Cross-Database Evaluation of a Multilead Heartbeat Classifier

Mariano Llamedo, Antoun Khawaja, and Juan Pablo Martínez

***Abstract*—In this paper,we studied the improvement in heartbeat classification achieved by including information from multilead ECG recordings in a previously developed and validated classification model. This model includes features from the RR interval series and morphology descriptors for each lead calculated from the wavelet transform. The experiments were carried out in the INCART database, available in Physionet, and the generalization was corroborated in private and public databases. In all databases, the AAMI recommendations for class labeling and results presentation were followed. Different strategies to integrate the additional information available in the 12-leads were studied. The best performing strategy consisted in performing principal component analysis to the wavelet transform of the available ECG leads. The performance indices obtained for normal beats were sensitivity (S) 98%, positive predictive value (P^+) 93%; for supraventricular beats, (S) 86%, (P^+) 91%; and for ventricular beats (S) 90%, (P^+) 90%. The generalization capability of the chosen strategy was confirmed by applying the classifier to other databases with different number of leads with comparable results. In conclusion, the performance of the reference two-lead classifier was improved by taking into account additional information from the 12-leads.**

***Index Terms*—Cross-database, heartbeat classification, linear classifier, multilead, principal component analysis (PCA), wavelet transform (WT).**

I. Introduction

CARDIOVASCULAR diseases are currently the biggest single cause of death in developed countries according to their public health agencies [1]. The analysis of the ECG signal is a noninvasive, inexpensive, and well-established technique to analyze the heart function. One important aspect that cardiologists must study in the ECG are arrhythmias, which 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 heartbeats is an important task for the study of arrhythmias, and could represent a challenge for cardiologists when the analysis is not restricted to short-term recordings. Detection and diagnosis of infrequent or subtle arrhythmias require careful inspection of long-term recordings. Note that this type of arrhythmias may represent a long-term threat without a proper treatment. Therefore, the aid in the analysis provided by automatic algorithms allows cardiologists to improve their diagnostics.

Many algorithms for ECG heartbeat classification were developed in the last decades (see [3] and [4]) using the available two-lead databases. Few of these works used the joint information present in both leads, as the vectocardiogram maximal vector (VCG_M) and angle (VCG_ϕ) used in [5]. Another multilead strategy can be seen in [3], where a final decision is taken from several posterior probabilities calculated from single-lead features. This last approach is not practical for multilead classification because of the need of a different model designed for each set of leads, and the consequent growth in features dimensionality. Some methodological key points in the development of heartbeat classifiers allowed results comparison of several works [3], [4], [6]. Probably, the most relevant aspects were the use of public and standard databases, the fulfillment of Association for the Advancement of Medical Instrumentation (AAMI) recommendations [7], the patient-oriented data division [3], and the generalization capability (at least to the available public databases) of the classifier [4], [8]. The room for improvement in the field of heartbeat classification, together with the availability of 3- and 12-lead Holter devices, makes necessary the development of algorithms capable of exploiting the increase of recorded information. Recently, moreover, the St. Petersburg Institute of Cardiological Technics 12-lead arrhythmia database (INCART) has become freely available on Physionet [9], making possible the evaluation of multilead heartbeat classifiers in a comparable way.

The objective of this paper is to find an effective way of accounting for morphologic information present in multilead ECG signals. For that purpose, we compare several multilead classification strategies against the reference two-lead classifier that we developed in [4]. We assess the improvement in classification performance as well as the generalization capability to other databases not considered during the development. The main novelty presented in this paper is the generalization of the model developed in [4] to an arbitrary number of leads.

II. Methods

A. ECG Databases

In this study, we used the well-known MITBIH arrhythmia database (MITBIH-AR) [10] and other public and private

Manuscript received June 30, 2011; revised October 21, 2011; accepted March 29, 2012. Date of publication April 18, 2012; date of current version July 5, 2012. This work was supported by the Spanish Ministry of Science and Innovation under project TEC2010-21703-C03-02 and the Diputación General de Aragón through Grupos Consolidados GTC T-30 (Spain). The Centro de Investigación Biomédica en Red of Bioengineering, Biomaterials and Nanomedicine is an initiative of Instituto de Salud Carlos III.

M. Llamedo is with the Electronic Department, National Technological University, Medrano 951, C1179AAQ, Buenos Aires, Argentina, and also with the Communications Technology Group, Aragón Institute of Engineering Research, IIS Aragón, University of Zaragoza. María de Luna, 1, 50018 Zaragoza, Spain, and also with the Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain (e-mail: llamedom@electron.frba.utn.edu.ar).

A Khawaja is with Biosigna GmbH, 80337 Munich, Germany (e-mail: khawaja.antoun@biosigna.de).

J. P. Martínez is with the Communications Technology Group, Aragón Institute of Engineering Research, IIS Aragón, University of Zaragoza. María de Luna, 1, 50018 Zaragoza, 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/TITB.2012.2193408

TABLE I
DATABASES USED IN THIS STUDY

Database	N	S	V	F	Q	#Rec
INCART	153651	1959	20005	219	6	75
Biosigna	286246	1326	2541	0	0	56
MITBIH-AR	90089	2779	7007	802	15	44
MITBIH-SUP	162271	12195	9940	23	79	78
Totals	692257	18259	39493	1044	100	253

Heart beats classes are N: Normal, S: Supraventricular, V: Ventricular, F: Fusion, and Q: Unknown.

databases which are briefly described as follows. All public databases are available on Physionet [9]. For all databases, the AAMI recommendations for class labeling were adopted (see [7], Sec. 4.2]). The AAMI Q class (unclassified and paced heartbeats) was discarded since it is marginally represented. This limitation occurs to a lesser extent with the fusion (F) AAMI class, but instead of discarding the heartbeats of this class, a class-labeling modification to the AAMI recommendation is proposed here and was adopted. It consists in merging fusion (of normal and ventricular beats) and ventricular classes, as the same ventricular class (V'). We will refer to this modification as AAMI2 labeling. The class distribution for the databases used is shown in Table I.

1) MITBIH Arrhythmia Database (MITBIH-AR): The database consists of 48 two-lead recordings of approximately 30 min and sampled at 360 Hz. The first 23 recordings were extracted from routine ambulatory practice, while the remaining 25 were selected because of the presence of less common complex ventricular, junctional, and supraventricular arrhythmias. The two recorded leads are not the same in all recordings, depending on the arrhythmia and physical limitation of the subject's body. The four recordings with paced beats were discarded in this study in accordance with AAMI [7].

2) MITBIH Supraventricular Arrhythmia Database (MITBIH-SUP): The database consists of 78 two-lead recordings of approximately 30 min and sampled at 128 Hz. The recordings were chosen to supplement the examples of supraventricular arrhythmias in the MITBIH arrhythmia database. The beat type annotations of the recordings were first automatically performed, by the Marquette Electronics 8000 Holter scanner and later reviewed and corrected by a medical student [11]. The original labeling was also adapted to the AAMI recommendations and to the AAMI2 modification.

3) St. Petersburg Institute of Cardiological Technics (INCART) 12-Lead Arrhythmia Database: This database consists of 75 beat type annotated recordings extracted from 32 Holter records. Each record is 30 min long and contains 12 standard leads, each sampled at 257 Hz. The beat annotations were produced by an automatic algorithm and then corrected manually, containing over 175 000 beat annotations in all. The original records were collected from patients undergoing tests for coronary artery disease (17 men and 15 women, aged 18–80; mean age: 58). None of the patients had pacemakers; most had ventricular ectopic beats. In selecting records to be included in the database, preference was given to subjects with ECG's consistent with ischemia, coronary artery disease, conduction abnormalities, and arrhythmias [9].

4) Biosigna Database: This is a private database developed at Biosigna GmbH, Munich, Germany, which consists of 56 recordings containing a broad set of pathologies. Each recording is 1-h length, sampled at 500 Hz with an amplitude resolution of 410 increments per millivolt using a 12-bit ADC, allowing a range of 10 mV approximately. The recordings were manually annotated by experienced annotators. More detailed information about this database can be found in [12].

B. Signal Processing

The recordings of all databases were first resampled to 360 Hz, after filtering with a tenth-order low-pass finite impulse response filter, without observing any notable distortion (resample function, Signal Processing Toolbox of MATLAB, The Mathworks Inc., Natick, MA). After this process, all databases were filtered to remove artifacts as described in [3].

1) Wavelet Transform (WT): The features that describes morphology (explained in following sections) are calculated from the WT of the ECG signal. The WT is defined for a continuous signal $s(t)$ as

$$W_s s(b) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} s(t)\psi\left(\frac{t-b}{s}\right) dt,\ s > 0. \qquad (1)$$

This transformation maps the ECG signal into a time-scale plane (understanding scale as a surrogate of frequency). The responsible of the mapping is the prototype WT $\psi(t)$, affected by both scaling (s) and translation (b) parameters. The WT allows location of details or fast transitions when scale parameter s is small, and coarser aspects or trends for higher values. The translation parameter b indicates the location of these finer or coarser details, resulting in a time-scale plane with the same sampling rate at each scale (Algorithme à trous). We used the derivative of a smoothing function (quadratic spline) as the prototype wavelet $\psi(t)$, so the discrete wavelet transform (DWT) components correspond to smoothed (or low pass) derivatives of the ECG at different scales, implemented as a digital filter bank. As a result, the DWT retains at certain scales the useful information present in the ECG in form of absolute maxima and zero crossings. In this study, we only used scale 4 of the WT since this scale (at a sampling frequency of 360 Hz) retains with good signal-to-noise ratio (SNR) the morphological features of the QRS complex that we desire to model. For background and implementation details, the interested reader is referred to [13] for a more detailed description of the WT and its implementation for ECG delineation.

C. Heartbeats Classification: Classifier and Features

We follow the results obtained in [4] where we developed a heartbeat classifier with good generalization capability, using rhythm and morphological features together with a linear classifier (compensated for the class imbalance).

1) Classifier: Regarding to 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 leads to the

TABLE II
FEATURES USED IN THE MODEL OBTAINED IN [4] ONLY FOR TWO-LEAD RECORDINGS

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^x	Zero-cross position of the WT autocorrelation sequence in lead 1
k_Z^y	Zero-cross position of the WT autocorrelation sequence in lead 2
k_M^x	Maximum position of the WT autocorrelation sequence in lead 1
k_M^y	Maximum position of the WT autocorrelation sequence in lead 2

linear classifier defined by the discriminant functions [14]

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

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

The values of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}$ in (2) were computed from the training data with the sample mean

$$\boldsymbol{\mu}_i = \frac{1}{M_i} \sum_{m=1}^{M_i} \mathbf{x}_m \qquad (3)$$

and weighted covariance matrix expressions

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

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 ith class, the rule assigns an unlabeled observation $\mathbf{x}$ to the class i which 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 classification tasks were performed using the PRtools toolbox [15] for MATLAB (The Mathworks Inc.).

2) Features: The features used in [4] are described in Table II. As the rhythm features of the model do not depend on the number of available leads, the first four features in Table II remain the same. Therefore, we will focus the analysis on those features describing heartbeat morphology, which are the ones that can be improved by the addition of new leads.

The morphology features calculated are the first zero crossing (k_Z^L) and maximum position of the autocorrelation sequence of the ECG WT at scale 4 (k_M^L) for each lead L (see [4] for more details). Both features were calculated in four sets of leads to study the most suitable way of including the additional information. 1) The first strategy consists of including k_Z^L and k_M^L from all available ECG leads and is referred as 12L (or 3L when only three leads are available), resulting in two morphology features per lead. 2) The second strategy computes k_Z^L and k_M^L from the three vectocardiogram (VCG) leads X, Y, and Z, transformed from 12L by the Dower matrix. This strategy can only be performed in 12L recordings. 3) In the third strategy, referred as ECG-PCA, we apply principal component analysis (PCA) to the available ECG leads; then, the morphology features are computed from the two most important components. 4) Finally, for the fourth strategy, called WT-PCA, the PCA is applied not to the ECG, but to the fourth scale of the WT ($W_4 s(k)$), and the two morphological features k_Z^L and k_M^L were calculated from the principal components.

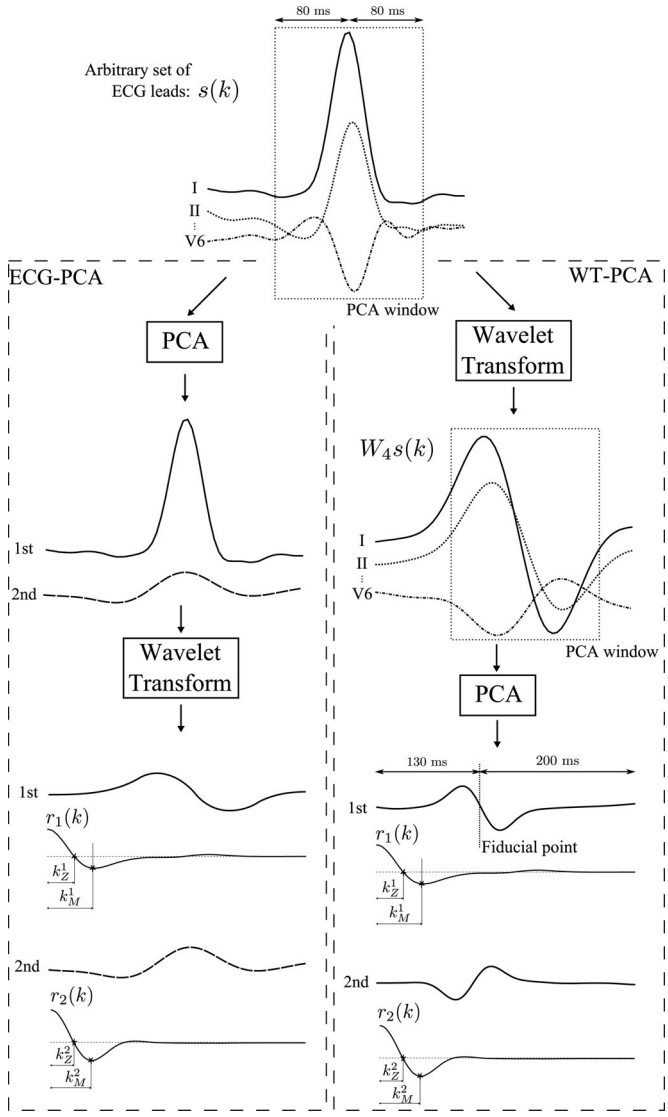

Fig. 1. Illustration of the features calculated from the wavelet correlation signals. The autocorrelation sequence of the QRS complex at scale 4 is shown for both 12L-PCA and WT-PCA strategy. The calculated features, zero crossings and peaks of the autocorrelation sequence, are indicated with an asterisk.

The PCA is performed for each heartbeat in a 160-ms window centered at the QRS complex detection sample, or fiducial point (PCA window in Fig. 1). Then, the multilead signal is projected into the PCA basis in a wider window starting 130 ms before and ending 200 ms after the QRS fiducial point. Then, one or two principal components are selected to compute in them the morphological features, as it is known from previous works that the first two PCA of the ECG retain most of the signal energy [16]. Note that the main difference between ECG-PCA and WT-PCA strategies, as it can be seen in Fig. 1, is where the PCA is calculated: in the first case, PCA is applied to the ECG

signal, while in the second case, PCA is applied to the $W_4 s(k)$ signal.

Then, for strategies 12/3L, VCG, and ECG-PCA, we calculated $W_4 s(k)$ for each ECG lead. Remember that for WT-PCA, the WT was calculated previous to the PCA. After that, the autocorrelation sequence of $W_4 s(k)$ is calculated, obtaining $r_L(k)$, where the final step consists in detecting the first zero crossing (k_Z^L) and the position of the first minimum (k_M^L), as shown in Fig. 1.

D. Performance Evaluation

We use two approaches to evaluate the performance of a classification experiment. One consists in estimating the parameters of the classifier in a training dataset, for the subsequent evaluation of its predictive performance in a disjoint test dataset. The other approach used is known as cross validation, and is adopted when there are not many examples to build the train and test datasets. It consists in dividing a dataset into k disjoint folds (we use ten folds), and use them as test sets, obtaining, therefore, k performance measures, one for each fold. The division is performed by patients to avoid the presence of heartbeats of the same patient in both training and test datasets. Note that each cross-validation fold implies training in $k-1/k$ of the database patients, and testing in the remaining $1/k$. The resulting performance is the mean of the k performances.

Now, we focus on how the performance is evaluated for both of the approaches described previously. In a multiclass classification problem, the confusion matrix shows the outcome achieved by a classifier and a detailed distribution of the misclassified events. For a *C* class problem, the confusion matrix is a square matrix of dimension *C*

$$\begin{array}{cc} & \text{Estimated classes} \\ & \begin{array}{ccccccc} & 1 & \dots & i & \dots & C & \\ 1 & n_{11}^T & \dots & n_{1i}^F & \dots & n_{1C}^F & N_1 \\ \vdots & \vdots & \ddots & \vdots & & & \vdots \\ \text{True classes } i & n_{i1}^F & \dots & n_{ii}^T & \dots & n_{iC}^F & N_i \\ \vdots & \vdots & & \vdots & \ddots & & \vdots \\ C & n_{C1}^F & \dots & n_{Ci}^F & \dots & n_{CC}^T & N_C \\ & P_1 & \dots & P_i & \dots & P_C & N_T. \end{array} \end{array}$$

For the ith class, n_{ii}^T is the number of correctly classified examples and n_{ij}^F is the number of examples of class i classified as class j; N_i is the total number of examples for class i, P_i is the number of examples classified as class i, and N_T is the total number of examples in the dataset

$$N_i = n_{ii}^T + \sum_{m \neq i} n_{im}^F$$

$$P_i = n_{ii}^T + \sum_{m \neq i} n_{mi}^F$$

$$N_T = \sum_{i=1}^{C} N_i = \sum_{i=1}^{C} P_i.$$

The performance is calculated from this matrix, 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 (P^+) as suggested in [7]. Then, S_i and P_i^+ for the ith class are defined as

$$S_i = \frac{n_{ii}^T}{N_i} \tag{4}$$

$$P_i^+ = \frac{n_{ii}^T}{P_i} \tag{5}$$

and the global accuracy (A), sensitivity (S), and positive predictive value (P^+) are calculated as

$$A = \frac{1}{N_T} \sum_{i=1}^{C} n_{ii}^T = \sum_{i=1}^{C} \frac{N_i}{N_T} S_i \tag{6}$$

$$S = \frac{1}{C} \sum_{i=1}^{C} S_i \tag{7}$$

$$P^+ = \frac{1}{C} \sum_{i=1}^{C} P_i^+. \tag{8}$$

From these equations, it is clear that any imbalance in the class representation affects P^+, P_i^+, and A calculation, but not S and S_i.

Although the AAMI recommendation does not suggest any measure to deal with the strong class size imbalance (see Table I), we considered weighting the classes previous to the calculation of P_i^+ and A in order not to neglect the performance of the less represented classes. The balancing approach used in this study consists in multiplying each row of the confusion matrix by a constant such that the sum of each row N_i is equal for all classes, or $N_i = N_j, \forall i \neq j$. This is equivalent to repeat examples of the less represented classes, in order to balance the class presence.

III. Results

As a first experiment, we compared the classification performance of the different multilead strategies (12L, VCG, ECG-PCA, and WT-PCA) in the INCART database. For each of the strategies, we tested the performance in different subsets of leads. Table III shows class and global classification performances using a $k$fold (k =10) cross-validation approach. In Table IV, we show the results of a similar experiment in the same database, but considering that only three pseudo-orthogonal leads were available (AvF, V2, and V5). In both cases, the best model resulted the one with features computed from the first two leads of the WT-PCA set. Therefore, this multilead strategy is considered for the next experiment.

In the last experiment, we validated the generalization capability of the classification model to other databases for different number of leads. For all databases available for a given number of leads, we assessed the performance using all possible pairs of different databases as train and test sets. We also evaluated the cross validated performance within each database. To have an upper bound reference, we additionally assessed the

TABLE III
PERFORMANCE COMPARISON BETWEEN THE DIFFERENT STRATEGIES SEPARATING AAMI2 CLASSES (N, S, AND V') IN INCART

Set of leads	Comments	Leads	# Features	Normal		Suprav.		Ventr.		Total		
				S	P^+	S	P^+	S	P^+	A	S	P^+
12L	best single lead	III	6	98	92	86	87	83	88	89	89	89
	Ref. model [4]	II-V1	8	97	93	87	87	84	89	90	90	90
	all leads	all	28	97	93	86	89	86	89	90	90	90
VCG	best lead	Y	6	98	93	83	83	81	85	87	87	87
		X Y Z	10	98	93	82	84	83	85	87	87	87
ECG-PCA		1	6	98	93	87	87	86	90	90	90	90
		1-2	8	98	93	82	87	86	86	89	89	88
WT-PCA		1	6	99	93	86	90	89	90	91	91	91
	Selected model	**1-2**	**8**	**98**	**93**	**86**	**91**	**90**	**90**	**92**	**92**	**91**

The selected strategy is in bold.

TABLE IV
PERFORMANCE COMPARISON BETWEEN THE DIFFERENT STRATEGIES SEPARATING AAMI2 CLASSES (N, S, V') IN THREE PSEUDO-ORTHOGONAL LEADS FROM INCART

Set of leads	Comments	Leads	# Features	Normal		Suprav.		Ventr.		Total		
				S	P^+	S	P^+	S	P^+	A	S	P^+
3L	best lead	V2	6	98	91	88	82	77	89	87	87	88
	all leads	all	10	98	93	86	86	84	88	89	89	89
ECG-PCA		1-2	8	98	93	87	87	86	91	90	90	90
WT-PCA		**1-2**	**8**	**98**	**93**	**87**	**89**	**88**	**91**	**91**	**91**	**91**

The selected strategy is in bold.

TABLE V
PERFORMANCE FOR ALL DATABASES WHERE THE GENERALIZATION OF THE WT-PCA STRATEGY WAS STUDIED

	Tested in	Obs.	Trained in	Normal		Suprav.		Ventr.		Total		
				S	P^+	S	P^+	S	P^+	A	S	P^+
12 leads	INCART	*biased*	*INCART*	*99*	*94*	*89*	*92*	*91*	*92*	*93*	*93*	*93*
			crossval.	98	93	86	91	90	90	92	92	91
			Biosigna	99	91	86	92	89	90	91	91	91
	Biosigna	*biased*	*Biosigna*	*98*	*86*	*92*	*86*	*77*	*97*	*89*	*89*	*90*
			crossval.	98	85	91	85	75	97	88	88	89
			INCART	97	86	89	84	77	96	88	88	89
3 leads	INCART	*biased*	*INCART*	*98*	*94*	*92*	*91*	*89*	*95*	*93*	*93*	*93*
			crossval.	98	93	87	89	88	91	91	91	91
			Biosigna	99	91	84	90	86	88	90	90	90
	Biosigna	*biased*	*Biosigna*	*98*	*85*	*93*	*84*	*73*	*97*	*88*	*88*	*89*
			crossval.	98	84	92	84	73	97	87	87	88
			INCART	96	86	89	84	76	92	87	87	87
2 leads	MITBIH-AR	*biased*	*MITBIH-AR*	*95*	*82*	*83*	*87*	*78*	*87*	*85*	*85*	*85*
			INCART	97	74	75	83	71	89	81	81	82
			crossval.	94	77	70	82	75	80	79	79	80
			MITBIH-SUP	91	68	67	70	65	93	75	75	77
			Biosigna	97	58	40	71	68	89	68	68	72
	MITBIH-SUP	*biased*	*MITBIH-SUP*	*94*	*83*	*78*	*75*	*75*	*90*	*82*	*82*	*83*
			crossval.	93	80	74	74	74	88	80	80	81
			Biosigna	97	66	49	77	76	85	74	74	76
			MITBIH-AR	94	77	45	75	81	69	73	73	73
			INCART	96	75	44	74	77	69	72	72	73
	INCART	*biased*	*INCART*	*98*	*93*	*91*	*88*	*84*	*93*	*91*	*91*	*91*
			MITBIH-AR	93	93	88	84	85	89	89	89	89
			Biosigna	99	89	84	88	83	88	89	89	88
			crossval.	98	91	82	83	79	84	86	86	86
			MITBIH-SUP	88	95	88	76	81	88	86	86	86
	Biosigna	*biased*	*Biosigna*	*98*	*84*	*91*	*85*	*74*	*97*	*88*	*88*	*88*
			crossval.	98	84	90	83	72	96	87	87	88
			MITBIH-SUP	85	91	94	69	68	96	82	82	85
			INCART	97	86	68	81	74	73	80	80	80
			MITBIH-AR	94	87	52	77	80	65	76	76	76

For the biased case train and test distributions are forced to be the same, this being the best performance achievable for this model.

performance of the model when trained and tested in the same database. This optimistically biased performance serves as an upper bound, and represents the performance of the model if the distributions of the examples in both training and test datasets were identical. These results, grouped by test database, are presented in Table V for databases with 12, 3, and 2 leads. Results show that the reference model extended with the selected WT-PCA multilead strategy presents good generalization properties for 3 and 12 leads, while a certain degradation is observed when using only two leads.

IV. Discussion and Conclusions

In this paper, we have adapted and improved a two-lead heartbeat classifier by including the additional morphology information present in multilead recordings, like those of INCART database. We followed the concept of the morphology features assessed in [4], but calculating these features in sets of leads obtained by following different lead transformation strategies. The simplest strategies consisted in computing the features k_Z^L and k_M^L in all available leads (12L/3L), and in the derived orthogonal leads (VCG). The other two strategies apply PCA to the ECG or its WT previously to the morphology feature computation. The results suggest that strategies using PCA performed better. Moreover, the WT-PCA strategy obtains the best improvements with respect to the two-lead classifier obtained in [4], either in recordings of 12 or 3 leads. This can be explained because in WT-PCA, the PCA is calculated in $W_4 s(k)$ where most of the noise and other components not related to the QRS have been filtered out [13], and therefore, PCA provides a better representation of the multilead evolution of the QRS complex. It must be remembered that although both PCA and WT are linear transformations, the eigenvector calculation is not linear, and therefore, sets ECG-PCA and WT-PCA differ in how the multilead signal is projected into the principal components. Tables III and IV show that only the WT-PCA strategy showed the largest observed performance improvement with respect to the two-lead reference model developed in [4].

Results in Table V confirm the generalization capability of the model using the selected WT-PCA strategy to the rest of the studied databases. For the case of 12L recordings, as those included in the INCART and Biosigna databases, results show very good generalization for both databases since the performance is slightly lower than the biased performance when training in the other database. In the same table, almost the same figures can be seen for the case of 3L. The last results presented in Table V are for databases of 2L. It can be noted that the interdatabase dispersion in the performance increased, probably because of the heterogeneity of the databases considered.

The results suggest that databases INCART and MITBIH-AR share similar distributions in the feature space, since both obtained the maximal reciprocal performance. Other interesting aspect regarding the MITBIH databases is the lower performance obtained even for the biased case. This fact evidences the diversity of patients and ECG contamination (noise, lead disconnection, and misplacement) included in these databases, and therefore the limitation of our classifier to model the data and achieve higher performances. Certainly, the biased performance can be thought as a metric of how difficult to classify is a database by a given classification model, the closer to 100%, the easier. This last result reinforces the importance of evaluating the performance of a classifier in several databases.

One advantage of the proposed approach is that it can be used for an arbitrary number of leads, because after the PCA we only retain the two most important components for the morphological feature calculation. These components are calculated specifically for the QRS complex, and in the $W_4 s(k)$ signal (with a band between 11.25 and 22.5 Hz), typically where the ECG presents high SNR. However, in the case of a large-scale artifact during the QRS complex (as a lead disconnection), the PCA calculation would be corrupted, being this the main limitation found for this approach.

The performance improvement with respect to [4] is however moderate, probably because the automatic classification approach is close to the performance limit achievable with the current classification model. The worst aspect of performance remains classification of supraventricular ectopic beats, where further study is needed. Regarding the ventricular class, techniques of patient adaptation as described in [17] are also under development.

This results represent an improvement in performance with respect to the previous two-lead classification model, concluding that the adequate addition of multilead information allows the performance improvement of a heartbeat classifier.

References

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

[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. P. 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] I. Christov, G. Gómez-Herrero, V. Krasteva, I. Jekova, A. Gotchev, and K. Egiazarian, "Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification," *Elsevier Med. Eng. Phys.*, vol. 28, pp. 876–887, 2006.

[6] K. Park, B. Cho, D. Lee, S. Song, and J. Lee, "Hierarchical support vector machine," in *Computers in Cardiology*. vol. 35, Washington, DC: IEEE Computer Society Press, 2008, pp. 229–232.

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

[8] V. Chudácek, G. Georgoulas, L. Lhotská, C. Stylios, M. Petrík, and M. Cepek. (2009). Examining cross-database global training to evaluate five different methods for ventricular beat classification. *Physiol. Meas.* [Online]. *30(7)*, p. 661. Available: http://stacks.iop.org/0967-3334/30/i=7/a=010.

[9] 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.

[10] G. Moody and R. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.

[11] S. Greenwald, "Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information," Ph.D.

dissertation, Harvard–MIT Division of Health Sciences and Technology, Cambridge, 1990.
[12] R. Fischer, M. Sinner, R. Petrovic, E. Tarita, and S. Kääb, T. K. Zywietz, "Testing the quality of 12 lead Holter analysis algorithms," in *Proc. Comput. Cardiol.*, 2008, vol. 35, pp. 453–456.
[13] J. P. Martínez, R. Almeida, S. Olmos, A. Rocha, and P. Laguna, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.
[14] F. van der Heijden, R. Duin, D. de Ridder, and D. Tax, *Classification, Parameter Estimation and State Estimation*. New York: Wiley, 2005.
[15] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. deRidder, D. Tax, and S. Verzakov. (2008). PR-tools, a MATLAB toolbox for pattern recognition. [Online]. Available: http://www.prtools.org.
[16] B. Acar and H. Köymen, "SVD-based on-line exercise ECG signal orthogonalization," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 3, pp. 311–321, Mar. 1999.
[17] 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.

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, in 2005.

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

Antoun Khawaja was born in Damascus, Syria, in 1977. He received the M.Sc. degree in biomedical engineering from Damascus University, Damascus, in 2001, and the Ph.D. degree in biomedical engineering from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2007.

Since 2007, he has been a Researcher and Scientific Developer in the industry. He is currently the Chief Scientist of Biosigna GmbH, Munich, Germany. His current research interests include the field of biosignal processing and advanced automatic ECG analysis.

Juan Pablo Martínez was born in Zaragoza, 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, 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, Zaragoza. His current research interests include biomedical signal processing, with main interest in signals of cardiovascular origin.