automatic algorithms can improve the diagnostic achieved by cardiologists.

Many algorithms for ECG heartbeats classification were developed in the last decades (see references in [3] and [4]), but due to the lack of standardization in the development and evaluation criteria, comparison of results across most of these works could not be performed fairly or is impossible. In order to overcome this problem, some methodological aspects in the development and evaluation of heartbeat classifiers were followed in recent works [3]–[6]. The most relevant key-points are as follows.

1) Use of public and standard databases, as the ones available in Physionet [7].
2) Fulfillment of AAMI recommendations for class labeling and results presentation [8].
3) Patient-oriented data division into training and testing sets, as described in [3].

Another aspect suggested in our previous works is the analysis of the capability of the classifier to retain its performance in other databases not considered during the development [4]. We refer to this property of a classifier as generalization capability, and its analysis provides a broader idea of the performance achieved. Up to the writing of this manuscript, only few of the reviewed works used more than one database either for the development [9], [10] or for a generalization assessment [11]–[13].

In the current state-of-the-art, it seems that the automatic classification approach has approximated to a performance upper bound, probably because the huge interpatient variability makes impossible that the probability distribution learned in a train set be representative of that found in a test set and during the normal operation of the classifier. The patient adaptation technique by means of expert assistance (i.e., manual beat annotation) was reported to be useful in two works to overcome this problem [14], [15], at the expense of sacrificing automaticity. Other works also reported better performances than the ones obtained by automatic classifiers, always taking advantage of the expert assistance [5], [6], [10], [16]. One aspect to study when adopting this technique is the efficient use of the assistance, in order to keep the classifier as much automatic as possible. It is interesting to note that some classifiers require from 2 to 5 min of manual annotations, which is equivalent to several hundred of

Manuscript received December 22, 2011; revised March 29, 2012 and May 14, 2012; accepted May 28, 2012. Date of publication June 6, 2012; date of current version July 18, 2012. This work was supported by the Ministry of Science and Innovation (MICINN), Spain, and DGA, Spain, under Project TEC2010-21703-C03-02 and Project GTC T-30, respectively. *Asterisk indicates corresponding author.*

*M. Llamedo is with the Aragon Institute of Engineering Research (I3A), University of Zaragoza, Zaragoza 50009, Spain (e-mail: llamedom@electron.frba.utn.edu.ar).

J. P. Martínez is with the Communications Technology Group, Aragón Institute of Engineering Research, University of Zaragoza, Zaragoza 50009, Spain, and also with the Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain (e-mail: jpmart@unizar.es).

Digital Object Identifier 10.1109/TBME.2012.2202662

0018-9294/$31.00 © 2012 IEEE

TABLE I
DATABASES USED IN THIS WORK AND ITS CLASS REPRESENTATION

Database	Purpose	Length	Leads	Fs (Hz)	N	S	V'		Q	#Rec
							V	F		
MITBIH Arrhythmia (MITBIH-AR) [7]	A	30'	2	360	90089	2779	7007	802	15	44
American Heart Association (AHA) [20]	A	30'	2	250	319125	0	32745	1266	0	155
MITBIH Supr. Arrhythmia (MITBIH-SUP) [7]	A	30'	2	128	162271	12195	9940	23	79	78
European ST-T (ESTTDB) [7]	ST	2h	2	250	784568	1095	4467	354	0	90
MITBIH ST Change (MITBIH-ST) [7]	ST	<1h	2	360	46215	798	319	0	0	18
MITBIH long-term database (MITBIH-LT) [7]	LT	14-22h	2	128	594672	1499	63584	2785	0	7
Long-Term ST (LTSTDB) [7]	LT/ST	21-24h	2/3	250	6422003	30905	37894	0	0	86
St. Petersburg Inst. of Card. Tech. (INCART) [7]	A	30'	12	257	153651	1959	20005	219	6	75
Total					8858840	52556	178502	5449	100	609

A: Arrhythmia/Heartbeat classification; ST: Ischemia detection; LT: Long-term detection/classification.
Heart beats classes are N: normal, S: supraventricular, V: ventricular, F: fusion, Q: unknown, and V':AAMI2 Ventricular.

expert labeled heartbeats [5], [6], [14], [15], while [10] requires the annotation of several heartbeats, depending on the number of arrhythmias present. One drawback of several patient-adaptable approaches is that they cannot operate without assistance [5], [6], [10], [16]. This is not the case of those developed as an evolution of a previous automatic classifier [14], [15].

In this paper, we propose an expert assistance approach pursuing two objectives: first, to be able to perform automatic classification, and second, if assistance is available, to take advantage of it efficiently. For this purpose, we suggest the integration of a well-known clustering algorithm based on mixture of Gaussians (MoG) [17], with the linear discriminant classifier (LDC) presented in [4]. In this solution, clustering is responsible of retaining all patient-specific data ordering, while the automatic classifier performs the cluster labeling, and can be assisted by an expert, as will be described in detail in the following sections.

The objective of this paper is to study how the classification performance of a previously developed multilead algorithm [4], [18] can be improved, by implementing a patient-adaptation technique based on clustering. Our working hypothesis is that within a given recording, classes are clearly separated and beats of the same class tend to be grouped in one or more homogeneous clusters. In other words, after clustering the beats from a recording, all beats grouped in the same cluster would likely belong to the same class. For that purpose, first we search for an appropriate set of features for intrarecording clustering, and compare several integration strategies in a development dataset, to finally assess the final performance and generalization capability to other databases not considered during the development. The performance will be compared with other state-of-the-art classifiers [5], [6], [10], [15], [19].

II. METHODOLOGY

A. ECG Databases

All experiments performed in this paper were carried out in several public databases available on Physionet [7], and the well-known American Heart Association database [20]; their relevant details are summarized in Table I. For all databases, the AAMI recommendations for class-labeling were adopted (in [8], Sec. 4.2]). The AAMI Q class (unclassified and paced heartbeats) was discarded since it is marginally represented in all databases. This limitation occurs to a lesser extent with the fusion (F) AAMI class, but instead of discarding the heartbeats of this class, we adopted an alternative labeling scheme already used in [4]. It consists in merging the fusion (of normal and ventricular beats) and ventricular classes, as the same ventricular class (V' in Table I). This labeling does not compromise the comparability with other AAMI compliant works, since F and Q classes are scarcely represented in the databases used. The databases used include different types of ECG recordings: some of them were recorded during routine ambulatory practice, but others were selected to include less common ventricular, junctional, or supraventricular arrhythmias, or baseline ST segment displacement or other ECG abnormalities. As a result, we use in this paper a dataset with a broad range of normal and pathological ECG recordings to evaluate the algorithm performance. Moreover, the different length of recordings will evidence the ability of the algorithm to handle the nonstationarities present in the ECG. Further details of each database can be found on Physionet [7].

We will refer as the "development dataset" to the union of the MITBIH-SUP database and the 22 recordings included in the DS1 subset of MITBIH-AR defined in [3], while the "evaluation dataset" includes the rest of databases described in Table I.

B. Heartbeats Classification

Following the scheme presented in Fig. 1, the patient-adaptable algorithm includes a LDC and an expectation-maximization clustering algorithm (EMC). Both LDC and EMC work independently and each performs a preliminary classification/clustering task in different feature spaces. The LDC was developed and trained as described in [4], while the EMC development will be described later. Finally, the heartbeat and cluster labels provided by the LDC and EMC, respectively, are integrated with a voting scheme into a final heartbeat label. Three modes of operation are proposed, depending on the degree of expert assistance available in the application scenario: 1) automatic, 2) slightly assisted, and 3) assisted. The algorithm performs the following procedures: 1) cluster and centroid identification, 2) LDC automatic classification, and 3) expert assistance.

For the *automatic mode*, in each record, K clusters and centroids are identified, corresponding to groups of similar heartbeats, while at the same time, the LDC computes the labels for each heartbeat. Then, for each cluster, the algorithm tests if any label obtains a qualified majority, meaning that the most represented label exceeds the α percent of the cluster population. In

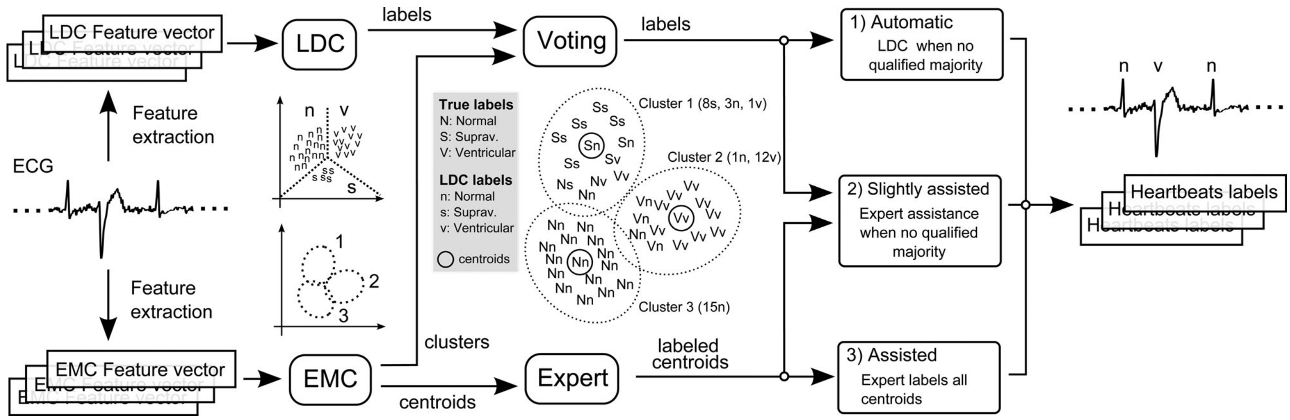

Fig. 1. Overview of the proposed algorithm. There is a graphical description in the center of the scheme about the task carried out by each block. The toy example in the middle is also commented in the text to better understand the three modes of operation.

case this label exists, it is assigned to the whole cluster, superseding the LDC labels. If the qualified majority is not reached, the uncertainty is considered to be too high to change the labels, an thus the LDC labels remain unchanged.

The *slightly assisted mode* is similar to the automatic, with the exception that in case of not finding a class qualified majority, expert assistance is required to label the cluster centroid and propagate it to the whole cluster, ignoring LDC labels. The procedure of expert assistance is simulated by inspecting the true labels provided with each database.

Finally in the *assisted mode*, K clusters and centroids are identified, and then the expert is required to label each centroid. The algorithm concludes assigning these labels to the rest of heartbeats in each cluster.

To better understand the three modes of operation, a toy example can be found in the center of Fig. 1. The LDC by itself makes four errors in cluster 1 and three in cluster 2. For the automatic mode, clusters 2 and 3 have majority of V and N classes, respectively. Then, votation propagates centroid labels to the rest of examples within both clusters, occurring one mistake in cluster 2. In cluster 1, there is no qualified majority for $\alpha = 50\%$, so the LDC labels remain unchanged and five mistakes happen. The automatic mode made 1 mistake less than the LDC. For the slightly assisted mode, only cluster 1 would be modified, by propagating the true label of the centroid S, resulting four errors for this cluster. Finally for the assisted mode, in this case the result is the same as in the previous mode, four errors in cluster 1, one error in cluster 2, and no errors in cluster 3. In summary, seven errors for the LDC, six for the automatic mode, and five for the slightly and assisted modes. As was shown, the algorithms rely heavily in the ability of the EMC to cluster the heartbeats adequately.

1) Automatic Classifier: We follow a scheme similar to the one in [4] and [18], where we developed a multilead heartbeat classifier with good generalization capability. We used a linear classifier compensated for the class imbalance, while as feature model we adopted rhythm and morphological features computed in a multilead manner. Regarding the classifier used, we found that linear discriminant functions were suitable for the heartbeat classification task in terms of performance and generalization capability. Under the assumption of independent and normally distributed data, the maximum *a posteriori* criterion (MAP) leads to the linear classifier defined by the discriminant functions [17]

$$g_i(\mathbf{x}) = \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \log(P(\omega_i)) \qquad (1)$$

for the ith class, where $\mathbf{x}$ represents the feature vector describing each heartbeat, $\boldsymbol{\mu}_i$ is the the mean vector, $\boldsymbol{\Sigma}$ is the covariance matrix, and $P(\omega_i)$ is the prior probability.

The model parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}$ in (1) were estimated from the training data as the sample mean

$$\boldsymbol{\mu}_i = \frac{1}{M_i}\sum_{m=1}^{M_i}\mathbf{x}_m \qquad (2)$$

and weighted covariance matrix expressions

$$\boldsymbol{\Sigma} = \frac{1}{\sum_{i=1}^{C} w_i}\sum_{i=1}^{C}\frac{w_i \sum_{m=1}^{M_i}(\mathbf{x}_m - \boldsymbol{\mu}_i)\cdot(\mathbf{x}_m - \boldsymbol{\mu}_i)^{\mathrm{T}}}{M_i} \qquad (3)$$

while the values for the prior probabilities $P(\omega_i)$ were considered the same for all classes. For the classification of C classes, where M_i is the number of examples $\mathbf{x}_m$ of the *i*th class, the rule assigns an unlabeled observation $\mathbf{x}$ to the class i that results in the maximum posterior probability $g_i(\mathbf{x})$. The class-weighting possibility with w_i is of much interest due to the heavy class-size imbalance inherent to this application, where the normal class is, in general, one order of magnitude more represented than other classes. The weights used in [4] and [18] were also used in this study, being w_S = 10; w_V=10; and w_N = 1. The classification tasks were performed using the PRtools toolbox [21] for MATLAB (The Mathworks Inc., MA).

The features used by the automatic classifier are described in Table II. The morphology features k_Z^L and k_M^L for lead L are calculated in the two principal ECG leads after integrating the multilead information with a principal component analysis (PCA). Therefore, these features account for a multilead morphological description of the QRS complex. For a detailed description of the features and the multilead strategy used see [4] and [18].

TABLE II
FEATURE MODEL USED BY THE AUTOMATIC CLASSIFIER FOR RECORDINGS OF TWO OR MORE LEADS

Feature	Description
$\ln(RR[i])$	Current RR interval
$\ln(RR[i+1])$	Next RR interval
$\ln(RR_1)$	Average RR interval in the last minute
$\ln(RR_{20})$	Average RR interval in the last 20 minutes
k_Z^1	First zero-cross position of the WT autocorrelation sequence in the 1st principal component
k_Z^2	First zero-cross position of the WT autocorrelation sequence in the 2nd principal component
k_M^1	First minimum position of the WT autocorrelation sequence in the 1st principal component
k_M^2	First minimum position of the WT autocorrelation sequence in the 2nd principal component

TABLE III
FEATURES USED WITH THE EMC ALGORITHM

Feature	Description
$\ln(RR[i])$	Current RR interval
$\ln(RR[i-1])$	Previous RR interval
$\ln(P_{RR})$	Prematurity of the heartbeat
$\ln(dRR_L)$	Local RR interval variation
$\ln(RR_{20})$	Mean RR interval within the last 20 minutes
$\ln(S^1_{\text{QRS}})$	QRS mean wavelet scale in the first principal component
k_M^1	First minimum position of the WT autocorrelation sequence in the 1st principal component
$r_{\text{QRST}}(k_M)$	Value of the first maximum in the QRST complex crosscorrelation sequence between WT scale 3 of the first two principal components

2) Clustering Algorithm: The EMC algorithm used in this paper is based on the MOG model [17]. It consists of estimating the parameters of a density function

$$p(\mathbf{x}|\Psi) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_\text{k})$$
$$= \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^m \, |\boldsymbol{\Sigma}_k|}} \exp^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}-\boldsymbol{\mu}_k)} \quad (4)$$

where the m-dimensional vector $\mathbf{x}$ is modeled by K Gaussians with mixing coefficients π_k, in order to retain a more realistic structure of the data. The parameter set $\Psi = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k | k = 1, \ldots, K\}$ is estimated by maximum likelihood criterion. We maximize the log likelihood

$$L(X|\Psi) = \ln \prod_{n=1}^{N} p(\mathbf{x}_n|\Psi) \quad (5)$$

for the N heartbeats in each recording named $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Since there is not a closed form solution for Ψ by maximizing $L(X|\Psi)$, the well-known expectation-maximization (EM) algorithm is used to obtain the estimation equations of the parameters Ψ at each iteration j, which are the mixing coefficient for each cluster

$$\hat{\pi}_k^j = \frac{1}{N} \sum_{m=1}^{N} \beta_{m,k}^{j-1} \quad (6)$$

the cluster mean

$$\hat{\boldsymbol{\mu}}_k^j = \frac{1}{N\hat{\pi}_k^j} \sum_{m=1}^{N} \beta_{m,k}^{j-1} \mathbf{x}_m \quad (7)$$

and cluster covariance matrix

$$\hat{\boldsymbol{\Sigma}}_k^j = \frac{1}{N\hat{\pi}_k^j} \sum_{m=1}^{N} \beta_{m,k}^{j-1} (\mathbf{x}_m - \hat{\boldsymbol{\mu}}_k^j) \cdot (\mathbf{x}_m - \hat{\boldsymbol{\mu}}_k^j)^\text{T} \quad (8)$$

where $\beta_{m,k}^j$ is known as the ownership variable, which indicates the probability of sample $\mathbf{x}_m$ to have been generated by the kth component at iteration j

$$\beta_{m,k}^j = \frac{\hat{\pi}_k^{j-1} \cdot \mathcal{N}(\mathbf{x}_m|\hat{\boldsymbol{\mu}}_k^{j-1}, \hat{\boldsymbol{\Sigma}}_k^{j-1})}{\sum_{n=1}^{K} \hat{\pi}_n^{j-1} \cdot \mathcal{N}(\mathbf{x}_m|\hat{\boldsymbol{\mu}}_n^{j-1}, \hat{\boldsymbol{\Sigma}}_n^{j-1})}. \quad (9)$$

The EM algorithm iteratively computes the weight, location, and dispersion for each of the K clusters [see (6)–(8)], until $\beta_{m,k}^j$ does not change significantly, which is equivalent to obtain stable clusters. The interested reader is referred to [17], [21] for details, equations, and the implementation used in this paper.

Regarding the feature model used with the EMC, we followed the same feature selection procedure described in [4], by means of a sequential floating feature selection algorithm (SFFS) [17], [22]. The complete pool of features consisted of 61 features, most of them described in the works cited in Section I. For the case of clustering, instead of looking for features with generalization capability or interpatient separability, we looked for those with high intrapatient separability. This criterion was achieved by modifying the SFFS' optimization criterion used in [4], in order to find a feature model that provides as much intrapatient class separability as possible, facilitating the clusters identification. The first modification consisted in evaluating our clustering algorithm in a patient by patient fashion since this is how this algorithm will be used in practice. The second is that the performance will be evaluated in an optimistically biased fashion, described in [18], assuming that we know *a priori* the true labels of the heartbeats. This is done in order to estimate a performance upper bound. The feature selection experiments were carried out in a dataset formed by the union of MITBIH-SUP with DS1 subset of MITBIH-AR [3]. As the SFFS performs thousands of model evaluations, this task is very demanding in processing power, specially for the random and iterative nature of the EMC. For this reason, we replaced only for the feature selection task, the EMC for a classifier based on MoG, which uses the same algorithm used for cluster discovery. The classifier based on MoG (MoGC) models each AAMI class with K Gaussian distributions, in contrast with the LDC that models each class with a mean vector and a pooled covariance matrix for all classes (see (2) and (3)). The MoGC uses during training the EM algorithm for the estimation of the Gaussian components. This modification results, first, in moving through more deterministic paths through the performance surface evaluated with the SFFS, and second in easing the EM iteration since the heartbeat labels are known *a priori*.

As a result, a model of eight features was obtained. This model also includes a description of the rhythm and morphology of heartbeats as shown in Table III.

TABLE IV
PERFORMANCE OBTAINED IN THE DEVELOPMENT DATASET FOR THE ELECTION OF K AND α PARAMETERS

			Normal		Supraventricular		Ventricular (V')		Total		
Operation mode	α	K	S	P^+	S	P^+	S	P^+	A	S	P^+
Slightly assisted	50	9	96 ± 0	97 ± 0	58 ± 2	53 ± 1	79 ± 1	67 ± 1	93 ± 0	77 ± 0	72 ± 0
		12	96 ± 0	97 ± 0	56 ± 2	51 ± 1	78 ± 1	66 ± 1	92 ± 0	77 ± 1	72 ± 0
	75	**9**	**97±0**	**98±0**	**60±1**	**61±1**	**84±1**	**71±1**	**94±0**	**80±1**	**77±1**
		12	97 ± 0	98 ± 0	61 ± 1	60 ± 1	84 ± 1	71 ± 1	94 ± 0	80 ± 0	76 ± 0
Automatic	0	9	96 ± 0	97 ± 0	52 ± 3	49 ± 1	78 ± 1	65 ± 1	92 ± 0	75 ± 1	71 ± 1
		12	95 ± 0	97 ± 0	50 ± 2	49 ± 1	77 ± 1	63 ± 1	92 ± 0	74 ± 1	70 ± 1
	50	**9**	**96±0**	**97±0**	**53±2**	**51±2**	**77±1**	**65±1**	**92±0**	**75±1**	**71±0**
		12	95 ± 0	97 ± 0	52 ± 1	48 ± 1	77 ± 1	64 ± 0	92 ± 0	75 ± 0	70 ± 0
	75	9	95 ± 0	97 ± 0	50 ± 1	47 ± 0	77 ± 0	61 ± 0	92 ± 0	74 ± 0	68 ± 0
		12	95 ± 0	97 ± 0	50 ± 1	47 ± 0	77 ± 0	61 ± 0	92 ± 0	74 ± 0	68 ± 0

K: number of clusters; α: majority threshold (in percent).

TABLE V
PERFORMANCE COMPARISON WITH REFERENCE ALGORITHMS

			Normal		Supraventricular		Ventricular (V')		Total		
Dataset	Observation	#MAHB	S	P^+	S	P^+	S	P^+	A	S	P^+
MITBIH-AR (subset [6])°	Hu *et al.* [14]	$\approx 300^\dagger$	–	–	–	–	79	76	–	–	–
	de Chazal *et al.* [15]	500	–	–	76	39	76	91	–	–	–
	Jiang *et al.* [5]	$\approx 300^\dagger$	–	–	75	79	94	96	–	–	–
	Ince *et al.* [6]	$\approx 300^\dagger$	–	–	82	63	90	92	–	–	–
	this work	12	99±0	99±0	89±2	88±3	93±1	97±1	98±0	85±1	85±2
MITBIH-AR (24 rec.)$^\triangledown$	Jiang *et al.* [5]	$\approx 300^\dagger$	99	96	51	68	85	93	95	78	86
	this work	12	99±0	98±0	91±1	88±2	90±1	96±1	98±0	93±1	94±1
MITBIH-AR (DS2)•	de Chazal *et al.* [15]	500	94	99	88	47	95	82	94	92	76
	this work	12	100±0	99±0	92±1	90±3	92±1	97±1	99±0	94±1	96±1
	de Chazal *et al.* [3]	0	87	99	76	39	87	43	87	83	60
	Mar *et al.* [19]	0	90	99	83	34	87	61	89	87	65
	this work	0	93±0	99±0	77±0	39±0	82±0	70±0	92±0	84±0	69±0
MITBIH-AR (DS1-DS2)•	Ince *et al.* [6]	$\approx 300^\dagger$	98	98	64	54	85	87	96	82	80
	this work	12	100±0	99±0	89±2	88±3	90±1	97±0	98±0	93±1	95±1
MITBIH-LT	Kiranyaz *et al.* [10]	$\leq 900^\star$	99	100	40	17	98	99	99	79	72
	this work	20	99±1	99±0	16±16	0±0	94±0	91±4	98±0	70±5	74±9

#MAHB: manually annotated heartbeats per recording.
• DS1 and DS2 datasets defined in [3].
° Comparison against results presented in Table II of [6].
$^\triangledown$24 recordings from 200 to 234 used in [5].
† Heartbeats in the first 5 min of each recording.
* Heartbeats in approximately 15 min of each recording.

Among the rhythm features used in the model, some of them have already been used in previous works. The prematurity of a heartbeat

$$P_{RR}[i] = \frac{RR[i]}{\sum_{k=i-1}^{i+1} RR[k]} \tag{10}$$

measures how anticipated is a heartbeat respect to the previous and next RR interval. The local RR interval variation is defined as $dRR_L[i] = \sum_{k=i-1}^{i+1} |dRR[k]|$, where $dRR[i] = RR[i] - RR[i-1]$. One of the morphology related features is the wavelet scale where the QRS complex is mostly projected. It is known that fast evolving signals, as a normal heartbeat, tend to be projected in lower wavelet scales or contains higher frequency components. The QRS center scale for each lead ($S_{\mathrm{QRS}}^{\mathrm{Lead}}$) is calculated as the weighted sum

$$S_{\mathrm{QRS}}^{L} = \frac{\sum_{s=1}^{6} A_s^L .s}{\sum_{s=1}^{6} A_s^L} \tag{11}$$

where A_s^L is the mean absolute amplitude of the QRS peaks at scale s of the DWT, and lead L

$$A_s^L = \frac{1}{D} \sum_{d=1}^{D} \left| W_s^L s(l_d) \right|, \quad s = 1, 2, \ldots 6 \tag{12}$$

being D the number of detected peaks (1 or 2) and l_d the positions of the peaks. The last morphology feature used is the maximum of the autocorrelation sequence of the ECG WT at scale 3 ($r_{\mathrm{QRST}}(k_M)$), which describes the QRST complex similarity between PCA leads at scale 3 of the WT. This feature is related to changes in the multilead morphology and the depolarization axis of the QRST complex. See [4], Figs. 2 and 3] for details about the calculation of all the morphology features used.

C. Performance Evaluation

The performance is calculated from the confusion matrix after performing a classification experiment, in terms of the class sensitivity S_i, class positive predictive value P_i^+, global accuracy A, global sensitivity S, and global positive predictive value

TABLE VI
GENERALIZATION EVALUATION OF THE PROPOSED ALGORITHM FOR ALL DATABASES USING THE 3 AAMI2 CLASSES

				Normal		Supraventricular		Ventricular (V')		Total		
Dataset	Observation	Ref.	#MAHB	S	P^+	S	P^+	S	P^+	A	S	P^+
AHA	*biased*		*all*	*100*	*100*	–	–	*99*	*97*	*100*	*100*	*98*
	Assisted		12	100±0	100±0	–	–	98±0	98±0	100±0	99±0	99±0
			9	100±0	100±0	–	–	97±0	98±0	99±0	98±0	99±0
	Slightly assisted		1.2	96±0	100±0	–	–	86±0	82±0	95±0	91±0	60±0
	Automatic		0	93±0	99±0	–	–	75±1	71±1	92±0	85±0	57±0
		[18]	0	91	99	–	–	75	66	89	83	55
ESTTDB	*biased*		*all*	*100*	*100*	*98*	*41*	*97*	*83*	*100*	*98*	*74*
	Assisted		12	100±0	100±0	42±2	75±5	91±2	90±1	100±0	77±1	88±3
			9	100±0	100±0	37±4	65±13	88±1	81±3	100±0	75±2	82±4
	Slightly assisted		1	97±0	100±0	43±3	2±0	81±1	43±3	96±0	73±1	48±1
	Automatic		0	92±0	100±0	28±12	1±0	77±1	32±4	92±0	66±4	44±1
		[18]	0	89	100	47	1	78	31	89	72	44
INCART	*biased*		*all*	*99*	*100*	*98*	*68*	*99*	*98*	*99*	*99*	*88*
	Assisted		12	100±0	100±0	85±2	91±2	98±0	98±0	99±0	94±1	96±1
			9	100±0	99±0	79±3	90±2	98±0	98±1	99±0	92±1	96±1
	Slightly assisted		1.1	96±0	99±0	74±3	18±1	93±1	97±1	95±0	88±1	72±0
	Automatic		0	89±0	99±0	74±4	8±0	88±0	96±1	89±0	84±1	68±0
		[18]	0	88	100	76	7	84	94	87	83	67
LTSTDB	*biased*		*all*	*99*	*100*	*96*	*49*	*98*	*84*	*99*	*98*	*78*
	Assisted		15	100±0	100±0	51±0	58±0	82±0	85±0	99±0	78±0	81±0
			9	100±0	100±0	52±1	58±5	78±2	84±3	99±0	77±1	80±3
	Slightly assisted		1.2	99±0	100±0	44±1	36±5	77±1	59±4	99±0	73±1	65±0
	Automatic		0	96±0	100±0	50±2	8±0	66±1	36±2	96±0	71±0	48±1
		[18]	0	92	100	57	4	70	29	91	73	44
MITBIH-AR	*biased*		*all*	*98*	*100*	*98*	*83*	*98*	*93*	*98*	*98*	*92*
	Assisted		12	100±0	99±0	89±2	88±3	90±1	97±0	98±0	93±1	95±1
			9	99±0	99±0	86±1	85±2	88±1	96±1	98±0	91±1	93±1
	Slightly assisted		1	98±0	99±0	79±3	59±3	83±2	92±1	97±0	87±1	83±1
	Automatic		0	96±0	98±0	76±2	43±2	80±2	82±3	94±0	84±1	74±1
		[18]	0	95	98	76	35	76	77	93	82	70
MITBIH-LT	*biased*		*all*	*97*	*100*	*95*	*11*	*92*	*91*	*97*	*95*	*67*
	Assisted		20	99±1	99±0	16±16	0±0	94±0	91±4	98±0	70±5	74±9
			9	99±0	99±0	1±1	0±0	89±6	92±2	98±0	63±2	64±1
	Slightly assisted		2.6	95±1	98±0	0±0	0±0	83±1	65±3	94±0	59±0	54±1
	Automatic		0	88±0	95±0	2±0	0±0	37±2	44±1	83±0	42±1	46±0
		[18]	0	86	95	33	1	38	44	81	52	47
MITBIH-ST	*biased*		*all*	*99*	*100*	*98*	*89*	*100*	*98*	*99*	*99*	*96*
	Assisted		12	100±0	100±0	89±1	87±2	95±1	99±0	100±0	94±0	95±1
			9	100±0	100±0	83±5	80±6	94±2	99±0	99±0	93±2	93±2
	Slightly assisted		1.4	71±1	99±0	76±6	5±0	93±3	25±1	71±1	79±3	43±0
	Automatic		0	67±1	99±0	57±6	3±0	43±3	11±1	66±1	56±3	38±0
		[18]	0	65	99	62	3	53	9	65	60	37
MITBIH-SUP	*biased*		*all*	*97*	*100*	*95*	*68*	*94*	*90*	*96*	*95*	*86*
	Assisted		12	99±0	98±0	74±1	79±1	88±1	90±1	96±0	87±0	89±0
			9	98±0	98±0	73±2	77±3	87±2	90±1	96±0	86±1	89±1
	Slightly assisted		1.3	96±0	98±0	62±1	61±1	85±1	65±1	93±0	81±1	74±1
	Automatic		0	94±0	97±0	47±3	50±1	82±0	54±1	91±0	75±1	67±1
		[18]	0	93	98	51	46	79	49	90	74	64

#MAHB: manually annotated heartbeats per recording.

P^+ as suggested in [8] and described in [4] and [18]. As the initialization of the EMC is random, the results of the clustering algorithm are not deterministic. Then, each experiment is repeated 30 times to evaluate the mean and standard deviation of the performance estimates. The amount of expert assistance required in the patient-adaptable modes of operation will be also accounted for each experiment.

III. RESULTS

We performed two experiments, in the first one we studied the values of the algorithm parameters that will be used in the second to evaluate its performance. The objective of the first experiment was to set up the number of clusters (K) and the qualified majority percentage used in votations (α), both parameters used in automatic and slightly assisted modes of operation. These parameters were assessed in the development dataset (MITBIH-SUP and DS1 subset of MITBIH-AR), and then used for the final performance evaluation in the remaining datasets. Table IV shows the results of this experiment for two values of the evaluated parameters. As a result of this experiment, we adopted $K = 9$ and $\alpha = 50\%$ for the automatic mode, and $K = 9$ and $\alpha = 75\%$ for the slightly assisted mode.

The final evaluation of the algorithm was performed in a broad set of databases in order to obtain a realistic estimation of its

TABLE VII
RESULTS AND CONFUSION MATRICES USING THE 5 AAMI CLASSES IN DS2 OF MITBIH-AR

Dataset	Observation	#MAHB	Normal		Supraventr.		Ventricular (V)		Total
			S	P^+	S	P^+	S	P^+	A
MITBIH-AR-DS2	Assisted	12	100±0	99±0	92±1	90±3	93±1	97±1	98±0
		9	99±0	99±0	92±1	86±3	91±3	96±1	98±0
	Slightly assisted	1	97±0	99±0	83±4	58±5	91±3	90±2	96±0
	Automatic	0	95±0	99±0	79±2	46±2	89±1	87±1	94±0

#MAHB: manually annotated heartbeats per recording.

performance, as done in [11], [18]. The three modes of operation were evaluated for each database with the parameter values obtained in the first experiment. The results of this experiment are presented in Tables V and VI grouped by dataset. Comparison with the most relevant algorithms found in the literature are presented separately in Table V. In Table VI, the performance obtained for all databases are presented. For each database, we present the performance of our previous classifier [4] at the bottom for comparison, and a biased performance estimation on top as an upper bound. This biased performance is obtained when a quadratic classifier [4], [17], [21] and the feature model presented in Table III is trained and tested in the same patient, for each patient in a database. This optimistically biased performance serves as an upper bound, and represents the performance of the model if it could be retrained for each patient. From the results presented in Table V, the proposed algorithm outperforms almost all reviewed algorithms, except the algorithms of Jiang [5] and Ince [6] in a small subset of MITBIH-AR, and the algorithm of Kiranyaz [10] in the MITBIH-LT. Finally, the results showed in Table VI evidence that the algorithm improves the baseline performances obtained by the LDC.

IV. Discussion and Conclusion

In this paper, we presented a versatile ECG heartbeat classification algorithm suitable for a broad range of scenarios, from automatic or unassisted to fully assisted mode. The automatic part of the algorithm relies on a previously developed automatic classifier with proven generalization capability [4], referred as LDC in Fig. 1. The main limitation of the LDC is the inability to handle large inter-patient rhythm and morphology variations. Many works overcame this limitation with the assistance of an expert [5], [6], [10], [14]–[16]. The approach to handle assistance presented in this paper is based on a cluster algorithm, responsible of retaining most of the patient specific characteristics of the heartbeats (EMC in Fig. 1). For this reason, the feature model used with the clustering algorithm pursues the maximum intrapatient class separability. This approach is different to the one used in the development of the LDC feature set in [4], which pursued the maximization of a generalization criterion. As a result, Tables II and III show the different feature models used in the algorithm. As can be seen, both feature models use rhythm and morphology features for heartbeats representation. Regarding the rhythm features, the EMC has the addition of features P_{RR} and dRR_L, which are related to the local RR interval variation. As for the morphology description, features S^1_{QRS} and k^1_M may together represent a robust surrogate of the QRS width; while feature $r_{\mathrm{QRST}}(k_M)$ describes the QRST complex similarity between PCA leads at scale 3 of the WT. This measure is related to morphologic and depolarization-axis changes in the QRST complex.

To operate in automatic and slightly assisted modes, the EMC algorithm uses K as the number of clusters to model, and the voting scheme uses α as a threshold to assume that a whole cluster belongs to a class. As can be seen in Table IV, the performance intervals are comparable for the selected configurations but a mild improvement can be seen for the highlighted configurations. These configurations will be used for the final evaluation and comparison of the algorithm.

The reference performances used for comparison purposes were in all cases AAMI [8] compliant. Given that F and Q classes are scarcely represented in the databases used, the AAMI2 alternative labeling used is numerically equivalent when calculating the S and V′ class performances, presented in Tables V and VI. The reader interested in this aspect is referred to section A.3.5.2 of [8]. Moreover, to ensure comparability, the results obtained with the same features and development databases, considering the five AAMI classes in DS2 are shown in Table VII.

The performance comparisons presented in Table V evidence the usefulness of the proposed algorithm. Without expert assistance the proposed algorithm performs slightly better than a recent algorithm of Mar *et al.* [19] and the automatic version of de Chazal *et al.* [3]. Furthermore, when assistance is available our algorithm outperforms the reviewed algorithms [5], [6], [14], [15] in several subsets of the MITBIH-AR, with the following clarification. The algorithms of Jiang and Kong [5], Ince *et al.* [6] and de Chazal and Reilly [15] outperform our algorithm in different subsets of MITBIH-AR, but using more expert assistance. However, the same algorithms perform worse in bigger subsets of the same database, as can be seen in Table V. This fact reinforces the importance of evaluating arrhythmia classifiers in a wide range of databases, to have a complete idea of its performance. Finally, the algorithm presented by Kiranyaz *et al.* [10] performed better than our algorithm in the MITBIH-LT, but with an increased effort in assistance of 900 MAHB per recording, respect to the 20 MAHB required by this algorithm. However, the differences in performance are moderate, and considering that the algorithm presented in [10] was specifically developed for long-term recordings. It is worth remarking that the MITBIH-LT presents the bigger class imbalance among the studied databases, showing some limitations of the EMC to detect scarcely represented classes, as the supraventricular.

An interesting aspect of the proposed algorithm is the improvement achieved in the amount of expert assistance required, 42 times less annotation effort than the algorithms of Hu *et al.* [14] and de Chazal and Reilly [15], 25 times less than Jiang and Kong [5] and Ince *et al.* [6], and 45 times less than Kiranyaz [10].

Regarding the comparison with the previous automatic multilead classifier, the fully automatic mode of the patient-adapted algorithm presented in this paper achieved performance figures higher than those obtained in [4] for all databases except in the two databases including long-term recordings (MITBIH-LT and MITBIH-ST), where the performance was slightly lower. The decrease in MITBIH-LT and ST shows a limitation of the clustering features to adequately account for changes in long recordings. Moreover, as both databases include a small number of recordings (7 and 18, respectively), a particular recording could have an exaggerated influence on the whole database performance. However, for the automatic mode the mean improvement across databases with respect to the LDC is of 2.3% in A, of -2% in S and of 1.8% in P^+. With a small degree of assistance, 1-2 MAHB per recording in the slightly assisted mode, we obtained a mean improvement of 6.9% in A, of 6.5% in S and of 8.9% in P^+. Furthermore, an assistance of just 12 MAHB per recording results in a mean improvement of 13.1% in A, of 13.9% in S and of 36.1% in P^+. The important improvement in performance achieved in assisted mode shows that our working hypothesis, separate and homogeneous clusters within a recording, is corroborated in most cases. The algorithm showed robust dealing with different types of noise present throughout the evaluated databases, as a result of using robust features. In addition, the results presented in Table VI are consistent, as more assistance is translated into larger performance improvement. This experiment evidences that the algorithm can handle properly the different degrees of assistance provided by an expert. Note that the development dataset, which includes DS1 of MITBIH-AR and MITBIH-SUP, is included in the results presented in Table VI. These results are optimistically biased and should be considered only as an additional description of the algorithm performance.

In slightly assisted mode, it is worth noting that the intra-cluster class-heterogeneity in the recordings analyzed is proportional to the assistance required. Remember that the algorithm ask for assistance in those clusters where a qualified majority is not reached. According to Table VI, from 11% to 16% of the clusters did not reach a qualified majority. For the particular case of the MITBIH-LT this figure raised to the 29%. This increase is reasonable since the nonstationarities of the ECG signal, and thus the cluster heterogeneity, are more evident in long-term recordings.

Assisted 12 MAHB

Truth	Algorithm: n	s	v	f	u	Total
N	44032±96	140±83	84±36	4±8	–	44259
S	133±33	1688±34	16±9	0	–	1837
V	139±53	58±38	2971±72	53±46	–	3221
F	107±43	8±5	61±52	213±59	–	388
Q	6±1	0	1±1	0	–	7
Total	44416±131	1893±122	3133±107	270±93	–	49712

Slightly assisted

Truth	Algorithm: n	s	v	f	q	Total
N	43049±262	1072±275	114±55	0±1	24±38	44259
S	79±28	1585±116	173±115	0	1±2	1837
V	172±105	65±32	2903±109	81±57	–	3221
F	153±69	10±5	39±36	187±98	–	388
Q	6±1	0	1±1	0	–	7
Total	43459±333	2731±272	3229±218	268±138	24±39	49712

Assisted 9 MAHB

Truth	Algorithm: n	s	v	f	u	Total
N	43942±109	208±91	106±57	4±13	–	44259
S	134±35	1684±37	20±10	0	–	1837
V	178±106	60±32	2902±109	81±57	–	3221
F	111±41	10±4	39±36	229±30	–	388
Q	6±1	0	1±1	0	–	7
Total	44369±190	1961±101	3068±152	314±54	–	49712

Automatic

Truth	Algorithm: n	s	v	f	q	Total
N	42186±298	1562±226	103±65	356±148	53±49	44259
S	61±17	1409±97	364±98	0±1	2±6	1837
V	224±73	94±31	2878±53	22±26	3±2	3221
F	317±40	0	50±39	21±24	0	388
Q	5±1	0	2±1	0	0±1	7
Total	42792±293	3065±232	3397±122	399±149	58±50	49712

The algorithm's computational efficiency was not analyzed in detail; however, it takes around 25 s to classify an MITBIH-AR recording (30 min of two-lead ECG) in a desktop PC (Intel Core2 E8500 CPU). The measurement was performed in a freely available implementation of the algorithm in MATLAB [23]. Although the execution time is not excessively large, there is room for improvement with an optimized implementation.

From the evaluation of the algorithm some limitations were found and need to be addressed in future improvements. The first is the inability of the clustering algorithm to find marginally represented classes. This problem slightly affects the global performance since the less represented classes have a mild effect in a database-aggregated performance estimates. However, in certain applications the misclassification of this kind of infrequently arrhythmias could limit the usefulness of algorithms based on clustering techniques. Other limitations are related to the feature model used by the EMC, presented in Table III. In certain recordings where the classes are reasonably represented to be clustered, the EMC fails to recognize the clusters probably due to the inability of the feature model to separate the classes. This problem is also evidenced in the biased evaluation performance, showed on top for each database in Table VI. Theoretically, if the features and classifier used could adequately model the data, the biased performance should be perfect for all classes (S_i and P_i^+ 100%). Since this is not true, it can be concluded that the presented model still has limitations. This could be improved with the development of better features or a most sophisticated classifier. The last limitation found during the evaluation appeared in long-term recordings. In these recordings the evident nonstationarities in the feature space make the algorithm performance to decrease considerably. For this reason, the assistance provided to the algorithm was increased for long-term recordings. Strategies to deal with nonstationarities, as the proposed in [10], will be studied in the future.

The results presented in this paper represent a performance improvement with respect to the published works in the field of automatic and patient-adaptable heartbeats classification. These results show that the performance of an automatic classifier can be improved with an efficient handling of the expert assistance. The authors freely distribute a MATLAB implementation of the algorithm for academic use [23].

ACKNOWLEDGMENT

The CIBER of Bioengineering, Biomaterials and Nanomedicine is an initiative of ISCIII.

REFERENCES

[1] World Health Organization. (2012) Cardiovascular diseases. World Health Organization. [Online]. Available: http://www.who.int/cardiovascular_diseases/en/ 2012.

[2] G. J. Taylor, *150 Practice ECGs: Interpretation and Review*. Oxford, U.K.: Blackwell, 2002.

[3] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[4] M. Llamedo and J. Martínez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 616–625, Mar. 2011.

[5] W. Jiang and S. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 18, no. 6, pp. 1750–1761, Nov. 2007.

[6] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[7] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[8] *Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms*, American National Standard, ANSI/AAMI/ISO EC57, American National Standard Std., AAMI-EC57, 1998–2008.

[9] R. Watrous and G. Towell, "A patient-adaptive neural network ECG patient monitoring algorithm," in *Proc. Comput. Cardiol.*, Sep. 1995, pp. 229–232.

[10] S. Kiranyaz, T. Ince, J. Pulkkinen, and M. Gabbouj, "Personalized long-term ECG classification: A systematic approach," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3220–3226, 2011.

[11] V. Chudácek, G. Georgoulas, L. Lhotská, C. Stylios, M. Petrík, and M. Cepek, "Examining cross-database global training to evaluate five different methods for ventricular beat classification," *Physiol. Meas.*, vol. 30, no. 7, p. 661, 2009. [Online]. Available: http://stacks.iop.org/0967-3334/30/i=7/a=010

[12] V. Krasteva and I. Jekova, "QRS template matching for recognition of ventricular ectopic beats," *Ann. Biomed. Eng.*, vol. 35, pp. 2065–2076, 2007.

[13] Z. Syed, J. Guttag, and C. Stultz. (2007, Jan.). Clustering and symbolic analysis of cardiovascular signals: Discovery and visualization of medically relevant patterns in long-term data using limited prior knowledge. *EURASIP J. Appl. Signal Process.* [Online]. *2007*, pp. 97–97. Available: http://dx.doi.org/10.1155/2007/67938

[14] Y. H. Hu, S. Palreddy, and W. Tompkins, "A patient-adaptable ECG beat classifier using mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–899, Sep. 1997.

[15] P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2535–2543, Dec. 2006.

[16] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ECG complexes using hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838–848, Jul. 2000.

[17] F. van der Heijden, R. Duin, D. de Ridder, and D. Tax, *Classification, Parameter Estimation and State Estimation*. New York: Wiley, 2005.

[18] M. Llamedo and J. P. Martínez, "Cross-database evaluation of a multilead heartbeat classifier," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, in press, 2012.

[19] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ECG classification by means of feature selection," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2168–2177, Aug. 2011.

[20] American Heart Association. (2010). American heart association ECG database. [Online]. Available:https://www.ecri.org.

[21] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. deRidder, D. Tax, and S. Verzakov. "PR-tools, a Matlab toolbox for pattern recognition," (2008). [Online]. Available: http://www.prtools.org.

[22] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.

[23] M. Llamedo. (2012). Argentine-aragonese heartbeat classifier (a2hbc). [Online]. Available: http://code.google.com/p/a2hbc/

Mariano Llamedo was born in Buenos Aires, Argentina, in 1979. He received the M.Sc. degree in electronic engineering from the National Technological University of Buenos Aires (UTN- FRBA), Buenos Aires, Argentina, in 2005.

From 2005 to 2008, he was with the Department of Electronic Engineering, UTN-FRBA, as an Assistant Professor and a Research Fellow. Since 2008, he has been a Researcher with the Aragon Institute of Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain. His professional research interest includes biomedical signal processing, with main interest in signals of cardiovascular origin.

Juan Pablo Martínez was born in Zaragoza, Aragon, Spain, in 1976. He received the M.S. degree in telecommunication engineering and the Ph.D. degree in biomedical engineering from the University of Zaragoza (UZ), Zaragoza, Spain, in 1999 and 2005, respectively.

In 2000, he was an Assistant Professor at the Aragon Institute of Engineering Research, UZ, where since 2007, he has been an Associate Professor. He is also with the Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN). His current research interests include biomedical signal processing, with main interest in signals of cardiovascular origin.