Autonomic Neuroscience: Basic and Clinical 178 (2013) 103–110

Contents lists available at ScienceDirect

Autonomic Neuroscience: Basic and Clinical

journal homepage: www.elsevier.com/locate/autneu

Classifying healthy women and preeclamptic patients from cardiovascular data using recurrence and complex network methods

G.M. Ramírez Ávila [a,b,c], A. Gapelyuk [a], N. Marwan [b], H. Stepan [d], J. Kurths [a,b,e], Th. Walther [f,g], N. Wessel [a,*]

[a] Department of Physics, Humboldt-Universität zu Berlin, Berlin, Germany
[b] Potsdam Institute for Climate Impact Research, Potsdam, Germany
[c] Instituto de Investigaciones Físicas, Universidad Mayor de San Andrés, La Paz, Bolivia
[d] Department of Obstetrics and Gynecology, University of Leipzig, Leipzig, Germany
[e] Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, United Kingdom
[f] Department of Pediatric Surgery and Department of Obstetrics, University of Leipzig, Leipzig, Germany
[g] Institute for Experimental and Clinical Pharmacology and Toxicology, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form 24 April 2013
Accepted 2 May 2013

Keywords:
Heart rate
Blood pressure
Cardiac dynamics
Heart rate
Preeclampsia
Recurrences
Networks
Time series analysis

ABSTRACT

It is urgently aimed in prenatal medicine to identify pregnancies, which develop life-threatening preeclampsia prior to the manifestation of the disease. Here, we use recurrence-based methods to distinguish such pregnancies already in the second trimester, using the following cardiovascular time series: the variability of heart rate and systolic and diastolic blood pressures. We perform recurrence quantification analysis (RQA), in addition to a novel approach, ε-recurrence networks, applied to a phase space constructed by means of these time series. We examine all possible coupling structures in a phase space constructed with the above-mentioned biosignals. Several measures including recurrence rate, determinism, laminarity, trapping time, and longest diagonal and vertical lines for the recurrence quantification analysis and average path length, mean coreness, global clustering coefficient, assortativity, and scale local transitivity dimension for the network measures are considered as parameters for our analysis. With these quantities, we perform a quadratic discriminant analysis that allows us to classify healthy pregnancies and upcoming preeclamptic patients with a sensitivity of 91.7% and a specificity of 45.8% in the case of RQA and 91.7% and 68% when using ε-recurrence networks, respectively.

1. Introduction

Nowadays, a severe pathology called preeclampsia (PE) affects healthy nulliparous women in a range between 2% and 7% worldwide (Sibai et al., 2005). The main features of PE are severe hypertension and proteinuria for which the pathophysiology is not well understood at present. Several strategies are used in order to predict PE, among which we can mention biochemical markers, such as fms-like tyrosine kinase 1 (sFlt-1), placental growth factor (PlGF), soluble endoglin (Ohkuchi et al., 2011; Rana et al., 2007), maternal autoantibody, angiotensin II type I receptor agonistic autoantibody (AT1-AA) (Siddiqui et al., 2010), urinary biomarkers (Carty et al., 2011), noninvasive cardiovascular (CV) indicators (Malberg et al., 2007; Walther et al., 2006), or a combination of the above (Stepan et al., 2008).

In recent years, recurrence methods based on recurrence plots (RP) have been successfully used in different fields of natural sciences as physics (Ngamga et al., 2012) and biology (Angus et al., 2012), but also to answer economic (Hirata and Aihara, 2012) or medical questions (Wessel et al., 2009). Recurrence quantification analysis (RQA), in particular, constitutes a very useful tool for the description and analysis of a systems diversity (Marwan, 2008; Marwan et al., 2007). More recently, the recurrence concept has been extended to networks and applied in novel time series analysis methods (Marwan et al., 2009), finding several applications such as in paleoclimate modeling (Donges et al., 2009).

The detection of cardiovascular disorders has been considerably improved due to both technological advances and new methods of time series analysis. Nevertheless, there are still unclear mechanisms that cannot be explained by standard data analysis. Nonlinear data analysis and modeling methods of CV physics allow to improve clinical diagnostics and also a better understanding of CV regulation. One of the most important aspects of these methods is that they focus on noninvasive measured biosignals. Among the biosignals that CV physics deals with are the heart rate variability (HRV) and the variabilities of systolic blood pressure (SBPV) and diastolic blood pressure (DBPV).

* Corresponding author.
E-mail address: wessel@physik.hu-berlin.de (N. Wessel).

1566-0702/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.autneu.2013.05.003

In this work, we apply the approach of RQA and ε-recurrence networks to analyze CV biosignals, obtained by noninvasive techniques, with the aim of developing a classification method to identify patients who develop PE in a pool of pregnancies within the second trimester.

2. Methods

2.1. Clinical aspects

We considered for this study 96 pregnancies with abnormal uterine perfusion (AUP), followed by means of Doppler sonography in the second trimester, between the 18th and the 26th week of gestation (WOG) of pregnancy, at the Department of Obstetrics and Gynecology of the University of Leipzig, Germany. Immediately after the Doppler examination, the blood pressure was measured noninvasively via finger cuff for 30 min (sampling rate: 100 Hz, Portapres device model 2, BMI-TNO, Amsterdam, The Netherlands). The continuous blood pressure curves were used to extract the time series of beat-to-beat intervals and systolic and diastolic blood pressures, allowing us to obtain the CV values (HRV, SBPV, and DBPV). The length of the dataset per variable is roughly of 1600 samples (heart beats). At the time of examination, the women were healthy, normotensive, without clinical signs of cervical incompetence, and on no medication. After the 30th WOG, 24 patients developed PE. Further details on the methodology can be found in Malberg et al. (2007). We point out that the root mean square errors of heart beats calculated from blood pressure curves (compared to ECG slope detection) is about 5–6 ms (Suhrbier et al., 2006). Therefore, the computation of the beat-to-beat-intervals from the distal pulse wave measurement as it has been performed in this paper is an acceptable alternative; however, this has to be confirmed in another comparative study.

2.2. Recurrence methods

The concept of recurrence applied to a single trajectory of the dynamical system allows us to obtain the recurrence matrix whose elements are given by $R_{i,j} = \Theta(\varepsilon - ||\mathbf{x}_i - \mathbf{x}_j||)$, where $\Theta(\cdot)$ represents the Heaviside function, $||\cdot||$ is a suitable norm, and ε is a threshold distance that should be chosen adequately according to the characteristics of the embedded attractor into the phase space. We use RQA and ε-recurrence networks with the aim of distinguishing between healthy individuals and patients with PE.

2.2.1. Recurrence quantification analysis

The RQA is a method of nonlinear data analysis that quantifies the number and duration of recurrences of a dynamical system presented by its state space trajectory. This method was developed by Zbilut and Webber (1992) and extended by Marwan et al. (2002). Several measures might be used to quantify the time series of a system when using RQA, such as the following: recurrence rate (RR), the percentage of recurrence points in an RP, corresponding to the correlation sum; determinism (DET), the percentage of recurrence points forming diagonal lines; laminarity (LAM), the percentage of recurrence points forming vertical lines; trapping time (TT), the average length of the vertical lines; and some other self-explanatory measures such as longest diagonal line (L_{MAX}) and longest vertical line (V_{MAX}). A more detailed description of these measures can be found in Marwan et al. (2007).

2.2.2. Recurrence networks

The basic idea of time series analysis based on complex network techniques relies on the fact that a time series may be transformed into a complex network from which we can extract the adjacency matrix, allowing us to obtain local and global network properties (Donner et al., 2011). We interpret the recurrence matrix **R** as the adjacency matrix of an unweighted and undirected complex network, commonly called the ε-recurrence network, which is associated with a given time series. Possible self-loops must be avoided in this network; thus, a Kronecker delta must be subtracted from the recurrence matrix. The elements of the adjacency matrix for an ε-recurrence network are thus

$$A_{i,j}(\varepsilon) = R_{i,j}(\varepsilon) - \delta_{i,j}, \tag{1}$$

where the ε-dependence is considered explicitly as in the case of RQA. There is no universal criterion for choosing ε, but the choice must be made avoiding too small values, which lead to a situation in which there are not enough recurrence points, or too large values, implying that every vertex is connected with many other vertices irrespective of their actual mutual proximity in phase space (Donner et al., 2010b). Having reconstructed the adjacency matrix **A** from a time series, we can apply appropriate network characteristics to analyze and obtain information on the underlying system (Donges et al., 2012). In Appendix A, there is an explanation of how to obtain the adjacency matrix, the associated network, and the 4-element motifs. In this work, we focus our interest on five global network measures: the average path length ($\mathcal{L}$), which is the mean value of the shortest geodetic path lengths $l_{i,j}$ considering all pair of vertices (i,j); the mean coreness ($\mathcal{C}\prime$), which is the average of the coreness (significance of a node and its "popularity" in the network) of all the vertices (Batagelj and Zaveršnik, 2002); the global clustering coefficient ($\mathcal{C}$), which is the average of the clustering coefficient of each vertex (ratio of triangles including vertex i and the number of triples centered on vertex i, where triple refers to a pair (j,k) of vertices that are both linked with i, but not necessarily mutually linked); the assortativity ($\mathcal{A}$), the tendency for vertices in networks to be connected to other vertices that are like (or unlike) them in some way (Newman, 2003); and the scale local transitivity dimension ($D_{\mathcal{T}}$), defined as $D_{\mathcal{T}} = \frac{\log \mathcal{T}}{\log(3/4)}$, where $\mathcal{T}$ is the transitivity (ratio of the number of triangles in the network times three and the number of linked triples of vertices). These four measures depend on ε and have a global character. A detailed description of networks and their properties can be found in Boccaletti et al. (2006).

3. Data processing and statistics

We use an algorithm that avoids artifacts such as extrasystolic beats. The original time series from consecutive R waves were filtered using a preprocessing algorithm that first removes obvious recognition errors, then applies an adaptive percent filter, and finally an adaptive controlling filter (Wessel et al., 2007). With the aim of using a recurrence approach, we consider the three CV indicators and several possible embeddings. An estimation of the coupling structure of CV indicators has been performed using nonlinear additive autoregressive models with external input, following the idea of Granger causality (Riedl et al., 2010). This coupling analysis shows that HRV, DBPV, and SBPV respond to respiration; SBPV respond to DBPV and the latter to HRV. In our case, we do not consider respiration; thus, the coupling structure may be represented as in Fig. 1(a), where, according to the coupling scheme, there is a delay between the HRV, the DBPV, and the SBPV. For simplicity, we write down the coupling structure as (HRV(t), DBPV ($t+1$), SBPV ($t+2$)), or simply $H(t)D(t+1)S(t+2) \equiv 012$.

We sought to predict whether or not a patient develops PE using the CV indicators embedded in a phase space determined by the structure of coupling. We consider a minimalist assumption in which the structure of coupling between HRV, DBPV, and SBPV is identical in each subject of a group and that this structure does not change during the measurement. In this study, we set out to test all the possible structures of coupling shown in Fig. 1 and a wide range of the threshold ε going from 0.01σ to 0.99σ, where σ is the standard deviation of the underlying process in the embedded phase space. From a simple CV time series corresponding to each patient, we construct a complex network for each possible structure of coupling and each value of ε. Then we compute the four network measures: ($\mathcal{C}, \mathcal{L}, \mathcal{C}\prime, D_{\mathcal{T}}$), and with these new measures, we perform an analysis to classify the groups of individuals: healthy and preeclamptic patients. For that purpose, we firstly verify