

# Combining Template-based and Feature-based Classification to Detect Atrial Fibrillation from a Short Single Lead ECG Recording

Matthieu Da Silva–Filarder<sup>1</sup>, Faezeh Marzbanrad<sup>2</sup>

<sup>1</sup> IPB, Bordeaux, France

<sup>2</sup> Monash University, Melbourne, Australia

## Abstract

*Automated diagnosis of Atrial fibrillation (AF) has remained imperfect despite the threat it represents to millions of people. The main issues which can lead to a misdiagnosis of AF include its episodic nature, disease diversity and noise. The aim of 2017 PhysioNet/CinC Challenge is to classify short single lead ECG recordings as normal sinus rhythm, atrial fibrillation, other rhythm, or noisy. We present a method using heart rate variability features and noise detection features coupled with template-based wave morphology features. The method originality lies in the use of special templates sensitive to the heart rate variability as well as wave morphology. These special templates showed significant results in AF detection performances. Based on Cross-validation, an F1 score of 0.84 on AF classification, and a general classification score of 0.76 were obtained on the training set.*

## 1. Introduction

Atrial fibrillation (AF) is a growing epidemiological problem on the world scale. It significantly increases the risk of cardiac failure and therefore has a substantial impact on morbidity and mortality [1]. Moreover, AF affects millions of people around the world and its risk increases with age [2].

Electrocardiography (ECG) diagnosis can provide efficient AF detection in clinical practice since the main characteristics of AF are the absence of P-wave and/or irregular heart rate variability. Numerous methods have been proposed to detect AF, which are based on either heart rate variability (HRV) features such as Poincare plot analysis [3], or wave morphology with P-wave absence features, or can even use a combination of HRV, atrial activity and P-wave analysis [4]. Despite the large number of existing methods to detect AF, the lack of gold-standard database limited the performance evaluation. With Physionet/CinC Challenge 2017 [5], a large database of labeled ECGs emerged and included a vast diversity of ECGs.

In this paper, a method combining ECG morphology fea-

tures extracted from a template-based comparison, along with RR interval features and noise detection features was proposed. Compared to existing model-based ECG analysis [4], the proposed template computation method reflects the heart rate variability as well as the lack of P-Wave. The remainder of this paper is organized as follows: the section 2 introduces the method characteristics, presenting the features and the classifier, section 3 tackles in-house training details, section 4 exposes the results obtained, followed by a discussion and the conclusion.

## 2. Methods

The block diagram presented on Figure 1. summarizes the proposed method.

### 2.1. Database and Pre-processing

The Physionet/CinC Challenge 2017 training dataset was constituted of 8528 short single lead ECG recordings (9s-60s) classified into four categories: normal sinus rhythm (N), AF rhythm (A), alternative rhythm (O) or too noisy to be classified (noisy). Each ECG was first digitally filtered with cutoff frequencies of 1 and 50Hz. A Pan and Tompkins QRS detector [6] with a refractory period of 250ms was then used with sliding windows of 3 seconds. RR intervals were extracted from the detected QRS complexes. Some ECG presented noise corrupted parts, or missing QRS detections, therefore RR time intervals longer than 3 times the mean RR time interval were discarded.

### 2.2. Heart rate variability features

To detect efficiently the heart rate variability, crucial in AF detection, diverse time-domain features as well as frequency-domain features were computed to study the heart rate variability (HRV). The time-domain features based on RR interval distribution were: Mean RR, Standard Deviation, Root Mean Square Standard Deviation of NN-intervals. Tone and Entropy [7] as well as the Poincare plot indices (width and length) were also estimated and

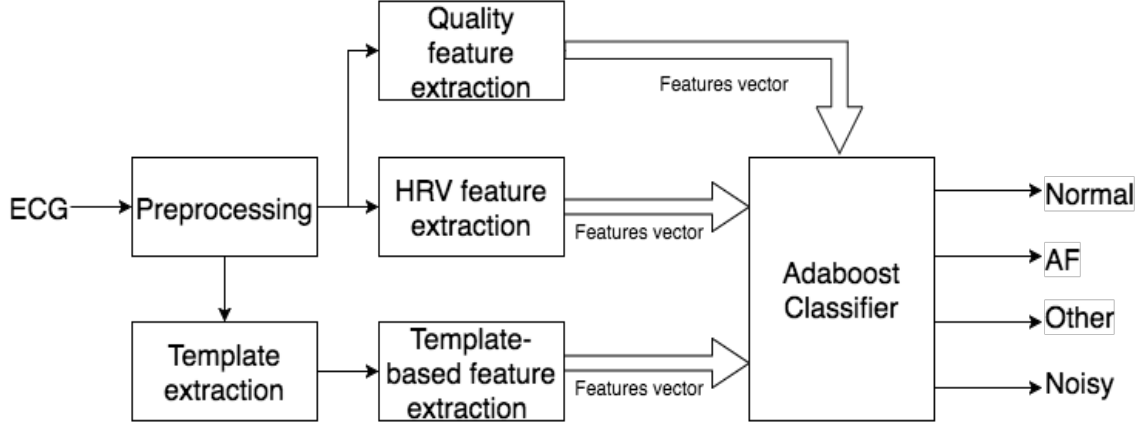


Figure 1. Block diagram summarizing the classification method

used as features. Frequency-domain HRV features such as preponderance of low frequency, high frequency component in the HRV power spectrum were included as features [8].

### 2.3. Noise detection features

Three noise features were extracted in order to classify ECG that were too noisy to be interpreted, in the noisy category :

1) ECG quality index : Percentage of RR intervals with at least a 0.4 correlation coefficient when compared to the mean ECG recording RR interval.

2) bSQI :This method is based on performing QRS detection with two methods having different sensitivity to noise, and comparing the results. It has demonstrated promising differentiation of good quality ECG from noisy signal [9]. In the proposed method, Pan-Tompkins detector and a length transform based detector (wqrs algorithm) [10] QRS references were compared. The F1 coefficient was used as the bSQI feature:

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (1)$$

TP being the number of identified QRS references, FN the number of missed QRS references, and FP the number of extra (incorrect) QRS detection.

3) Karhunen-Loeve noise index: Karhunen-Loeve transform is a common numerical tool in signal processing to isolate noise from a signal, and can be used to perform noise power estimation. Karhunen-Loeve transforms of QRS complexes were computed and the following index

served as feature :

$$KLI = \frac{\sum_{i=1}^P \lambda_i}{\sum_{j=1}^N \lambda_j} \quad (2)$$

Where  $\lambda_i$  is the  $i^{th}$  eigenvalue associated to the  $i^{th}$  eigenvector computed with Karhunen-Loeve transform,  $N$  is the total number of eigenvalues computed, and  $P$  represents the number of eigenvalues corresponding to a QRS signal, chosen at 11 according to the reference paper [11]. This coefficient estimates the denoised QRS signal power relative to the actual signal power.

### 2.4. Template-based features

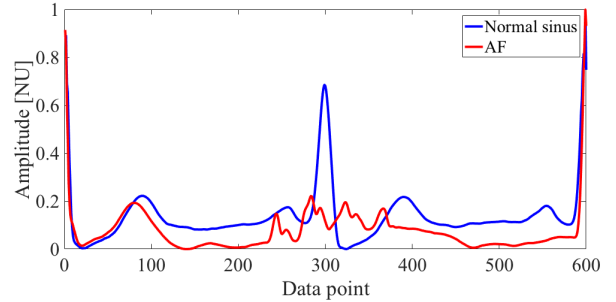


Figure 2. Comparison of two templates with a length of 2 RR intervals: AF (red) template and Normal sinus rhythm (blue) template

In order to accurately classify AF, a template for each ECG recording was computed with the purpose of exploiting the AF P-wave absence characteristic as well as the cardiac period variability. Figure 2 shows that the proposed template has a length of two cardiac cycles, resulting from the entire set sum of consecutive two R-R intervals which

satisfied a 0.2 Pearson's correlation threshold when compared to the ECG recording mean cardiac cycle.

The interest of such a template relied on 1) the P-waves remained absent in the AF template (noticeable on the Figure 2 around the 560<sup>th</sup> point) 2) AF R-R time interval variability induced an absence of distinctive R peak in the middle since the average penalized the non-stationary middle R peak location. Each template was resampled so that every template had the same length of 600 points, and amplitude was normalized between 0 and 1. Two closeness coefficients to the four classification categories were finally defined:

- 1) Pearson's correlation coefficient between the test and train template constituted the first coefficient
- 2) Spectrogram distance coefficient: template spectrograms acquired using discrete Fourier transform with rectangular window and overlapping (window size: 40 points, overlap : 20 points) were computed to perform a spectrogram distance coefficient presented below:

$$C = \sum_i \sum_j |x_{ij} - y_{ij}| \quad (3)$$

Where  $x_{ij}$  and  $y_{ij}$  are coefficients on the  $i^{th}$  row and  $j^{th}$  column of test and train template spectrogram matrix. Several window types, window sizes and overlaps were tested, the presented configuration showed best overall performances after cross-validation tests.

The best coefficients of the two methods obtained in each category (N, A, O and noisy) were used as (eight) features. These coefficients reflected the extent to which the ECG template was similar to the ECG templates of each category with a "nearest neighbor" logic.

### 2.5. AdaBoost Classifier

AdaBoost is a machine learning algorithm combining linearly an ensemble of weak learners to construct a strong classifier [9]. The number of learning cycles was limited to 70 to avoid over-fitting. Although AdaBoost direct classification into 4 categories is possible, AdaBoost binary tree classification showed better results. Figure 3 represents the decision tree used for AdaBoost binary tree classification. First the noisy records were discarded and a new training phase was conducted using only Normal rhythms, AF and Other rhythms records. Then the AF rhythm records were classified leaving only Normal and Other rhythm categories. A final training phase involving exclusively records from these last two categories led to the final classification.

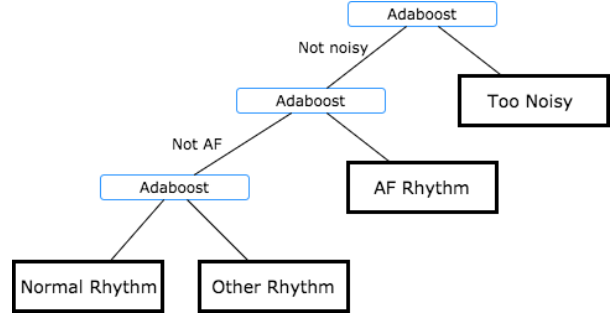


Figure 3. Binary tree used for classification with AdaBoost

### 3. Training and Validation

Algorithm development required training and validation sets to evaluate method performances. Training set contained 1600 ECG records (400 ECG records randomly selected from each category N,A,O and noisy), and validation set contained 400 ECG records (100 ECG records randomly selected from each category). Considering the database imbalanced categories (only 3% of the ECGs in the database were in the noisy category), bootstrapping was implemented for training and validation sets. Repeating training and validation phases 50 times over different sets, and averaging the results, led to stable and reliable results for evaluating method performances.

For the final submission, 40 training sets of 200 ECG records (50 N, 50 A, 50 O and 50 noisy) were prepared. The method proposed in this paper was applied using each training set independently, and led to 40 classifications. The final decision was made by voting.

### 4. Results

Table 1 shows the confusion matrix averaged on 50 validation sets and presents the proposed method results during training:

Table 1. Confusion table shows the performance (%) of our classification method based on cross-validation.

| ground truth \ prediction | prediction |      |      |       |
|---------------------------|------------|------|------|-------|
|                           | N          | AF   | O    | noisy |
| N                         | 81.1       | 1.9  | 12.8 | 4.2   |
| AF                        | 2.0        | 87.5 | 6.6  | 3.8   |
| O                         | 18.2       | 13.3 | 62.4 | 6.1   |
| noisy                     | 3.9        | 5.9  | 12.5 | 77.7  |

Specificity, sensitivity and F1 are presented in the table 2.

Table 2. Specificity, Sensitivity and F1 scores for N, A, O and noisy categories

| Measure     | N     | A     | O     | noisy |
|-------------|-------|-------|-------|-------|
| Sensitivity | 81.1% | 87.5% | 62.4% | 77.7% |
| Specificity | 92.6% | 93.4% | 90.4% | 95.6% |
| F1          | 79.0% | 83.9% | 64.2% | 81.0% |

Sensitivity/Specificity measures are the one generally found in the literature relating AF detection method results. F1 measure was also used by the Physionet/CinC Challenge 2017 to assess the algorithm performances on the hidden test set. Final F1 score acquired with in-house validation was 0,76.

## 5. Discussion and Conclusion

The results show a good differentiation between AF rhythm and Normal rhythm since our method AF true positive rate, true negative rate and F1 measures are respectively the highest, second highest and highest performances among the four categories. The confusion matrix Table 2 shows strong robustness of classification between AF and the remaining categories, with a score as low as 2% of AF ECG wrongly classified as normal ECGs. Normal sinus rhythm classification has the second F1 score with 79%. However confusion matrix exhibits that 12,8% of normal recordings are misclassified as other rhythms. Our method main limitation is the 'Other rhythm' classification since it scored the three worst specificity, sensitivity and F1 scores. Indeed, our template approach is not specifically suitable for Other rhythms category as it may loose valuable features when averaging. A practical example of misclassification can occur with a short ECG recording showing premature ventricular contractions: this type of recordings can exhibit solely one or two cardiac cycles with obvious abnormalities, the remaining cardiac cycles showing normal sinus rhythm. These few abnormal cardiac cycles are likely to be mistaken as noisy segments during R-R quality checking, leading to a classification as normal sinus rhythm. Therefore, the present method might be improved if, instead of systematically computing a single general template, secondaries templates could also be computed. These secondaries templates would contain cardiac cycles with poor correlation to the cardiac cycle majority, and good correlation with one or two other cardiac cycles.

Nonetheless, the other rhythm category constitutes the database innovation. Methods implemented prior to the challenge mainly used databases containing exclusively AF rhythm and normal rhythm ECG records. The database used for this challenge provides more accurate performance analysis thanks to its better equivalence to real-

world applications.

## References

- [1] Munger, TM; Wu, LQ; Shen, WK (January 2014). "Atrial fibrillation.". *Journal of biomedical research*. 28 (1): 117. PMC 3904170Freely accessible. PMID 24474959. doi:10.7555/JBR.28.20130191
- [2] A.J. Camm, P. Kirchhof, G.Y. Lip, U. Schotten, I. Savelieva, S. Ernst, I.C. Van Gelder, N. Al-Attar, G. Hindricks, B. Prendergast, H. Heidbuchel, O. Alfieri, A. Angelini, D. Atar, P. Colonna, R. De Caterina, J. De Sutter, A. Goette, B. Gorenek, M. Haldal, S.H. Hohloser, P. Kolh, J.Y. Le Heuzey, P. Ponikowski, F.H. Rutten, Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC), *Eur Heart J* 31 (19) (2010) 23692429
- [3] Jinho Park, Sangwook Lee, and Moongu Jeon. Atrial fibrillation detection by heart rate variability in Poincare plot *Biomed Eng Online* 2009. doi: 10.1186/1475-925X-8-38, PMID: PMC280
- [4] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, J. Habetha. *Detection of Atrial Fibrillation Using Model-based ECG Analysis*, IEEE, 2008.
- [5] Gari Clifford, Chengyu Liu, Benjamin Moody, Ikaro Silva, Qiao Li, Alistair Johnson, Roger Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017", *Computing in Cardiology (Rennes: IEEE)*, Vol 44, 2017 (In Press)
- [6] Pan J, Tompkins W. A Read-Time QRS Detection Algorithm. *IEEE T Biomed Eng* 1985;32(3):23036.
- [7] Ahsan H. Khandoker, Herbert F. Jelinek. *Evaluating Cardiovascular Risk Using the Tone-Entropy algorithm*, 35th Annual International Conference of the IEEE EMBS, 2013.
- [8] Task Force of the European Society of Cardiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *circulation* 93 (1996): 1043-1065.
- [9] Joachim Behar, Julien Oster, Qiao Li, Gari D. Clifford. *ECG Signal Quality During Arrhythmia and Its Application to False Alarm Reduction* IEEE Transaction on Biomedical Engineering.
- [10] W Zong, GB Moody, D Jiang. *A Robust Open-source Algorithm to Detect Onset and Duration of QRS Complexes*, CinC 2003.
- [11] George B. Moody and Roger G. Mark, *QRS Morphology Representation and Noise Estimation using the Karhunen-Loeve Transform*, *Computers in Cardiology* 1989, pp. 269-272.
- [12] Yoav Freund, Robert E. Schapire. *A Short Introduction to Boosting*. *Journal of Japanese Society for Artificial Intelligence*, September, 1999. 14(5):771-780

Address for correspondence:

Matthieu DA SILVA-FILARDER  
ENSEIRB-MATMECA, 1 avenue du Docteur Schweitzer, 33400  
Talence, FRANCE  
mdasilvafil@ipb.fr