Screening for Paroxysmal Atrial Fibrillation using Atrial Premature Contractions and Spectral Measures

B Hickey, C Heneghan

University College Dublin, Dublin, Ireland

Abstract

We present a technique for screening for imminent onset of paroxysmal atrial fibrillation (PAF) through automated assessment of 30-minute segments of electrocardiogram (ECG), which do not contain any episodes of atrial fibrillation. Algorithmic development was carried out using a training database of 75 half-hour records drawn from two subject groups. Subjects in the first group provided segments with PAF in the five minutes after the 30-minute recording; the second group do not have PAF (control subjects or subjects with non-PAF cardiac pathology). To differentiate between pre-PAF segments and non-PAF segments a linear discriminant classifier was developed, using the number of Atrial Premature Contractions (APCs) and two spectral measures as features. An independent test set of 72 recordings (28 pre-PAF and 44 non-PAF) was then classified, with an accuracy of 75% (sensitivity 79%, specificity 72%). When tested against a second database of subjects with no known cardiac pathology, the specificity rose to 95%.

1. Introduction

Atrial Fibrillation (AF) is the most common abnormal heart rhythm in clinical practice, and has serious morbidity and mortality [1]. AF is a significant risk factor for stroke; about 15% of strokes occur in people with AF. AF can either be chronic or intermittent; intermittent AF is referred to as Paroxysmal AF. AF is the result of chaotic current flow within the atria, and is usually triggered by a very early Premature Atrial Contraction (PAC), or by atrial tachycardia.

This paper was motivated by the 2001 Computers in Cardiology Challenge [2], in which researchers were asked to determine whether segments of surface ECG contained information sufficient (a) to distinguish subjects at risk of PAF from others not at risk, and (b) to predict imminent PAF in at-risk subjects. Researchers responded to this challenge by developing classification strategies based on detection of atrial premature contractions, P-wave morphology, and a range of standard heart rate variability (HRV) measures [3,4]. In this paper we set out to develop a new classifier, which combined elements of the most successful approaches, and to report results on other independent test data.

The development of predictors for atrial fibrillation onset may have a range of clinical utility. Firstly, means for counteracting atrial fibrillation through atrial pacing techniques have been developed in recent years. Such systems however require good predictors of onset in order to function optimally. Secondly, the definitive diagnosis of paroxysmal atrial fibrillation from 24 hour Holter monitoring can be challenging, since PAF is by definition intermittent. Holter recordings may often yield a negative result, without definitively ruling out the presence of PAF. For example, various reports have demonstrated that the diagnostic yield of event monitors is significantly higher than Holter monitors in establishing the presence of PAF [5]. If our classifier indicates that the person has a high probability of experiencing an episode of AF, but does not actually have an episode, that probability figure may be used as a useful clinical measure of the overall AF diagnosis for that subject.

2. Methods and data

The specific aim of this study was to distinguish between 30-minute segments of ECG data from PAF subjects *immediately* prior to an episode of AF (i.e., where AF occurs within five minutes of the end of the recording), and equivalent segments from non-PAF subjects (either normal subjects or those with pathologies other than PAF). This probably represents an easier task than identifying an arbitrary 30-minute segment to be from a PAF or non-PAF subject, since presumably the 30 minutes immediately prior to the AF episode is likely to contain more information than segments distant from the AF. This is equivalent to Event 1 in the CinC 2001 Challenge, with the modification that distant PAF segments are removed from the training and test data sets. Classification is to be based solely on the electrocardiogram (ECG) recording. ECG records from three databases were used:

(1) The Physionet Atrial Fibrillation Prediction Database, (AFPDB) database consists of 200 30-minute records, extracted from two-channel ECGs (Leads I and II), sampled at 128 Hz, and 12-bit

0276–6547/02 $17.00 © 2002 IEEE

Computers in Cardiology 2002;29:217–220.

resolution. The database contains 53 PAF patients, and 47 patients who do not suffer from PAF. These 47 patients include healthy controls, patients referred for long-term ambulatory ECG monitoring, and patients in intensive care units.

(2) Measurements taken at UCD (UCDDB), consisting of ten overnight records of approximately 8-hours duration from subjects with no known cardiac pathology, aged between 23 to 40 (modified V5), sampled at 128 Hz, and 12 bit resolution. Two 30-minute segments were chosen at random from each of these 10 recordings, to form a test set of 20 records

(3) The MIT-BIH AF database (AFDB) was also used to test the performance of the classifier. This database includes 25 long-term ECG recordings of human subjects with atrial fibrillation (mostly paroxysmal). The individual recordings are each 10 hours in duration, and contain two ECG signals each sampled at 250 Hz with 12-bit resolution.

To train the classifier, 25 pre-PAF and 50 non-PAF 30-minute ECG segments were extracted from the AFPDB. The remaining AFPDB test records (28 pre-PAF and 44 non-PAF), the UCDDB, and the AFDB were used for independent assessment of the classifier.

2.1. Data preprocessing

For the AFPDB and AFDB data, the raw ECG signals were processed using a linear phase bandpass FIR filter windowed using a Kaiser-Bessel window to reduce baseline wander and EMG artefact. The band stop frequencies were set at 0.25 Hz and 20 Hz. QRS detection was done using a Hilbert Transform method [6]. From this set of R peaks, the RR intervals were calculated. Plots of the RR intervals defined in this manner showed that some records had physiologically unreasonable RR intervals. A simple correction algorithm was applied. A new R wave was inserted if an RR interval of greater than twice the average RR interval was found. The number of inserted R waves was directly proportional to the length of the RR interval. i.e., two R waves are inserted if an RR interval three times the mean is found. For the UCDDB data, the Holter monitor returned corrected RR intervals directly.

2.2. Feature extraction

In this study, we searched for distinguishing features in the data. Entrants to the CinC 2001 Challenge, and other investigators have considered various changes in the ECG prior to PAF. Vikman *et al.* suggested that there may be an alteration in the complexity of RR interval dynamics prior to the onset of PAF [7]. Also, AF is often triggered by Atrial Tachycardia or by Atrial Premature Complexes (APCs), and was used by Zong *et al.* to detect onset of PAF [3]. Consequently RR intervals and PACs form the basis of our classifier.

The first feature group (FG1) was derived from the Power Spectral Density of the RR interval sequence [7]. To obtain a zero-mean sequence the mean RR interval was subtracted from each segment. The sequence was zero padded to the nearest power of two exceeding the length of the sequence, and the Fast Fourier Transform (FFT) was taken of the entire sequence. The absolute values of the FFT coefficients were squared to yield a periodogram estimate of the power spectral density. Adjacent frequency bins were then combined to result in a 64 point PSD estimate (of which only bins 0-32 are relevant since 33-63 provide identical information as 1-31). The magnitudes of these PSD bins were used as features. The above analysis was carried out not only on the full 30-minute recording, but also on the last 10, 5 and 1 minutes, to yield four groups of feature sets $FG1_{30}$,$FG1_{10}$,$FG1_5$, and $FG1_1$.

The second feature group (FG2) consists of the number of APCs present in each 30-minute recording. A two-stage algorithm was developed to characterise APCs. The first stage is based on the RR intervals, and the second on QRS morphology. Two types of APCs were defined: Compensated APC, (CAPC) and Non-Compensated APC (NAPC). The first stage works as follows: A beat is flagged as a suspected APC if the RR interval preceding the complex falls 20% short of a moving average of the RR intervals:

$$\overline{RR}[n] = \frac{1}{10}\sum_{i=n-10}^{n-1} RR[i] \qquad (1)$$

If RR_2 is found to vary by less than 10% of the moving average RR interval, a suspected NAPC was identified, i.e.,

$$if \quad |RR_2 - \overline{RR}[n]| < \frac{1}{10} RR[n] \Rightarrow NAPC \qquad (2)$$

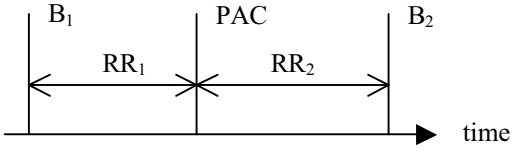

Figure 1: 3 beats represented over time

RR_1 = RR interval between B_1 and PAC
RR_2 = RR interval between PAC and B_2
$RR_3 = (RR_1 + RR_2)/2$

However, if RR_2 is found to be greater than 10% of the moving average, and RR_3 varies by less than 15% of the moving average RR interval, a suspected CAPC is identified, i.e.

$$\text{if } RR_2 - \overline{RR}[n] > \frac{\overline{RR}[n]}{10} \text{ and } RR_3 - \overline{RR}[n] < 0.15.\overline{RR}[n] \Rightarrow CAPC$$

The second stage of the detector calculates the width, amplitude and area of the QRS complex. If the QRS

parameters, as defined above, of the APC vary from those of a normal sinus beat by less than 10%, the beat is confirmed as an APC. The parameters of a normal sinus beat are determined by finding the average parameters of the first 100 beats that show a regular sinus rhythm. This APC detector was tested against the MIT Supraventricular Arrhythmia Database [8]. The algorithm's performance was comparable with that of Zong *et al*. [7], achieving a sensitivity of 59% and a positive predictive accuracy of 73%.

This feature group was also calculated over 30 minutes, final 10 minutes, and final 5 minutes. Surprisingly, only the 30-minute values yielded any significant results. The distribution of PACs also yielded little or no information.

2.3. Classification techniques

A linear discriminant classifier, based on Bayes' rule, was used. The classifier outputs a set of numbers representing the probability estimate of each class, in response to a set of input features. Linear discriminants partition the feature space into different classes using a set of hyper-planes. Optimization of the model is achieved through direct calculation and is extremely fast relative to other models such as neural networks [8]

2.4. Feature selection

An exhaustive search was carried out to find the best 2 features from each feature set, which optimized the classifier performance on the training data set. This included two spectral measures of the RR intervals, the number of CAPCs, and the number of NAPCs. The two optimal spectral measures corresponded approximately to spectral features at frequencies of 0.07Hz and 0.4 Hz. These four features were fed into the classifier and a probability output was calculated. These probabilities represented the likelihood of the recording coming from a pre-PAF segment.

2.5. Performance estimation

To estimate the classifier performance cross-fold validation was used [4]. As the amount of data in this study was quite small, multiple runs of cross validation were used to help improve the performance estimate of the classifier. The 75 records were divided into five folds of 15 records each. Five runs of 5-fold cross validation were used to estimate the performance of the classifier. There is a large variance in the results due to the small size of the training set. The accuracy achieved on the training set varied from 60% to 90%, and gave an expected accuracy of 76%.

3. Results and discussion

Table 1 shows the results obtained on both the test data set and the training set using the two feature groups. Results for feature group 1 (FG1) over four different time periods are presented. The final ten-minute window yielded the best results, followed by the final 5 minutes, final 30 minutes, and lastly the final minute before PAF occurs. However it did not appear that the choice of time period was particularly critical. We observed that FG1 performed better than feature group 2 (FG2) on the training set, whereas on the test set FG2 outperformed FG1 by 2%. To achieve the best results in each case, both sets of features were fed into one classifier. This method proved to be the most robust classifier, yielding the best results. The overall classification accuracy is comparable to that achieved by participants in the 2001 Computers in Cardiology challenge. However, 75% accuracy is probably not sufficient to provide real clinical utility. Accuracy is limited by both relatively low sensitivity (79%) and specificity (72%). One potential

	Testing set			***Training set***	**Test set 1**		
	Accuracy %	*Sensitivity %*	*Specificity %*	*Accuracy %*	*Sensitivity %*	*Specificity %*	*Accuracy %*
FG1 30	70	61	75	78	50	76	66
FG1 5	72	62	77	80	51	79	68
FG1 1	68	60	72	77	50	76	66
FG1 10	72	65	75	80	53	80	70
FG2	70	90	60	76	86	63	72
Combined	76	96	66	82	79	72	75

Table 1: The results for the columns under testing set are obtained by cross validation.
30 - represent the features from 30 minutes of ECG recording;
10,5,1 – features from final 10,5 and 1minutes of ECG recording

explanation is that specificity against non-PAF cardiac pathologies is poor. To test this, we assessed the specificity of the classifier by testing it against a data set of subjects with no known cardiac pathology (the UCD database). Since no incidences of AF occurred in these recordings, the sensitivity cannot be assessed. The specificity of the $FG1_{10}$ feature set was 67%, and the specificity of the FG2 feature set was 95%, with a combined feature set specificity of 95%. This indicates that atrial premature complexes counts are the most useful technique for distinguishing normals from non-normals.

Further testing of the combined-feature classifier was then carried out on the MIT-BIH AF database, which contains records from patients with atrial fibrillation (mostly PAF but some chronic). The atrial fibrillation episodes are annotated in the database. We applied our classifier to a rolling 30-minute segment through these records. Ideally, our classifier should return a probability of 1 for half-hour segments within five minutes of the onset of AF, and 0 otherwise. Figure 2 shows a typical plot of the actual probabilities for one of the 10-hour records, and for a subject with no PAF from the UCD database. The arrows correspond to the points at which an episode of AF occurred. In this case, assuming a threshold of 0.5, both episodes of AF were correctly predicted, but four other episodes were wrongly predicted. The classifier returns values close to zero for the normal subject. The classifier was tested on all files in this long-term database, and achieved a sensitivity of 77%, and positive predictive accuracy of 40% for the chosen threshold value of 0.5. In the AF cases, the classifier often returns quite a high probability of an imminent AF episode, without that episode actually occurring. However, it suggests that a useful strategy for overall assessment of AF is to use such a classifier, and look at the average probability for AF imminence as a diagnostic measure.

4. Conclusion

A classifier based on features derived from the RR interval spectrum and APC counts distinguishes 30-minute pre-PAF segments from 30-minute non-PAF segments with an accuracy of 75% on the AFPDB database. The classifier with combined features was more robust than either feature set alone. Tests on subjects with no cardiac pathology show that APC counts provide good specificity. However, we do not believe our classifier (or any other in the literature) can identify imminent PAF onset with sufficient accuracy to be clinically useful. However, we believe that by using the pre-PAF probability estimates over a long record, it may be possible to provide a useful diagnostic measure of whether a person has PAF, and we will explore this further. Further improvement to the system could be obtained by looking at the type of APCs. APCs can occur in isolation, as part of a bigeminal rhythm, as couplets, or as runs. No attempt was made to distinguish between these occurrences; therefore all APCs were weighted equally in level of importance.

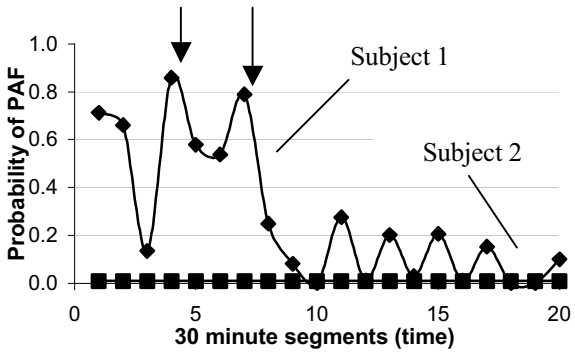

Figure 2: Probability estimates of PAF onset over a 10-hour recording for two subjects (Subject1=AF, Subject2=control). The arrows indicate actual episodes of AF in the AF subject.

References

[1] Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death. Circulation 1998; 98: 946-952.

[2] Moody, GB, Goldberger AL, McClennen S, Swiryn SP. Predicting the onset of paroxysmal atrial fibrillation: the Computers in Cardiology Challenge 2001. Computers in Cardiology 2001; 28:113-116.

[3] Zong W, Mukkamala R, Mark RG. A Methology for predicting atrial fibrillation based on ECG arrthymia feature analysis. Computers in Cardiology 2001; 28:125-128.

[4] de Chazal P, Heneghan C. Automated assessment of atrial fibrillation. Computers in Cardiology 2001; 28:117-120

[5] Kinlay S, Leitch JW, Neil A, Chapman BL, Hardy DB, Fletcher PJ. Cardiac event recorders yield more diagnoses and are more cost-effective than 48-hour holter monitoring in patients with palpitations - A controlled clinical trial. Annals Intern. Med. 1996; 124: 16-20.

[6] D Benitez, P.A. Gaydecki, A Zaidi, A.P. Fitzpatrick. The use of the Hilbert transform in ECG signal analysis. Computers in Biology and Medicine 2001; 31: 399-406.

[7] Vikman S, Makikallio T, Yli-Mayry S, Pikkujamsa S, Koivisto AM, Reinikainen P, Johani Airaksinen KE, and Huikuri HV. Altered complexity and correlation properties of RR interval dynamics before the spontaneous onset of paroxysmal atrial fibrillation. Circulation 199; 100: 2079-2084.

[8] Ripley BD. Pattern Recognition and neural networks. Cambridge University Press 1996.

Address for correspondence.

Conor Heneghan
Department of Electronic and Electrical Engineering, University College Dublin, Belfield, Dublin 4, IRELAND
conor.heneghan@ucd.ie