Comparative Study of Algorithms for Atrial Fibrillation Detection

N Larburu[1,2], T Lopetegi[2], I Romero[1]

[1]Body Area Networks, IMEC, Eindhoven, the Netherlands
[2] Department of Electrical & Electronic Engineering, Public University of Navarra, Pamplona, Spain

Abstract

Automatic detection of Atrial Fibrillation (AF) is necessary for the long-term monitoring of patients who are suspected to have AF. Several methods for AF detection exist in the literature. These methods are mainly based on two different characteristics of AF ECGs: the irregularity of RR intervals (RRI) and the fibrillatory electrical Atrial Activity (AA). The electrical AA is characterized by the absence of the P-wave (PWA) and special frequency properties (FSA).

Nine AF detection algorithms were selected from literature and evaluated with the same protocol in order to study their performance under different conditions.

Results showed that the highest sensitivity (Se=97.64%) and specificity (Sp=96.08%) was achieved with methods based on analysis of irregularity of RR interval, while combining RR and atrial activity analysis gave the highest positive predictive value (PPV=92.75%). Algorithms based on RR irregularity were also the most robust against noise (Se=85.79% and Sp=81.90% for SNR=0dB; and Se=82.52% and Sp=40.47% for SNR=-5dB).

1. Introduction

Atrial Fibrillation (AF) is the most common chronic arrhythmia associated with an adverse prognosis. The incidence of AF increases with advancing age and can have a prevalence of above 10% on the population older than 80 [1]. Some risk factors for AF include hypertension, congestive heart failure, myocardial infarction, and stroke [1]. Moreover the mortality rate of patients with AF is higher than that of patients with normal sinus rhythm (NSR) and the risk of stroke is higher [2]. Accurate detection of AF is crucial since a treatment could offer significant benefits. Unfortunately, AF is often asymptomatic and patients are unaware of their rapid and irregular heart rate due to AF. In such cases, the diagnosis may only be established during a fortuitous doctor visit. The challenge is even greater when episodes of AF are intermittent [3].

AF is characterized by two main ECG features. Firstly, during AF, electrical discharges conducted from the atrium into the ventricles are irregular and as a result, the heart rate becomes irregular and, usually rapid. Secondly, electrical atrial activity is disorganized. These two characteristics can be easily detected in an ECG by noticing the irregularity of the RR intervals (RRI) and the absence of P-waves [3]. Atrial activity (AA) can be analyzed in both time and frequency domain. Time domain analysis is based on the P wave absence (PWA), and frequency spectrum analysis (FSA) consists of the cancellation of ventricular activity (QRS complex and T wave) followed by the spectral computation of the remaining atrial signal. In the literature, several algorithms can be found for the automatic AF detection based on these two characteristics.

However, the algorithms existing in the literature were not always evaluated with the same datasets or followed the same evaluation protocol. Consequently the results reported are difficult to compare. In this research, nine methods from the literature were selected and evaluated by using the same method in order to study their performance under different conditions.

2. Methods

2.1. Databases

Two databases were used throughout this study: MIT-BIH Arrhythmia Database and MIT-BIH AF Database.

The MIT-BIH Arrhythmia Database consists of 48 fully annotated half-hour two-channel ambulatory ECG signals at 360 samples per second (sps) with 11-bit resolution over a range of ±10 mV. However, the number of AF signals is relatively small since only 8 out of the 48 signals contain some AF episodes, which is approximately 126 minutes and 54 seconds in duration [4]. Therefore MIT-BIH Arrhythmia Database was used as a development or training dataset for some of the algorithms.

The MIT-BIH AF Database is a larger data set that consists of 23 two-channel records of approximately ten-hour duration channels sampled at 250 sps with 12-bit resolution over a range of ±10 mV [4]. It contains a larger

number of AF episodes than MIT-BIH Arrhythmia Database and it was used for the evaluation of the algorithms.

2.2.   Evaluation Protocol

Performance Criteria

The results are assessed using four parameters: true positive (TP): AF is classified as AF; true negative (TN): non-AF is classified as non-AF; false negative (FN): AF is classified as non-AF; false positive (FP): non-AF is classified as AF. The performance of the algorithms was reported in terms of sensitivity (Se = TP/(TP+FN)), specificity (Sp = TN/(TN+ FP)), positive predictive value (PPV = TP/(TP+FP)), and the error rate (Err = (FP+FN)/(TP+TN+FP+FN)).

Algorithm Time Resolution Test

This test evaluates the time resolution of the algorithms by using different window lengths (WL) as input to the algorithms. Each window is classified as AF or non-AF independently. WL affects performance, as shorter segments could be more vulnerable to noise. On the other hand, longer segments result in lower time resolution of the algorithm.

Several WL were used in this test, ranging from 10 seconds, until 300 seconds. The WL that gave the highest performance for each technique in terms of maximum value of Se+Sp was identified and used for the remaining tests.

Initial Signal Length

This test was performed in algorithms that required a minimum input signal length to adjust some initial parameters in order to obtain the expected performance. Once a necessary time has elapsed, these parameters are stored in memory.

Noise Test

 The effect of noise in each algorithm was studied by adding noise to 10 clean ECG signals of 1 hour duration. Motion artifact noise was amplified by scaling it by a factor in order to obtain a range of SNR values from -30 to 30 dB. For each signal to noise ratio (SNR), the performance of the algorithms was evaluated and computed in terms of Se and Sp (%).

Computation Time

The computation time was calculated for each signal and averaged over the whole database. This parameter is highly dependent on the hardware used for the algorithm evaluation. The computation time is also affected by the resolution or WL used by each algorithm.

2.3.   AF Detection Algorithms

From the different algorithms existing in the literature, nine were selected. These methods were based on the analysis of both the RRI and AA.

RR Irregularity

The R wave is the most prominent characteristic within the ECG, making it relatively simple to detect. Therefore, algorithms that detect AF based on RRI are the most common in the literature.

Five algorithms based on RRI were selected for evaluation: Moody et al. [5], based on Markov Models (MM); Logan et al. [6] using a simple variance parameter; Linker et al. [7], that used a statistical framework combination; Tatento et al. [8], which applied Kolmogorov Smirnov test; and Cerutti et al. [9], which used an autoregressive modeling and compares RRI with white noise. R peak detection was done using Romero et al.'s algorithm [10].

Atrial Activity

P wave is absent in the AA within AF ECGs and replaced by a fibrillatory wave. AA can be analyzed in both time and frequency domains. Time domain consists of detecting the P wave or finding the P wave absence (PWA). Frequency spectrum analysis (FSA) requires cancellation of ventricular activity (QRS complex and T wave) and Fourier analysis of the remaining AA. Electrical AA on AF ECG is characterized by higher energy concentration in the band of 4-10 Hz as compared to normal ECGs [11].

Slocum et al.'s algorithm [11] was the one based only on the AA analysis to identify AF. Firstly, QRS-T cancellation was done and the remaining atrial signal is studied. In case P wave was detected, the signals were classified as non-AF. Otherwise, FSA was calculated and it was considered as AF if the total spectral power in the band of 4-9 Hz was more than 32% of the total spectrum.

QRST complex was removed using Romero et al.'s algorithm [12].

Combination of RRI and AA

Combination of both RRI and AA (PWA and/or FSA) are used to enhance detection performance.

Three algorithms were selected in this category for evaluation. Schmidt et al. [13] combined RRI using MM, with PWA and FSA. Babaezaideh et al. [2] added to RRI PWA based on the position and morphology of the P

wave. Finally, Couceiro et al. [14] combined the three main physiological characteristics of AF (RRI, PWA and FSA) and classified using Neural Networks model created previously.

3. Results

Table 1 shows the results obtained for the MIT-BIH AF database when the best WL was used for each algorithm. Linker et al. (RRI) had the highest Se=97.64%. Tatento et al. (RRI) algorithm evaluated with Kolmogorov Smirnov test for DRR study showed the highest Sp=96.08% and the lowest error, Err= 5.32%. Babaezaideh et al. (RRI+AA) had the highest PPV=92.75%. On the other hand, AF detection depending only on AA analysis (Slocum et al.) showed the lowest performance: Se=62.80%, Sp=77.46% and the highest error values of 28.39%.

The method with the highest time resolution was Linker et al., which obtained its highest performance with a WL of only 10 seconds. On the other hand, Slocum et al. was the algorithm that used the longest WL (180 seconds).

Some algorithms, such as the one proposed by Logan et al., require the initial calculation of one or some of their parameters. Therefore, an initial input signal is required to calculate those parameters and obtain the expected performance. In the case of Logan et al., a minimum initial input signal length of at least 1000 seconds was required.

Cerutti et al. showed the lowest computation time (0.36 seconds per hour of data analyzed). This algorithm was tested with the longest window length segment, and it is based only on RRI. The algorithms which combine several techniques, and also those that use a classifier require higher computation times, e.g. Couceiro et al. (11.35 seconds per hour of data analyzed).

For the robustness tests, as the level of noise increased, the algorithms that obtained the highest Se and Sp values were those based on RRI techniques. Tatento et al. obtained a Se=85.79%, Sp=81.90% with 0dB of SNR and Cerutti et al. obtained Se=82.52%, Sp=40.47% with -5dB of SNR (Figure 1).
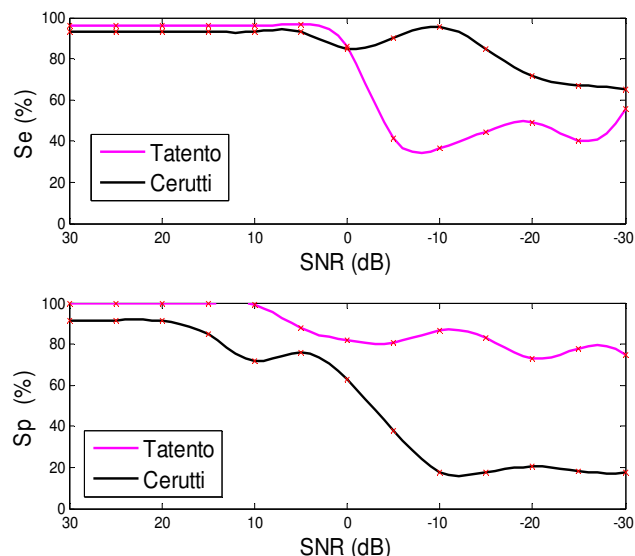

Figure 1. Noise effect in Tatento et al. and Cerutti et al. algorithms: Se (%) above and Sp (%) below, against SNR from 30dB to -30dB

Another common trend noticed during the noise test was the increase in Se coupled with decrease in Sp. Se increased with higher levels of noise in several algorithms (Logan et al., Linker et al., and Slocum et al.). This is because noise can show a chaotic baseline similar to AF, which is detected as AF, and thus resulting in an increment in Se while Sp decreases significantly.

Table 1. Results of the comparative study

Algorithm	Method	WL (seconds)	Se (%)	Sp (%)	PPV (%)	Err (%)
Moody et al.[5]	RRI	60	87.54	95.14	92.29	7.88
Logan et al. [6]	RRI	120	87.30	90.31	85.72	10.89
Linker et al. [7]	RRI	10	97.64	85.55	81.81	9.61
Tatento et al. [8]	RRI	50	91.20	96.08	90.32	5.32
Cerutti et al. [9]	RRI	90	96.10	81.55	75.76	16.62
Slocum et al. [11]	AA (PWA/FSA)	180	62.80	77.46	64.90	28.39
Schmidt et al. [13]	RRI/AA(PWA/FSA)	60	89.20	94.58	91.62	7.57
Babaezaideh et al. [2]	RRI/AA(FSA)	40	87.27	95.47	92.75	7.80
Couceiro et al. [14]	RRI/AA(PWA/FSA)	60	96.58	82.66	78.76	11.77

4.   Conclusions

The decision on which algorithm to use depends on the application in which it is used.

When an algorithm with high sensitivity is required, Linker et al., based on RRI analysis (Se=97.64%) is the best. When specificity is more important, Tatento et al. also based on RRI analysis, (Sp=96.08%) is preferred. This algorithm also gave the lowest error (5.32%), yet still having a high Se, making it a good choice.

In ambulatory conditions, the level of noise is high and therefore, the algorithms based on RRI are preferred as they are more robust against noise. Cerutti et al. showed Se=82.52%, Sp=40.47% with SNR=-5dB, and Tatento et al. showed Se=85.79%, Sp=81.90% with SNR=0dB. On the other hand, algorithms based on AA performed poorly since P wave detection was masked when the noise level was increased.

For real time AF detection, small window lengths are preferred. Linker's algorithm required a relatively short signal length of 10 seconds, giving Se=97.64%, Sp=85.55%, PPV=81.81% and Err=9.61%. Most of the algorithms did not show a good performance with short window lengths and required 1 minute of data in order to have a good performance. For most of the methods, analysis showed a good performance for window lengths of around 1 minute.

Since the QRS complex is the most prominent feature of an ECG and AA (P wave) is difficult to analyze in the surface ECG, most robust techniques for automated AF detection are those based on RRI.

In this study, the algorithms were chosen on the basis of the performance reported by the authors. Preference was also given to algorithms based on different approaches, thus discarding methods which were based on similar principles to those considered in this work. As a limitation to the study, it is important to note that some of the algorithms that were not implemented in this study could contribute significantly to a more extensive study.

References

[1]   Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation, 1998; 98(10):946-52.

[2]   Babaeizadeh S, Gregg RE , Helfenbein ED, Lindauer JM, Zhou SH. Improvements in atrial fibrillation detection for real-time monitoring. Journal of Electrocardiology 2009; 42: 522-526

[3]   Dash S, Chon KH, Lu S, Raeder EA. Automatic Real Time Detection of Atrial fibrillation. Annals of Biomedical Engineering. NY, USA : Biomedical Engineering Society 2009; 37: 1701-1709

[4]   Physionet: the reseach resource for complex physiologic signals. www.physionet.org [online]

[5]   Moody GB, Mark RG. A new method for detecting atrial fibrillation using R-R intervals. Computers in Cardiology 1983; 10:227-230

[6]   Logan B, Healey J. Robust Detection of Atrial Fibrillation for a Long Term Telemonitoring System. Computers in Cardiology 2005; 32: 619-622

[7]   Linker DT. Long-Term Monitoring for detection of Atrial Fibrillation. Patent Application Publication 2006. Seattle, US

[8]   Tatento K, Glass L. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals. Medical & Biological Engineering & Computing 2001; 39: 664-671.

[9]   Cerutti S, Mainardi LT, Porta A, Bianchi AM. Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes. Computers in Cardiology, 1997;24: 77-80

[10]   Romero I, Addison PS, Reed MJ, Grubb N, Clegg GR, Robertson CE, Watson JN. Continuous wavelet modulus maxima analysis of the electrocardiogram: Beat characterisation and beat-to-beat measurement. Int J Wavelets, Multiresolution Inf Process 2005; 3(1): 19-42.

[11]   Slocum J, Sahakian A, Swiryn S. Diagnosis of Atrial Fibrillation From Surface Electrocardiograms Based on Computer-detected Atrial Activity. Journal of Electrocardiology, 1992; 25: 1-8.

[12]   Romero I, Wubbeler G, Elster C. Common spatial pattern: an improved method for atrial fibrillation wave extraction based on the principal component analysis. Computers in Cardiology 2007; 34: 501-504.

[13]   Schmidt R, Harris M, Novac D, Perkhun M. Atrial Fibrillation Detection. Patent Cooperation Treaty 2008. Eindhoven, Netherlands

[14]   Couceiro R, Carvalho P, Henriques J, Antunes M, Harris M, Habetha J. Detection of Atrial Fibrillation using model-based ECG analysis. 19th International Conference on Pattern Recognition, 2008. Tampa, FL. 1-5

Address for correspondence.

Nekane Larburu Rubio
IMEC/Holst Cetre
High Tech Campus 31
5605 KN Eindhoven
The Netherlands
nekanelarburu@hotmail.com