2001 Proceedings of the 23rd Annual EMBS International Conference, October 25-28, Istanbul, Turkey

ROBUST FEATURE EXTRACTION FROM ECG SIGNALS BASED ON NONLINEAR DYNAMICAL MODELING

Mohamed I. Owis, Ahmed H. Abou-Zied, Abou-Bakr M. Youssef and Yasser M. Kadah
Biomedical Engineering Department, Cairo University, Giza, Egypt, E-mail: ymk@internetegypt.com

Abstract-**The early detection of abnormal heart conditions is vital for intensive care unit patients. The detection of such conditions is possible through continuous monitoring of electrocardiographic (ECG) signals to detect the presence of arrhythmia. Conventional methods of arrhythmia detection rely on observing morphological features of the signal in the time domain or after applying a certain transformation. Even though these techniques have been fairly successful in detecting such conditions, they are limited by the fact that they treat the heart as a linear system. In this paper, we present a comprehensive study of the nonlinear dynamics of ECG signals. The correlation dimension and largest Lyapunov exponent are used to model the chaotic nature of five different classes of ECG signals. The model parameters are evaluated for a large number of real ECG signals within each class and the results are reported. The proposed algorithms allow automatic calculation of the features. The statistical analysis of the calculated features indicates that they differ significantly among different arrhythmia types and hence can be rather useful in ECG signal classification. The results of this work show the potential of such features for use in arrhythmia detection in clinical cardiac monitoring.**
Keywords - **chaos theory, arrhythmia detection, ECG.**

I. INTRODUCTION

Sudden cardiac death remains a major unresolved clinical and public health problem. There are more than 300,000 sudden cardiac deaths each year of which ventricular fibrillation (VF) is implicated in the vast majority. In general, cardiac arrhythmia can be diagnosed by monitoring the Electrocardiographic (ECG) signals of the patient. The accurate detection of such conditions at an early stage is essential for timely management of the case.

The conventional method of diagnosing arrhythmia relies on detecting the presence of particular signal features by a human observer. Due to the large number of patients in intensive care units and the need for continuous watch for such conditions, automated arrhythmia detection systems have been developed to perform this task. Such techniques try to transform the mostly-qualitative diagnostic criteria into a more objective quantitative signal features. Classical techniques such as cross-correlation, spectrum analysis, sequential hypothesis testing, wavelets, as well as morphological features have been used to address this problem. Nevertheless, such techniques provide only limited information about the signal because they ignore the underlying nonlinear signal dynamics. Consequently, such techniques met a limited success in solving this problem.

In the last two decades, there has been an increasing interest in applying techniques from the domains of nonlinear analysis and chaos theory in studying biological systems [1]. For example, more complete dynamical information can be obtained when we model the heart as a multivariate, non-linear pumping system that sometimes exhibits unpredictable (chaotic) ECG pattern. This is a direct consequence of the complex nature of ECG signals, which involves interactions between several physiological variables including autonomic and central nervous regulation, hemodynamic effects, and other extrinsic variables.

In the field of chaotic dynamical system theory, several features can be used to describe system dynamics including correlation dimension (D_2), Lyapunov exponents (λ_k), approximate entropy, etc. These features have been used to explain ECG signal behavior by several studies (e.g., [2]). Nevertheless, these studies applied such techniques only to a few sample ECG signals. Due to the randomness of such signals, such studies did not allow the extraction of a general description of the dynamics of different arrhythmia types. Moreover, the details of implementation of feature extraction techniques were not discussed. Given that such techniques are particularly sensitive to parameter variations, it is impossible for other researchers to directly use the results or draw conclusions from these studies for their implementations. Therefore, the analysis of ECG chaotic behavior that includes a large number of signals from different types using a more detailed implementation of the feature extraction steps would be rather useful.

In this work, we address the problem of characterizing the nonlinear dynamics of the ECG signal and its variation with different arrhythmia types. The implementation details to automatically compute two important chaotic system parameters, namely the correlation dimension and largest Lyapunov exponent, are discussed. In particular, the algorithms used, parameter values in each technique, and their selection criteria are given and explained. The proposed implementations were used to compute these features from a large number of independent ECG signals belonging to five different ECG signal types from the MIT-BIH Arrhythmia Database [3]. The statistical mean and standard deviation of the computed parameters for each pathology are computed. The resultant statistics are compared between different pathologies to detect statistically significant differences among different pathologies. The results suggest the potential and robustness of using such features in ECG arrhythmia detection.

II. CORRELATION DIMENSION ESTIMATION

The mathematical description of a dynamical system consists of two parts: the *state* which is a snapshot of the process at a given instant in time, and the *dynamics* which is the set of rules by which the states evolve over time. In the case of the heart as a dynamical system, all the available information about the system is a set of ECG measurements from skinmounted sensors. There is no mathematical description of the dynamics of the heart. That is, we deal only with observables whose mathematical formulation and total number of variables are not known. Therefore, to study the dynamics of such system, we first need to reconstruct the state space trajectory. The most common method to do this is using delay time embedding theorem.

II.1. *Delay Time Embedding*

Consider a single variable, digitized, ECG time series v(.) (voltages), that consists of N data points evenly-spaced in time: v(1), v(2), v(3), ... , v(t), ..., v(N). To create larger dimensional geometric object out of these observables, the time series is embedded into a larger m-dimensional embedding space. The rows of the matrix X of reconstructed state vectors of length m is defined as follows [4],

$$x(k)= [v(1+(k-1)J),v(1+(k-1)J+L),\ldots, v(1+(k-1)J+(m-1)L)],\quad k=1,2,\ldots,((N-(m-1)L-1)/J)+1 \qquad (1)$$

Here, x(.) is a vector that constitutes a row in the matrix X, m is the embedding dimension, L is the lag time that is equal to the number of data points between components of each vector, J is the number of data points between vectors (e.g., if vectors are formed from each data point, then J=1), and ((N-(m-1)L-1)/J)+1 is the number of vectors that could be formed from N data points.

The value of m must be large enough for delay time embedding to work. When a suitable m value is used, the orbits of the system do not cross each other. This condition is tested using the false nearest neighbor (FNN) algorithm [1]. The dimension m in which false neighbors disappear is the smallest dimension that can be used for the given data.

Various algorithms for estimating a suitable time lag (L) for the reconstruction procedure have been proposed. For example, the time lag L can be chosen to be the value at which the autocorrelation function reaches zero, 1/e, 0.5, or 0.1 [4]. It can also be selected as the value at which the first minimum of the mutual information function occurs [1]. We used another approach where we used the time window length to calculate L [5]. The time window length (W) is defined by the time spanned by each embedding vector,

$$W=(m-1)L. \qquad (2)$$

After determining m using FFN, we need to select the optimal time window length (W). The selection procedure will become apparent later in this section.

II.2. Dimension Estimation using Grassberger-Procaccia (G-P) Algorithm.

The simplest way to think about the dimension D of an object is as the exponent that scales the bulk b of an object with linear distance r; i.e. b α r^D. The G-P algorithm uses the correlation integral C(r) to represent the bulk b. C(r) is the average number of neighbors each point has within a given distance r, given as,

$$C(r)=\frac{1}{N_p}\sum_{i,j}\theta\left[r-\|x(i)-x(j)\|\right]. \qquad (3)$$

Here: $\|\cdot\|$ symbolizes the Euclidean distance (2-norm) between reconstructed state vectors x(i) and x(j), $N_p = k(k-1)/2$ is the number of distinct pairs of reconstructed state vectors, θ symbolizes the Heaviside unit step function (i.e., θ(x)=0 when x < 0 and θ(x)=1 when x ≥ 0). The correlation dimension D_2 is defined as the slope of the linear region of the plot of log(C(r)) versus log(r) for small values of r. That is,

$$D_2=\lim_{r\to 0}\frac{\log[C(r)]}{\log(r)}. \qquad (4)$$

Unlike the calculation of C(r), the determination of the linear scaling region is not an easy task in practice. Because of the presence of noise, it may not be practical to compute the slope for very small values of r. Several different regions may appear visually to be equally valid, which makes this determination not repeatable for manual computation. In our implementation, we tried this approach combined with computerized regression and the results were not satisfactory. Then, we improved our implementation using a second order regression for the whole curve. The linear regression was then obtained for the part of this curve that appeared linear by vision. More consistent values for D_2 were obtained. Finally, we developed an automatic algorithm to determine the linear region to eliminate the need for human interaction. The algorithm can be summarized as follows:

1. Calculate the first derivative of the curve S_1 using the following approximation of differentiation,
$$S_1=\frac{\partial\log C(r)}{\partial\log r}\cong\frac{\Delta\log C(r)}{\Delta\log r} \qquad (5)$$
2. Differentiate S_1 once again to obtain the second derivative S_2. The linear region of the curve appears as a number of consecutive points with very small values in S_2.
3. Threshold the absolute value of S_2 to determine the extent of linear region using a small threshold of 0.1.
4. Examine the resultant contiguous linear regions. Discard short linear regions composed of a sequence of 5 points or less on the curve.
5. If more than one linear region satisfies the above conditions, select the one that yields the maximum D_2 value (i.e., smaller values of r as per its definition in (4)).

In order to select the optimal window length W for a given m value in the above algorithm (and consequently the time lag for delay time reconstruction), the window length is selected to maximize the plateau in the slope S_1 vs log(r) curve [5].

To choose the best embedding dimension m, the FNN criterion was applied and the first zero has been observed at m=8. The optimal window length was found to be around 583 ms (i.e., 210 samples at 360 samples per sec). Consequently, the time lag (L) was estimated to be 83 ms using (2). To evaluate the discrimination ability of D_2, pooled t-test is applied among the results of each category.

III. LYAPUNOV EXPONENTS

Lyapunov exponents quantify the sensitivity of the system to initial conditions, which is an important feature of chaotic systems. Sensitivity to initial conditions means that small changes in the state of a system will grow at an exponential rate and eventually dominate the behavior. Lyapunov exponents are defined as the long time average exponential rates of divergence of nearby states. If a system has at least one positive Lyapunov exponent, then the system is chaotic. The larger the positive exponent, the more chaotic the system becomes (i.e., the shorter the time scale of system predictability). Lyapunov exponents will be arranged such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, where λ_1 and λ_n correspond to the most rapidly expanding and contracting principal axes, respectively. Therefore, λ_1 may be regarded as an estimator of the dominant chaotic behavior of a system.

The presence of a positive exponent is sufficient for diagnosing chaos and represents local instability in a particular direction. It is important to notice that for the existence of an attractor (a stable regime), the overall dynamics must be dissipative (i.e., globally stable) and the total rate of contraction must dominate the total rate of expansion.

Now, consider the case of n-dimensional space where n is the number of state variables used to describe the system. A small n-dimensional hyper-sphere of initial conditions evolves into a hyper-ellipsoid as time progresses. In particular, its principal axes expand (or contract) at rates given by the Lyapunov exponents. Measuring the separation of nearby initial conditions is done along the Lyapunov directions that correspond to those principal axes. The Lyapunov directions depend on the system flow and are defined using the Jacobian matrix (i.e., the tangent map) at each point of interest along the flow.

III.1. Calculation of the largest Lyapunov exponent using Wolf's algorithm

In this work, the largest Lyapunov exponent, λ_1, is calculated as a measure of the chaotic behavior of the system using the Wolf algorithm [6]. Consider two trajectories with nearby initial conditions on an attracting manifold. When the attractor is chaotic, the trajectories diverge, on the average, at an exponential rate characterized by the largest Lyapunov exponent λ_1. The algorithm used is as follows,

1. Compute the distance d_0 of two, very close, points in the reconstructed phase space orbit.
2. Follow both points as they travel a short distance along the orbit. The distance d_1 between them is calculated.
3. If d_1 becomes too large, one of the points is kept and an appropriate replacement for the other point is chosen.
4. The two points are now allowed to evolve again following steps 1-3.
5. After s propagation steps, the largest Lyapunov exponent λ_1 is estimated as,

$$\lambda_1 = \frac{1}{t_s - t_0} \sum_{k=1}^{s} \log_2 \left(\frac{d_1(t_k)}{d_0(t_{k-1})} \right) \quad (6)$$

III.2. Practical Implementation

We used a software implementation of Wolf's algorithm [6]. This software is divided into two programs: BASGEN and FET. BASGEN stands for "dataBASe GENerator". It is considered as a preprocessing step for the main program FET. It generates a database that is used by FET to determine the closest points to any specific point. FET stands for "Fixed Evolution Time". It does the main job of calculating the average exponential rate of divergence of short segments of the reconstructed orbit. There are a lot of parameters that need to be defined for the two programs. We selected those parameters for BASGEN as follows: embedding dimension m=4, time delay L=60, and grid resolution ires= 20. It should be noted that for Lyapunov exponent calculations, the embedding dimension m was chosen as D_2 rounded to the next highest integer [2]. Also, the grid resolution refers to the fact that BASGEN places the reconstructed data into a grid of dimension m, with a resolution of ires cells per side. This grid will be used later by FET to efficiently find nearest neighbors to any point.

The parameters for FET were set as follows. The time step was chosen as the sampling period. The evolution time (evolve) was chosen as 25. That is, each pair of points is followed through the phase space for this number of steps at which the local contribution to orbital divergence is calculated, and a replacement is attempted if necessary. The minimum separation at replacement (dismin) was selected to be 0.01. When a replacement is decided, points whose distance from the kept point is less than dismin are rejected. The maximum separation at replacement (dismax) determines the distance between the pair of points beyond which a replacement is decided and was chosen as 15% of the data range. Finally, the maximum orientation error (thmax) is selected to be 30 to define the maximum allowed angular deviation from the identical orientation between two points when a replacement is decided.

To evaluate the discrimination ability of λ_1, pooled t-test is applied among the results of each category.

IV. RESULTS

The proposed chaotic feature estimation techniques were implemented and applied to a large number of real ECG signals. The ECG signals used in this research were obtained from the MIT-BIH Arrhythmia Database [3]. The data set used for this work was composed of 5 different types including normal (NR), ventricular couplet (VC), ventricular tachycardia (VT), ventricular bigeminy (VB), and ventricular fibrillation (VF). Each type was represented by 64 independent signals with each signal 3 seconds long. The VF signals were sampled at 250 sample/sec, while the others were sampled at 360 sample/sec.

The results for computing D_2 and λ_1 for different ECG signal classes are shown in table (I). P-values of the t-test based on D_2 are shown in table (II). The P-values of the t-test based on λ_1 are shown in table III.

V. DISCUSSION

The signal length for this analysis can be arbitrarily chosen provided that it is less than 10 sec. This is to satisfy the ANSI/AAMI EC13-1992 standard which requires alarms for abnormal ECG signals to be activated within 10 seconds of their onset. The variation of the number of points within this duration was not found to be crucial as long as the ECG signal is sufficiently sampled.

From table I, we observe non-integer correlation dimension D_2 values for all types indicating the presence of strange attractor. Also, the positive sign of λ_1 confirms the chaotic behavior of the ECG signal. The results generally support the hypothesis that cardiac electrical activity reflects a low-dimensional dynamic system behavior.

As indicated from Tables II and III, the results confirm that normal ECG signals can be statistically differentiated from abnormal by both dynamical system features. The very low p-values suggest the rejection of the null hypothesis and hence the presence of a significant difference. For example, the first row of these tables show that normal ECG signals can be differentiated from all other arrhythmia types. On the other hand, these measures are not successful in discriminating between some of the abnormal signals. In particular, when using D_2, there is significant difference between all pairs at 5% level except between VB and VF, which are significant at the 10% level. Moreover, there was no statistically significant difference between VT and VF. This may somewhat be explained by the presence of similarities in dynamics between these types. Similarly for λ_1, it is not possible to find statistically significant difference between VT, VF, and VB (shown in boldface inside the table). These statistically insignificant differences represent a limitation of the dynamical features in classifying abnormal arrhythmia types.

VI. CONCLUSIONS

The use of ECG signal features from nonlinear dynamical modeling was studied. The detailed implementation of automatic feature extraction algorithms was demonstrated. The results of applying this program on a large data set of actual ECG signals from five different classes were presented. The statistical analysis of the results suggests that the use of such features can be advantageous to ECG signal classification. They also illustrate the limitations of such features.

ACKNOWLEDGMENT

This work was supported in part by IBE Technologies, Egypt.

REFERENCES

[1] H.D.I. Abarbanel, T.W. Frison, and L. Tsimring, "Obtaining order in a world of chaos, time domain analysis of nonlinear and chaotic signals," *IEEE Sig. Proc. Mag.*, pp. 49-65, May 1998.

[2] R.B. Govindan, K. Narayanan, and M.S. Gopinathan, "On the evidence of deterministic chaos in ECG: surrogate and predictability analysis," *Chaos*, vol. 8, no. 2, pp.495-502, 1998.

[3] The MIT-BIH Arrhythmia Database, 3rd *ed.*, Harvard-MIT Division of Health Sciences and Technology, May 1997.

[4] W. S. Pritchard, and D. W. Duke, "Measuring chaos in the brain: a tutorial review of EEG dimension estimation," *Brain and Cognition,* vol. 27, no. 3, pp. 353-397, 1995.

[5] A.M. Albano, J. Munech, and C. Schwartz, "Singular-value decomposition and the Grassberger-Procaccia algorithm," *Phys. Rev. A*, vol. 38, no. 6, pp.3017-3026, September 1988.

[6] Http://www.users.iterport.net/~wolf.

TABLE I
COMPUTED VALUES FOR DYNAMICAL SYSTEM FEATURES
(mean ± standard deviation)

Type	Parameter	
	D_2	λ_1
NR	3.27 ± 0.42	8.18 ± 3.63
VC	2.54 ± 0.39	17.36 ± 3.68
VT	3.07 ± 0.52	13.55 ± 7.24
VB	2.71 ± 0.40	12.11 ± 5.08
VF	2.93 ± 0.71	13.2 ± 4.45

TABLE II
P-VALUES OF t-TEST FOR D_2

Type	VC	VT	VB	VF
NR	<1.0e-16	0.0071	1.7e-14	0.0006
VC		1.96e-9	0.0148	0.0002
VT			3.01e-5	**0.2201**
VB				0.0309

TABLE III
P-VALUES OF t-TEST FOR λ_1

Type	VC	VT	VB	VF
NR	<1.0e-16	4.7e-7	1.16e-6	1.38e-10
VC		2.7e-4	6.0e-10	5.82e-8
VT			**0.1929**	**0.7396**
VB				**0.1976**