

Analysis of Multidomain Features for ECG Classification

M Llamedo Soria^{1,2,3}, JP Martínez^{1,3}

¹Aragon Inst of Eng Research, Univ of Zaragoza, Aragon, Spain

²Universidad Tecnológica Nacional, Buenos Aires, Argentina

³CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Spain

Abstract

In this work we studied the classification performance of models based on intervals, angles and amplitudes. These features were extracted from both ECG leads and different scales of the wavelet decomposition. The MIT-BIH Arrhythmia database was used, following AAMI recommendations for class labeling and results presentation. The training and testing set and any cross-validation division of the database was made patient-oriented. A floating feature selection algorithm was used to obtain best performing models in the training set. This model was evaluated in the test set obtaining a global accuracy of 90%; for normal beats, sensitivity (Se) 92%, positive predictive value (+P) 85%; for supraventricular beats, Se 88%, +P 93%; for ventricular beats Se 90%, +P 92%. This classifier model based on multidomain features performs better than other state of the art methods, with a fraction of the features.

1. Introduction

The analysis of the electrocardiographic signal (ECG) provides a noninvasive technique to analyze the heart function for different cardiac conditions. Particularly, automatic classification algorithms focus on rhythm and morphology analysis of the ECG. Disturbances in the rate, regularity, site of origin or conduction of the electrical impulses are known as arrhythmias [1]. While some types of arrhythmias represent a life threat in the short term (e.g. ventricular fibrillation), there are other types that appear less frequently and represent a long-term threat without proper treatment. It is in those later cases, which require carefully inspection of long ECG recording, where the use of automatic algorithms represents a significant help for diagnostic.

Many algorithms for ECG classification were developed in the last decade (e.g. [2, 3, 4]), but only few of them have completely comparable methodologies and therefore results, although AAMI recommendations [5] have been available since 1998 for easing this problem. From those articles, some of them can classify without any local expert

(LE) assistance, but others can take advantage from a LE to improve the performance. The database used without exception by all the groups was the MIT-BIH arrhythmia database [6].

Regarding to the features used, the RR intervals were used by almost all groups. Other typical choices were the decimated ECG samples (mostly from the QRS complex), or transformed by Hermite polynomials or wavelet decomposition (WT). Some groups use features that integrates information present in both leads, like the vectocardiogram (VCG) maximum value (VCG_{max}) and VCG angle (VCG_{angle}).

Since the methodology for feature extraction generally requires previous ECG delineation, we propose an ECG classifier that uses a set of features already calculated in the delineation stage. These features should preferably be calculated from the WT in a multilead (ML) approach, resulting in a classification model robust to the typical noise present in ECG signals. The objective of this work is to develop and evaluate a model for fully automatic ECG classification (without LE intervention), under the hypothesis that wavelet and ML features should provide a better classification performance.

2. Methods

2.1. ECG database

In this work we used the MIT-BIH Arrhythmia database [6] for training and evaluating the classifier. The database consists of 48 two-lead recordings of approximately 30 minutes 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 annotations provided with the database were used for training and testing purposes, following the recommendations and class-labeling of AAMI. We adopted the training (DS1) and

test (DS2) set division scheme used in [4] for comparative purposes of the results. We also studied the result of training in a subset of recordings of DS1 ($DS1_M$), given that recordings 201 and 207 have atypical beat morphologies that decrease the generalization ability desirable from training process. Also AAMI unclassified class (Q) was discarded since it is poorly represented in the database. Finally, a class-labeling modification to the AAMI standard was evaluated, considering fusion (of normal and ventricular beats) and ventricular classes, as the same ventricular class. We will refer to this modification as AAMI2 labeling. The division scheme is summarized in Table 1.

Table 1. Scheme of the division of the MIT-BIH database into training ($DS1$ and $DS1_M$) and testing (DS2) sets. Recordings with paced beats were excluded. Heart beats classes are N: normal, S: supraventricular, V: ventricular and F: fusion.

	N	S	V	F	#Rec
$DS1$	45673	929	3755	412	22
$DS1_M$	42502	706	3366	410	20
$DS2$	44053	1833	3202	388	22
Full MIT-BIH	88175	1635	7121	822	44

$DS1$ comprises recordings 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230.

$DS2$ comprises recordings 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234.

$DS1_M$ is $DS1$ without 201 and 207. Recordings 102, 104, 107, 217 have paced beats and were excluded.

2.2. Signal processing

All recordings in the MIT-BIH database were first pre-processed to remove artifacts like described in [4]. No energy or amplitude normalization was done, as we were interested in some amplitude related features. As our objective is the evaluation of a classification model, the QRS location is assumed to be known and used from the annotations included with the MIT-BIH. Many of the considered features are based on the wavelet transform (WT) of the ECG signal. The WT is defined for a continuous signal $x(t)$ as

$$W_a x(b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi \left(\frac{t-b}{a} \right) dt, a > 0. \quad (1)$$

The WT can be discretized using a dyadic scheme where the scale factor is $a = 2^k$ for $k \in \mathbb{Z}^+$, with the same sampling rate at each scale (Algorithme à trous) and is easily implemented as a filter bank. We used a quadratic spline as the prototype wavelet $\psi(t)$, being the derivative of a smoothing function. As a result of this transformation, the original ECG signal can be analyzed as a smoothed derivative at different scales (frequency bands) and translations (times). See [7] for a detailed description of WT and its implementation for ECG delineation.

2.3. Classification features

Following the conclusions of previous works [2, 4], we included in our model both interval and morphological features. As interval features we used the typical features from the RR sequence $RR[i-1]$, $RR[i]$, $RR[i+1]$, the median of the last 10 and 20 RR intervals as an estimate of the local rhythm, and the mean RR interval of the last 10 minutes (RR_{avg}) as a global rhythm.

As morphological features we considered the QRS width (QRS_W), the maximum modulus of the QRS loop (VCG_{max}) and the angle of the loop at this position (VCG_{angle}). Other morphology related feature is the wavelet scale where the QRS complex is mostly projected, since fast evolving signals (like a normal beat) tend to be projected in lower wavelet scales (higher frequency content). This feature is calculated as a weighted sum, where the

$$A_i = \frac{1}{M} \sum_{m=1}^M |W_i x(t_m)| \quad (2)$$

are the mean absolute peak amplitude for scales $i = 1, 2, \dots, 6$ and $M \leq 2$, being M the number of peaks detected at each scale at times t_m . Then is calculated the QRS projected scale for each lead (S_{QRS}^{lead}) as

$$S_{QRS} = \frac{\sum_{i=1}^6 A_{i.i}}{\sum_{i=1}^6 A_i}. \quad (3)$$

As the features to include in our model belong to diverse domains, like \mathbb{R} , \mathbb{R}^+ and S^2 (angular or directional domain) we have to transform or deal with them in order to perform classification tasks. In our case, we assume that each feature is normally distributed and therefore valid in the \mathbb{R} domain. According to this, all interval and morphological features defined in \mathbb{R}^+ should first being transformed to the \mathbb{R} domain by a log operation. Also circular (or S^2) features requires a special treatment that will be briefly described, the interested reader is referred to [8] for details. First consider the circular feature ϑ , as the argument of a complex number of unitary module $e^{j\vartheta}$, being $j^2 = -1$. Then, the expectation of this modified feature defines the mean direction and circular variance[8], counterparts of the linear mean and variance.

$$E[\exp^{j\vartheta}] = \rho_\vartheta e^{j\mu_\vartheta^c} \quad (4)$$

$$\mu_\vartheta^c = \arg(E[\exp^{j\vartheta}]) \quad (5)$$

$$V_\vartheta^c = 1 - \|E[\exp^{j\vartheta}]\| = 1 - \rho_\vartheta \quad (6)$$

Where ρ_ϑ is also known as the resultant length. Then for a multivariate F -dimensional model, where Fw are the indexes of the circular features, the maximum likelihood parameter estimates are

$$(\hat{\mu}_x)_f = \begin{cases} \frac{1}{M} \sum_{m=1}^M x_f(m) & \text{if } f \notin Fw \\ \arg \left(\frac{1}{M} \sum_{m=1}^M \exp(jx_f(m)) \right) & \text{else} \end{cases}$$

$$\hat{\Sigma}_x = \frac{1}{M-1} \sum_{m=1}^M \mathbf{x}'(m) \cdot \mathbf{x}'(m)^T \quad (7)$$

being

$$\mathbf{x}'(m) = \left[x'_1(m) \dots x'_F(m) \right]^T \quad (8)$$

$$x'_f(m) = \begin{cases} x_f(m) - (\hat{\mu}_x)_f & \text{if } f \notin F_w \\ (x_f(m) - (\hat{\mu}_x)_f)_{\text{mod } 2\pi} & \text{else} \end{cases} \quad (9)$$

As can be noted from equation 7, $\hat{\Sigma}_x$ can be directly calculated from the circular mean $(\hat{\mu}_x)_f$ and the raw data without any other calculation, even with the same algorithm used in linear features.

2.4. ECG classification

We used both linear and quadratic discriminants for classification purposes. The general quadratic discriminant functions for feature vectors \mathbf{x} , and the i -th class can be written as

$$g_i(\mathbf{x}) = -\frac{1}{2} \mathbf{x}^T \Sigma_i^{-1} \mathbf{x} + \mu_i^T \Sigma_i^{-1} \mathbf{x} - \frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \log(|\Sigma_i|) + \log(P(\omega_i)) \quad (10)$$

where μ_i , Σ_i and $P(\omega_i)$ are the mean vector, covariance matrix and *a priori* probability of the i -th class. The values of μ_i and Σ_i were computed from $DS1/DS1_M$, and equal *a priori* probabilities were considered. The classification rule assigns \mathbf{x} to the class i which results in the maximum $g_i(\mathbf{x})$.

When covariance matrices Σ_i are assumed equal for all classes, the discriminant functions become linear. In this case, a pooled covariance matrix Σ is calculated as in [4]. All classification tasks were performed with the PRtools toolbox [9] for Matlab.

2.5. Model selection and dimensionality reduction

It is well known that low dimensional models generalize better to examples not presented during the training phase, resulting in a more robust and realistic classifier. In order to obtain the smallest, best performing model a floating feature selection algorithm was used [10]. Also many models were explored empirically, following the results obtained from the floating search.

2.6. Performance evaluation

All models were evaluated in $DS1/DS1_M$ using a k-fold cross-validation scheme, where each fold is one of the

22/20 recordings present in each division scheme. This is motivated by the fact that the automatic classifier will have to deal with new subjects instead of new heart beats from the same subjects. It is clear from [4] that this later scheme tends to optimistically bias the results.

The performance was measured in terms of the global accuracy, class sensitivity (Se), and the class positive predictive value (+P) as suggested in [5] for both, training and testing datasets. Although the AAMI recommendation does not suggest any measure to deal with the strong class size unbalance in the MIT-BIH, we equalize class sizes previous to calculate the class +P. Up to the moment we have no indication that any other group addressed this problem in any way.

Finally $DS1/DS1_M$ were used to train the best performing model and $DS2$ exclusively for testing purposes.

3. Results

Results of the model selection procedure are shown in Table 2. The best model for each combination of dataset and labeling are showed for comparison. As we are interested in evaluating the dataset reformulation and the labeling scheme, we selected two models for the final test in $DS2$. Results and confusion matrix in the test set are shown on table 3 and 4 respectively. The best models highlighted in table 3 includes the features $RR[i]$, RR_{avg} , S_{QRS}^1 , S_{QRS}^2 for the AAMI2 labeling, and for the original AAMI labeling QRS_{angle} was added.

4. Discussion and conclusions

The floating feature selection algorithm did not perform as well as expected, obtaining suboptimal models that were modified to achieve the performance detailed in table 2. Although this limitation, the algorithm provided a coarse estimation of the feature's performance, allowing a posterior fine tuning. From table 2 we evaluated the generalization performance of the best performing models trained in both $DS1$ and $DS1_M$. When testing in $DS2$, all our selected models performed as good or better than the suggested in [4]. Other interesting aspect about these models is the considerable smaller size than previous works, thus improving the parameter estimation in smaller datasets and favoring the generalization. The final decision was made considering both performances in $DS1/DS1_M$ and $DS2$, resulting our choice in favor of a model of 4 features: $RR[i]$, RR_{avg} , S_{QRS}^1 and S_{QRS}^2 .

The alternative labeling suggested in this work (AAMI2) improved the classification performance of the resulting classes, this could be produced because of the subtle differences between V and F classes and the low representation of the F class in MIT-BIH. Other remarkable aspect is the careful design of the training set, avoid-

Table 2. Performance for each model on $DS1/DS1_M$ separating all classes. The results are expressed in percentages.

Trained in	labeling	size	Normal		Suprav.		Ventr.		Fusion		Total		
			Se	+P	Se	+P	Se	+P	Se	+P	Acc	Se	+P
$DS1$	AAMI	5	64	76	74	69	31	89	93	54	75	75	67
$DS1$	AAMI2	7	88	78	80	91	86	87	NA ¹		85	85	85
$DS1_M$	AAMI	5	83	87	81	93	79	88	94	74	84	84	86
$DS1_M$	AAMI2	4	92	86	86	93	88	88	NA ¹		89	89	89

¹AAMI2 labeling consider Fusion and Ventricular classes as Ventricular.

Table 3. Final evaluation for the best performing models, trained in $DS1/DS1_M$ and tested in DS2 separating AAMI/2 classes. All results are expressed in percentages.

Trained in	labeling	size	Normal		Suprav.		Ventr.		Fusion		Total			
			Se	+P	Se	+P	Se	+P	Se	+P	Acc	Se	+P	
This work de Chazal et al. [4]	$DS1_M$	AAMI	5	72	90	90	94	89	86	91	75	86	86	86
	$DS1_M$	AAMI2	4	92	85	88	93	90	92	NA ¹		90	90	90
	$DS1$	AAMI	48	87	52	76	39	78	82	89	86	80	80	68

¹AAMI2 labeling consider Fusion and Ventricular classes as Ventricular.

Table 4. Confusion matrices for the final evaluation of highlighted models in table 3. On the left the AAMI labeling model, and on the right the AAMI2 model.

		Algorithm				Totals			Algorithm				Totals
		f	n	s	v				n	s	v		
Reference	F	353	18	1	16	388	Reference	N	40532	2434	1220	44186	
	N	9053	31918	1884	1331	44186		S	126	1622	88	1836	
	S	19	43	1648	126	1836		V	319	46	3237	3602	
	V	268	24	47	2875	3214		Totals	40977	4102	4545	49624	
Totals		9693	32003	3580	4348	49624							

ing the inclusion of rare examples, like in rec. 201 and 207 that degrades generalization.

Future improvements should corroborate the generalization capability of the models studied in this work in databases like AHA. Another improvement is the adaptation to the patient under study by means of a local expert.

Acknowledgments

This work was supported by projects TEC-2007-68076-C02-02 from CICYT and GTC T-30 from DGA (Spain). The CIBER of Bioengineering, Biomaterials and Nanomedicine is an initiative of ISCIII.

References

[1] Taylor GJ. 150 Practice ECGs: Interpretation and Review. Blackwell Science, 2002. ISBN 0-632-04623-6.
 [2] Hu YH, Palreddy S, Tompkins W. A patient-adaptable ecg beat classifier using mixture of experts approach. IEEE Transactions on Biomedical Engineering 1997;44:891–899.
 [3] Lagerholm M, Peterson C, Braccini G, Edenbrandt L, Sörnmo L. Clustering ecg complexes using hermite functions and self-organizing maps. IEEE Transactions on Biomedical Engineering 2000;47:838–848.

[4] de Chazal P, O’Dwyer M, Reilly RB. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering 2004;51:1196–1206.
 [5] Testing and reporting performance results of cardiac rhythm and st-segment measurement algorithms. American National Standard, ANSI/AAMI/ISO EC57, 1998–(R)2008.
 [6] Mark R, Moody G. Mit-bih arrhythmia database 1997. <http://ecg.mit.edu/dbinfo.html>, 1997.
 [7] Martínez J, Almeida R, Olmos S, Rocha A, Laguna P. A wavelet-based ecg delineator: Evaluation on standard databases. IEEE Transactions on Biomedical Engineering 2004;51:570–581.
 [8] Bahlmann C. Directional features in online handwriting recognition. Pattern Recognition 2006;39(1):115 – 125. ISSN 0031-3203.
 [9] Duin R, Juszczak P, Paclik P, Pekalska E, deRidder D, Tax D, Verzakov S. Pr-tools, a matlab toolbox for pattern recognition, 2008. URL <http://www.prtools.org>.
 [10] Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. Pattern Recognition Letters 1994; 15(11):1119–1125.

Address for correspondence:

Mariano Llamedo Soria, llamedom@unizar.es
 Universidad de Zaragoza. María de Luna 1, Edificio Ada Byron
 Lab. 2.05 – CP: 50018, Zaragoza, España.