engineering proceedings

MDPI

Proceeding Paper

Statistical Classification for the Screening of Cardiological Patients by Means of a Sensorized T-Shirt †

Martina Ferrazza [1,4], Noemi D'Abbondanza [1,4], Emanuele Piuzzi [2] and Antonio Pallotti [3,4,*]

[1] Sapienza University of Rome, Rome, Italy; ferrazza.martina@gmail.com (M.F.); email@gmail (N.D.)
[2] Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Rome, Italy; emanuele.piuzzi@uniroma1.it
[3] Department of Management and Law, University of Rome Tor Vergata, Rome, Italy
[4] Technoscience
* Correspondence: antonio.pallotti@uniroma2.it
† Presented at the 2nd International Electronic Conference on Applied Sciences, 15–31 October 2021; Available online: https://asec2021.sciforum.net/.

Abstract: This paper fits into the contest of telemonitoring systems and wearable devices in the field of cardiology and aims to present an algorithm for the automatic detection of atrial fibrillation (AF), the most widespread arrhythmia, that can lead to stroke and heart failure if not detected in time. The developed algorithm performs features extraction from bio signals collected with a sensorized T-shirt equipped with a single-lead ECG, a temperature sensor, a pulse oximeter, and an accelerometer for breathing rate. Two PhysioNet databases are also employed to test ECG signals analysis. Kolmogorov-Smirnov test is taken to establish the significance (p = 0.05) of the ECG features extracted and different classification models are evaluated and compared for the automatic detection of AF.

Keywords: Atrial Fibrillation; Heart Rate Variability; wearable sensors; telemonitoring; statistical classification; ECG signal; pulse oximeter signal; body temperature; respiratory signal

Citation: Ferrazza, M.; D'Abbondanza, N.; Piuzzi, E.; Pallotti, A. Statistical Classification for the Screening of Cardiological Patients by Means of a Sensorized T-Shirt. *Engineering Proceedings* **2021**, *3*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: date

1. Introduction

In the last two decades life expectancy increased markedly and the chronicization of diseases enhanced along with it. It has been estimated that noncommunicable disease are the leading cause of death in Europe and America, especially cardiovascular ones [1]. In this contest, abnormalities of cardiac rhythm are prevalent in community-dwelling adults (affecting >2% of individuals) [2] and atrial fibrillation (AF) is considered the most widespread arrhythmia. With an estimated growing prevalence of 0.4% to 1% in the general population, AF accounts for about 1/3 of hospitalization for cardiac rhythm disorders causing an extremely costly public health problem [3]. AF is associated with high cardiovascular morbidity and mortality and suffering from this disease may lead to stroke and to heart failure [3]. Because of its impact on health, an early detection and a recurring monitoring are determining factors in the disease management and in the prevention of disease exacerbation. In clinical practice, the diagnosis of AF requires at least a single-lead ECG recording during the arrhythmia and it may be facilitated by a Holter monitoring [3]. Unfortunately, because AF may be paroxysmal and/or asymptomatic, its diagnosis can be challenging, and it may not be detected even on a standard 12-lead ECG or on a Holter monitoring system. Since a portable single-lead ECG recording device may help establish the diagnosis in cases of paroxysmal AF, the answer to this challenge may be represented by the more and more widespread wearable devices, that provide a low-cost and user-friendly way for patients' health continuous monitoring [3,4]. On this back-

ground, this paper aims to present an algorithm for the screening and monitoring of patients suffering from heart disease by implementing the analysis of different bio signals obtained with a sensorized T-shirt. With the development and the circulation of wearable and portable devices and telemonitoring systems, mobile and personalized medicine has reached a new level of improvement and the number of algorithms for automatic detection of heart disease increased along with it. On the ECG, AF is characterized by the absence of P wave and an irregular ventricular response, when atrioventricular conduction is intact [3,5]. Since the morphological analysis of P-waves is complex, especially in wearable devices field where the collected data are often exposed to different kind of artifacts, and since ventricular activity is influenced by atrial conductivity in AF, the AF automatic detection algorithms are mainly based on Heart Rate Variability (HRV) [6] analysis to provide a higher robustness of the results.

A large amount of works focuses on the analysis of ECG records in order to extract statistical parameters or features for classification, mainly based on HRV analysis in different domains: [5] and [7] review features and classification models used for AF automatic detection and in [8] three different algorithms to extract time and frequency domain features for AF detection are presented and compared; in [9] the density histograms of RR and dRR intervals are used for coefficients of variation (CV) test and in the Kolmogorov-Smirnov test in order to perform AF detection; in [10] frequency domain features (PSD features estimated with Welch's method) are extracted in order to compare two classification models, one based on an evolutionary neural system approach and the other based on the combination of the convolutional neural network (CNN) with the bidirectional long short-term memory (BiLSTM); Refs [11–13] based their detection system on the analysis of Poincare plot and in particular on geometric features and on features linked to points distribution in the plot; in [14] two probability density functions are employed to model the histograms of RR differences and Neyman-Pearson (NP) detection approach is used to obtain criteria for AF detection; in [15] the RR intervals sequence PSD is estimated in order to compute the LF/HF correlation as a feature for a linear discriminant classifier; the work [16] estimates PSD of ECG signals (using Welch's method and a discrete Fourier transform) in order to perform machine learning algorithms; in [17] HRV time and frequency domain features are extracted and data are compared by Mann Whitney test and by Kolmogorov-Smirnov test; in [18] a simple AF detection algorithm is implemented based on the number of RR interval combinations differing more than x seconds in a sliding window of length N; in [19] the number of non-empty cells in RdR (RR-dRR) plot is computed and compared with a threshold; [20] deals with the analysis of HRV in time, frequency and statistical domain and gives each parameter a different weight based on its accuracy of detecting arrhythmia; in [21] AF automatic detection is performed through HRV-based features (non-linear and in time domain) and frequency analysis. All the algorithms developed in these works are tested on PhysioNet databases [22].

Our proposed system combines elements of these works, providing an algorithm for AF automatic detection based on HRV features in time and frequency, but instead of limiting the analysis to ECG records, it implements the analysis of different bio signals recorded through a sensorized T-shirt equipped with a single-lead ECG, a temperature sensor, a pulse oximeter and a 3-axis accelerometer for the respiratory rate [23]. In fact, since changes in atrial electrophysiology may be associated with alterations in respiratory physiology [24], it may be important for the management of the disease to estimate respiratory parameters such as breathing rate (BR). In this way, through ECG analysis and with the support of the other recorded bio signals to monitor patient's physiological parameters, it is possible to perform an automatic detection of arrhythmias (specifically AF) to help in the diagnosis and management of the disease.

2. Materials and Methods

This work is based on the analysis of four different bio signals recorded with a sensorized T-shirt equipped with a single-lead ECG (AD8232 SparkFun Single Lead Heart

Rate Monitor), a temperature sensor (LM35 by Texas Instrument), a pulse-oximeter (SparkFun Pulse Oximeter and Heart Rate Sensor) and a 3-axis accelerometer (ADXL335 by Analog Devices). We collected data of 20 healthy subjects, that act as a control group, and for each one we recorded two minutes of a resting trial during which the subject remained seated. Of these 20 subjects, 11 belong to an early measurement phase during which chest movement was not recorded, so breathing rate is not available, while all the signals are collected for the rest of the subjects. The T-shirt recorded the following signals: single-lead ECG sampled at 250 Hz; the trend of temperature over the time with a sampling frequency of 1 Hz; the value of blood oxygen saturation (SpO2) sampled at 1 Hz; chest acceleration along the normal axis over the time sampled at 5 Hz. The signals are analyzed in Matlab to extract physiological parameters and features for the monitoring of the subjects and the detection of AF.

The performance of the algorithm for the analysis of the ECG signals was also evaluated on two PhysioNet databases [22], in particular:

- MIT-BIH Atrial Fibrillation Database (AFDB): it includes 25 long-term ECG recordings (10 h) of human subjects with AF and each recording contains two ECG signals sampled at 250 Hz.
- MIT-BIH Normal Sinus Rhythm Database (NSDB): it includes 18 long-term (24 h) ECG recordings of subject with no particular arrhythmias and each recording is composed of two ECG traces sampled at 128 Hz.

2.1. Data Organization and Preprocessing

Signals collected for each subject with the T-shirt are organized in distinct vectors. Since data recorded with wearable devices are often characterized by different kind of artifacts, to improve the signal to noise ratio the ECG signal is bandpass filtered between 0.05 and 30 Hz with an eight order Butterworth filter, and the accelerometer's output is high-pass filtered ($f_{highpass} = 0.01$ Hz by spectral evidence), to remove the baseline, and low-pass filtered with a moving average, to remove high frequency noise. Because of the lower level of noise and a less complex post-processing, temperature sensor and pulse oximeter's outputs are not preprocessed, and the signals are kept raw for the further analysis.

Data obtained from PhysioNet databases are organized into two different structures to allow the analysis of both long-term and short-term ECG recordings. For the analysis of long-term recordings, the length of AF recordings remains unchanged (10 h), while normal sinus rhythm (NS) ECGs are shortened from 24 h to 10 h, to allow the comparison between the two class. To simulate short-term recordings, the length of the signals from both databases is reduced to two minutes, to match T-shirt recordings duration. For both databases only the first ECG trace of each recording is considered for the analysis.

2.2. Analysis and Features Extraction of ECG data

Both ECG signals from databases and collected with the T-shirt are analyzed with the same algorithm. The first step in the analysis of ECG signals is the detection of R-peaks to be able to extract HRV-based features. Pan and Tompkins algorithm is adopted to perform R-peaks detection, because of its high accuracy (99.3%) and its simplicity in the implementation [25]. Since HRV is conventionally described as the variation of both instantaneous HR and RR intervals, the time intervals between consecutive heartbeats (RRIs) are computed, to extract RRI sequence. Once HRV is computed, time and frequency domain features for AF detection are extracted.

One of the features is based on the Poincare plot, where, given consecutive RRI pairs, RRi are plotted against RRi + 1 (i = 1,…, N−1; N = number of RR intervals). Different studies [11,12,21] agree on the evidence that Poincare plots constructed with non-AF data show some patterns, while plots obtained from AF data show irregular shapes. In this work, we use data points to fit an ellipse and to extract the ratio between the long (SD1) and the short (SD2) axis as a descriptor about Poincare plot, since SD1 and SD2 are believed to

reflect the long-term and short-term dispersion of HRV and so the dynamics of heart activity.

The following other HRV- based features are extracted in time domain: the mean and the standard deviation of RR intervals and of instantaneous HR (HR = 60/RRI); the standard deviation (SD) and the root mean square (RMS) of successive difference (SD), where the successive differences are the temporal differences between two consecutive RR intervals (SD = RRIi + 1 − RRIi); the percentage of RR intervals above 50 milliseconds (PRR50).

To compute HRV-based features in frequency domain, HRV power spectrum is computed with FFT algorithm. Since the RRI sequence taken as a digital signal is irregularly sampled in time domain, the HRV signal is resampled to obtain a signal equally spaced in time, with a sampling frequency of 4 Hz. The following features are extracted: the low frequency (LF) power, calculated as the PSD (area under the PS curve) at low frequencies (0.04–0.15 Hz); the high frequency (HF) power, calculated as the PSD in the frequency band 0.1–0.4 Hz; the ratio of LF and HF power; the total power, considered as the sum of LF and HF power; the percentage of power affecting the LF band and the percentage in HF band.

Features extracted from ECG signals are employed in the development of machine learning models to classify healthy subjects and AF patients. To test the significance of the features extracted through ECG signals analysis, the two sample Kolmogorov-Smirnov (KS) test is taken, based on studies [9,17].

2.3. Analysis of Bio Signals Collected With the T-Shirt

As well as ECG signal, also the other bio signals collected by means of the T-shirt are analyzed to extract additional features for the characterization and monitoring of patients with AF. Since subjects were asked to perform a resting trial which does not affect the value of body temperature and oxygen in time, temperature sensor and pulse oximeter outputs are processed to extract the mean value of the temperature (°C) and of the blood oxygen saturation (%) during the trial. To extract the respiratory rate, the output signal of the accelerometer sensor is processed to compute the power spectrum through the Fast Fourier Transform algorithm. The frequency value at the maximum peak of the spectrum corresponds to the respiratory rate (Hz).

3. Results

Mean and standard deviation values of the control group are computed for pulse oximeter signals, body temperature and breathing rate (Table 1).

For ECG Signals, three different pairs of samples are tested in multiple classification models. The first pair of sample is composed of AF patients from PhysioNet AFDB (short-term records) and healthy subjects from PhysioNet NSDB (short-term records); the second pair is composed of long-term records from AFDB for the sample of AF patients and long-term records from NSDB for the sample of healthy subjects; the last couple of samples has the same unhealthy subjects of the first case, but the control group is composed of short-term records from NSDB added to records collected with the T-shirt.

For each pair of samples, classification models are trained and validated using both all the features computed, and the features selected with KS test, whose results are shortly shown in Table 2 as the number of significant features (with a p-value = 0.05) on the number of total features.

Table 1. Mean and std dev. of SpO2, T and BR of the control group.

Features	*Number of Subjects*	*Mean*	*Std*
$SpO2_{mean}$ *[%]*	*20*	*96.939*	*0.8777*
T_{mean} *[°C]*	*20*	*36.273*	*0.632*
BR [Hz]	*9*	*0.213*	*0.081*

Table 2. Kolmogorov-Smirnov test results.

Samples	*Test Results**
short-term database records	*11/14*
Long-term database records	*10/14*
Samples with mixed control group	*11/14*

* Reported as the number of significant features (with a *p*-value = 0.05) on the number of total features.

Table 3 shows the results of the classification for the three pairs of samples using a 5-fold Cross-Validation. For each test are reported the three classificators with the best performances in terms of validation accuracy, sensitivity, specificity and area under the curve (AUC).

Table 3. Results of classification for the three different samples: the samples with the mixed control groups, the samples composed of short-term records from databases and the samples formed of long-term records of the databases.

Samples with Mixed Control Group						
	All-Features			***Features Selected with KS Test***		
Classification model	Bagged trees	Subspace Discriminant	Kernel Naïve Bayes	Bagged Trees	Gaussian Naïve Bayes	Fine Tree
Validation Accuracy (%)	75	75	70	75	75	71.7
Sensitivity (%)	54.5	50	68.2	63.6	50	78.9
Specificity (%)	86.8	89.5	71	84.2	89.5	59
AUC	0.84	0.79	0.77	0.86	0.83	0.74
Short-term database records						
	All-features			***features selected with KS test***		
Classification model	Kernel Naïve Bayes	Bagged Trees	Fine Tree	Bagged trees	Gaussian Naïve Bayes	Kernel Naïve Bayes
Validation Accuracy (%)	90	90	87.5	95	92.5	90
Sensitivity (%)	95.5	90	90	100	95.5	95.5
Specificity (%)	83.3	88.9	83.3	88.9	88.9	83.3
AUC	0.96	0.92	0.86	0.94	0.96	0.94
Long-term database records						
	All-features			***features selected with KS test***		
Classification model	Fine Tree	Kernel Naïve Bayes	Bagged Trees	Kernel Naïve Bayes	Cubic SVM	Bagged Trees
Validation Accuracy (%)	95	95	92.5	95	95	92.5
Sensitivity (%)	100	100	95.5	100	100	95.5
Specificity (%)	88.9	88.9	88.9	88.9	88.9	88.9
AUC	0.94	0.94	0.95	0.94	0.95	0.96

4. Discussion and Conclusions

Results show that this algorithm achieves competitive performances on databases records, both on long and short term ECGs, reaching high level of accuracy, sensitivity and specificity.

In the present work, the algorithm for ECG analysis aims to extract and select features for ECG signals statistical classification. On the contrary, the classification algorithm of the dataset with the mixed control group reports less performing results, reaching a maximum accuracy of 75%, probably because of data heterogeneity.

In all the three different datasets it is possible to observe how feature selection with KS test either helped improving performances or didn't affect them, but never worsened the results.

This study fits into the contest of machine learning applications in the filed of cardiology [26] proposing an algorithm for the automatic detection of AF that can be employed in the telemonitoring systems through wearable devices, such as the sensorized shirt used in this study to record different bio signals, including ECG. Future developments include the enlargement of the dataset of the shirt recordings, by including more healthy subjects and a sample of patients suffering from AF, in order to develop a classification model, based on the presented algorithm, to be used in telemonitoring of cardiopathic patients.

Institutional Revie w Board Statement:

Informed Consent Statement:

Data Availability Statement:

References

1. World Health Organization. *World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals 2021*; World Health Organization: 2021; Licence: 24 CC BY-NC-SA 3.0 IGO.
2. Khurshid, S.; Choi, S.H.; Weng, L.C.; Wang, E.Y.; Trinquart, L.; Benjamin, E.J.; Ellinor, P.T.; Lubitz, S.A. Frequency of Cardiac Rhythm Abnormalities in a Half Million Adults. *Circ. Arrhythmia Electrophysiol.* **2018**, *11*, 1–9, doi:10.1161/CIRCEP.118.006273.
3. Fuster, V.; et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: 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. *Circulation* **2006**, *114*, e257–e354, doi:10.1161/CIRCULATIONAHA.106.177292.
4. Kaasenbrood, F.; Hollander, M.; Rutten, F.H.; Gerhards, L.J.; Hoes, A.W.; Tieleman, R.G. Yield of screening for atrial fibrillation in primary care with a hand-held, single-lead elec-trocardiogram device during influenza vaccination. *Europace* **2016**, *18*, 1514–1520, doi:10.1093/europace/euv426.
5. Lim, H.W.; Hau, Y.W.; Lim, C.W.; Othman, M.A. Artificial intelligence classification methods of atrial fibrillation with implementation technology. *Comput. Assist. Surg.* **2016**, *21*, 155–162, doi:10.1080/24699322.2016.1240303.
6. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart Rate Variability: Standards of Measurements, Physiological Interpretation, and Clinical Use. *Circulation* **1996**, *93*, 1043–1065, doi:10.1161/01.CIR.93.5.1043.
7. Sahoo, S.K.; Lu, W.; Teddy, S.D.; Kim, D.; Feng, M. Detection of Atrial fibrillation from non-episodic ECG data: A review of methods. In Proceedings of the Annual International Conference of the IEEE EMBS 2011, Boston, MA, USA, 30 August–3 September 2011; pp. 4992–4995, doi:10.1109/IEMBS.2011.6091237.
8. Kikillus, N.; Hammer, G.; Lentz, N.; Stockwald, F.; Bolz, A. Three different algorithms for identifying patients suffering from atrial fibrillation during atrial fibrillation free phases of the ECG. *Comput. Cardiol.* **2007**, *34*, 801–804, doi:10.1109/CIC.2007.4745607.
9. Tateno, k.; Glass, L. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and deltaRR intervals. *Med Biol. Eng. Comput.* **2001**, *39*, 664–671, doi:10.1007/BF02345439.
10. Liang, Y.; Yin, S.; Tang, Q.; Zheng, Z.; Elgendi, M.; Chen, Z. Deep Learning Algorithm Classifies Heartbeat Events Based on Electrocardiogram Signals. *Front. Physiol.* **2020**, *11*, 569050, doi:10.3389/fphys.2020.569050.
11. Park, J.; Lee, S.; Jeon, M. Atrial fibrillation detection by heart rate variability in Poincare plot. *BioMedical Eng. Online* **2009**, *8*, 1–12, doi:10.1186/1475-925X-8-38.
12. Kikillus, N.; Hammer, G.; Wieland, S.; Bolz, A. Algorithm for identifying patients with paroxysmal atrial fibrillation without appearance on the ECG. In Proceedings of the Annual International Conference of the IEEE EMBS—Proceedings 2007, Lyon, France, 23–26 August 2007; pp. 275–278, doi:10.1109/IEMBS.2007.4352277.
13. Luo, C.; Li, Q.; Rao, H.; Huang, X.; Jiang, H.; Rao, N. An improved Poincaré plot-based method to detect atrial fibrillation from short single-lead ECG. *Biomed. Signal. Process. Control.* **2021**, *64*, 102264, doi:10.1016/j.bspc.2020.102264.
14. Ghodrati, A.; Marinello, S. Statistical analysis of RR interval irregularities for detection of atrial fibrillation. *Comput. Cardiol.* **2008**, *35*, 1057–1060, doi:10.1109/CIC.2008.4749227.
15. Hickey, B.; Heneghan, C.; De Chazal, P. Non-episode-dependent assessment of paroxysmal Atrial Fibrillation through measurement of RR interval dynamics and atrial premature contractions. *Ann. Biomed. Eng.* **2004**, *32*, 677–687, doi:10.1023/B:ABME.0000030233.39769.a4.
16. Pławiak, P. Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Syst. Appl.* **2018**, *92*, 334–349, doi:10.1016/j.eswa.2017.09.022.

17. Kim, D.; Seo, Y.; Jung, W.R.; Youn, C.H. Detection of long-term variations of heart rate variability in normal sinus rhythm and atrial fibrillation ECG data. BioMedical Engineer-ing and Informatics: New Development and the Future. In Proceedings of the 1st International Conference BMEI, Sanya, China, 28–30 May 2008; Volume2, pp. 404–408, doi:10.1109/BMEI.2008.273.
18. Petrenas, A.; Marozas, V.; Sörnmo, L. Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Comput. Biol. Med.* **2015**, *65*, 184–191, doi:10.1016/j.compbiomed.2015.01.019.
19. Lian, J.; Wang, L.; Muessig, D. A simple method to detect atrial fibrillation using RR intervals. *Am. J. Cardiol.* **2011**, *107*, 1494–1497, doi:10.1016/j.amjcard.2011.01.028.
20. Sadiq, I.; Khan, S.A. Fuzzification of the analysis of heart rate variability using ECG in time, frequency and statistical domains. In Proceedings of the 2010 2nd ICCEA, Bali Island, Indonesia, 26–29 March 2010; Volume 1, pp. 481–485, doi:10.1109/ICCEA.2010.99.
21. Mei, Z.; Gu, X.; Chen, H.; Chen, W. Automatic atrial fibrillation detection based on heart rate variability and spectral features. *IEEE Access* **2018**, *6*, 53566–53575, doi:10.1109/AC-CESS.2018.2871220.
22. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physio-logic Signals. *Circulation* **2000**, *101*, e215–e220, doi:10.1161/01.CIR.101.23.e215.
23. D'Abbondanza, N.; Ferrazza, M.; Piuzzi, E.; Pallotti, A. Sensorized T-shirt for cardiological patients in telemonitoring. In Proceedings of the Second International Electronic Conference on Applied Science 2021, online, 15–31 October 2021; under review.
24. Shah, V.; Desai, T.; Agrawal, A. The Association between Chronic Obstructive Pulmonary Disease (COPD) and Atrial Fibrillation: A Review. *Chron Obs. Pulmon Dis.* **2016**, doi:10.21767/2572-5548.100002.
25. Pan, J.; Tompkins, W.J. A Real-Time QRS Detection Algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *BME-32*, 230–236, doi:10.1109/TBME.1985.325532.
26. D'Angelantonio, E.; Lucangeli, L.; Camomilla, V.; Mari, F.; Mascia, G.; Pallotti, A. Classification-Based Screening of Phlebopathic Patients using Smart 66 Socks. In Proceedings of the 2021 IEEE MeMeA, Lausanne, Switzerland, 23–25 June 2021; doi:10.1109/MeMeA52024.2021.9478688.