Screening and Prediction of Paroxysmal Atrial Fibrillation by Analysis of Heart Rate Variability Parameters

C Maier, M Bauch, H Dickhaus

University of Heidelberg / University of Applied Sciences Heilbronn, Heilbronn, Germany

Abstract

This study has been performed within the scope of the CinC-2001 challenge on detection and prediction of paroxysmal atrial fibrillation (PAF) from 200 paired two channel ECGs of 30 minutes duration. Different features of heart rate variability (HRV) describing the magnitude as well as the regularity of heart rate fluctuations and the number of supraventricular (SVPCs) and ventricular (VPCs) premature beats were investigated for their suitability with respect to the classification task using ROC-analysis, classification by ranks and linear polynomial classifiers with jackknife validation. Moreover, time courses of mean parameter values were calculated to identify possible trends. Although promising results of up to more than 80 % accuracy in screening and 92 % in prediction were achieved on training data, these were not reproducible on an independent test set.

1. Introduction

Atrial fibrillation (AF) is one of the most common arrhythmias observed in clinical practice. Although not immediately life-threatening for itself, secondary complications, especially thromboembolism, can imply dramatic consequences and pose a major risk of stroke. Mapping studies have suggested multiple reentrant wavelets within the atrial tissue as a basic mechanism.

Often, sustained forms of AF are preceded by paroxysms of AF. Since there are ways to electrically stabilize or circumvent atrial arrhythmias, a reliable possibility of predicting the spontaneous onset of PAF would be of high clinical interest. The CinC challenge 2001 aims at identifying changes in the surface ECG which might be suitable as markers for a prediction algorithm or for the identification of patients prone to PAF.

Studies have shown many different modes in the initiation of PAF [1], however there are hints on the autonomic nerve system (ANS) playing an important role as a trigger for its spontaneous onset [1, 2].

The aim of this study is to investigate the utility of parameters characterising heart rate variability (HRV), which are known to reflect ANS control [3], in identifying patients prone to PAF and predicting its onset.

2. Material and methods

The data set used in this study consists of 200 2-channel ECG signals of 30 minutes in duration, recorded from 99 different persons with a sampling frequency of 128 Hz and 16 bit amplitude resolution. It is divided into two groups (training set and tests set) each containing records of patients suffering from PAF (A-group), and probands (N-group) which are either healthy or have cardiac disorders different from PAF. For each proband, two ECG records are provided. In the A-group, one record (AA-record) immediately precedes the onset of atrial fibrillation (AF) whereas 45 minutes before and after the second record no AF is present (AN-record). In the N-group, inevitably, both records are distant from AF.

For the training set, information was available whether a proband belongs to the A- or the N-group (each 25 persons in size) and which record is the one preceding AF. In the test set, no clues were given for the record pairs. The only fact known was that there are between 20 and 30 A-group patients in the tests set, the actual number, as announced later, is 28.

The challenge is divided into two events: Event 1 (PAF-screening) aims at identifying the PAF-patients in the test set, whereas in event 2 (PAF-prediction) the one record out of each pair preceding the onset of AF had to be identified.

Our approach is based on the application of standard pattern recognition techniques on features that mainly quantify heart rate variability (HRV). Prior to QRS-detection we increased the sampling rate to 1024 Hz by means of cubic spline interpolation and applied a median highpass filter (width: 501 samples) to reduce baseline wander. To assess whether more relevant information on the classification problem is contained in the ectopic beats or normal beats, we performed separate analyses on the interval series containing ectopics (RR-series) and on the series consisting only of normal-normal (NN) intervals. Identification of ectopic beats was based on prematurity- and delay-thresholds that were automatically derived individually for each record by analysis of the relative increase of percentile-values calculated from the difference of successive RR-intervals and from deviations of an estimated tachogram baseline. Each beat suspected

0276-6547/01 $17.00 © 2001 IEEE

Computers in Cardiology 2001;28:129-132.

to be ectopic was then submitted to a correlation analysis where it was classified as being either of supraventricular origin when the cross correlation coefficient with the dominant QRS-waveform exceeded the value of 0.9 or of ventricular origin otherwise.

We were interested to see whether classification results would improve when shorter data segments temporally closer to the target event are used as basis for feature extraction. Therefore, we calculated parameters on the whole 30 minutes segment as well as on intervals of 10, 5, and 2 minutes in duration at the end of each record, which in case would immediately precede the onset of AF. To possibly identify typical time courses, we plotted average parameter values calculated on a 5 minutes basis for the whole duration of the record.

A part of the features used in this study are statistical time domain measures common to HRV analysis [3], such as the standard deviation (SD) of all RR intervals (SDRR) or between successive beats of normal origin, (SDNN), the relative (pNN50) number of successive pairs of NN-intervals that differ more than 50 ms and the square root of the mean of the summed squares of differences between adjacent NN-intervals (RMSSD). These parameters mainly quantify the magnitude of HRV.

To be able to analyze variations on different timescales, we applied the discrete wavelet transform (DWT) with the Daubechies10-wavelet on scales 1 to 10 to the interbeat interval series. Standard deviations of the wavelet coefficients [5] were calculated on the available scales and used as absolute (DWT1 .. DWT10) and normalized (DWT_N1 .. DWT_N10) numbers, with respect to the maximal value.

The final set of features aims at quantifying the structure or regularity of heart rate fluctuations [4] and is calculated from vectors of dimension m obtained by time-delay embedding the interbeat interval time series x_i.

$$\vec{x}_i = (x_i \quad x_{i+1} \quad \Lambda \quad x_{i+(m-1)})^T$$

The sorted eigenvalues l_i of the embedding vector's covariance-matrix are the basis for those parameters

$$l_1 \Lambda\, l_m = Eigenvalues(\mathrm{cov}(\vec{x}_i))$$

$$\text{where} \quad l_i > l_{i+1} \qquad \text{for } i = 1\Lambda\, m-1$$

The magnitude of each eigenvalue is normalized with respect to the sum of all eigenvalues:

$$\lambda_i^m = \frac{l_i}{\sum_{i=1}^{m} l_i}$$

and the normalized maximal eigenvalue (NME) serves as classification feature:

$$NME\_m = \lambda_1^m$$

A second feature is calculated as

$$EEV\_m = 2^{H_m}$$

where H_m denotes the Entropy of the embedding space eigenspectrum at embedding dimension m:

$$H_m = -\sum_{i=1}^{m} \lambda_i \cdot ld(\lambda_i)$$

Moreover, four-parameter models were fitted (Levenberg-Marquard algorithm) to the scaling curves obtained by calculation of NME_m and EEV_m over varying embedding dimension m from 2 to 50:

$$EEV(m) = a + b \cdot m + c \cdot m^2 + \frac{d}{m}$$

$$NME(m) = a + b \cdot \log(m) + c \cdot \log^2(m) + d \cdot \log^3(m)$$

The respective model parameters a, b, c and d served as classification features EEV_M1..EEV_M4 and NME_M1 .. NME_M4.

To assess the suitability of single features in the screening task, we determined an optimal threshold from the training set by means of ROC-Analysis and applied it to the test set. Thresholding results in separate decisions c_1 and c_2 for the records 1 and 2 of each patient's record pair. Therefore, post-processing (pp) strategies had to be employed to guarantee consistent decisions $c^{pp}{}_1 = c^{pp}{}_2$ (according to formatting constraints in the challenge) for both records of each patient in screening and contrary decisions $c^{pp}{}_1 \neq c^{pp}{}_2$ in prediction. In the screening event, we tested AND and OR as strategies do derive a conjoint decision $c^{pp} = c^{pp}{}_1 = c^{pp}{}_2$:

$$c^{pp} = \begin{cases} A & \text{if } (c_1 = A) \wedge (c_2 = A) \\ N & \text{otherwise} \end{cases} \qquad \text{AND}$$

$$c^{pp} = \begin{cases} A & \text{if } (c_1 = A) \vee (c_2 = A) \\ N & \text{otherwise} \end{cases} \qquad \text{OR}$$

For the prediction event, classification by comparison of ranks was employed, i.e. we looked how often a parameter calculated from the AA-record immediately preceding the onset of PAF shows consistently higher (or lower) values than the same parameter calculated on the corresponding record of each pair. One may criticize that this approach is highly adopted to the structure of the data set, however the primary aim was to identify features that reflect consistent changes prior to the onset of PAF and in this respect it makes sense to take advantage of an intra-patient analysis.

To evaluate feature combinations, in both events a first order polynomial classifier was trained on the training set and validated using the jackknife or leave-one-out method.

3. Results

Table 1 gives results for the best single features in the prediction event, based on comparison of ranks. All parameters were calculated from the entire 30 minutes, containing all RR intervals. The interpretation of the first line is, that the standard deviation of the wavelet coefficients on scale 9, was greater when calculated from

Table 1: Rank-Distribution for single features (30 minutes, RR intervals) in PAF prediction.

Parameter	PAF-Rank	Training set	Test set
DWT9	>	22/24	14/28
DWT3	>	19/24	
pRR50	>	19/24	16/28
SDRR	>	18/24	
NoSVPCs	>	17/24	
DWT_N3	<	13/24	

the record preceding the onset of AF in 22 out of 24 training set patients, compared to the control record. However, this was reproducible in only 14 out of 28 test cases. Even simple pNN50 performs better here. A higher number of supraventricular ectopic beats in the 30 minutes before AF onset was found in 17/24 patients.

Although, on training data, several features show remarkably consistent rank distributions, a discrepancy between training and test set results is clearly evident, which continues in recognition of patients prone to PAF. Table 2 summarizes sensitivity and specificity for the best single features. The results were obtained by ROC analysis using the indicated threshold and post-processing strategy. Again, several parameters achieve satisfactory results on the training set with an accuracy around 80%, however their sensitivity on the test set is so bad that the results could not be submitted to CinC due to constraint violations. This performance decrease is at least partly due to differing distributions within learning and test set, as can be seen from figure 1 for parameter EEV_M3, one of the model parameters of the eigenspectrum entropy scaling curve. The indicated threshold yields a sensitivity of 79% and specificity of 84% (table 2, figure 1 upper). However, applied to the test set (figure 1 lower), only 13 values, which we would interpret as PAF positive, are found below that threshold. Facing those deviations, it becomes clear that feature combinations using a polynomial classifier must experience the same problem on test data, although jackknife-validated results of up to 22/24 in event 2 and 88%/76% in event 1 using three features suggest reproducibility.

Table 2: Classification results by thresholding of single features in event 1 (30 minutes, RR intervals).

Parameter	Thresh	PP	Training set sens./spec.	Test set
EEV_M3		or	79% / 84%	13 PAF
NME_M3		and	88% / 71%	16 PAF
EEV_M1		and	83% / 71%	13 PAF
NoSVPCs	> 3	and	88% / 69%	
NoSVPCs	> 14	or	88% / 71%	
EEV_7		or	75% / 76%	33/50
DWT_N1	> 0.08	and	79% / 71%	
DWT_1	> 20.88	and	63% / 63%	

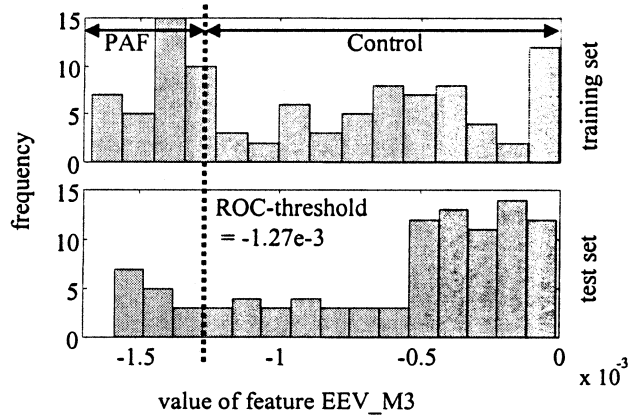

Figure 1. Distribution of parameter EEV_M3 in the training (upper) and test (lower) set. Optimal threshold (dotted line) obtained by ROC-analysis.

If in spite of those disillusioning results we try to identify some trends from the training set, we can see from table 1 that the more successful features in prediction rather quantify the magnitude of heart rate fluctuations, whereas the best features for screening (table 2) are based on properties of the embedding space eigenspectrum and relative magnitudes (DWT_N1 vs. DWT1 in table 1 and table 2). In both events, increased supraventricular ectopic activity is related to PAF, which agrees with findings reported in other studies [1, 2].

Trend plots showed on average an increasing magnitude of HRV in AA-records (figure 2). Therefore, features like SDSD (figure 2 upper), pRR50, SDRR, and the absolute DWT-parameters, yielded slightly better results in the prediction event when calculated on shorter time segments closer to the onset of AF. However, the improvement normally only comprised 1 to 3 better classified pairs, which might not be statistically significant.

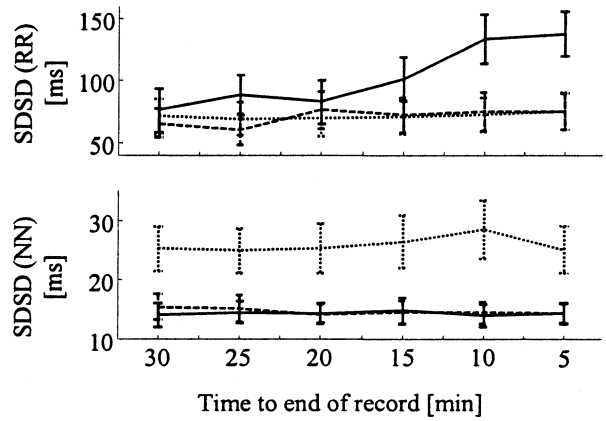

Figure 2. Time course of 5-min mean of parameter SDSD, calculated from RR (upper) and NN (lower) data for AA- (solid), AN- (dashed) and N-records (dotted). Intervals indicate ± standard error of mean.

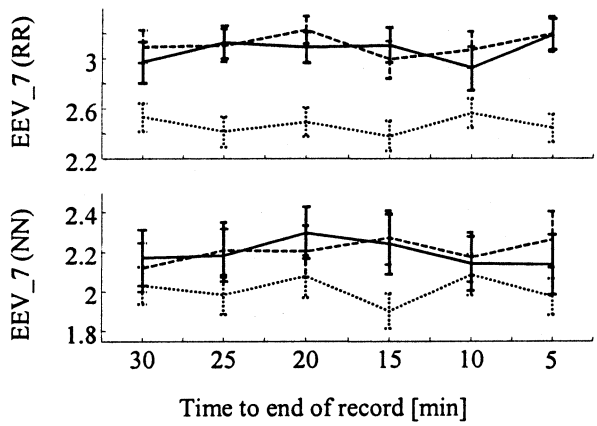

Figure 3. Time course of 5-min mean of parameter EEV with m=7, calculated from RR (upper) and NN (lower) data for AA- (solid), AN- (dashed) and N-records (dotted). Intervals indicate ± standard error of mean.

This increase in magnitude of interval fluctuations turned out to be closely linked to ectopic activity. When calculated on NN intervals, the observed trends and differences between AA- and AN-records vanished for virtually all features (figure 2 lower). Therefore, generally classification performance decreased markedly in event 2 on NN data. On the other hand, for the time domain features, differences between A- and N-groups became more distinct and revealed reduced sinus HRV in PAF-patients. SDNN reached 71% sensitivity and specificity in separating the groups using a threshold of 27 ms.

Remarkably, for the eigenspectrum parameters, neither a trend in the AA-record nor a difference between AA- and AN-record is visible on RR data (figure 3 upper), which explains their worse results in event 2. However, there are time-independent differences between A- and N- group that are reflected in good results using the AND-pp strategy in table 2. Interestingly, although these features don't show the trend of figure 2, they are definitely influenced by ectopics, since differences between A- and N-group are significantly less marked on NN data (figure 3 lower).

4. Discussion

Clearly, the discrepancy between training and test set results renders any conclusions from this study questionable. Looking only at the test set, the most obvious interpretation is that rhythm features alone are not sufficient for the problem of PAF recognition. This however, is inconsistent with the promising jackknife validation results obtained on the training set. Overall, this rather suggests either a sample size too small, or a difference between training and test data, the origin of which remains unclear. Considering the manifold mechanisms underlying AF-initiation [1] and the possible heterogeneity of the groups with respect to cardiac diseases, ectopic activity and medication, both alternatives are plausible.

Well aware of the uncertain validity of any generalization, we would nevertheless like to summarize some interesting observations on the training data:

As expected, we found that most information on the problem is contained in the occurrence of ectopic beats, when analysis is restricted to rhythm features alone. In agreement with previous studies [1, 2], on average, there seems to be a progressive increase in atrial ectopic activity prior to the onset of AF, which – due to prematurity and compensatory pauses - is directly reflected in higher mean values of parameters quantifying the magnitude of HRV as SDRR, pRR50 and RMSSD or SDSD (figure 2, table 1). Due to this temporal trend, features quantifying the extent of HRV were generally more successful in PAF prediction on RR-data, especially when calculated shortly before PAF-onset. The fact that this superiority completely vanishes on NN-data (figure 2 lower), emphasizes the crucial role of SVPCs. In this context, it is highly remarkable, that parameters calculated from the embedding space eigenspectrum, although definitely influenced by ectopics, do not reflect this time course (figure 3). Their concordant behaviour within the PAF-group, consistently different from the controls, rather indicates a possible existence of more permanent HRV changes in PAF-patients, independent of the imminent onset of AF, which might be useful in screening.

The significance of those findings, however, will not be assessable until more information on the composition of the training set is available.

References

[1] Hnatkova K, Waktare JEP, Murgatroyd FD, Guo X, Baiyan X, Camm AJ, Malik M. Analysis of the cardiac rhythm preceding episodes of paroxysmal atrial fibrillation. Am Heart J 1998;135:1010-1019.

[2] Vikman S, Makikallio TH, Yli-Mayry S, Pikkujamsa S, Koivisto AM, Reinikainen P, Airaksinen KEJ, Huikuri HV. Altered complexity and correlation properties of R-R interval dynamics before the spontaneous onset of paroxysmal atrial fibrillation. Circulation 1999;100:2079-2084.

[3] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use. Circulation 1996;93:1043-1065.

[4] Maier C, Bauch M, Dickhaus H. Recognition and Quantification of Sleep Apnea by Analysis of Heart Rate Variability Parameters. In: Computers in Cardiology 2000. IEEE Press, 2000;27:741-744.

[5] Teich MC, Lowen SB, Jost BM, Vibe-Rheymer K. Heart rate variability: Measures and models. In: Akay M, editor. Nonlinear Biomedical Signal Processing Vol. II 2000. Piscataway, NJ: IEEE Press, 2000:159-213.