Characterization and classification of patients with different levels of cardiac death risk by using Poincaré plot analysis

Javier Rodriguez, Andreas Voss, Pere Caminal, Antoni Bayés-Genis and Beatriz F. Giraldo *Senior Member, IEEE*

Abstract*— Cardiac death risk is still a big problem by an important part of the population, especially in elderly patients. In this study, we propose to characterize and analyze the cardiovascular and cardiorespiratory systems using the Poincaré plot. A total of 46 cardiomyopathy patients and 36 healthy subjets were analyzed. Left ventricular ejection fraction (LVEF) was used to stratify patients with low risk (LR: LVEF $>$ 35%, 16 patients), and high risk (HR: LVEF $\leq$ 35%, 30 patients) of heart attack. RR, SBP and T_{Tot} time series were extracted from the ECG, blood pressure and respiratory flow signals, respectively. Parameters that describe the scatterplott of Poincaré method, related to short- and long-term variabilities, acceleration and deceleration of the dynamic system, and the complex correlation index were extracted. The linear discriminant analysis (LDA) and the support vector machines (SVM) classification methods were used to analyze the results of the extracted parameters. The results showed that cardiac parameters were the best to discriminate between HR and LR groups, especially the complex correlation index ($p = 0.009$). Analising the interaction, the best result was obtained with the relation between the difference of the standard deviation of the cardiac and respiratory system ($p = 0.003$). When comparing HR vs LR groups, the best classification was obtained applying SVM method, using an ANOVA kernel, with an accuracy of 98.12%. An accuracy of 97.01% was obtained by comparing patients versus healthy, with a SVM classifier and Laplacian kernel. The morphology of Poincaré plot introduces parameters that allow the characterization of the cardiorespiratory system dynamics.

I. Introduction

Elderly patients are predisposed to developing heart failure as a result of age-related changes in the cardiovascular system. Diseases related to cardiovascular system are one of the most common cause of death, especially in elderly people [1], [2]. The problem of identifying cardiomyopathy patients who are at risk of sudden cardiac death (SCD) is still unsolved. Left ventricular ejection fraction (LVEF) is the most common indicator of heart failure. Patients with values above 55% are considered normal, values between 35% and 55% are considered below normal, and patients with values below 35% are classified pathological [3], [4].

Some of the conventional signal-processing techniques such as time-domain, power spectral density and cross-correlation analysis are often not enough to characterize their complex dynamics. The Poincaré plot is a graphic method used to evaluate the self-similarity in a given system. It also allows to identify if its behavior is chaotic or random [5], [6]. This technique is a useful visual tool to summarize time series derived from signals as ECG, respiratory flow or blood pressure, in one picture, and a quantitative technique which gives information on the long- and short-term variabilities of these signals. Poincaré plot is a widely used method for analysing ECG signals in different pathologies associated with the cardiac system [7]–[9].

In our previous works, we have proposed to analysis cardiorespiratory and cardiovascular interactions in patients with cardiomyopathy, through time- and frequency domain parameters, or using joint symbolic dynamics analysis [4], [10]. In this work we propose new parameters to classify patients with ischemic and dilated cardiomyopathies, according to low and high cardiac risk in function of the left ventricular ejection fraction (LVEF). We analyze the interactions between time series extracted from the ECG, the respiratory flow and the blood pressure signals using Poincaré plot method. We have obtained the best characeristics that describe the dynamics of the cardiorespiratory and cardiovascular systems, and classified the patients into two different risk levels, applying linear discriminant analysis and support vector machine methods.

II. Database

The noninvasive continuous electrocardiographic (ECG), blood pressure (BP) and respiratory flow (RF) signals were recorded for 30 minutes from 46 cardiomyopathy patients, at the Santa Creu i Sant Pau Hospital in Barcelona, Spain. All patients were studied according to a protocol, previously approved by the Hospital ethics committee. Information about cardiomyopathy patients database is presented in [4], [10]. The control group, 36 healthy subjects, were recorded at the Institute for Bioengineering of Catalonia, Barcelona,

This work was supported in part by the Secretariat of Universities and Research of the Department of Economy and Knowledge of the Government of Catalonia (Consolidated research group GRC 2014 SGR 1569), by CERCA Programme, Generalitat de Catalunya, and by the Ministry of Economy and Competitiveness of the Spanish Government and the European Regional Development Fund, ref. DPI2015-68820-R (MINECO/FEDER).

J. Rodriguez is with Institut de Bioenginyeria de Catalunya (IBEC), Universitat Politècnica de Catalunya (UPC). C/ Baldiri Reixac, 4, Torre I, 9th floor, 08028, Barcelona, Spain.

A. Voss is with Dept. of Medical Engineering and Biotechnology, University of Applied Sciences Jena, Germany.

P. Caminal is with with Dept. of ESAII, UPC, Centre for Biomedical Engineering Research (CREB) and CIBER de Bioenginiería, Biomateriales y Nanomedicina (CIBER-BBN), Spain.

A. Bayés Genis is with Cardiology Service, Hospital Universitari Germans Trias i Pujol, Badalona, Spain.

*B.F. Giraldo is with IBEC, Dept. ESAII, Barcelona East School of Engineering (EEBE), UPC, CIBER-BBN, Spain. (e-mail: beatriz.giraldo@upc.edu)

Spain (according to the Institutional Review Board), using BIOPAC Systems Inc. MP150 for the ECG and RF signals, and Finometer-Finapres Medical Systems for BP signal, for 15 minutes, at a sampling frequency of 500 Hz, in supine position.

According to clinical diagnostic, the patients were classified in function of their LVEF in High (HR: LVEF $\leq$ 35%, 30 patients) and Low (LR: LVEF $>$ 35%, 16 patients) risk. Table I shows the clinical data information. Parameters related to left ventricular diastolic dimension, auricular diameter and left ventricular ejection fraction were obtained using an echocardiogram.

TABLE I: Clinical Parameters

	LR	HR
Patients	16	30
Age [years]	62.25 $\pm$ 13.28	66.56 $\pm$ 9.20
Weight [kg]	75.31 $\pm$ 15.10	80.83 $\pm$ 15.98
BMI [kg/m^2]	27.25 $\pm$ 3.92	28.46 $\pm$ 4.89
NYHA	2.25 $\pm$ 0.57	2.03 $\pm$ 0.31
LVDD [mm]	56.43 $\pm$ 6.67	63.5 $\pm$ 6.03
AD [mm]	44.43 $\pm$ 5.04	47.43 $\pm$ 6.58
ProBNP	1137.56 $\pm$ 906.09	2298.3 $\pm$ 3649.45
LVEF [%]	43.81 $\pm$ 8.79	29.23 $\pm$ 5.42

BMI =Body Mass Index; NYHA = New York Heart Association functional classification; LVDD = Left Ventricular Diastolic Dimension; AD = Auricular Diameter; ProBNP = Brain Natriuretic Peptide; LVEF = Left Ventricular Ejection Fraction.

III. Methodology

A. Signal Processing

All signals were analyzed, linear trend was removed and pre-processing tools were used to reduce artifacts and spikes, and eliminate outliers. From ECG signal, the cardiac interbeat interval $\{(RR(t_c)\}$ time series was extracted automatically using a custom algorithm, based on wavelet analysis. Systolic blood pressure $\{(SBP(t_b)\}$ time series was obtained from the BP signal as the difference between two consecutive absolute maximums. The time series of the breath duration $\{(T_{Tot}(t_r)\}$ was extracted using an algorithm based on the zero-crossing of the RF signal. Thereafter, all time series were visually inspected and corrected, if necessary. Figure 1 shows an excerpt of ECG, BP and RF signals with their corresponding marks used to obtain their time series. Finally, new signals $\{(RR(n_c)\}$, $\{(SBP(n_b)\}$ and $\{(T_{Tot}(n_r)\}$ were obtained by 1 Hz resampling of the interpolated time series, and normalized (mean 0 and standard deviation 1).

B. Poincaré plot

Poincaré plot is a scatterplot of the current interval plotted against the preceding interval. It is also a quantitative technique that describe its dispersion through parameters based on the morphology of their plot. The short- and

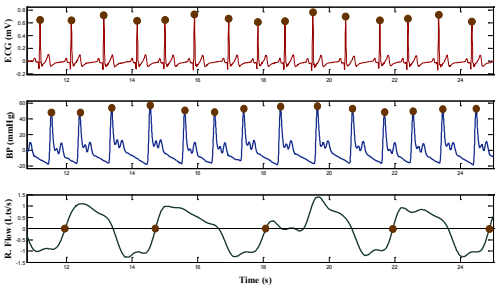

Fig. 1: Excerpt of (a) ECG signal with RR time series marks, (b) BP signal with SBP time series marks, and (c) RF signal and its T_{Tot} time series marks.

long-term variabilities are quantified by (SD_1) and (SD_2), respectively. These are calculated fitting an ellipse to the shape of the Poincaré plot and measuring the dispersion along the major and minor axis of the ellipse, around the identity line (Id). The values of the (SD_1) and (SD_2) are based on the standard deviation of the ellipse around the minor and major axis, respectively, given by equations 1 and 2. Figure 2 presents an example of the Poincaré plot for a patient of the LR, HR and control group.

$$SD_1 = \frac{\sqrt{2}}{2} SD(x_n - x_{n+1}) \tag{1}$$

$$SD_2 = \sqrt{2SD(x_n)^2 - \frac{1}{2}SD(x_n - x_{n+1})^2} \tag{2}$$

Other two parameters are obtained with the ratio between SD_1 and SD_2, and the difference between them (ΔSD). The information above the Id line represents a reduction in the variability of the system, while below it represents an increase. These variabilities are calculated measuring the distance of each point to the Id line, above (D_{UP}) and below (D_D). The parameters SD_{UP} and SD_D represent the variability of the system around the Id line (above and below, respectively), according to the equations 3 and 4,

$$SD_{UP} = \frac{1}{nUP} \sum_{i=1}^{nUP} (D_{UP_i})^2 \tag{3}$$

$$SD_D = \frac{1}{nD} \sum_{i=1}^{nD} (D_{D_i})^2 \tag{4}$$

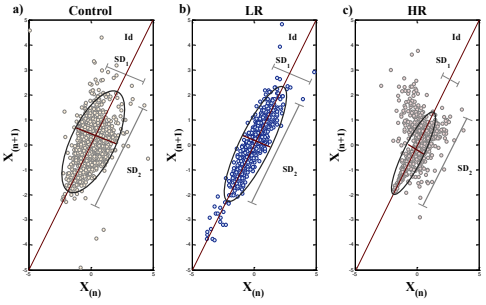

Fig. 2: Poincaré plot example for (a) a subject of control group, (b) a patient of LR group, and (c) a patient of HR group.

being nUP and nD the number of points in each case.

Lastly, the acceleration and deceleration of the system is represented by C_{UP} and C_D parameters, defined as

$$C_{UP} = \frac{SD_{UP}^2}{SD_1^2} \tag{5}$$

$$C_D = \frac{SD_D^2}{SD_1^2}. \tag{6}$$

If the Id line is changed by a linear regression line (Lr), the parameters LrC_{UP} and LrC_D are obtained. The last parameter calculated is the complex correlation index (CCI) that represents the time domain changes of the signals [11]. This parameter is based on the measurement of the area A_i defined by each three consecutive points, and a normalization coefficient C_n, calculated as $C_n = \pi \cdot SD_1 \cdot SD_2$ [7]. CCI is shown in equation 7

$$CCI = \frac{1}{C_n(N-2)} \sum_{i=1}^{N-2} ||A_i||. \tag{7}$$

Table II summarizes all types of parameters analyzed. These parameters were analyzed for each ECG, BP and RF signals, corresponding to indices c, b and r, respectively. The same parameters were analyzed considering the interactions between each two signals. Finally, another group of parameters were analyzed considering the interaction between the differences into each signal (Δc, Δb and Δr). A total of 81 parameters were obtained.

TABLE II: Parameters summary

Parameter	Description
xSD_1	Short time standard deviation
xSD_2	Long time standard deviation
xSD_1/SD_2	Short and long deviations ratio
$x\Delta SD$	Standard deviation difference
xC_{UP}	Normalized deceleration
xC_D	Normalized acceleration
$xLrC_{UP}$	Linear regression based normalized deceleration
$xLrC_D$	Linear regression based normalized acceleration
$xCCI$	Complex correlation index

x represent: c for ECG signal, b for BP signal, and r for RF signal.

The Kolmogorov-Smirnov test was used to determine parameters with statistical significant differences between groups (p-value < 0.05). Additionally, parameters with high correlation ($\rho \geq 0.7$) were discarted.

C. Classification Methods

Once the most relevant parameters were obtained, the next classification methods were applied. Different kernel functions were analyzed. The leave-one-out cross-validation procedure was used to validate the results. These were presented in terms of accuracy, sensitivity, and specificity.

- *Linear discriminant analysis*

Linear discriminant analysis (LDA) was used to find a linear combination of predictors that characterize two or more classes. This method is based on the maximizing the distance between variances of classes [12]. Given a set of parameters, the linear discrimination is defined as (equation 8)

$$Y = \mu_0 + \mu_i \cdot x_i \tag{8}$$

where μ_0 and x_i are the independent term and independent parameters, and μ_i is the discriminant function coefficient. The distances between groups is calculated using Mahalanobis method.

- *Support Vector Machines*

Support Vector Machines (SVM) are based on transforming data into a higher dimensional space. It can to transform a complex classification problem into a simpler one that can be solved by linear discriminant function, known as a hyperplane. Given a set of data vectors $X = \{x_1, \ldots, x_L\}$, where $x_i \in \Re^n$, and their corresponding labels $Y = \{y_1, \ldots, y_L\}$, where $y_i \in \{1, -1\}$, SVM function is defined by [10]

$$f(x) = wz + b = \sum_i^L \alpha_i y_i K(x_i y_i) + b \tag{9}$$

where α_i and b define the efficiency of the classifier based on the optimal values, and K represent the Kernel function used. For this study, Gaussian, Laplacian and ANOVA Kernel were proposed.

The Gaussian Kernel is commonly used when the data is distributed radially, and is defined by

$$K(x, y) = e^{-(\frac{\|x-y\|^2}{2\sigma^2})} \tag{10}$$

where σ is the penalty term.

The Laplacian Kernel is similar to the Gaussian kernel but less influenced by the σ parameter, and is given by

$$K(x, y) = e^{-(\frac{\|x-y\|}{2\sigma})}. \tag{11}$$

The ANOVA Kernel is defined by

$$K(x, y) = \sum_{k=1}^{n} e^{(-\sigma(x^k - y^k)^2)^d} \tag{12}$$

where σ and d are the parameters to optimize this kernel.

IV. Results

A total of three groups were analyzed: patients with low and high cardiac death risk in function of the LVEF, and control group. There were not parameters with statistically significant differences when comparing control group with LR and HR groups.

When comparing patients of HR and LR groups, 11 parameters presented statistically significant differences. In each case, after applied the correlation analysis, the parameters most correlated were discarted. Finally, 5 parameters were selected when comparing HR and LR groups (Table III), and 13 when comparing patients and control group (Table IV).

Tables V and VI shows the results of obtained using LDA and SVM classification methods. The best results were

TABLE III: Mean and standard deviation (SD) and p-value of the best parameters that classified HR vs LR patients groups

Parameter	HR	LR	p-value
cSD_1	0.81 ± 0.22	0.66 ± 0.32	0.015
cSD_1/SD_2	0.43 ± 0.13	0.35 ± 0.18	0.015
$cCCI$	27611± 20448	37083± 19032	0.009
$\Delta c\Delta rSD_1/SD_2$	0.38 ± 0.02	0.41 ± 0.04	0.003
$\Delta c\Delta rC_{UP}$	0.74 ± 1.83	0.89 ± 0.98	0.041

TABLE IV: Mean and standard deviation (SD) and p-value of the best parameters that classified patients vs control groups

Parameter	Control	Patients	p-value
cSD_2	150.31±66.52	109.62 ± 61.93	0.010
cSD_1/SD_2	0.35±0.07	0.40 ± 0.15	0.017
$c\Delta SD$	0.02±0.04	0.02 ± 0.06	0.015
$cCCI$	22916± 8415	30906± 20273	0.0017
$cLrC_{UP}$	0.47± 0.04	0.52±0.13	0.0002
$cLrC_D$	0.53± 0.04	0.48± 0.13	0.0002
bSD_1/SD_2	0.37±0.05	0.41±0.09	0.0020
$bLrC_{UP}$	0.52±0.04	0.57±0.11	0.0035
$bLrC_D$	0.48±0.04	0.43 ±0.11	0.0004
rSD_1	0.11±0.03	0.21±0.15	0.0004
$rLrC_{UP}$	0.48±0.04	0.50± 0.04	0.030
$rCCI$	7.95±5.02	4.24±4.22	0.00003
$crCCI$	16243± 5307	20056±7579	0.019

obtained using ANOVA kernel to compare HR vs LR groups, with an accuracy of 98.12%, and using Laplacian kernel an accuracy of 97.01% was obtained.

TABLE V: Accuracy (Acc), sensitivity (Sn) and specificity (Sp) obtained with LDA and SVM classifiers of HR vs LR groups

Method	C	σ	D	Acc	Sn	Sp
SVM - Gaussian	3	0.1	-	76.04	43.44	93.33
SVM -Laplace	3	1	-	97.87	93.89	100
SVM -ANOVA	0.3	100	1	98.12	94.57	100
LDA	-	-	-	71.5	25.41	96.08

TABLE VI: Accuracy (Acc), sensitivity (Sn) and specificity (Sp) obtained with LDA and SVM classifiers of patients vs control group

Method	C	σ	D	Acc	Sn	Sp
SVM - Gaussian	1	0.1	-	95.06	91.3	100
SVM - Laplace	1	0.3	-	97.01	94.72	100
SVM - ANOVA	0.3	0.3	1	92.7	91.3	94.54
LDA	-	-	-	92.25	91.01	93.89

V. Discussion and Conclusion

The objective of this study was to characterize the nonlinear dynamics of the cardiovascular and cardiorespiratory variability system, and to achieve classification based on the extracted information. The statistical analysis of the morphology of the Poincaré plot allowed classification of the patients with different levels of risk and between patients and control subjects.

The best discrimination between the groups of patients HR and LR was given by cardiac parameters associated with the shor-term (cSD_1) variability, the ratio with the long-term variability, and the $cCCI$ index. The main interaction between the systems was given by the relation between cardiac and respiratory differences, in the ratio of the short- and long-term variability ($\Delta c\Delta rSD_1/SD_2$), and their deceleration ($\Delta c\Delta rC_{UP}$). The patients of the HR group presented more significant variability on their cardiac activity. On the other hand, the deceleration of the cardiorespiratory interaction was faster and more stable in LR patients group. The SVM classification method was shown to obtain the best results with accuracies above 98%.

The differences exposed by cSD_1 parameter suggests that there is an increased parasympathetic activity on HR patients as a consequence of an abnormal performance of the vagal system. This malfunction is also reflected in the cSD_1/SD_2 parameter, showing a greater imbalance between the sympathetic / parasympathetic activity in high risk patients.

In conclusion the analysis of the Poincaré plots allow to find parameters useful on the classification of patients with different levels of cardiac death risk. However, these results should be validated with a greater number of patients.

References

[1] R. Mentz, S. Broderick, L. Shaw, M. Fiuzat, and C. O'Connor, "Heart failure with preserved ejection fraction," *Journal of the American College of Cardiology*, vol. 63, 2014.

[2] M. Nichols, N. Townsend, P. Scarborough, and M. Rayner, "Cardiovascular disease in europe 2014: epidemiological update," *European Heart Journal*, pp. 1–10, 2014.

[3] N. El-Sherif, A. Khan, J. Savarese, and G. Turitto, "athophysiology, risk stratification, and management of sudden cardiac death in coronary artery disease," *Cardiology Journal*, vol. 17, pp. 4–10, 2010.

[4] A. Arcentales, A. Voss, P. Caminal, A. Bayes-Genis, M. Domingo, and B. Giraldo, "Characterization of blood pressure signal considfering low and high risk stratifiaction in cardiomyopathy patients," *Computing in Cardiology*, vol. 40, pp. 795–798, 2013.

[5] A. Golinska, "Poincaré plots in analysis of selected biomedical signals." *Studies in logic, grammar and rhetoric*, pp. 117–127, 2013.

[6] A. Drkosova, J. Kozumplík, and Z. Nováková, "Heart rate variability expressed by poincaré plot in metabolic syndrome," *International Congress on Electrocardiology*, pp. 221–224, 2014.

[7] G. P. Piskorski J, "Geometry of the poincaré plot of rr intervals and its asymmetry in healthy adults," *Physiological Measurement*, pp. 287–300, 2007.

[8] C. Karmakar, A. Khandoker, J. Gubbi, and M. Palaniswami, "Complex correlation measure: a novel descriptor for poincaré plot," *BioMedical Engineering OnLine*, pp. 8–17, 2009.

[9] R. Hoshi, C. Pastre, L. Vanderlei, and M. Godoy, "Poincaré plots indexes of heart rate variability: relationship with other nonlinear variables," *Autonomic Neuroscience*, pp. 271–274, 2013.

[10] B. Giraldo, J. Rodriguez, P. Caminal, A. Bayes-Genis, and A. Voss, "Cardiorespiratory and cardiovascular interactions in cardiomyopathy patients using joint symbolic dynamic analysis," *7th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBC)*, pp. 306–309, 2015.

[11] C. Karmakar, A. Khandoker, and M. Voss, A amd Palaniswami, "Sensitivity of temporal heart rate variability in poincaré plot to changes in parasympathetic nervous system activity," *BioMedical Engineering OnLine*, pp. 10–17, 2011.

[12] C. Huberty, "Applied discriminant analysis, whiley series in probability and mathematical statistics," *Editorial Jhon Wiley & Sons Inc*, 1994.