

# ECG-derived Markers to Identify Patients Prone to Atrial Fibrillation

Adrian Luca<sup>1</sup>, Sasan Yazdani<sup>1</sup>, Alain Viso<sup>1</sup>, Jean-Marc Vesin<sup>1</sup>, Giulio Conti<sup>2</sup>, Angelo Auricchio<sup>2,3</sup>

<sup>1</sup>Swiss Federal Institute of Technology, Lausanne, Switzerland

<sup>2</sup>Division of Cardiology, Cardiocentro Ticino, Lugano, Switzerland

<sup>3</sup>Center for Computational Medicine in Cardiology, Lugano, Switzerland

## Abstract

*This study was undertaken to determine the ability of different markers extracted from single lead ECG recorded in sinus rhythm to identify patients prone to atrial fibrillation (AF). For this purpose, 5-minute ECGs recorded in sinus rhythm from two populations were compared: patients with a history of AF and healthy subjects without any history of AF. Several features based on P-waves and RR-intervals were extracted from the ECG. Among the extracted features, the most discriminative ones to identify the AF susceptibility were the P-wave duration, the standard deviation of the beat-to-beat Euclidean distance between successive P-waves and the sample entropy of the RR-intervals. The discriminative power of the aforementioned features was assessed using a classification tree approach. The results showed that the combination of P-wave duration, beat-to-beat Euclidean distance between P-waves and sample entropy could efficiently separate the two populations and therefore be used as an effective detection tool of patients at risk to develop AF.*

## 1. Introduction

Atrial fibrillation (AF) is the most common type of sustained arrhythmia with a very complex behavior and a high incidence rate compared to other cardiac arrhythmias. Thereby, developing methods for the early detection of patients at risk of AF, *i.e.* patients prone to develop AF in the near future, from ECG recordings is of clinical importance.

Atrial fibrillation is a reentrant rhythm and atrial conduction delay is an important precondition for reentrant activity. Previous studies have shown that P-wave duration reflects abnormal conduction in the atria and consequently P-wave indices such as the maximum duration, dispersion, area or axis, have been used to identify patients with an increased risk of AF development [1, 2].

Many other studies have pointed out that the addition of the P-wave morphological analysis, to the traditional time-domain P-wave indices, may help in identifying patients prone to develop AF. The morphological variability of the P-wave has been efficiently used for early prediction of paroxysmal AF (PAF) events [3-4]. Prediction of the PAF onset by the analysis of RR-intervals in ECG segments distant from the PAF event has been also addressed in several studies [5-7].

This study aims to investigate whether electrographic P-wave characteristics and heart rate variability (HRV) features, extracted from a single lead ECG recorded during sinus rhythm, are able to discriminate between subjects with a documented history of PAF and individuals with no history of AF.

## 2. Methods

### 2.1. Data

This study consisted of 76 consecutive patients with no overt cardiac abnormalities divided into two groups: the study group including 36 patients with a history of AF ( $59 \pm 9$  years, 29 male) and the control group including 40 patients ( $54 \pm 10$  years, 20 male) without any history of AF. Overall, the two groups did not present significant differences in terms of clinical characteristics except for the body mass index that was significantly higher for the patients with AF history when compared to the patients without a history of AF ( $29 \pm 4.6$  kg/m<sup>2</sup> vs.  $26.4 \pm 3.8$  kg/m<sup>2</sup>,  $p < 0.01$ ).

Table 1. Clinical characteristics of the study population.

	Patients with AF history (N = 36)	Patients without AF history (N = 40)
Age, years	$59.3 \pm 9.5$	$54.1 \pm 10.4$
BMI (kg/m <sup>2</sup> )	$29.04 \pm 4.6$	$26.44 \pm 3.8^*$
Heart rate, beats/min	$63.48 \pm 8.4$	$62.78 \pm 9.5$
Male n (%)	29 (80%)	23 (57%)

\*  $p < 0.01$

## 2.2. Signal processing and P-wave extraction

The recorded 12-lead ECGs lasted 5 minutes and were acquired in sinus rhythm at a sampling frequency of 1 KHz. In what follows, only the lead II was considered as it exhibits the highest P-wave amplitude.

From each ECG signal, QRS complexes were detected after correction of the baseline by two consecutive median filters. The high-frequency noise and the power line interferences were also removed by a five-order low-pass Butterworth filter with a cutoff frequency of 40 Hz. A search window preceding each detected R-wave was created to identify the P-wave (the width and position of the search window were fixed for all beats, but dependent on the subject). The onsets and offsets of the P-waves were obtained using the first- and second-derivative approximations of the ECG signal. Premature beats were removed from analysis.

## 2.3. P-wave and heart rate variability features

In addition to the traditional electrocardiographic characteristics (P-wave duration, PR-interval), we computed the beat-to-beat Euclidean distance between P-waves as a measure of P-wave morphological variability over time. Note that, the Euclidean distance between beat-to-beat P-waves was already efficiently used in our group to compare ECG segments from healthy individuals with that of patients subject to paroxysmal AF selected for catheter ablation [8].

P-wave duration was computed as the difference between the offset and the onset of the P-wave. PR-interval was defined as the interval between the P-wave onset and the onset of the QRS complex. After a resynchronization of P-waves with regard to their centers of gravity, the beat-to-beat Euclidean distance was computed as:

$$EDist_i = \frac{\sqrt{\sum_{k=1}^N (P_{i+1}[k] - P_i[k])^2}}{\sqrt{\sum_{k=1}^N (P_{i+1}[k])^2}}, \text{ for } i = 1, 2, \dots, L - 1$$

where  $P_i = \{P_i[1], P_i[2], \dots, P_i[k], \dots, P_i[N]\}$  represents the  $i^{\text{th}}$  P-wave ( $N$  being its total length in samples) and  $L$  is the number of P-waves extracted from the lead II of one subject. It is noteworthy that the assessment of P-wave morphological variability, using Euclidean distance, is carried out without any a prior assumption about the P-wave pattern/template [4, 9].

Figure 1 shows a representative example of beat-to-beat Euclidean distance estimation for a healthy subject (top) and for a patient with AF history (bottom), respectively. P-wave morphological variability over time was quantified using the mean and the standard deviation of beat-to-beat Euclidean distance.

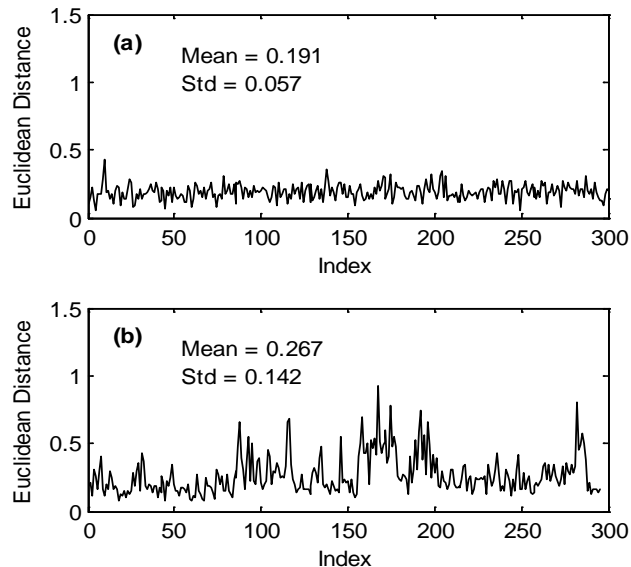


Figure 1. Beat-to-beat Euclidean distance between extracted P-waves for a healthy subject (a) and a patient with history of AF (b). Mean and standard deviation of the beat-to-beat Euclidean distance are also indicated.

Alongside P-wave features, we aimed at examining more high-level heart rate variability (HRV) features in order to find out if they were capable of separating the two population of our study. To this end, R-waves were extracted from the ECG, ectopic beats were removed, and normal-to-normal beat intervals (NNI) were extracted. Afterwards, time-domain, frequency domain and nonlinear HRV indices were extracted. Table 2 reports HRV features extracted for our analysis. Comprehensive information about these features can be found in [10-12].

Table 2. HRV features extracted.

HRV Feature type	Feature Name	Description/Parameter
Time Domain	Mean/std HR	-
	Mean/std NNI	-
	RMSSD	-
Frequency Domain *	PNNx	x=[10, 20, ..., 90]
	NP** ULF	ULF = (0,0.003)
	NP VLF	VLF = (0.003,0.04)
	NP LF	LF = (0.04,0.15)
Nonlinear	NP HF	HF = (0.15, 0.4)
	Sample Entropy	Dimension( $m$ ) = 2
	Approximate Entropy	Distance( $r$ ) = 0.2

\* NNI was regularly sampled at 4 Hz.

\*\*NP = Normalized Power

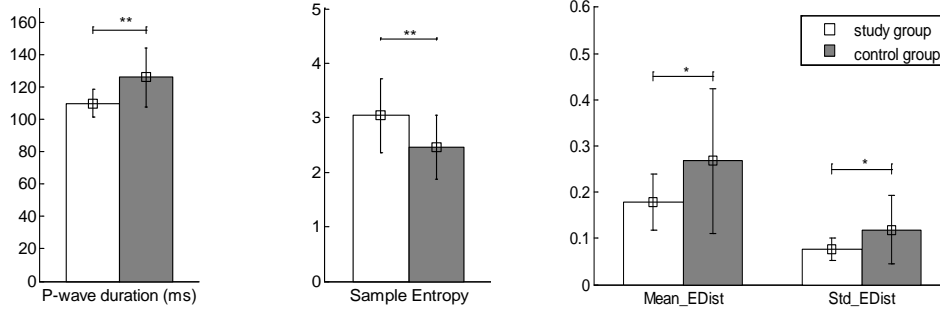


Figure 2. ECG characteristics features extracted from the lead II. The mean (square) of each group is shown. The vertical lines delimit the mean  $\pm$  the standard deviation (\* $p < 0.01$ , \*\* $p < 0.001$ ).

## 2.4. Statistical analysis

All values were expressed as a mean  $\pm$  standard deviation. The statistical comparison between the two groups was performed by using the Mann-Whitney U test for unpaired data. Results were considered to be statistically significant at  $p < 0.01$ .

The discriminative capability of the extracted features was assessed using correlation-based feature selection [13]. Individual predictive ability of each feature together with the correlation between features were used to select the best subset of features.

With the most powerful features selected, the global classification accuracy, defined as number of patients correctly classified between the two population, was obtained using classification trees. Finally, in order to assess the robustness and to generalize the performance of the created classifier, a 10-fold cross-validation approach was used. More specifically, the database was randomly divided into 10 subsets, with subsets roughly having the same sizes. Each subset was used as a test set while the classifier was trained on the remaining subsets. Finally, the overall accuracy of the classifier was obtained as the average of all trained classifiers.

## 3. Results

Individual feature performance analysis showed the most representative features for the two populations to be the P-wave duration, mean and standard deviation of the beat-to-beat Euclidean distance, and sample entropy of the NNI (Figure 2; mean and standard deviation of each feature are also indicated).

The patients with a history of AF presented significantly longer P-wave duration ( $125 \pm 18$  ms vs  $110 \pm 8$  ms,  $p < 0.001$ ) and higher mean and standard variation of the beat-to-beat Euclidean distance (higher variability of P-wave morphology over time) when compared to the control group.

On the other hand, HRV analysis showed that patients with a history of AF presented lower sample entropy compared to patients without previous AF history

( $2.4 \pm 0.59$  vs  $3.0 \pm 0.67$ ,  $p < 0.001$ ). In other words, RR-intervals presented lower complexity in subjects with a history of AF.

The classification between patients with a documented AF and healthy subjects was performed on P-wave duration, beat-to-beat Euclidean distance, sample entropy, and a combination between these features. The numerical results are reported in Table 3.

Table 3. Classification results.

	Sensitivity	Specificity	Accuracy
F1	55%	95%	76%
F2	41%	95%	70%
F3	63%	75%	70%
F1 + F2	61%	92%	78%
F1 + F3	63%	92%	78%
F1 + F2 + F3	81%	92%	86%

F1: P-wave duration.

F2: Standard deviation of beat-to-beat Euclidean Distance.

F3: Sample Entropy.

The numerical results showed that the P-wave duration presented limited ability to detect the altered atrial conduction in patients with documented AF, using ECG segments distal from the paroxysmal AF event. The accuracy (the ratio between the correctly classified subjects and the total number of analyzed patients) was 76%. 121 ms was found as the optimal threshold (53% sensitivity and 95% specificity) for the P-wave duration above which patients presented an increased risk of AF. With regard to the classification performance using the P-wave morphological time-changes, standard deviation of the beat-to-beat Euclidean distance presented a global accuracy of 70%, but with a low sensitivity. In addition, RR-intervals variability provided good ability to identify patients at risk of AF development, but in contrast showed a low specificity.

In a multivariate classification analysis based on P-wave duration, standard deviation of beat-to-beat Euclidean distance and sample entropy, the global

accuracy was 86%, with sensitivity and specificity equal to 81% and 92%, respectively.

#### 4. Conclusion

The present study showed that combination of different electrocardiographic characteristics driven from P-wave and RR-intervals analysis could lead to an effective tool to identify patients at risk of AF. Concretely, the combination of P-wave duration, standard deviation of the beat-to-beat Euclidean distance and sample entropy significantly improved the classification rate into patients without/with AF history as compared to the results based only on individual features.

Moreover, our findings are in line with previous statements that P-wave duration could present limited abilities to detect the loss of homogeneity in interatrial contraction and the P-wave morphological variability should be also taken into account.

#### Acknowledgements

The work has been supported by the Theo-Rossi di Montelera Foundation (Lausanne, Switzerland).

This study was performed in the framework of the Nano-Tera initiative supported by the Swiss National Science Foundation (SNSF).

#### References

- [1] Nielsen JB, et al. P-wave duration and the risk of atrial fibrillation: results from the Copenhagen ECG Study. *Heart Rhythm* 2015; 12(9):1887-1895.
- [2] German DM, Kabir MM, Dewland TA, Henrikson CA, Tereshchenko LG. Atrial Fibrillation Predictors: Importance of the Electrocardiogram. *Annals of Noninvasive Electrocardiology* 2016; 21: 20-29.
- [3] Alcaraz R, Rieta JJ. Morphological variability of the P-wave for premature envision of paroxysmal atrial fibrillation events. *Physiological Measurement* 2014; 35:1-14.
- [4] Martínez A, Alcaraz R, Rieta JJ. Gaussian modeling of the P-wave morphology time course applied to anticipate paroxysmal atrial fibrillation. *Computer Methods in Biomechanics and Biomedical Engineering* 2015; 18:1775-1784.
- [5] Chesnokov YV. Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artificial Intelligence in Medicine* 2008; 43: 151-165.
- [6] Mohebbi M, Ghassemian H. Prediction of paroxysmal atrial fibrillation using recurrence plot-based features of the RR-interval signal. *Physiological Measurement* 2011; 32:1147.
- [7] Boon KH, Khalil-Hani M, Malarvili MB, Sia CW. Paroxysmal atrial fibrillation prediction method with shorter HRV sequences. *Computer Methods and Programs in Biomedicine* 2016; 134: 187-196.
- [8] Cabasson A, Dang L, Vesin JM, Buttu A, Abächerli R, Leber R, Kappenberger L. P-wave indices to detect susceptibility to atrial fibrillation. *Computing in Cardiology* 2011; 257-260.
- [9] Censi, F., Calcagnini, G., Ricci, C., Ricci, R.P., Santini, M., Grammatico, A. and Bartolini, P., 2007. P-wave morphology assessment by a gaussian functions-based model in atrial fibrillation patients. *Biomedical Engineering, IEEE Transactions on*, 54(4), 663-672.
- [10] Task Force of the European Society of Cardiology. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, 1996; 17: 354-381.
- [11] Pincus SM, Viscarello RR. Approximate entropy: a regularity measure for heart rate analysis. *Obstet Gynecol* 1992; 79:249-55.
- [12] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 2000; 278:H2039-H2049.
- [13] Hall MA. Correlation-based feature selection of discrete and numeric class machine learning 2000.

Address for correspondence.

Adrian Luca  
EPFL SCI STI JMV - Station 11  
1015, Lausanne, Switzerland  
adrian.luca@epfl.ch