Experimental analysis of heart rate variability of long-recording electrocardiograms in normal subjects and patients with coronary artery disease and normal left ventricular function

S. Nikolopoulos, A. Alexandridi, S. Nikolakeas, and G. Manis

Abstract

The heart rate signal contains valuable information about cardiac health, which cannot be extracted without the use of appropriate computerized methods. This paper presents an analysis of various electrocardiograms, the aim of which is to categorize them into two distinct groups. Group A represents young male subjects with no prior occurrence of coronary disease events and Group B represents middle-aged male subjects who have symptomatic coronary artery disease without myocardial infarction and whose 12-lead ECGs do not contain any abnormalities, thus wrongly indicating a normal subject. Electrocardiographic recordings are approximately 2h in length and acquired under conditions that favor the stationarity of collected data. Linear and nonlinear characteristics are studied by applying several techniques including Fourier analysis, Correlation Dimension Estimation, Approximate Entropy, and the Discrete Wavelet Transform. The small variations of the diagnostic information given by each one of the methods as well as the slightly different conclusions among similar studies indicate the necessity of further investigation, combined use, and complementary application of different approaches.

Keywords: Heart rate variability; Delay Times; Approximate Entropy; Fourier transform; Discrete Wavelet Transform

1. Introduction

Heart rate variability (HRV) refers to the beat-to-beat alterations in heart rate and a decrease in its value has been used as a measure for predicting future coronary events [1–3]. Abnormalities present in the time interval between R-wave peaks in the ECG serve as a means for exposing cardiac dysfunction. The main objective of related studies is to find a reliable clinical method that differentiates between patients that are likely to show a decrease in HRV from those that are not.

Two of the nervous subsystems that comprise the human nervous system are the Central Nervous System and the Autonomic Nervous System. The central nervous system is composed of the brain and spinal chord. The autonomic nervous system, which functions at a subconscious level, controls many functions of the internal organs, including the level of pumping activity by the heart, movements of the gastrointestinal tract, and glandular secretion [4]. The Autonomic Nervous System is then further subdivided into the sympathetic and the parasympathetic branches.

Physiological function and control of the entire body is maintained by both the sympathetic and parasympathetic sections, which act in opposite directions. An example of this opposite control is the coordination of the heart rate [5,6], which reflects the balance of the instantaneous activity (tone) of the sympathetic and parasympathetic nervous systems. During exercise, the sympathetic nerves serve to quicken the rate of the heartbeat, while during rest the vagus nerve of the...
parasympathetic division slows it down. The muscle tone in the arterial walls is affected similarly. When the parasympathetic system is activated, vessel walls are relaxed and blood pressure drops, as opposed to the sympathetic constricting of blood vessels and raising of blood pressure [4,5].

The study of heart rate variability is important because it provides a means for observing the heart’s ability to respond to normal regulatory signals that affect its rhythm. There are many influences on the regulation of cardiac rhythm, since it is affected by almost every system that modulates the Autonomic Nervous System. The analysis of HRV has proven useful in the diagnosis and monitoring of a number of pathologies, in predicting mortality after myocardial infarction [6], and in estimating the risk of rejection after cardiac transplantation [7]. Recent research has also suggested a potential association between emotional states and HRV in studies that have identified physiological illnesses such as depression and panic disorders [8]. There are also several prospective studies that have shown that HRV independently predicts mortality within the initial two years following a heart attack [8].

It is known throughout the medical community that age and cardiac health affect heart rate variability directly. In applications related to a range of medical settings, findings [9–11] have associated sickness and aging with more regular patterned sinus rhythm, resulting in less complex signals and higher variability than those of youth or cardiac robustness. It is thought that HRV is the result of interaction among complex feedback mechanisms in the cardiovascular system. In this view, as these feedback mechanisms are degraded by disease(s), HRV diminishes.

In this study, specific linear and nonlinear methods are used in order to differentiate between two subject categories. Category A (Group A), is composed of young males with no coronary disease history. Category B (Group B) subjects had angina pectoris or angina-equivalent symptoms. It is important to note that the electrocardiograms done on Group B subjects were normal, showing no signs of coronary artery disease. Since coronary patients had either angina or positive myocardial scintigraphy with Th-201 for ischemia, they all underwent coronary angiography.

A timeseries was extracted from the ECG files and the Fourier Transform and Autocorrelation functions were used to determine whether or not each series is linear. The HRV is then analyzed using the well-known Delay Times, Approximate Entropy, Signal Values Decomposition, and the Discrete Wavelet Transform methods. All of the selected methods are combined in an original way with HRV analysis and it was found that by applying various methods to the same data sets, we could scrutinize the validity of each method as well as comparing all of them to the extent that their intricacies allow. The goal of this paper is not only to serve as a report on an experiment, but also to contribute to existing methods for discriminating between the aforementioned subject categories.

The paper is structured as follows. In the next section, “Background,” we briefly describe the algorithms that were compared in this study. Their main characteristics as well as the reasons they were chosen are discussed. The following section, “Methods,” describes the clinical subjects, the methods of data acquisition, and how the algorithms were applied to our data. The results are presented in section four. The last two sections offer a discussion and some main conclusions of the experimental analysis.

2. Background

2.1. Linear methods

2.1.1. Fourier transforms

Spectral analysis is a widely used method for identifying the most significant HRV signal characteristics and can also be used to determine whether they are linear or nonlinear. According to Sayers et al. [12], spontaneous fluctuations in heart rate have been separated into three spectral bands: Very Low Frequencies (VLF: <0.05 Hz), Low Frequencies (LF: 0.05–0.15 Hz) and High Frequencies (HF: >0.15 Hz). The Power Spectral Density of a normal heart rate timeseries exhibits a continuous broadband spectrum with three distinct peaks superimposed, respectively, in the VLF, LF, and HF band. Chess et al. [13] found that the HF band is mediated entirely by the parasympathetic division of the central nervous system. Akselrod et al. [14] found vagal modulation and heart rate oscillations below 0.15 Hz and established the importance of the sympathetic division of the central nervous system in the generation of LF oscillations.

In general, it is expected that normal ECGs, as more irregular, will demonstrate more wide and overlapped broadband frequency peaks in contrast to abnormal ECGs, where the spectrum is expected to be thinner, exposing a more regular behavior.

2.1.2. Autocorrelation

The Autocorrelation function of a signal can identify and extract signal periodicities. Qualitatively, Autocorrelation compares a signal with itself with respect to time, thus providing information on whether this signal is periodic. This comparison is achieved by delaying its onset by \( \tau \), known as the time lag. Therefore the aim is to find how alike the signal \( X(t) \) is with the signal \( X(t - \tau) \).

In practice, this involves the computation of the Autocorrelation function, \( R(\tau) \), for various values of the time lag, \( \tau \), and the construction of the plot of \( R(\tau) \) vs \( \tau \). Details
of the mathematical description of the Autocorrelation function for a signal is included in Appendix A.1.

The Autocorrelation of a periodic signal is itself a periodic function, while the Autocorrelation of a random signal is a function which tends asymptotically to zero. Aside from the ability to discern periodicities in a signal, the Autocorrelation also provides an estimation of the decorrelation time, which is the time lag for which the Autocorrelation function begins to tend to zero. The decorrelation time produces valuable information that is used in nonlinear analysis of HRV, since it represents the moment in time beyond which the signal no longer exhibits periodicities. A random signal is expected to produce decorrelation times that are very close to zero.

2.1.3. Discrete Wavelet Transform

Wavelet analysis provides a means for simultaneously analyzing both the time and frequency characteristics of a signal. It is well-suited to ECG analysis since it analyzes signals at multiple scales and is not affected by the presence of discontinuities.

The Discrete Wavelet Transform (DWT) is a linear operation, which transforms a linear vector whose length is originally a power of two, to another vector of the same length [15].

Just as Fourier analysis breaks up a signal into sine waves of various frequencies, wavelet analysis breaks up a signal into various shifted and scaled version of the original signal, termed the mother wavelet. Scaling simply implies stretching or compressing the signal, where shifting denotes the delaying or hastening its onset. There are many wavelet families that have been used the simplest of which is the Haar wavelet, the mother function of which is simply a step function. Its description is found in Appendix A.2.

The Discrete Wavelet Transform is applied to the signal with the Haar wavelet as the mother function and the result is two sets of coefficients, the approximation coefficients, which are the high-scale (low frequency) signal components and the detail coefficients, which are the low-scale (high frequency) components of the signal. These sets of coefficients are then downsampled (dyadic decimation) that the resulting length of the signal, incorporating both the detail and approximation coefficients, is the same as the original signal length. Thus, the first scale of analysis is obtained. In order to compute subsequent scales, the same process is applied but this time only to the approximation coefficients of each respective previous scale. Essentially, the signal is convolved with a low-pass and a high-pass filter; the results of both filters are downsampled. The downsampled, low-passed output is then once again applied through both types of filters and downsampled, until all that remains (in the last scale) are detail coefficients. Finally, the standard deviation of the detail coefficients of each scale is computed and plotted.

2.2. Non-linear methods

2.2.1. The delay times method

The Delay Times method is an important tool in nonlinear analysis and gives both a qualitative and quantitative measure of the complexity of the timeseries under examination. It was first established by Grassberger and Procaccia [16] and is based on the Takens Theorem [17]. A timeseries is constructed from a set of successive and experimentally derived values. From the original timeseries, we then construct a new series, which in this case is composed of vectors. For the construction of each of the vectors the estimation of two parameters, the embedding dimension, \( m \), and the time lag, \( \tau \) is required. The time lag represents the window that is used for the computation of the coordinates of these vectors. It is estimated from the decorrelation time, which, as mentioned earlier, is the window beyond which the signal ceases to present periodicities. The decorrelation time is calculated either from the first zero-value of the Autocorrelation function, or from the first value of the mutual information function [18] that is close to zero. The mutual information function is a widely accepted method that computes nonlinear and linear correlation of a signal. The parameter \( m \) is assigned increasing integer values, in a range that satisfies both the Takens criterion and the maximum admitted window length, according to basic nonlinear dynamics theory. Appendix A.2 includes analytical information on these parameters, as well as a detailed description of the entire Delay Times method.

Once the above is completed, the correlation integral, \( C(r) \) is computed for increasing values of \( r \). This integral basically computes how many of the above vectors have a distance between them less than \( r \), where \( r \) is a ray in the vector space. We are then able to plot \( \ln(C) \) vs. \( \ln(r) \), where \( \ln \) is the natural logarithm function. From this plot, we select a scaling region and compute the slope of the curve in that region. This process is repeated for increasing values of the embedding dimension, \( m \), and if the values of the slopes converge, then we have found the Correlation Dimension of the timeseries \( X(i) \). The convergence value of the slope is an estimation of the Correlation Dimension. A timeseries that results from a complex nonlinear dynamic system, yields a larger value for the Correlation Dimension, as opposed to a timeseries which results from a regular and linear dynamic system, lower Correlation Dimension values. Generally, the Correlation Dimension represents the independent degrees of freedom that are required for the proper description of a system or for the construction of its model.

2.2.2. Approximate Entropy

There are various definitions of entropy, most of which usually arise from entropy computation such as the Shannon entropy or the Kolmogorov–Sinai entropy. From all known methods, Approximate Entropy
(ApEn) is chosen, since it has been introduced as a quantification of regularity in data and as the natural information parameter for an approximating Markov Chain to a process [11].

Given the original timeseries \(X(t)\), we construct a series of vectors, and then we find the heuristic estimation of an integer parameter, \(m\), which in this case represents a window size. We then, one again, heuristically estimate a threshold, \(r\), which arises from the product of the standard deviation of the time series and an arbitrary constant form 0 to 1, which is kept the same for all timeseries. We then apply an iterative procedure which finally produces an approximation of ApEn\((m, r)\).

Generally, random timeseries produce increasing values of ApEn\((m, r)\), compared to regular timeseries, a property which we exploit here. More details as well as a more analytical description of the method is included in Appendix A.4.

2.2.3. The SVD Karhunen–Loeve algorithm

The Singular Values Decomposition (SVD) Karhunen–Loeve method is also mentioned in the literature as Principal Component Analysis, Empirical Component Analysis [19], Singular Values Decomposition, or Empirical Eigenfunction Decomposition and is closely related to Factor Analysis, which is used in psychology and economics [19–22]. This method is used for the estimation of the complexity of a dynamic system.

Given the original timeseries, we once again construct a series of vectors similar to that of the Delay Times method, and estimate two parameters, \(m\) and \(\tau\), which are the embedding dimension and time lag, respectively. This series is made up of a two-dimensional array, which is termed \(\Lambda\). The covariance of this array is computed, serving a purpose much like the Autocorrelation of signals as discussed in the previous section. We then compute the eigenvalues of the Covariance array. In the absence of noise, the rank (i.e., the number of non-zero eigenvectors) of the covariance matrix is equal to the number of non-zero eigenvalues. The number of eigenvectors produces the smallest space dimension that is required to reconstruct the array \(\Lambda\). The presence of noise introduces infinitesimal values to the normally zero-valued eigenvalues. Therefore, this method is constrained to the use of large eigenvalues. Practically, it is expected that random timeseries will produce a larger number or eigenvalues compared to regular timeseries.

A more detailed description of the method is included in Appendix A.5.

2.3. Discriminating chaotic process – the method of surrogate data

The Surrogate Data algorithm [23] is used to ensure that data complicating dynamics come out of a complex dynamic process and not from randomness. According to this method, each original HRV data file is sorted based on its statistical priorities rather than its dynamics, so that the resulting signal does not contain any other dynamics thus yielding a random timeseries. In this way, new timeseries are constructed from the original timeseries and are called Surrogate Data.

The same nonlinear methods described above are applied to the Surrogate Data and if the same results are obtained, then the original data do not contain any dynamics and thus result from a stochastic process. On the other hand, if the results of the Surrogates are different from those of the original set, then the presence of system dynamics is confirmed and the data result from a chaotic process. The use of Surrogate Data is a statistical hypothesis testing method that involves two ingredients, a null hypothesis against which observations are tested, and a discriminating statistic. The null hypothesis attempts to describe the data while the discriminating statistic is a number that quantifies some aspects of the data. The discriminating statistic is a parameter that when used in combination with the Surrogate Data method may result in values that are different from those using the original data. If this occurs, then the null hypothesis is rejected.

3. Methods

This section describes the selection of subjects, data collection methodology, and details of the applications of the aforementioned methods. Subjects were selected by a cardiologist based on the medical record. The 10 Group A subjects were normal young males aged 25–29 yrs, with unremarkable medical histories and normal physical examinations. All subjects belonging to Group A were non-smokers, received no drugs and abstained from caffeine for 24 hours prior to acquisition. The 10 Group B subjects were patients hospitalised with coronary artery disease. All Group B subjects had 1 or 2 vessel coronary disease, which was angiographically confirmed, and normal left ventricular function (defined as an ejection fraction greater than or equal to 50%).

Subjects with a history of myocardial infarction, coronary angioplasty or bypass grafting, cardiac rhythm disturbances, left ventricular dysfunction (defined as an ejection fraction less than 50%), severe arterial hypertension, and medical conditions affecting heart rate variability (e.g., diabetes, mellitus, hormonal disturbances, treatment with psychotropic drugs, and respiration diseases) were excluded. All Group B subjects were undergoing treatment with nitrates, angiotensin converting enzyme inhibitors, salicylics and calcium antagonists (nifedipine 4 patients and nisoldipine 1 patient). No one had a history of stroke, peripheral vascular disease or clinically significant valvular abnormalities. Their ages were 42, 44 (4 pts), 49 (3 pts), 52 (2 pts) years old.
3.1. Data acquisition

To guarantee that valid and precise data were acquired, a cardiologist was present to ensure that all preparation and procedure details during electrocardiogram acquisition were followed properly. All recordings were performed in a quiet room, between the hours of 15.00 and 17.00, in the supine position under continuous monitoring by the cardiologist who confirmed the absence of any cardiac rhythm disturbances throughout the recording. Subjects were told to breathe normally and an attempt was made to maintain the respiratory rate at around 12/min.

A significant factor in ensuring the accuracy of the recordings is the technical characteristics of the electrocardiograph. In this study, a 12-lead digital electrocardiograph (“Cardioperfect”) with 8 channels and 12-bits quantization (satisfactory accuracy) was used. All recordings averaged at 5000–6000 R–R intervals. By R–R interval we mean the time elapsed between consecutive R-peaks in the QRS complex. Data were acquired with a sampling rate of 300 Hz. The acquired signal was sent to a PC via an RS232 port and stored on hard disk. Long ECG recordings were preferred for more accurate analysis and since large amounts of data ensure higher precision of the results; especially in nonlinear methods, the amount of data is a crucial issue. An alternative solution would have been to concatenate short recordings, but this would almost certainly produce inaccurate results.

By inclusion criteria, no subject, neither healthy nor unhealthy, had cardiac rhythm disturbances. Throughout ECG recordings, the procedure was closely supervised by a cardiologist who confirmed that no rhythm or conduction abnormalities had occurred. The appearance of sporadic isolated atrial complexes was addressed using the following procedure: The sporadic ectopic systoles were removed and replaced by the local-linear approximation of the previous five samples. However, this procedure had to be performed in only two of the coronary subjects, who presented rare atrial premature complexes (≤ 1/min).

The methods that we chose to use in this paper are multi-parameter and sensitive to parameter selection. The literature is still hazy on the identification of those parameters which are more suitable for his type of research. In this work, we present a set of parameter estimation approaches which, when applied, seems to optimize the experimental results. The methods that were applied to the acquired data are described in detail below.

3.2. Timeseries construction

Using QRS detection algorithms, the R points of the QRS complexes were extracted and the R–R interval signals were stored in data files, ready for analysis.

3.3. Proof of stationarity

Stationarity is an essential property of a timeseries since it ensures that all statistical quantities of a process are independent of absolute time and can be examined by splitting the timeseries into subsets and comparing critical statistics of these subsets with those of the entire set. The term critical statistics refers to a selection of magnitudes that describe basic statistical quantities of the series. These include the Fourier Power Spectrum, the mean, the standard deviation, etc. The issue of finding methods that examine timeseries stationarity is one of great discussion in relevant literature.

The stationarity of the data used in this study was determined in two ways. In the first, the Fourier spectrum of subsets of a timeseries was compared to that of the entire series and in the second, the $\chi^2$ test was applied.

3.4. Identifying nonlinearity

The nature of the signals as linear or nonlinear is identified with the application of linear methods such as the Fourier Transform and the Autocorrelation function. Proof of linearity/nonlinearity in the signal helps determine whether linear or nonlinear signal analysis methods would best describe system behavior. The application of the Fourier Transform produces a wide spectrum of overlapped frequencies, proving the existence of nonlinear or stochastic dynamics in the signal. The Autocorrelation function also reveals nonlinearity or stochasticity in HRV signals since, after a small time lag, it tends to zero.

3.5. Validity of acquired data

The validity of our acquired data was established through the application of the Fourier Transform and the Autocorrelation functions. The data sets were analyzed using linear methods both in the time and frequency domains in order to draw general conclusions about the condition of each subject. The Fast Fourier Transform was applied to every timeseries and the Power Spectrum and periodogram were calculated. The Autocorrelation function was computed for time lags up to 500 of the signals in order to find domain periodicities and decorrelation time lags. The results obtained using our data sets proved analogous to those of similar studies. This ensures that the cardiograms we acquired are equivalent to those used by other researchers.

3.6. The discrete wavelet transform

In this study, an improvement of the known DWT method was used, as described in [2]. This new method led to a clearer separation between the two subject
groups over a larger number of scales of analysis and includes a new way of evaluating the detail coefficients and enhances the diagnostic accuracy. It introduces the use of the mean length for the interval analysis, which for two dimensions is defined as:

\[ L = \frac{1}{n-1} \sum_{i=1}^{n-1} |w_{i+1} - w_i|, \]

where \( n \) is the total number of R–R intervals analyzed and \( w \) is the value of each R–R interval.

In the calculation of the mean length, neighboring wavelet coefficients are computed from neighboring samples or sample areas. This results in higher accuracy, since the standard deviation does not take advantage of the order of the wavelet coefficients, as opposed to the mean length, which does.

3.7. Application of the delay times method

Through the use of the Autocorrelation and mutual information functions, a value was determined for the time lag, \( \tau \), that is most suitable for this study and that was \( \tau = 5 \). The dimensions chosen for the phase space reconstruction started at \( m = 3 \) and went to \( m = 20 \), based on the fact that after several trials these values yield the best reconstruction and thus lead to more accurate results and subject discrimination.

The correlation integral was then calculated for an extended range of \( r \) (up to \( 10^8 \), experimentally determined). This correlation integral was used to estimate the Correlation Dimension values.

3.8. Approximate Entropy

The Approximate Entropy was computed for a variety of \( r \)-values proposed by previous researchers and it was found that the optimum value yielding clearest discrimination was \( r = 0.65\text{STD} \), where STD is the standard deviation of the timeseries.

3.9. Singular values decomposition

The SVD algorithm was applied to the data sets and the largest eight eigenvalues were found for each set.

3.10. Surrogate data

Surrogate Data sets were produced from respective timeseries. Three different techniques were applied to all signals: the Unwindowed Fourier Transform Algorithm, the Windowed Fourier Transform, and the Amplitude Adjusted Fourier Transform [11]. All of the above methods were then applied to the Surrogate Data, in order to ascertain whether cardiac function is the result of a random mechanism. Surrogate time series were produced for each HRV data file both of Group A and of Group B. The nonlinear methods were applied to these Surrogate sets and a mean value was computed for each of the parameters to be compared (Correlation Dimension and Approximate Entropy). It is this average that is illustrated in the figures so that it may be compared to the original (that of the pure data sets).

4. Results

4.1. Fourier transform

Fig. 1 illustrates a typical spectrum shape of a Group A subject, while Fig. 2 depicts that of a Group B subject. In the first case, the three expected bands of frequencies are easily identified. Careful observation of these two figures indicates that in the case of the Group B subject, the high frequencies are missing. Most of the ECG files belonging to Group B demonstrated similar spectrum characteristics. However, this form is not imperative for all cardiac diseases and is not always clear as a discrimination mechanism. Indeed there are some data sets from Group B where this discrimination was not as clear. Issues of linearity and nonlinearity are addressed in the next paragraph.

4.2. Autocorrelation

Fig. 3 depicts the Autocorrelation function for two data sets, one of a Group A and one of an Group B subject. This figure confirms that the Autocorrelation of signals belonging to Group A steeply decreases and tends to zero for short decorrelation times. In contrast,
the Autocorrelation of the Group B subject ECGs is more regular and tends to zero much more slowly, demonstrating increased decorrelation time lags. The decorrelation time can be quite large, reaching up to 25% of the total length of the timeseries. This length serves as the upper limit for a satisfactory estimation of the Autocorrelation function. Both methods indicate strong nonlinearity in Group A data sets, contrary to the ones of Group B, where occasional periodicities are found. Generally, the spectrum of a linear signal contains isolated narrow bands and its Autocorrelation does not tend to zero for short time lags. The decorrelation time shown in the graphs resulting from the application of the Autocorrelation function varies depending on possible cardiac episodes that may have occurred during recordings. There are also cases where occasionally noisy recordings distort the figures, giving them the appearance of stochastic signals. An inspection of the results of the Fourier spectrums shows that they exhibited a spread spectrum of overlapped frequencies without any domain periodicities. In addition to this, the Autocorrelation function plots indicate a steep descent. The above outcomes denote nonlinear or random behavior, thus a nonlinear analysis is required.

4.3. Discrete Wavelet Transform

Fig. 4 presents the results for the application of the DWT using the mean length as a measure of variability. As shown in [2], the mean length shows a clear separation between the two groups over nine scales, as opposed to the more common standard deviation, in which a clear separation between the two groups occurs at scales four and five.

As shown in Fig. 4, the differentiation between the two groups begins with the first scale and is very clear up to the ninth scale and even at the tenth. Similar figures result with the use of either of the Haar or Daubechies wavelet families. The separation between the two groups is very distinct, occurs over nine scales for both families, and begins at the first, as opposed to that provided by the standard deviation. The results of this method are illustrated in Table 1. The last column of the table shows the $t$-test statistical error probability, which represents the probability of a false categorization.
4.4. Delay Times

The distinction between the two groups of subjects is clear, as shown in Fig. 5. The Group A (normal) signals have higher Correlation Dimensions depicting a more complicated dynamic system with more degrees of freedom than the ones of Group B. This is expected, according to our initial hypothesis. The dynamic system of the subjects of Group B tends to behave more normally and linearly, represented by models with 4–5 degrees of freedom for non-severe cases or 3 to 4 for worse ones. In contrast, the Group A signals depict a dynamic system with strong nonlinearity, which requires 6 to 7 degrees of freedom for describing cardiac behavior. The mean values for the estimated dimensions are: The mean Correlation Dimension, \( D_2 = 7.865686 \) for Group A subjects with standard deviation of 0.480169 and for Group B subjects, \( D_2 = 4.657146 \) and \( \text{STD} = 0.827082 \).

The results of the Delay Times method are depicted in Table 2. Once again, the last row of the table shows the \( t \)-test statistical error probability, which represents the probability of a false categorization.

Previous work that uses the Delay Times method can be categorized into approaches that analyze only the R–R intervals and those that examine the entire ECG. Included in the latter is the work of Bezerianos et. al. [24], which claims that \( 2 < D_2 < 4 \) for subjects such as those belonging to Group A and Bogaert et. al. [25], which suggests slightly higher values of \( D_2 \approx 5.78 \). The research activities that belong to the first category are summarized in Table 3. The values of \( D_2 \) shown in Table 3 range from 3.6 to 9.62, which indicates the necessity of further investigation of the problem. Our contribution is a vote to those that believe in high Correlation Dimension values, especially for healthy subjects. A physical explanation for the expected high Correlation Dimension was not provided in the text, which suggests that this might be due to the complexity of cardiac function or other factors.

<table>
<thead>
<tr>
<th>Disease present</th>
<th>Delay Times Method Mean</th>
<th>STD</th>
<th>Approximate Entropy Mean</th>
<th>STD</th>
<th>Singular Values Decomposition Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>4.657146</td>
<td>0.827082</td>
<td>0.241085</td>
<td>0.091374</td>
<td>7.7176</td>
<td>3.35302</td>
</tr>
<tr>
<td>No</td>
<td>7.865686</td>
<td>0.480169</td>
<td>1.220074</td>
<td>0.156503</td>
<td>10.4457</td>
<td>1.914205</td>
</tr>
<tr>
<td>Statistical error probability</td>
<td>( 1.94 \times 10^{-8} )</td>
<td>( 9.11 \times 10^{-12} )</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dimension is the high complexity of the cardiac system whose function is affected by many different and independent factors including sympathetic-vagal balance, respiration, hormones, blood pressure, posture and chemicals. The large number of contributing factors clearly leads to a highly complex system.

4.5. Approximate Entropy

The results of applying the ApEn to the data sets are presented in Fig. 6. The discrimination between the two categories is evident. The Group A set presents increased values for the ApEn as opposed to the decreased values for the Group B set. As mentioned, the best distinction is obtained with \( r = 0.65 \text{STD} \). Over this threshold the mean Approximate Entropy Value for Group A is 1.22e-074, where for Group B it is 0.241085. For ApEn < 1, there is evidence of regularity in the Group B subject data set. This method produces the clearest discrimination between the two categories as compared to the other nonlinear methods discussed thus far. The comparative results of the application of this method are shown in Table 4.

If this application is based on the Pincus [11] proposition, the Group A has an Approximate Entropy value of 2.730843 with a standard deviation of 0.30942 and the Group B has a mean Approximate Entropy value of 0.721386 with a standard deviation of 0.237473. In previous publications [26] the range of values starts at 0.8 and goes up to 1.2 for subjects with atrial fibrillation episodes, and in [11] the ApEn is found to be 0.742 for aborted Sudden Infant Death Syndrome (SIDS) infants and 1.457 for normal SIDS infants.

Table 3
Delay Times method

<table>
<thead>
<tr>
<th></th>
<th>Number of subjects</th>
<th>Sampling points</th>
<th>Duration</th>
<th>Correlation Dimension ( D_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ganz and Lenz (1996)</td>
<td>79</td>
<td>1024</td>
<td>24h</td>
<td>5.37–9.62</td>
</tr>
<tr>
<td>Guzzetti et al. (1996)</td>
<td>7</td>
<td>7</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Otsuka et al. (1997)</td>
<td>116</td>
<td>4h</td>
<td></td>
<td>5.52–7.49</td>
</tr>
<tr>
<td>Nashoni et al. (1998)</td>
<td>10</td>
<td>20 min</td>
<td></td>
<td>3.29–5.16</td>
</tr>
<tr>
<td>Babloyantz and Destexhe [27]</td>
<td>4</td>
<td>4 min</td>
<td></td>
<td>3.6–5.2</td>
</tr>
<tr>
<td>Fojt and Holcik (1998)</td>
<td>7</td>
<td>15,000</td>
<td></td>
<td>5.3–7.6</td>
</tr>
<tr>
<td>Our suggestion</td>
<td>10</td>
<td>6000</td>
<td>20 min</td>
<td>7.87</td>
</tr>
</tbody>
</table>

Our Delay Times method results, along with those of previous researches.

Table 4
Approximate Entropy

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Approximate Entropy</th>
<th>Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pincus 1991</td>
<td>Healthy</td>
<td>2.730843</td>
<td>0.30942</td>
</tr>
<tr>
<td>Pincus 1991</td>
<td>patient</td>
<td>0.721386</td>
<td>0.237473</td>
</tr>
<tr>
<td>Pincus et. al. 1993</td>
<td>normal SIDS infants</td>
<td>0.742</td>
<td></td>
</tr>
<tr>
<td>Pincus et. al. 1993</td>
<td>aborted SIDS infants</td>
<td>1.457</td>
<td></td>
</tr>
<tr>
<td>Makikallio et al. 2000</td>
<td>atrial fibrillation episodes</td>
<td>0.8–1.2</td>
<td></td>
</tr>
<tr>
<td>Our suggestion</td>
<td>Group A</td>
<td>1.220074</td>
<td>0.156503</td>
</tr>
<tr>
<td>Our suggestion</td>
<td>Group B</td>
<td>0.241085</td>
<td>0.091374</td>
</tr>
</tbody>
</table>

Our Approximate Entropy method results, along with those of previous researches.
4.6. Singular values decomposition

The results for the SVD method are depicted in Fig. 7. The respective values of the results shown in the figure are depicted in Table 2, which were found by taking the sum of the three larger eigenvalues as a measure. In Fig. 7, overlapping between the healthy and unhealthy subject results is apparent, thus deeming the SVD method as non-satisfactory. In addition, the standard deviation of the values shown in Table 2 is too large and confuses the SVD outcome.

4.7. Surrogate data

The Surrogate Data analysis was done using all methods. Fig. 8 presents only the results of the Correlation Dimension estimation, since it is sufficient enough to discover any underlying dynamics. It is important to note that each value that is presented with respect to each Surrogate Data set is an average of closely computed values which in turn corresponds to an original HRV data set. It would be unnecessarily redundant to include all data set outcomes, since this illustration is representative of the kind of results obtained from all the methods.

Fig. 8 clearly depicts an increase in the Correlation Dimension values for the Surrogate Data sets compared to those of the original signals. This increase declares the presence of underlying dynamics in the data sets. If the data were a noise bounded or purely stochastic processes, then the Correlation Dimension estimation would not have been increased, preserving all statistical properties, as stated in the null hypothesis. Therefore the application of the above nonlinear methods to the Surrogate Data classifies the data sets correctly and indicates that the signals do not come from a stochastic process.

5. Discussion

The most widely used method of those discussed above is the Fourier Transform, which has even been used widely in commercial systems. The Fourier Transform is a linear method and fails to extract the nonlinear dynamics of the heart or to provide any information on the nature of these dynamics. It is significantly affected by noise and does not provide clear quantitative categorization of the heart rate signals.

The Autocorrelation function is accepted as better suited for drawing conclusions in medical applications but also fails in the accurate classification of the subjects of this study. The Autocorrelation function and the Fourier Transform are related through the Weiner-Khintchine theorem, which states that the spectral density function of a signal can be expressed as the Fourier Transform of the Autocorrelation function. Therefore, information extracted from the signal using the Fourier Transform is coherent to that obtained from the Autocorrelation function, with the difference that it is in the time domain. Thus, both methods may be used in characterizing the signals for nonlinearity, but they cannot be used alone for discerning stochastic or noisy signals from chaotic ones.

Wavelet analysis is also a linear method. It performs analysis in the time domain and clearly separates the two groups of subjects. Contrary to the Fourier Transform, it allows quantitative classification of electrocardiograms.
and seems to provide additional medical information, something that has not been satisfactorily investigated yet. When noise-free electrocardiograms are used, it succeeds in discriminating between the two groups of subjects with extremely high accuracy.

The Delay Times method clearly distinguishes Group A from Group B and accurately describes the system's stochasticity in conjunction with the Surrogate Data method. Since it is a nonlinear dynamics method, it provides accurate information about the signal's nature and its underlying dynamics. Indeed, in this study, a high dimensional underlying dynamic system was obtained for Group A, as opposed to the relatively low dimensional system of Group B.

One point must be made clear, regarding the Delay Times method. From its introduction and initial application in 1987 by Babloyantz and Destexhe [27] and Goldberger [28] until its last appearance in bibliography [25,29], an apparent battle has broken out over whether the cardiac dynamic system demonstrates the behavior of chaotic dynamics or if it is completely stochastic. Even those who denounce the methods discussed here [29] are careful and thus they rule out only low dimensional chaos in cardiac dynamics. Indeed it is difficult to compute a precise Correlation Dimension and it is questionable if this would be very beneficial. The value of the application of this method lies precisely in the fact that it produces a very clear differentiation between the two subject groups. The Correlation Dimension 7.5309 does not yield as much information since it is only an estimation, but it does provide a clear indication of high dimensional dynamics. Another formulation for high dimensional dynamics is complex dynamics, where complexity does not necessarily imply chaotic. All of the above arguments lead us to the use of complexity for healthy subjects and the strong discrimination it provides. Non-linear methods are undoubtedly sensitive, but when combined with other methods, they yield useful results.

Approximate Entropy is even more accurate that all those described and is also computationally simple. Since the ApEn is a measure of regularity rather than a magnitude statistic, its use is most effective when applied in conjunction with other methods or statistics and not as a sole indicator of system characteristics. For example, if the system contains large amounts of noise, then the ApEn will show an increase in irregularity, which is an inaccurate dynamics characterization in this case. This erroneous outcome can be corrected by using the ApEn in combination with other methods, such as the Delay Times method, and reinforcing the results with Surrogate Data. Therefore, the use of ApEn is most valuable when it is used along with other methods and statistics.

Singular Values Decomposition lacks accuracy, since the discrimination it provides is not so clear. However, it does allow for physical conclusions. Indeed, this method gives information about the systems state space eigenvalues in relation to a time delay domain analysis.

All of the above indicate the necessity of using more than one of these methods for classifying the electrocardiograms with high accuracy. A complementary study and experimental analysis of all presented methods allow for more accurate discrimination and safer conclusions.

It has been shown that the Wavelet Transform, they Delay Times method and the Approximate Entropy successfully provide a clear separation. It was therefore of interest to group the results obtained from these methods and apply them in a way that would produce a procedure useful in the clinical setting. The conventional approach is to introduce an index, using simple algebraic variable equations, which would result from the application of these three methods. This approach is outdated and would also require a very large number of samples for a proper statistical study, the results of which would be used to construct the index. On the other hand, the grouping of samples and their characteristics belong to a different logic than that of the conventional approaches. A new, non-conventional scheme for the evaluation of indices, based on more current methods such as fuzzy logic or neural networks, is quite possibly a better and simpler approach that the algebraic evaluation of method parameters. Therefore, we used the simpler insertion of method results to an Learning Vector Quantization (known as the LVQ) neural network.

Training of this LVQ network was performed using heuristic values that were chosen based on the inspection of the results of the experimental analysis that were deemed difficult to classify. These were apen = [0.8, 2.1, 3, 1.1, 1.2, 0.3, 0.6, 0.7, 0.3, 0.75] with corresponding Delay Times method = [8, 9, 10, 11, 12, 3, 5, 4.7, 4, 3.3] and wavelets = [38, 36, 37, 36.6, 35.6, 34, 31, 33.7, 33.8, 34.2]. The first five values belonged to Group A while the last five values to Group B. Once the network was trained, the results of our methods were given to it as input. In every case, the network made a correct classification, yielding 100% success. We also tried using fictional values as input, some of which were fuzzy. Once again, the results were satisfactory, and these values were also correctly classified. The LVQ network proves successful in its use as a simple evaluator of the results of our methods.

6. Conclusions

We have studied several methods which have been used for the categorization of two subject groups, one which represents subjects with no prior occurrence of coronary disease events and another group who have
had a coronary disease event but who’s ECGs appear normal.

The Fourier Transform and the Autocorrelation function are linear methods and thus are fail to extract nonlinear signal dynamics, which expose much information that can be quite useful in signal processing applications. Furthermore, these two methods are sensitive to noise and thus cannot provide an accurate and clear quantitative categorization between healthy and unhealthy subject cardiograms, especially when the ECGs of the latter group do not present any clear signs of cardiac pathology.

Contrary to the Fourier and Autocorrelation functions, it has been shown that wavelet analysis, the Delay Times method and the computation of the Approximate Entropy present coherent results and succeed in clearly and accurately differentiating healthy subject ECGs from those of unhealthy subjects and coronary patients. The methods discussed in this paper are applied in order to construct an automated tool utilizing an LVQ neural network that is useful and reliable in clinical settings.

Acknowledgments

We would like to thank Dr Poralis for his permission to use the Airforce and Veterans Hospital of Athens cardiology clinic, cardiogram data and for always being willing and enlightening in aiding our research. We would also like to thank Professors Papakonstantinou and Hizanidis for their invaluable guidance and support.

Appendix A

A.1. Autocorrelation function

The Autocorrelation function is primarily a result of the correlation coefficient. In general, the correlation coefficient of two arbitrary signals, \(X\) and \(Y\), is defined as

\[
r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}
\]

(A.1)

where \(\text{Cov}(X, Y)\) is the covariance of the signal \(X\) with \(Y\), and \(\text{Var}(X), \text{Var}(Y)\) is the variance of the signals \(X\) and \(Y\), respectively.

Autocorrelation coefficient \(R\) is an adaptation of \(r\), which considers the signal \(Y\) as a time delay of the signal \(X\), so that \(Y(i) = X(i + \tau)\), where \(\tau\) is the time lag and \(i\) is the time for which the signal \(X\) or \(Y\) is considered. Here, \(\tau_{\text{lag}} = \tau \cdot \tau_s\), where \(\tau_s\) is the sampling time for the timeseries. After the timeseries has been extracted, \(\tau\) is the number of points in the window \((i + \tau)\). Due to the discrete nature of this signal, it suffices to use only \(\tau\).

If \(X\) is a discrete timeseries of size \(N\), the Autocorrelation function is computed as

\[
r(\tau) = \frac{\sum_{i=1}^{N-\tau} X_i X_{i+\tau} - \frac{1}{N} \left( \sum_{i=1}^{N} X_i \right) \left( \sum_{i=1}^{N} X_{i+\tau} \right)}{\sqrt{\left( \sum_{i=1}^{N} X_i^2 - \frac{1}{N} \left( \sum_{i=1}^{N} X_i \right)^2 \right)^{\frac{1}{2}} \sqrt{\left( \sum_{i=1}^{N} X_{i+\tau}^2 - \frac{1}{N} \left( \sum_{i=1}^{N} X_{i+\tau} \right)^2 \right)^{\frac{1}{2}}}}}
\]

(A.2)

where \(X_i = X(i)\) for \(i = 1, \ldots, N\) for and \(X_{i+\tau} = X(i + \tau)\) for \(\tau \in N\).

A.2. Wavelets

The wavelet application on which this section is based is the Discrete Wavelet Transform (DWT). The DWT is a linear operation, which transforms a linear vector whose length is originally a power of 2, to another vector of the same length \([1,2]\). Given a signal \(X\) of length \(n\) (where \(n\) must be a power of 2), the DWT consists of at most \(\log_2 n\) stages \((\log_2 n - 1, \text{ if } N \text{ is not a power of } 2)\). The first step produces, starting from \(X\), two sets of coefficients, the approximation coefficients (scaling coefficients) and the detail coefficients (wavelet coefficients). The vectors are obtained by convolving \(\tau\) with a low-pass filter for approximation and a high-pass filter for detail, followed by dyadic decimation. This process is applied repeatedly only to the resulting approximation coefficient vector, and then continues until the \(\log_2 n\)th stage. The nature of the filters is dependent upon the family of wavelets used.

For every signal processed, it is the scale-dependent statistics of the detail coefficients that provide the graphical diagnostic output. Specifically, once the detail coefficient vector has been computed at each scale, its standard deviation is calculated. The same is repeated for every scale, until the \(\log_2 n\)th scale is reached. It is important to note that with decrease in scale there is subsequent increase in detail obtained from the Wavelet Transform. The standard deviation vectors are displayed on the same graph so that separation may be discerned.

In this case, the detail coefficients of the Wavelet Transform are calculated as in (A.3), while the approximation coefficients may be calculated with (A.4),

\[
\psi_{m,n}^{\text{wav}} = 2^{-(m/2)} \sum_{i=0}^{M-1} X_i \psi_m(i - n2^m),
\]

(A.3)

\[
S_{m,n}^{\text{wav}} = 2^{-(m/2)} \sum_{i=0}^{M-1} X_i \phi_m(i - n2^m).
\]

(A.4)

In both equations, \(m\) is the scale variable \((m = 1, 2, \ldots)\) and \(n\) is the shift variable \((n = \ldots, -1, 0, 1, \ldots)\). In (A.3), \(\psi_m(i)\) is the discrete-time mother wavelet for the \(m\)th scale in the calculation of the detail coefficients and in (A.4), \(\phi_m(i)\) is that for the calculation of the
approximation coefficients. \( M \) is the total number of samples being analyzed. The \( m \)-scale DWT can be obtained by applying the one-scale DWT \( m \) times subsequently to the approximation components. The coefficients that are statistically manipulated and graphed are the detail coefficients. The choice of wavelet family applied to the signal through the DWT is usually dependent upon the application. We chose the simplest family, the Haar, since it is most efficient and has been proven as sufficient for the reliable classification of healthy and unhealthy subjects [2].

The discrete-time Haar wavelet is defined as in (A.5) and the discrete-time scaling function as in (A.6),

\[
\psi(i) = \begin{cases} 
1 & \text{for } i = 0, \\
-1 & \text{for } i = 0, \\
0 & \text{otherwise},
\end{cases} 
\]

(\text{A.5})

\[
\phi(i) = \begin{cases} 
1, & i = 0, 1, \\
0, & \text{otherwise}.
\end{cases} 
\]

(A.6)

Therefore, the discrete time sequence obtained from the original signal is transformed into a space of wavelet coefficients resulting in the creation of scale-dependent statistics. Smaller scales correspond to more rapid variations and therefore to higher frequencies [2], as discussed in the analogous section of the paper.

### A.3. Delay Times method

Given a timeseries, \( X(t), t \) is an integer, \( t \in (1, N) \) and \( N \) is the total number of timeseries points. The Delay Times method was first established by Grassberger and Procaccia [16] and based on the Takens Theorem [17]. According to this method, the timeseries \( x(t) \) is a measure of a single coordinate of an \( m \)-dimensional system’s underlying dynamics. Assuming \( m \) is the embedding dimension (the dimension of space in which the assumed system’s trajectory is unfolded) and \( \tau \) is the time lag, then phase space reconstruction (described below) is performed with time delays and the following \( m \)-dimensional vectors are constructed:

\[
\mathbf{x}(t) = [X(t), X(t+\tau), X(t+2\tau), \ldots, X(t+(m-1)\tau)].
\]

(A.7)

In this way, using the original timeseries, \( X(t) \), we are able to construct a new vector timeseries, \( \mathbf{x}(t) \), which represents the trajectory from \( \mathbf{x}(0) \) up to and including \( \mathbf{x}(t) \) within the reconstructed phase space.

These vectors are defined in an \( m \)-dimensional phase space and are used in constructing the trajectory of the signal dynamics to this space. If the original phase space of the dynamics produce the attractor \( A \), then the reconstruction of the phase space with the Delay Times method produces the reconstructed attractor, \( A' \). If the reconstruction is accurate, then \( A' \) is the topological conjugate of the original attractor, \( A \). Consequently, all dynamic properties of \( A \) are projected to \( A' \). The criterion of the Takens Theorem [17,30] for a precise phase space reconstruction of an experimental trajectory dictates that \( m \) must be greater then \( [2mc + 1] \), where \( mc \) is the estimated dimension of the attractor.

According to the Takens Theorem, this is efficient when the number of points of the timeseries, \( N \), is infinite, meaning that for an infinite number of points, \( A \) and \( A' \) have the same properties. However, for most experimental methods, \( N \) is a finite number and in many cases is confined to 3000–4000 points. Therefore, only \( A' \) is estimated in the reconstructed space and retains only some of the properties of \( A \) (not all). Essential to phase space reconstruction, especially for the Delay Times method, is the estimation of the time lag, \( \tau \). There is a range of methods for estimating \( \tau \), the most popular being the calculation of the decorrelation time.

The decorrelation time is calculated either from the first zero-value of the Autocorrelation function, or from the first minimum value of the mutual information function. The Autocorrelation function has been described in the previous section. The mutual information method is widely accepted and it computes the nonlinear and linear correlation of the Autocorrelation function. Once the Autocorrelation function has been normalized, the decorrelation time is found from the smallest time lag for which the function tends to zero. Similarly, the decorrelation time can also be found from the smallest time lag for which the mutual information function tends to zero.

According to [31–34] the results of a timeseries analysis depends on the window length \((m-1)\tau\), which incorporates both the embedding dimension \( m \) and the time lag \( \tau \). Therefore, the constraint to the above methods is the limit on the size of the window, \((m-1)\tau\). A proper value for the window size provides good phase space reconstruction and ensures that all the points of the reconstructed phase space come from the same trajectory. As mentioned above, the Takens theorem dictates that proper phase space reconstruction is achieved when \( m \) is greater than \( [2mc + 1] \). This criterion is difficult to satisfy for increased values of \( \tau \) due to the subsequently larger values of \((m-1)\tau\). A consistent window arises from the decorrelation time, seen as the time needed for the first decay of the Autocorrelation function. A time lag, \( \tau \), is chosen and the reconstructed dynamics are embedded in the \( m \)-dimensional phase space.

After the phase space reconstruction of the system’s assumed dynamics, nonlinear dynamics algorithms are developed for the experimental analysis of a timeseries. The most popular algorithmic method is the Delay Times method, also known as the Algorithm of Grassberger and Procaccia [16], which estimates the Correlation Dimension from the computation of the correlation integral.
The Grassberger–Procaccia Algorithm [16] assumes a timeseries, $X(i)$, which is a measure over time $i$ of a parameter of an $m$-dimensional dynamic system, for $i$ in [1]. The phase space reconstruction of this system is done according to the Takens theorem. Once again, the vector coordinates are constructed as in (A.7) and it is assumed that this vector is the trajectory vector of the $ith$ time point of the reconstructed phase space of the dynamic system. The whole trajectory is $\tilde{x}(1), \tilde{x}(2), \ldots, \tilde{x}(i), \ldots \tilde{x}(\rho)$ where $\rho = N - (m - 1)\tau$. As mentioned and according to the Takens theorem, $A'$ is the attractor to the reconstructed system dynamics and the topological conjugate to the original attractor $A$. Properties such as the Correlation Dimension are maintained after the projection of $A$ to $A'$. The Correlation Dimension is defined as
\[
D_2 = \lim_{N \to \infty} \frac{\log(C(m, r, \tau))}{\log(r)},
\]
where $r$ is a distance radius in the reconstructed phase space. The index 2 in $D_2$ is used because the Correlation Dimension is a special case of the generalized dimension $D_q$ where $q$ integer. $C(m, r, \tau)$ is the correlation integral and is defined as
\[
C(m, r, \tau) = \frac{2}{N - 1} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta\left[ r - \| \tilde{x}_i - \tilde{x}_j \| \right],
\]
where $\tilde{x}_i$ and $\tilde{x}_j$ are as in (A.7). $\Theta$ is the Heavyside function:
\[
\Theta(i) = \begin{cases} 
1 & \text{if } i \geq 0, \\
0 & \text{if } i < 0.
\end{cases}
\]
The Euclidean norm used in the above equation states that the difference between $\tilde{x}_i$ and $\tilde{x}_j$ is the maximum difference among their coordinates:
\[
\| \tilde{x}_i - \tilde{x}_j \| = \left\{ \begin{array}{l}
|X(i) - X(j)|^2 + |X(i + \tau) - X(j + \tau)|^2 \\
+ \cdots + |X(i + (m - 1)\tau) - X(j + (m - 1)\tau)|^2
\end{array} \right\}^{1/2}.
\]
The formula (A.9) simply says: for specific $m, r, \tau$ find all pairs of $\tilde{x}_i$ and $\tilde{x}_j$ in the reconstructed timeseries $\tilde{x}(i)$ for which the distance $\| \tilde{x}_i - \tilde{x}_j \|$ is smaller than $r$.

According to this algorithm, if $C(m, r, \tau)$ convergences, then increasing the dimension of the phase space reconstruction, $m$, results in a scaling region for $\log(C(m, r, \tau))$ versus $\log(r)$. Therefore, the scaling region is found from
\[
\frac{\log(C(m, r, \tau))}{\log(r)} \approx a \Rightarrow C \approx r^a,
\]
where $a$ is an arbitrary constant real value, indicating the linear connection of $\log(C)$ vs $\log(r)$ inside a hypothesized scaling region of the $\log(C)$ vs $\log(r)$ plot.

This means that $D_2 \approx a$, indicating an estimation of the Correlation Dimension of $A'$ and thus of the original system dynamics attractor, $A$.

### A.4. Approximate Entropy

Given $N$ data points, $X(1), X(2), X(3), \ldots, X(N)$, the ApEn($m, r, N$) is estimated, where $r$ is a threshold and $m$ a window size. The vector sequences necessary for phase space reconstruction, $\tilde{x}(i)$, are constructed with $\tilde{x}(N - m + 1)$, defined by $\tilde{x}(i) = [X(i), \ldots, X(i + m - 1)]$. These vectors represent $m$ consecutive X values, using the $ith$ point as the starting point. The distance $\| \tilde{x}(i), \tilde{x}(j) \|$ is defined between the vectors $\tilde{x}(i)$ and $\tilde{x}(j)$ as the infinity norm
\[
\| X(i) - X(j) \| = \max \{|X(i) - X(j)|, |X(i + 1) - X(j + 1)|, \ldots, |X(i + m - 1) - X(j + m - 1)| \}.
\]
The probability that $|X(i + m - 1) - X(j + m - 1)| \leq r$ given that $|X(i) - X(j)| \leq r$ and $|X(i + 1) - X(j + 1)| \leq r$ and $|X(i + 2) - X(j + 2)| \leq r$ and – is true is termed $C^m_{\#}(i)$, where, once again, $r$ is the a threshold and $m$ the window size. For example, if $m = 2$, $C^2_{\#}(i)$ for $i = 1, \ldots, N$ is the probability that $|X(i + 1) - X(j + 1)| \leq r$ given that $|X(i) - X(j)| \leq r$.

The sequence in (A.13) is used to construct the $C^m_{\#}(r)$ for each $i \leq N - m + 1$ as in
\[
C^m_{\#}(i) = \left\{ \begin{array}{l}
\text{no. of } j \leq N - m + 1, \text{ such that } \| \tilde{x}(i) - \tilde{x}(j) \| \leq r \\
N - m + 1
\end{array} \right\}.
\]
$\Phi^m(r)$ is defined as
\[
\Phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C^m_{\#}(r),
\]
where $\ln$ is the natural logarithm.

Then Approximate Entropy is defined as
\[
\text{ApEn} = \lim_{N \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)].
\]
It is therefore found that $-\text{ApEn} = \Phi^{m+1}(r) - \Phi^m(r)$ and is equal to the average over $i$ of the natural log of the conditional probability that $|X(j + m) - X(i + m)| \leq r$, given that $|X(j + k) - X(i + k)| \leq r$, for $k = 0, 1, 2, \ldots, m - 1$.

Several trials of this algorithm were run on the HRV data and it was adjusted accordingly in order to obtain a better distinction between the two subject groups. The first step in computing the Approximate Entropy is finding the length vector for $m = 2$, which is $[X(i), X(i + 1)]$, denoted $\tilde{x}(i)$. All vectors that are close to $\tilde{x}(i)$, $\tilde{x}(j) = [X(j), X(j + 1)]$, are identified. As has already been stated, the vector $\tilde{x}(j)$ is close to $\tilde{x}(i)$ if $\| \tilde{x}(i), \tilde{x}(j) \| \leq r$. This, by definition, means that both
$|X(i) - X(j)| \leq r$ and $|X(i + 1) - X(j + 1)| \leq r$ apply. A count of all the vectors $\bar{x}(j)$ close to $\bar{x}(i)$ is found and called $B$. The next step is to compute the rest of the $\bar{x}(j)$ vectors for which $|X(i + 2) - X(j + 2)| \leq r$, and call it $A$. The ratio of $A/B$ represents the conditional probability that $X(j + 2)$ is close to $X(i + 2)$, given that the vector $\bar{x}(j)$ is close to $\bar{x}(i)$.

The above process is repeated for each length 2 vector $\bar{x}(i)$, calculating the conditional probability. The ApEn is found by calculating the average of the logarithm of these conditional probabilities and taking its negative (to make it positive), as seen in (A.17)

$$-\text{ApEn} = \Phi^m_r - \Phi^{m+1}_r$$

$$= \left[ \frac{1}{N-m} \sum_{i=1}^{N-m} \ln \left( C^m_r(i) \right) \right]$$

$$- \left[ \frac{1}{N-m} \sum_{i=1}^{N-m} \ln \left( C^{m+1}_r(i) \right) \right]$$

$$\approx \frac{1}{N-m} \sum_{i=1}^{N-m} \left[ \ln \left( C^{m+1}_r(i) \right) - \ln \left( C^m_r(i) \right) \right]$$

$$= \frac{1}{N-m} \sum_{i=1}^{N-m} \ln \left( \frac{C^{m+1}_r(i)}{C^m_r(i)} \right).$$

The calculation of the conditional probabilities will result in values between 0 and 1. If the timeseries is regular, the values $X(i), X(i+1), X(i+2)$ are expected to be close to each other, as are $X(j), X(j+1), X(j+2)$. Therefore, the differences $|X(i) - X(j)|, |X(i + 1) - X(j + 1)|$ and $|X(i + 2) - X(j + 2)|$ will be close to each other for many values of $i, j$. This means that the conditional probabilities are expected to be closer to 1 for timeseries coming from more regular processes. The negative logarithm of such a value will be closer to 0.

Conversely, random processes will produce conditional probabilities closer to 0, the negative logarithms of which will be closer to 1. The comparison of subsequent vectors in a random signal will result in different values in the successive vector distances. Thus, the ApEn values for signals coming from regular processes will be lower than the ApEn values coming from random signals. In this application, this implies that low ApEn values are to be clinically associated with cardiac pathology, while high values indicate a healthy and robust heart.

The previous algorithm calculates an estimation of the value of the Approximate Entropy, which is equal to the theoretical one, when the $N$ tends to infinity. An examination of this algorithm reveals that it is analogous to the Grassberger and Procaccia Algorithm (the Delay Times method) for the Correlation Dimension estimation. Theoretical calculations by Wolf et al. [35], indicate that reasonable estimations are achieved with an $N$ value of at least 10 m and preferably 30 m. In the experimental analysis, $N = 2000$ and $m = 2$ were used, producing satisfactory statistical ApEn validity.

However, the ApEn is a biased statistic. The expected value of $\text{ApEn}(m, r, N)$ increases asymptotically with $N$ to $\text{ApEn}(m, r)$ for all processes. The choice of window for each vector $\bar{x}$ is also important for the ApEn estimation. However, the interest in this method is not in the reconstructed space, but rather in having a sufficient number of vectors in close proximity to each other, so that accurate conditional probabilities can be found.

A.5. SVD

Once again, the phase space vectors are constructed from the timeseries of R–R intervals, and then the trajectory matrix is found. The eigenvalues of the trajectory matrix are calculated from the covariance matrix, which, given the fact that it is square, simplifies their computation. Thus, the eigenvalues of the trajectory matrix are mathematically related to the eigenvalues of the covariance matrix.

Assume that the timeseries, $x(i)$, where $i = 0, \ldots, N$, has been constructed by sampling an ECG and extracting the R–R intervals. We construct the phase space embedded vectors, $\bar{x}(i) = [x(i), x(i + \tau), x(i + 2 \tau), \ldots, x(i + (m - 1) \tau)]$, where $m$ is the embedding dimension and $\tau$ is the time lag. Values of $\tau$ and $m$ are selected experimentally, in a way similar as those selected for the aforementioned algorithms. The trajectory matrix is then constructed as a column of the $m$-dimensional embedding vectors derived from the timeseries phase space reconstruction. The definition of the trajectory matrix is

$$\bar{A} = \frac{1}{\sqrt{N}} \begin{bmatrix} \bar{x}(1) \\ \bar{x}(2) \\ \vdots \\ \bar{x}(N) \end{bmatrix}$$

where $\rho = N - (m - 1) \tau$. (A.18)

The covariance matrix is defined as

$$C_{i,j} = \frac{1}{N} \sum_k x(k + i \tau) \cdot x(k + j \tau)$$

$$\equiv \langle A \cdot A \rangle_{i,j}$$

$$\equiv \bar{A} \cdot \bar{A}.$$ (A.19)

The covariance matrix can be transformed to a diagonal matrix $U' \cdot C \cdot U = S$, where $S$ is the matrix that contains the eigenvalues of $C$. The eigenvalues of the two matrices are associated as: $s_i(l) = s(l)^2$, where $s_i$ are the eigenvalues of $C$ and $s$ are the eigenvalues of $A$.

References


