Biomedical Signal Processing and Control 7 (2012) 295–302

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Linear and non-linear analysis of cardiac health in diabetic subjects

Oliver Faust [a,*], U.Rajendra Acharya [a], Filippo Molinari [b], Subhagata Chattopadhyay [c], Toshiyo Tamura [d]

[a] Ngee Ann Polytechnic, 535 Clementi Road, Singapore 599489, Singapore
[b] Biolab, Department of Electronics, Politecnico di Torino, Torino, Italy
[c] School of Computer Studies, Department of Computer Science and Engineering, National Institute of Science and Technology, Berhampur 761008, Orissa, India
[d] Department of Medical System Engineering, Chiba University, Chiba 263-8522, Japan

ARTICLE INFO

Article history:
Received 5 January 2011
Received in revised form 21 May 2011
Accepted 3 June 2011
Available online 13 July 2011

Keywords:
Cardiomyopathy
Non-linear methods
Linear methods
Correlation dimension
Approximate entropy
Sample entropy
Recurrence plot properties
Poincare plot

ABSTRACT

Diabetes is a chronic disease characterized by hyperglycaemia, which leads to specific long-term complications: retinopathy, neuropathy, nephropathy and cardiomyopathy. Analysis of cardiac health using heart rate variation (HRV) has become a popular method to assess the activities of the autonomic nervous system (ANS). It is beneficial in the assessment of cardiac abnormalities, because of its ability to capture fast fluctuations that may be an indication of sympathetic and vagal activity.

This paper documents work on the analysis of both normal and diabetic heart rate signals using time domain, frequency domain and nonlinear techniques. The study is based on data from 15 patients with diabetes and 15 healthy volunteers. Our results show that non-linear analysis of HRV is superior compared to time and frequency methods. Non-linear parameters namely, correlation dimension (*CD*), approximate entropy (*ApEn*), sample entropy (*SampEn*) and recurrence plot properties (*REC* and *DET*), are clinically significant.

1. Introduction

Diabetes mellitus, or diabetes, is a chronic disease, which is characterized by hyperglycaemia. Hyperglycaemia, is a metabolic disorder, were excess glucose is present in the blood. This results in an elevated blood glucose level, which leads to serious detrimental consequences. The disease affects eye (retinopathy) [1,2], nerves (neuropathy) [3], kidney (nephropathy) [4] and heart (cardiomyopathy) [5].

According to World Health Organization (WHO) more than 220 million people worldwide had diabetes, in 2009. It is estimated that this figure will increase to 440 million by the year 2030 [6]. In 2008, the American Diabetes Association reported 23.6 million (approximately 7.8% of the population) children and adults in the United States have diabetes [7].

Cardiovascular disease (CVDs) is the number one cause of death globally. WHO estimated about 29% of deaths were due to CVDs (totaling 17.1 million) in 2004. They projected about 23.6 million people will succumb to the disease by 2030 [8].

Heart rate variation (HRV) is the name of a biological time series signal which indicates the variation of heart rate between two consecutive heart beats [9]. HRV is a non-invasive tool to assess the autonomic nervous system (ANS). HRV may take precedence over the situation where loads and loads of data are to be collected for several hours in order to understand and identify abnormalities. Thus, HRV can be seen as highly effective diagnostic tool. In recent years, there has been much work by various researchers on the analysis of heart rate variability [10–12]. HRV also gives information about the sympathetic–parasympathetic autonomic balance and about the risk of sudden cardiac death in these patients [13]. HRV measurements are easy to obtain and they are reproducible, if measured under standardized conditions [14,15].

Biological time series analysis can be done in time and frequency domain as well as with non-linear methods. The aim of the analysis is to detect important dynamical properties of the physiological phenomena which are hidden in the data. A particular problem of biological time series analysis comes from the fact that statistical characteristics can vary with time. This makes time domain analysis unreliable. Frequency domain parameters give better assessment of the autonomic function, but the reliability of spectral power diminishes with the decrease in power signal and signal-to-noise ratio [16].

Non-linear dynamical techniques are used in many areas including biology and medicine, because they can overcome the shortcomings of time and frequency domain methods [17]. These techniques yield useful indicators of pathologies, because many biological systems, such as the cardiovascular system, are complex and can never be linear in nature. Schumacher [18] have explained

* Corresponding author. Tel.: +65 6460 6602.
E-mail address: fol2@np.edu.sg (O. Faust).

1746-8094/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2011.06.002

the use of linear and non-linear techniques in the analysis of HR signals. Furthermore, non-linear methods have been applied in tracking HRV signals and predicting the onset of dangerous cardiovascular related events, such as Ventricular Tachycardia and congestive heart failure [19].

The decreased beat-to-beat variability during deep breathing of patients with diabetic neuropathy was first reported by Wheeler and Watkins [20] and confirmed by many others. Studies which compared cardiac autonomic function tests and HRV indices (based on both short (5-min) and 24-h electrocardiogram (ECG) recordings), showed that in diabetic patients without abnormal function tests HRV was lowered [21]. It was concluded that cardiac (parasympathetic) autonomic activity was diminished in diabetic patients before clinical symptoms of neuropathy became evident [22,23].

In this paper we analyze HRV signals with both linear and non-linear methods. The non-linear methods uncover subtle changes in the beat to beat variation of the signal. This makes them more robust against systematic (linear) measurement errors. Furthermore, these methods aim to reveal the underlying physiological processes directly and not, as in linear methods, measure secondary or indirect signal phenomena. Therefore, we promote non-linear methods to determine the cardiac health of diabetic patients. These measures can be used for treatment monitoring and to assess the results of clinical trials.

2. Materials and methods

2.1. Data

The electrocardiograms (ECG) of 15 patients (10 male and 5 female) with diabetes and of 15 healthy volunteers (8 male and 7 female)) were recorded, with the patients in a relaxed supine position for 60 min. The subjects under study of diabetes were in the age group of 50 ± 70 years (mean $\pm$ standard deviation = 58.5 ± 6.42 years) and the duration of diabetes for the patient groups was 5 ± 15 years. The normal subjects were in the age group of 40 ± 60 years (mean $\pm$ standard deviation = 50 ± 8.8 years). ECG signals were taken at the Kasturba Medical Hospital, Manipal, India. The ethics committee, which consists of senior doctors, has approved the use of this data for research purposes.

The recording was done with the BIOPAC^{TM} equipment, in corporation with the AcqKnowledge software, using a sampling rate of 500 Hz [24]. The ECG signal is passed through a low pass filter with a cut-off frequency of 15 Hz to remove unwanted high frequency components which are present in ECG signals. Then it is passed through a highpass filter with a cut-off frequency of 0.3 Hz to remove the so called baseline wander [25]. A Notch filter with a cut-off frequency of 50 Hz is used to remove power-line interference noise. Finally, a median filter is used to extract the baseline wander and then subtracted from the original ECG to get the ECG signal without baseline wander. The *R* peaks of the ECG signal were detected using Tompkins's algorithm [26,27]. A total of 73 data sets from 15 normal subjects and 71 data sets from 15 diabetic subjects were used in this study. Each data set contained 1000 samples.

A specific *RR* interval n within the ECG signal ($T_{RR}(n)$ seconds) is defined as the interval of two successive QRS complexes and the heart rate (beats per minute) is given as:

$$HR(n) = \frac{60}{T_{RR}(n)} \tag{1}$$

The graphs in Fig. 1(a) and (b) show the respective plots of sample HR signals for normal and diabetic subjects. All subsequent plots in this paper are based on the same signals.

2.2. Time domain analysis

In this section we discuss the time domain parameter extraction. The extracted parameters are $\overline{HR}$, $\overline{HF}$, *NN*50, *PNN*50, and *HRV* $\triangle$ *Index*. All these parameters were selected because they are statistically significant.

2.2.1. Mean heart rate ($\overline{HR}$)

The heart rate mean can be calculated as:

$$\overline{HR} = \frac{1}{N}\sum_{n=0}^{N-1} HR(n) \tag{2}$$

where N is the observation interval.

2.2.2. Statistical parameters (NN50, PNN50)

In this study we have used the statistical parameters *NN*50 and *PNN*50. *NN*50 is the number of successive *RR* interval pairs that differ by more than 50 ms. *PNN*50 is the number of successive difference of intervals which differ by more than 50 ms (*NN*50) expressed as a percentage of the total number of ECG cycles analyzed. In other words, *NN*50 divided by the total number of *RR* intervals, expressed as a percentage gives *PNN*50.

$$PNN50 = \frac{NN50}{N-1} \times 100 \tag{3}$$

2.2.3. Histogram parameter (HRV $\triangle$ Index)

Besides the statistical parameters mentioned above, we used the *RR* interval histogram, shown in Fig. 2, to calculate the geometric parameter *HRV* $\triangle$ *Index*. Triangular interpolation approximates the *RR* interval distribution by a linear function and the baseline width of this approximation is used as a measure of the HRV index. [28,10]. This index possesses a high correlation with the standard deviation of all *RR* intervals [10]. The *HRV* $\triangle$ *Index* is calculated as the *RR* interval histogram integral D, which indicates the area under the curve, divided by the height Y of the histogram. It is a measure where the length of the base of the triangle is used as an approximate by the main peak of the *RR* interval frequency distribution diagram.

2.3. Frequency domain analysis

Although time domain methods are straight forward and easy to use, they lack the ability to differentiate if the HRV is sympathetic or para-sympathetic. To overcome these shortcomings the power spectrum density (PSD) estimate can be used to analyze HRV signals.

We used an autoregressive (AR) method to estimate the PSD. The necessary AR coefficients ($a(k)$) were calculated using linear equations, therefore PSD estimation is a linear method. The data is modeled as the output of a causal all pole discrete filter with white noise as input.

$$x(n) = -\sum_{k=1}^{p} a(k)\,x(n-k) + w(n) \tag{4}$$

where p is the filter order and $w(n)$ is white noise with a variance of σ^2. A specific model can be characterized by the following parameters: $a[1]$, $a[2]$, ..., $a[p]$, and σ^2.

Akaike and other researchers have stressed the importance of the model order p on the accuracy of the PSD estimation [29,30]. In this work we have taken the AR model order: $p = 16$, because this order yields the most significant features. Fig. 3(a) and (b) shows the Frequency Domain (AR Spectrum) of a Normal and a Diabetic subject, respectively. The frequency range, depicted on the x axis of these plots is partitioned into individual frequency bands which