PHYSCON 2009, Catania, Italy, September, 1–September, 4 2009

MULTIDIMENSIONAL ANALYSIS TOWARD THE IDENTIFICATION OF ECG NONLINEAR DYNAMICS

M.Bucolo, F. Di Grazia, F.Sapuppo
Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi
Universit degli Studi di Catania
V.le A, Doria 6, 95125 Catania, Italy
maide.bucolo@diees.unict.it

D. Nikolic, B. Vuksanovic
Department of Electronic and Computer Engineering
Faculty of Technology
University of Portsmouth, UK

Abstract

Using ECG it is possible to detect the rate and regularity of heartbeats and identify possible irregularities to the heart activity. In this paper, a method to classify normal and two types of abnormal ECG signals is introduced. In the first stage of described process, techniques used to extract a number of potential classification parameters evaluated from 2 minutes long ECG signal epochs are described. The extracted parameters can be generally divided into three groups: (i) standard statistical signal parameters, (ii) nonlinear parameters and (iii) specific heart rate variability parameters. Two dimensionality reduction algorithms, principal component analysis (PCA) and linear discriminant analysis (LDA), have been employed in order to reduce the size of dataset containing ECG parameters and followed by a clustering algorithm. The results show the ability of this method to detect different pathologies and to distinguish normal ECG behaviour from pathological ones. Furthermore, this approach could be implemented in real time applications and embedded in a portable device and this effort is the first step towards the final realization of an automatic solution for the real-time characterization and monitoring of heart signals.

Key words

Nonlinear indicators, clustering approach, dimensional reduction algorithms

1 Introduction

HERE is a vast literature on the analysis of Electrocardiogram (ECG) signals and in general of biological systems from the classical linear approaches, autocorrelation [S. G. Guilln et al, 1989], frequency domain features [K. Minami et al, 1999] [L. Khadra et al, 1997], time frequency domain, to the nonlinear analysis and chaos theory. In particular in ECG it was shown evidence that the dynamics underlying the cardiac signals is nonlinear and indicate the possibility of deterministic chaos [H. Kantz and T. Schreiber, 1998]. Several features can be used to describe system dynamics including correlation dimension (D2) [M. I. Owis et al, 2002], Lyapunov exponents (λ), approximate entropy, etc. These features have been used to explain ECG signal behavior by several studies. Nevertheless, these studies applied such techniques only to a few sample ECG signals that did not allow the extraction of a general statistical description of the dynamics of different arrhythmia types. Given that such techniques are particularly sensitive to parameter variations, it is not possible to directly utilize these results or attempt to draw conclusions based on these studies about the robustness of their implementations. Therefore, the need arises for a study of the ECG behavior based on a large number of signals using a suitable algorithm for the characterization of time series coming from physiological systems with the effort of realize a portable device that implements novel and well-known methodologies towards the realization of an automatic solution for the real-time characterization of the nonlinear dynamics of the ECG signal and its variation with different arrhythmia types. In this paper the LDA method [A.M Martinez, A.C. Kak, 2001] combined with the PCA approach [I. Jolliffe, 1986] have been used to classify nonlinear indicators as the asymptotic distance d (d-infinite) and the Lyapunov exponents and other well known statistical values as the mean, the standard deviation, the range, and linear parameters as the power and HRV parameters [V.Afonso et al, 1999], extracted from an huge dataset of ECG signals. Furthermore the problem of characterize, the linear and nonlinear dynamics of the ECG signal and its variation with different arrhythmia types has been addressed. The proposed implementations were used to compute these features for a large number of independent ECG signals belonging to three different ECG signal types from the MIT-BIH Arrhythmia Database [MIT-BIH Arrhythmia Database] also including normal subjects [Politecnico Biosignals Archives]. The results show interesting advantages due to the potentiality of the proposed methods to discriminates different pathological states readable for future clinical applications.

2 Case Study

The ECG signal data set consists of three different types including normal (NR) from the Politecnico of Milano VCG/ECG Database on Young Normal Subject [Politecnico Biosignals Archives], arrhythmia (AR) from the MIT-BIH Arrhythmia Database [MIT-BIH Arrhythmia Database], ventricular arrhythmia (VAR) from the MIT-BIH Malignant Ventricular Arrhythmia Database (Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center). Each type was represented by 20 half-hour excerpts of two-channel ambulatory ECG recordings, but 10 minutes per patient have been considered. The time series relative to the normal subjects were acquired with a sampling frequency Fs = 500Hz, while the time series relative to arrhythmic patients have Fs = 250Hz or 360Hz for Ventricular Arrhythmia (both Ventricular Tachycardia and Ventricular fibrillation). The Divergence Algorithm was implemented on the three different data types (NR, AR, VAR). In Figure 2 and 2 two examples of arrhythmia and of ventricular arrhythmia are shown. Each heartbeat is normally represented as 5 major waves: P, Q, R, S, and T. The Q, R, and S waves all represent the same portion of the heartbeat and are referred to as a unit: QRS complex. Occasionally, a 6^{th} wave will appear. It is referred to as the U wave. Although it does not always appear, its presence is perfectly normal, (see Figure 2).

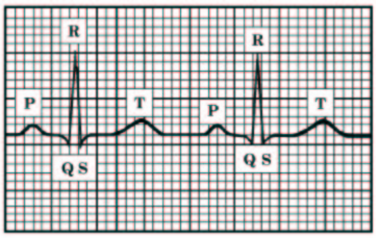

Figure 1. Normal heartbeat.

An irregular heartbeat is an arrhythmia (also called dysrhythmia). Heart rates can also be irregular. A normal heart rate is 50 to 100 beats per minute. Arrhythmias and abnormal heart rates don't necessarily occur together. Arrhythmias can occur with a normal heart rate, or with heart rates that are slow (called bradyarrhythmias – less than 60 beats per minute). arrhythmias can also occur with rapid heart rates (called tachyarrhythmias – faster than 100 beats per minute).

One of the most serious arrhythmias is sustained ventricular tachycardia (see Figure 2). In sustained ventricular tachycardia, there are consecutive impulses that arise from the ventricles at a heart rate of 100 beats or more per minute until stopped by drug treatment or electrical conversion. This condition is very dangerous. It is dangerous because it may degenerate further into a totally disorganized electrical activity known as

Figure 2. ECG in a case of arrhythmia.

ventricular fibrillation (see Figure 2). In ventricular fibrillation, heart's action is so disorganized that it quivers and does not contract, thus failing to pump blood.

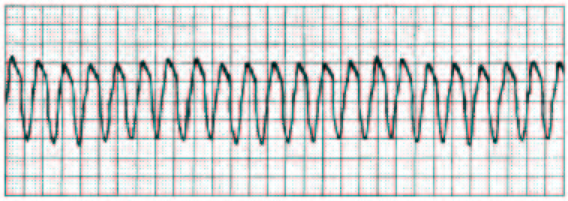

Figure 3. ECG in a case of Ventricular Tachycardia.

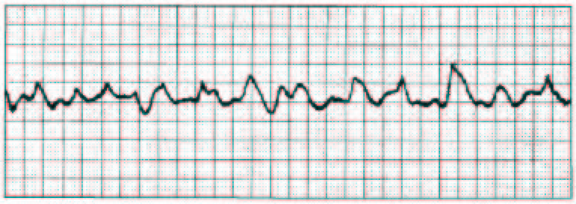

Figure 4. ECG in a case of Ventricular fibrillation.

3 Methods

The ECG signals used for the signal analysis were acquired from 30 subjects: 10 normal subjects, 10 subjects suffering from arrhythmia and 10 subjects suffering from ventricular arrhythmia. The procedure adopted to classify those three different ECG types is schematized in Figure 3. First various linear and non-linear parameters have been extracted from the data, then after a dimensional reduction through PCA the classification has been performed.

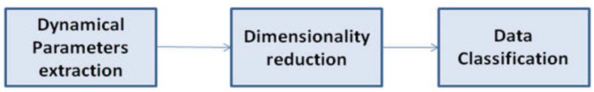

Figure 5. Block scheme of the adopted procedure.

3.1 Parameters extraction

For each subject 10 minutes of registration have been considered and specifically slots of 2 minute each have been analyzed using the Divergence Algorithm (DivA) [Bucolo M. et al, 2008] to extract two nonlinear signal parameters of asymptotic divergence (d_∞) and maximum Lyapunov exponent (λ). In addition, other parameters have been calculated: the mean (μ), standard deviation (σ), range, power (P), a control parameter (Π) and heart rate variability (HRV) parameters. In order to compute these HRV parameters, it is required to determine precise timing of heart beats (R peaks). This is done by using the ECG beat detection algorithm which employs filter banks [V.Afonso et al, 1999]. Once when the R peaks are detected, heart rate fluctuations are assessed by calculating parameters based on the statistics of RR intervals i.e. time intervals between two successive heart beats. As a result, the final set of 15 parameters (dimensions) is formed holding the statistical, nonlinear and HRV parameters extracted from the five consecutive 2 minutes long epochs of the ECG signals. The detailed list of these parameters is given in Table 1.

Table 1. Description of the ECG parameters.

No	Symbol	Description
Statistical Parameters		
1	μ	mean
2	Σ	Standard Deviation
3	D	Data Range
4	P	Signal Power
5	Π	Control Parameter
Nonlinear parameters		
6	d_∞	Asymptotic Divergence
7	λ	maximum Lyapunov exponent
HVR parameters		
8	μ_{HR}	Average Heart Rate
9	μ_{HR}	Average RR Interval
10	RR_{max}	Maximum Length of RR Interval
11	RR_{min}	Minimum Length of RR Interval
12	d_{RR}	Difference between Maximum and Minimum RR Interval Length
13	σ_{RR}	Standard Deviation of RR Intervals
14	RMSSD	Square Root of the Mean Squared Differences of Successive RR Intervals
15	N_{RR}	Number of Beats Occurring in a Chosen Time Window

3.2 Nonlinear parameters

The maximum Lyapunov exponents is often fully representative of the sensitivity to initial condition (stretching phase) of a given dynamics, being positive for chaotic behaviours [S. H. Strogatz, 1998]. The asymptotic distance between trajectories starting from the very close initial conditions is, called d-infinite (d_∞) and it is also exploited in order to characterise chaotic behaviours; this parameter not only characterizes the stretching but also the folding phase. The d_∞ parameter represents therefore a complementary and alternative parameter characterizing complex dynamics when Lyapunov exponents are either very difficult, or impossible to work out. These two measures are widely used in literature to analyse different type of physiological signals as indicator of nonlinear phenomenon or, generally speaking to study nonlinear dynamics. In the theoretical determination of the λ, and d_∞, the knowledge of the finite difference equations in the discrete domain as well as the differential equations in the continuous domain is fundamental. However, in the experimental field the laws that describe the dynamical evolution of the nonlinear system variables are not known, thus making difficult the computation of those parameters. The DivA (acronym of Divergence Algorithm) algorithm, an alternative methodology for the extraction of the asymptotic distance d_∞ in experimental data, has been used. This implementation results to be computationally less onerous than the conventional ones, since it is not based on the time-delay embedding concept and also no intermediate computational steps are needed to obtain the final result. Therefore, the procedure developed here evaluates the d_∞ as the asymptotic value of the average distance between trajectories that are directly extracted from the time series. This algorithm is particularly suitable when coping with extended, both in time and in space, datasets.

3.3 Dimensionality reduction: Principal Component Analysis and Linear Discriminant Analysis

Two methods - Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) - have been used to try to reduce the dimensionality of the set of classification parameters obtained from the ECG signals. Both of those methods are searching for the linear combinations of correlated parameters which can best explain the obtained data set. PCA [I. Jolliffe, 1986], [R. O. Duda et al, 2000] reduces the complex set of data by projecting it to a space with smaller dimensions whilst preserving as much of the variability in the parameter set as possible. Principal components are uncorrelated so it is possible to consider one principal component without taking into account the other obtained principal components. However, PCA has its drawbacks - obtained principal components are usually difficult to interpret or assign any physical meaning since they represent the linear combination of all variables from the original data set. Similarly to PCA, LDA [R. A. Fisher, 1936]-[A.M Martinez, A.C. Kak, 2001] also tries to find a linear combination of variables that best explains the obtained data set. Under the presumption that the data in the set comes from a number of different classes, LDA is trying to perform the dimensionality reduction while trying to preserve as much of class discriminatory information as possible. LDA seeks to find the direction along which the classes are best separated by taking into account variability within the classes as well as variability between the classes. From the above discussion, it is clear that the PCA method deals with the data in its entirety - it is trying to characterize variability in the data set without taking into account class membership of the data. LDA on the other side is trying to take into account as much of class discriminatory information as possible. For this reason, principal components obtained from the PCA might not provide a very good separation between the data in the set. In many classification tasks, LDA based algorithms might therefore be superior to PCA based ones. This is proven by the initial classification results obtained from the ECG signals and described in the paper. It should however be noted that there exist a number of classification problems where PCA achieves better dimensionality reduction results compared to LDA. LDA is a parametric method since it assumes unimodal Gaussian distributions of data within each class. If discriminatory information is in the data variance rather than in the mean of data, performance of LDA will be poor. As pointed in [A.M Martinez, A.C. Kak, 2001], performance of PCA can be superior compared to LDA when the data sets are small. PCA is also less sensitive to different data training sets.

4 Results and Discussions

According to the previous section LDA has been applied on the full 15-D data set to reduce its dimensionality (Figure 4d) and it is possible to see a clear separation of the three different types of ECG. The other plots shown in Figure 4 justify also the use of both nonlinear and HRV parameters in this analysis showing the results of LDA implementation on the statistical and nonlinear (Figure 4b) or statistical and HRV parameters (Figure 4c) in comparison to only statistical parameters (Figure 4a) where the discrimination among the three different types of ECG is not performing.

Preliminary results obtained after implementing PCA and LDA methods show the ability of both methods to distinguish between the class representing normal subjects and the other two classes with data acquired from the subjects suffering from arrhythmia and ventricular arrhythmia. However, LDA outperforms PCA (Figure 4) being more efficient in the separation of the classes related to two pathologies, thus for further studies the LDA approach has been preferred. Thus, as shown in Figure 6 the combination by LDA approach of all 15 parameters has been adopted as the best solution for the classification of the three different classes.

The possibility of discriminate different types of ECG by using dimensionality reduction methods has encourage the possibility of trying clustering approach to a better classification of various ECG behaviours. Two measures used to assess the classification accuracy achieved when using different parameters or parameter extraction approaches are sensitivity and specificity. Sensitivity refers to the ability of a test to detect a disease when it is present (the rate at which this occurs is called the false-negative rate), and specificity refers to the ability of a test to indicate nondisease when no disease is present (the rate at which this occurs is called the false-positive rate). These two measures could be used for the decision on each classification parameter. Here a preliminary clustering algorithm, the k-means,

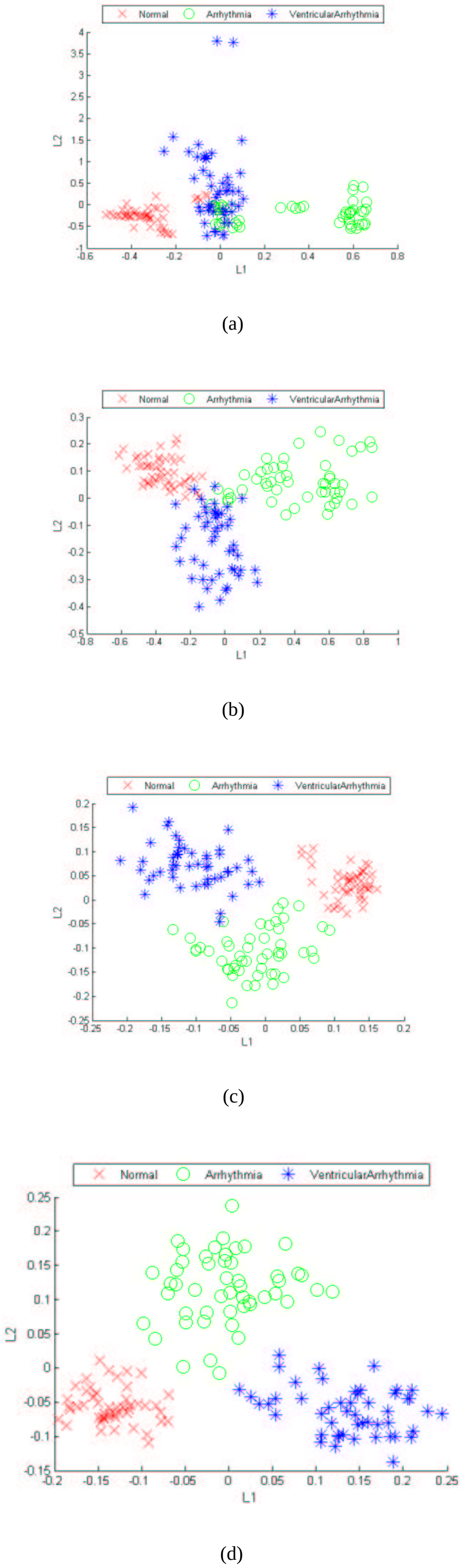

Figure 6. Results obtained after applying LDA on (a) statistical parameters only, (b) the statistical and nonlinear parameters, (c) the statistical and HRV parameters, (d) all 15 parameters.

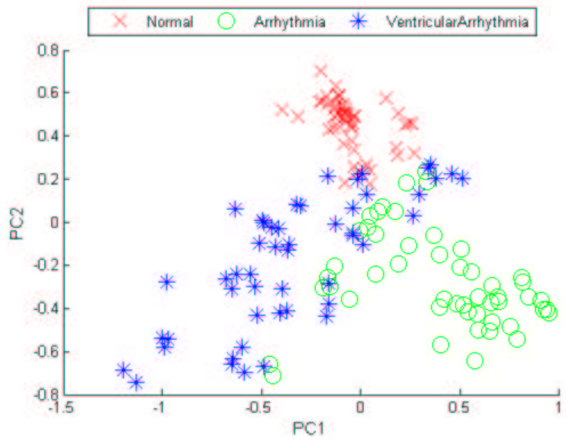

Figure 7. Classification results obtained using PCA.

has been used and the results shown in Figure 4 encourages further clustering studies.

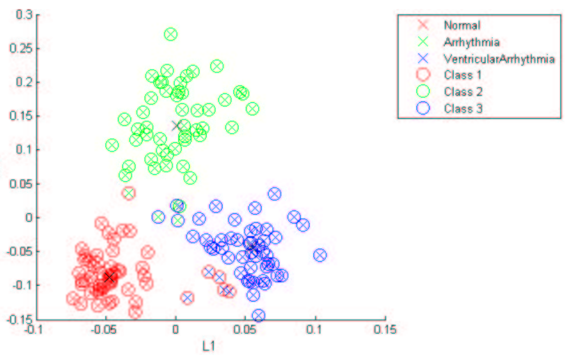

Figure 8. Result of data classification using k-means algorithm.

5 Conclusions

The paper aimed to both shows the potentialities of linear and nonlinear dynamical indicators to classify the different cardiac pathologies recorded through ECG technique. It has been found that the Linear Discriminant Analysis is able to distinguish different type of ECG signals and also to perform the classification of pathologies to be modeled with the same mathematical paradigm. Furthermore a large amount of heterogeneous results have to be compared through multivariate analysis and thus, in order to avoid the use of useless information, the dimensional reduction approaches has been performed. They allow a representation of the featured indexes in a lower-dimensional space and to enhance the class-discriminatory information in the lower-dimensional space. Thus the extraction of linear and nonlinear indicators combined with method for dimensionality reduction and clustering approach have shown consistent results that encourage the realization of a portable device with those methods embedded towards the realization of an automatic solution for the real-time characterization of the dynamics of the ECG signal and its variation with different pathological states. Furthermore for a full experimentation a large dataset could be analyzed, by extracting those indicators from long ECG recordings.

References

Afonso V., Tompkins W., Nguyen T., and Luo S., ECG beat detection using filter banks, *IEEE Transactions on Biomedical Engineering*, **46**, No. 2, (1999) 192-202

I. Jolliffe, *Principal component analysis*, Springer Verlag, (1986)

R. O. Duda, P. E. Hart, and D. Stork. Pattern Classification. Wiley, (2000)

R. A. Fisher, The use of multiple measurements in taxonomic problems,*Annals of Eugenics*,**7**, Part II (1936) 179-188.

A.M Martinez, A.C. Kak, PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, No. 2, (2001) 228-233

S. G. Guilln, M. T. Arredondo, G. Martin, and J. M. F. Corral, Ventricular fibrillation detection by autocorrelation function peak analysis, *J. Electrocardiol.*, **suppl. 22**, pp. 253-262, 1989.

K. Minami, H. Nakajima, and T. Toyoshima, Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network, *IEEE Trans. Biomed. Eng.*, **46**, pp. 179-185, Feb. 1999.

L. Khadra, A. S. Al-Fahoum, and H. Al-Nashash, Detection of lifethreatening cardiac arrhythmias using thewavelet transformation, *Med. Biolo. Eng. Comput.*, **35**, pp. 626-632, Nov. 1997.

H. Kantz and T. Schreiber, Human ECG: Nonlinear deterministic versus stochastic aspects, *Inst. Elect. Eng. Proc. Sci. Meas., Technol.*,**145**, no. 6, pp. 279-284, 1998.

M. I. Owis, A. H. Abou-Zied, A.-B. M. Youssef, and Y. M. Kadah, Study of Features Based on Nonlinear Dynamical Modeling in ECG Arrhythmia Detection and Classification, *IEEE Transaction on Biomedical Engineering*, textbf 49, N. 7, pp. 733-736, July 2002.

S. H. Strogatz, *Nonlinear Dynamics and Chaos Perseus* Book, Cambridge, Massachusetts.Webster J.G. (1998), Medical Instrumentation, Wiley.

The MIT-BIH Arrhythmia Database, 3rd ed., Harvard-MIT Div. HealthSci. Technol., Cambridge, MA, 1997.

Politecnico Biosignals Archives on CD-ROM, Copyright (C) Politecnico di Milano 1992.

Bucolo M., Di Grazia F, Sapuppo F, Virz M.C (2008). *A new Approach for Nonlinear Time Series Characterization, DivA* . The 16th Mediterranean Conference on Control and Automation (MED08). Ajaccio, Corsica, France, June 25-27, p. 1-4.