Fusion of Multiple Univariate Data Analysis-based Detectors to Build a Specific Fingerprint of Atrial Fibrillation

Zouhair Haddi[1,2], Bouchra Ananou[1], Youssef Trardi[1], Stéphane Delliaux[3], Jean-Claude Deharo[4], Mustapha Ouladsine[1]

[1]Aix-Marseille University, University of Toulon, CNRS, LIS, Marseille, France
[2] NVISION Systems & Technologies, Barcelona, Spain
[3]Aix-Marseille University, INSERM, INRA, C2VN, Marseille, France
[4] Centre Hospitalier Universitaire de la Timone, Marseille, France

Abstract

Automatic and fast atrial fibrillation (AF) diagnosis is still a major concern for the healthcare professional. Several algorithms based on univariate and multivariate analysis have been developed to detect AF. Although the published results do show satisfactory detection accuracy, computational complexity of such methods is still questionable. This study proposes an alternative way to diagnosis AF arrhythmia which is based on the combination of seven univariate data analysis-based detectors followed by a majority voting in order to build a digital fingerprint of AF. Four publicly-accessible sets of clinical data were used for AF assessment. The time series were segmented in 10 s RR interval window. The features of the four databases were merged in order to give rise huge variability and therefore to better characterize AF arrhythmia. Afterwards, a receiver operating characteristic curve analysis has been conducted to fix optimal thresholds for AF detection. Finally, the seven obtained detectors have been concatenated and then a majority rule was applied to yield a final decision on AF diagnosis. The results showed that this strategy performed better than some existing algorithms do, with 98.50% for sensitivity and 95.1 % specificity.

1. Introduction

According to the World Health Organization (WHO), cardiovascular diseases are among the top 10 causes of death in older adults [1] and classified as chronic non-communicable diseases (NCDs) [2]. NCDs are associated with a slow, long-term, or even lifelong advancement, and are usually silent, thereby affecting the quality of life [3]. Atrial fibrillation (AF) is the most common cardiac arrhythmia diagnosed in clinical practice [4]. It is well-known by a rapid, irregular and heterogeneous electrical activity of the heart that is initiated by atrial contractions disorder [5]. Even though AF is affecting more than 6 million people in Europe and 3 million people in the US [6], the true prevalence of AF is still unknown as many patients are asymptomatic and therefore AF remains undetected [7]. For example, in China, approximately 10 million of people suffer from AF [8], and the estimated AF prevalence is probably 1% in the general population, increasing with age, more so in men than in women [9]. In fact, it is estimated that due to averaging the total number of AF patients will double by the year 2050 [10].

Inter-beat interval time series analysis coupled with rapid technological developments have led to many new commercially available tools intended to automatically detect arrhythmias, from smartphone smart watches applications to noncontact detection methods [11–13]. In this context, a smart phone RR-based application has been conducted on moderately sized prospective cohort study regarding 76 patients with AF undergoing cardioversion [14]. Findings revealed that an algorithm combining normalized root mean square of successive 64-RR difference and Shannon entropy yielded 96.2% and 97.5% for sensitivity and specificity respectively. The authors assumed that exposure patients with a high burden of premature beats and/or atrial tachyarrhythmias might affect the performance characteristics of their iPhone-based AF detection application. Lahdenoja et al. [15] have attempted to detect AF via accelerometer and gyroscope sensors (Inertial Measurement Unit, IMU) of a Smartphone equipped with Google Android OS. Features such as approximate entropy, spectral entropy, turning-point-ratios, heart rate and heart rate variability serve as input to Kernel support vector machine classifier with cross-validation and majority voting strategies. An accuracy of 97.4% in AF vs. healthy classification (a sensitivity of 93.8% and a specificity of 100%) is reached. Andersson et al., have designed a real-time detector for AF episodes as an application specific integrated circuit (ASIC) [16]. The basis for detection was a set of three

 ISSN: 2325-887X DOI: 10.22489/CinC.2020.449

parameters extracted from RR time series, i.e., turning point ratio, root mean square of successive differences, and Shannon entropy. The obtained sensitivity and specificity were more than 94% and 95% respectively when where tested on MIT-BIH Atrial Fibrillation Database. Furthermore, deep learning techniques, as a new generation of machine learning pipelines, began attracting interest to AF screening. In this sense, O. Faust et al., [17] have applied a deep Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) on 100-RR interval. The sensitivity and specificity reached 98.32% and 98.67% respectively with 10- fold cross-validation process. Although the significant values of the sensitivity and specificity reached throughout the development of hand-held and/or cumbersome classifiers, their integration into portable electronic devices is still challenging. Herein, a very simple strategy coupling multiple univariate data analysis and voting majority approach is proposed to detect AF of episode as small as 10 s.

2. Methods

2.1. Databases

The proposed approach was evaluated on four publicly accessible sets of clinical data: AF Termination Challenge Database, MIT-BIH AF, Normal Sinus Rhythm RR Interval Database, and MIT-BIH Normal Sinus Rhythm Databases. The time series were segmented in 10 s RR interval windows (total of 493813 and 25169 time series for NSR and AF respectively after removing outliers). Usually, authors trained their algorithms on one of these datasets and tested them on the remaining. We believe that if the four datasets are merged (and categorized into AF and NSR groups), this could give rise to huge variability which results in good AF characterisation as previously demonstrated by the group research works [18-20].

We have also demonstrated that classifiers accuracy can significantly enhanced through the investigation of time series dynamical state. Indeed, most of reported studies focused on unique sources (i.e., R-peaks). In classical mechanics, it is well-known that a dynamical state is better characterized on the basis of its position, displacement, speed, and acceleration rather than a single position [21]. Thus, the application of time derivatives of RR interval time series can generate additional information sources and could result in rapid and efficient approach for clinical decision support. The forward finite-difference formula, as defined in equation 1 and 2, was used to compute the time derivatives of the time series and the absolute derivatives of the time series:

$$\frac{dx}{dt}[n] \approx \frac{x[n+1]-x[n]}{t_{n+1}-t_n} \qquad (1)$$

$$abs(\frac{dx}{dt}[n]) \approx \left|\frac{x[n+1]-x[n]}{t_{n+1}-t_n}\right| \qquad (2)$$

2.2. Feature extraction

The most used features used to characterise NSR and AF arrhythmia have been calculated. The mean, median, standard deviation and the features listed in table 1 have been extracted from the RR times series and its derivatives.

Table 1. Formulas of the extracted features

Features	Formulas
VAI	$VAI = \left(\sum_{i=1}^{N-1} \lvert\theta_i - 45\rvert\right) \Big/ N-1$ where θ_i is the angle between the line plotted from every scatter point to the original point and the x-axis
VLI	$VLI = \sqrt{\sum_{i=1}^{N-1} (l_i - L)^2 \Big/ N-1}$ where : $L = \sum_{i=1}^{N-1} l_i \Big/ (N-1)$ l_i is length between every scatter point and the original point, L is the mean of all the li, N is the number of scatter points
SD1	$SD1 = std\left(\lvert RR_{n+1} - RR_n\rvert \Big/ \sqrt{2}\right)$
SD2	$SD2 = std\left(\left\lvert (RR_{n+1} + RR_n) \Big/ \sqrt{2} - 2\overline{RR}\right\rvert\right)$
SD1/SD2	Ratio $= \frac{\text{SD1}}{\text{SD2}}$
RMSSD	$RMSSD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(RR_{j+1} - RR_j)^2}$
SDSD1	$SDSD1 = std\left(\lvert J_{n+1} - J_n\rvert \Big/ \sqrt{2}\right)$ where: $J_n = \lvert RR_{n+1} - RR_n\rvert \Big/ \sqrt{2}$
Recurrence Rate	$REC = \frac{1}{(N-m+1)^2}\sum_{j,k=1}^{N-m+1} RP(j,k)$ where: $RP(j,k) = \begin{cases} 1, & d(u_j,u_k) \le r \\ 0, & Autrement \end{cases}$
L_{max}	$Div = \frac{1}{l_{max}}$
L_{mean}	$L_{mean} = \frac{\sum_{l=l_{min}}^{l_{max}} lN_l}{\sum_{l=l_{min}}^{l_{max}} N_l}$

Determinism	$L_{mean} = \frac{\sum_{l=l_{min}}^{l_{max}} lN_l}{\sum_{j,k=1}^{N-m+1} RP(j,k)}$
Shannon Entropy	$ShanEn = -\sum_{l=l_{min}}^{l_{max}} n_l \ln n_l$ where: $n_l = \frac{N_l}{\sum_{l=l_{min}}^{l_{max}} N_l}$
Approximate Entropy	$ApEn(m,r,N) = \phi^m(r) - \phi^{m+1}(r)$ where: $\phi^m(r) = \frac{1}{N-m+1}\sum_{j=1}^{N-m+1} \ln C_j^m(r)$ $C_j^m(r) = \frac{nbr\ \{u_k \| d(u_j,u_k) \leq r\}}{N-m+1} \quad \forall k$ $d(u_j,u_k) = max\{\|RR_{j+n} - RR_{k+n}\| \| n = 0,\ldots,m-1\}$
Sample Entropy	$SampEn(m,r,N) = ln(C^m(r)/C^{m+1}(r))$ where: $C^m(r) = \frac{1}{N-m+1}\sum_{j=1}^{N-m+1} C_j^m(r)$ $C_j^m(r) = \frac{nbr\ \{u_k \| d(u_j,u_k) \leq r\}}{N-m+1} \quad \forall k \neq j$
Correlation Dimension	$D_2(m) = \lim_{r\to 0}\lim_{N\to\infty} \frac{\log\ C^m(r)}{\log\ r}$ where: $C^m(r) = \frac{1}{N-m+1}\sum_{j=1}^{N-m+1} C_j^m(r)$ $C_j^m(r) = \frac{nbr\ \{u_k \| d(u_j,u_k) \leq r\}}{N-m+1} \quad \forall k$ $d(u_j,u_k) = \sqrt{\sum_{l=1}^{m}(u_j(l) - u_k(l))^2}$
Kurtosis	$\beta = \mathbb{E}\left[\left(\frac{u-\overline{RR}}{\sigma_{RR}}\right)^4\right]$
Skewness	$\gamma = \mathbb{E}\left[\left(\frac{u-\overline{RR}}{\sigma_{RR}}\right)^3\right]$

3. Results and discussion

The analysis of the area under ROC has permitted to order the features according to their discriminant power. Features with an area under the ROC curve between 0.9981 and 0.9985 were considered to be an excellent diagnostic accuracy test. Figure 1 shows an example of the standard deviation of the second derivative of 10s RR time series. The application of these thresholds has resulted in the selection of seven features:

- $\text{STD\_d2RR} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\text{d2RR}_j - \overline{\text{d2RR}})^2}$
- $\text{RMSSD\_d2RR} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(\text{d2RR}_{j+1} - \text{d2RR}_j)^2}$
- $\text{M\_abs\_G} = \overline{\text{abs\_G}} = \frac{1}{N}\sum_{j=1}^{N} \text{abs\_G}_j$
- $\text{M\_abs\_d2RR} = \overline{\text{abs\_d2RR}} = \frac{1}{N}\sum_{j=1}^{N} \text{abs\_d2RR}_j$
- $\text{STD\_abs\_d2RR} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\text{abs\_d2RR}_j - \overline{\text{abs\_d2RR}})^2}$
- $\text{RMSSD\_abs\_d2RR} =$ $\sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(\text{abs\_d2RR}_{j+1} - \text{abs\_d2RR}_j)^2}$
- $SD2\text{\_abs\_d2RR} =$ $std\left(\left|{(\text{abs\_d2RR}_{n+1} + \text{abs\_d2RR}_n)}/{\sqrt{2}} - 2\overline{\text{abs\_d2RR}}\right|\right)$

Where:

- $\text{d2RR}_i = \frac{\left(\frac{\partial RR}{\partial t}\right)_{i+1} - \left(\frac{\partial RR}{\partial t}\right)_i}{RR_{i+1}} = \frac{RR_i RR_{i+2} - RR_{i+1}^2}{RR_{i+1}^2 RR_{i+2}}$
- $\partial RR_i = RR_i - RR_{i-1}$
- $abs\_G = |G_i| = \left|\frac{\left|1-\frac{RR_{i+1}}{RR_{i+2}}\right| - \left|1-\frac{RR_i}{RR_{i+1}}\right|}{RR_{i+1}}\right|$;
- $abs\_d2RR = \left|\left(\frac{\partial^2 RR}{\partial t^2}\right)_i\right| = \frac{|RR_i RR_{i+2} - RR_{i+1}^2|}{RR_{i+1}^2 RR_{i+2}}$

As one can see, the selected features belong to the dynamical state of RR time series which is in line with our previous studies. Besides, instead of training the data with well-known machine learning classifiers [22,23], herein, we opted for voting majority approach using the selected features in order to reduce the computational complexity as much as possible. This approach has a rapid execution time since all univariate based classifiers must be executed to make the final decision which is well adapted for real time execution [24] and portable electronic devices. Although its simplicity, the proposed approach has attained satisfactory results in terms of sensitivity and specificity by reaching 98.50% and 95.1 % respectively.

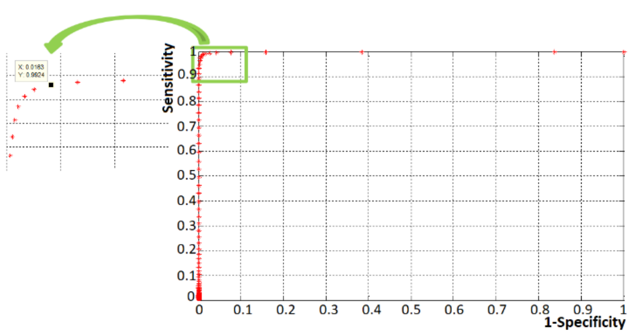

Figure 1. ROC curve for the standard deviation of the second derivative of RR time series.

4. Conclusion

This work aimed at developing AF classifier based on simple strategy combining univariate date analysis through ROC curve and majority voting approach. The first two derivatives of RR time series along with their absolute values have been subjected to a widely feature extraction process including linear, geometrical and non-linear parameters. Seven features from the dynamical space have been selected according their high discriminant power. The application of majority voting approach has resulted in 98.50% for sensitivity and 95.1 % specificity.

References

[1] WHO, "Top 10 causes of death: Situation and trends", 2018, https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/ (accessed 30 July 2020)

[2] L. Allen, Are we facing a noncommunicable disease pandemic? Journal of Epidemiology and Global Health, 7, pp. 5-9, 2017

[3] WHO, "Noncommunicable diseases", 2018, https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed 3 September 2020)

[4] G. Aarabi, R. Schnabel, G. Heydecke, U. Seedorf, Potential Impact of Oral Inflammations on Cardiac Functions and Atrial Fibrillation, Biomolecules, vol. 8, pp. 66, 2018.

[5] A Report of the American College of Cardiology/ American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy. Circulation, vol. 104, pp. 2118–2150, 2001.

[6] CA. Morillo, A. Banerjee, P. Perel, D. Wood, X. Jouven, Atrial fibrillation: The Current Epidemic. Journal of Geriatric Cardiology, vol. 14, pp. 195–203, 2017.

[7] D. Jabaudon, J. Sztajzel, K. Sievert, T. Landis, R. Sztajzel, Usefulness of Ambulatory 7-day ECG Monitoring for the Detection of AF and Flutter after Acute Stroke and Transient Ischemic Attack. Stroke, vol. 35, pp. 1647–1651, 2004.

[8] LSC. Li, CC. Wang, YL. Xia, G. Wu, F. Wang, CQ. Xu, et al., Assessment of Association of RS2200733 on Chromosome 4q25 with Atrial Fibrillation and Ischemic Stroke in a Chinese Han population. Human Genet., vol. 126, pp. 843–849, 2009.

[9] NF. Murphy, CR. Simpson, PS. Jhund, S. Stewart, M. Kirkpatrick, J. Chalmers, K. MacIntyre, JJV. McMurray, A National Survey of the Prevalence, Incidence, Primary Care Burden and Treatment of Atrial Fibrillation in Scotland. Heart, vol. 93, pp. 606–612, 2007.

[10] G.V. Naccarelli, H. Varker, J. Lin, K.L. Schulman, Increasing Prevalence of Atrial Fibrillation and Flutter in the United States. The American Journal of Cardiology, vol. 104, pp. 1534-1539, 2009.

[11] P.-H. Chanet al. Diagnostic Performance of a Smartphone-Based Photoplethysmographic Application for Atrial Fibrillation Screening in a Primary Care Setting. J. Am. Heart Assoc. vol. 5, 2016.

[12] J. Lee, Y. Nam, DD. McManus, KH. Chon, Time-Varying Coherence Function for Atrial Fibrillation Detection. IEEE Trans. Bio. Eng., vol. 60, pp. 2783–2793, 2013.

[13] A. Carpenter, A. Frontera, Smart-watches: A Potential Challenger to the Implantable Loop Recorder? Europace, vol. 18, 791 LP–793, 2016.

[14] DD. McManus, J. Lee, O. Maitas, N. Esa, R. Pidikiti, A. Carlucci, J. Harrington, E. Mick, K.H. Chon. A Novel Application for the Detection of an Irregular Pulse using an iPhone 4S in Patients with Atrial Fibrillation. Heart Rhythm, vol. 10, pp. 315-319, 2013.

[15] O. Lahdenoja, T. Hurnanen, Z. Iftikhar, S. Nieminen, T. Knuutila, A. Saraste, T. Kiviniemi, T. Vasankari, J. Airaksinen, M. Pänkäälä, T. Koivisto, Atrial Fibrillation Detection via Accelerometer and Gyroscope of a Smartphone. IEEE Journal of Biomedical & Health Informatics, vol. 22, pp. 108–118, 2018.

[16] O. Andersson, K. H. Chon, L. Sörnmo, and J. N. Rodrigues, A 290 mv Sub-asic for Real-time Atrial Fibrillation Detection. IEEE transactions on biomedical circuits and systems, vol. 9, pp. 377–386, 2015.

[17] O. Faust, A. Shenfield, M. Kareem, TR. San, H Fujita, UR. Acharya, Automated Detection of Atrial Fibrillation using Long Short-Term Memory Network with RR Interval Signals. Computers in Biology and Medicine, vol. 102, pp. 327–335. 2018.

[18] Z. Haddi, J-F Pons, S. Delliaux, B. Ananou, J-C Deharo, A. Charaï, R. Bouchakour, M. Ouladsine, A Robust Detection Method of Short Atrial Fibrillation Episodes. Computing in Cardiology, vol. 44, pp. 1–4, Sept. 2017.

[19] Z. Haddi, B. Ananou, Y. Trardi, J-F. Pons, S. Delliaux, M. Ouladsine, J-C. Deharo, Relevance Vector Machine as Data-Driven Method for Medical Decision Making. IEEE proceeding of the European Control Conference (ECC), pp. 1011–1016, 2019.

[20] Z. Haddi, B. Ananou, Y. Trardi, J-F. Pons, S. Delliaux, J-C. Deharo, M. Ouladsine, Advanced Machine Learning Coupled with Heart-Inter-beat derivatives for Cardiac Arrhythmia Detection. IEEE proceeding of the American Control Conference (ACC), pp. 5433–5438, 2020.

[21] J-F. Pons, Z. Haddi, J-C. Deharo, A. Charaï, R. Bouchakour M. Ouladsine, S. Delliaux, Heart Rhythm Characterization Through Induced Physiological Variables. Scientific Reports, vol. 7, no. 5059, 2017.

[22] Z. Haddi, A. Amari, F. E. Annanouch, A, Ould Ali, N. El Bari, B. Bouchikhi, Potential of a Portable Electronic Nose for Control Quality of Moroccan Traditional Fresh Cheeses. Sensor Letters, vol. 9, pp. 2229–2231, 2011.

[23] Z. Haddi, et al., Application of a Portable Electronic Nose Device to Discriminate and Identify Cheeses with Known Percentages of Cow's and Goat's Milk. IEEE Sensors Conference, pp. 771–774, 2010.

[24] A. Mjahad et al., Detection of Ventricular Fibrillation Using the Image from Time-Frequency Representation and Combined Classifiers without Feature Extraction, Appl. Sci., vol. 8, 2057, 2018.

Current Address :

NVISION Systems & Technologies, S.L.,
Avenida Barcelona, 105, Edificio Ig-No
08700, Igualada
Barcelona, Spain
zouhair.haddi@nvision.es