A new approach to detect Life threatening cardiac arrhythmias using Sequential spectrum of Electrocardiogram signals

Chandrakar Kamath
Ex-Professor, Electronics and Communication Dept., Manipal Institute of Technology, Manipal-576104.
chandrakar.kamath@gmail.com

Received: 2011/04/12; Accepted: 2011/09/17 Pages: 13-28

Abstract

This study evaluates the discriminative power of sequential spectrum analysis of the short-term electrocardiogram (ECG) time series in separating normal and subjects with life threatening arrhythmias like, ventricular tachycardia/fibrillation (VT/VF). The raw ECG time series is transformed into a series of binary symbols and the binary occupancy or relative distribution of mono-sequences (i.e. tuples containing only one type of symbol '0' or '1') is computed. The quantified approximate entropies (ApEn0 and ApEn1) of the binary occupancies in the sequential spectra are found to have potential in discriminating normal and VT/VF subjects and thus can significantly add to the prognostic value of traditional cardiac analysis. The receiver operating characteristic curve (ROC) analysis confirms the robustness of this new approach and exhibits an average sensitivity of about 98.0% (97.1%), specificity of about 93.3% (93.3%), positive predictivity of around 94.7% (89.3%), and accuracy of about 95.9% (94.7%) with ApEn0 to distinguish between normal and VT (VF) subjects.

Keywords: *Ventricular tachycardia, Ventricular fibrillation, Sequential spectrum, Symbolic dynamics*

1. Introduction

Ventricular tachycardia (VT) and ventricular fibrillation (VF) are life threatening cardiac arrhythmias [1]. Despite numerous recent advances in the field of medicine, Ventricular tachycardia/fibrillation (VT/VF) has been difficult to manage with in clinical practice and mortality rate has remained high. As a consequence the development of new noninvasive methods and measures of mortality risk in VT/VF, including sudden cardiac death, is still a major challenge. For this reason, a number of quantitative analysis techniques for ECG arrhythmia detection have been proposed [1-3]. Besides this, there is a need to reach remote and under served communities with life saving healthcare. A reliable automated classification system combined with high-speed communication can resolve this issue. This work is an attempt to develop such an automated system to discriminate between normal and Ventricular tachycardia/fibrillation subjects.

Physiological data more often show complex structures which can not be quantified or interpreted using linear methods. The classical nonlinear methods suffer from the disadvantage of dimensionality. Further, there are not enough samples in the time series to arrive at a reasonable estimate of the nonlinear measures. From this point of view it is

advisable to resort to methods which can quantify system dynamics even for short time series, like the symbolic dynamics. The prime advantages of symbolic dynamics are the following: If the fluctuations of the two data series are governed by different dynamics then the evolution of the symbolic sequences is not related. The resulting symbolic sequences histograms give a reconstruction of their respective histories and provide a visual representation of the dynamic patterns. In addition, they may be used as a basis to build statistics to compare the regions that show different dynamical properties and indicate which patterns are predominant. Thus methods of symbolic dynamics are useful approaches for classifying the underlying dynamics of a time series. Parameters of time domain and frequency domain often leave these dynamics out of consideration. Besides computational efficiency, symbolic methods are also robust when noise is present. The process of symbolization can be used to represent any possible variation over time, depending on the number of symbols and the sequence lengths used. This is a very powerful property because it does not make any assumptions about the nature of the signals/patterns (e.g., it works equally well for both linear and nonlinear phenomena).

Symbolic time series analysis has found application for the past few decades in the field of complexity analysis, including astrophysics, geomagnetism, geophysics, classical mechanics, chemistry, medicine and biology, mechanical systems, fluid flow, plasma physics, robotics, communication, and linguistics [4]. To be specific, in medicine, various implementations of symbolic sequences have been used to characterize encephalography (EEG) signals to understand the interaction between brain structures during seizures [5]. Under mechanical systems, symbolic methods were applied to combustion data from internal combustion engines to study the onset of combustion instabilities [6] and in multiphase flow data-symbolization were found to be useful in characterizing and monitoring fluidized-bed measurement signals [7]. Symbolic dynamics, as an approach to investigate complex systems, has found profound use in the analysis of heart rate variability signals [8-12]. However, there is hardly any literature where sequential spectrum is applied for analysis of the first difference of raw ECG signals.

There are many ways symbolic dynamics can be used for analysis of time series and all of them require coding i.e. converting the time series into symbolic series. The differences in symbolic methods are usually in their coding procedure or used complexity indices. In this contribution we employ sequential spectrum [13] as the symbolic method and approximate entropy as a measure of complexity of the sequential spectrum [14-15] to classify ECG signals obtained from standard Holter recordings from PhysioNet database [16] into normal and VT/VF subjects. The rationale behind the application of sequential spectrum and approximate entropy measures is that both are suitable for short-term segments of the ECG signal. Receiver operating characteristic (ROC) plots were used to evaluate the ability of these complexity measures to discriminate normal from VT/VF subjects. Both the approximate entropy measures ($ApEn_0$: approximate entropy of the binary occupancies of symbol ‘0’ and $ApEn_1$: approximate entropy of the binary occupancies of symbol ‘1’) yielded excellent results. However, it is found that $ApEn_0$ performs better than $ApEn_1$ with an average sensitivity of about 98.0% (97.1%), specificity of about 93.3% (93.3%), positive predictivity of around 94.7% (89.3%), and accuracy of about 95.9% (94.7%) to distinguish between normal and VT (VF) subjects.

2. Methods and materials

We first explain the ECG data used and then the symbolic dynamics of the time series.

2.1 ECG records

All the ECG records used are from the benchmark PhysioNet databases [16]. The work involved 18 ECG records from normal sinus rhythm (NSR) database (nsrdb) and ECG records of 35 subjects who experienced episodes of sustained ventricular tachycardia, ventricular flutter and ventricular fibrillation (VT/VF) from Creighton University ventricular tachyarrhythmia database (cudb). The NSR database includes 5 men, aged 26 to 45 years, and 13 women, aged 20 to 50 years. The age and gender of subjects in VT/VF database are not available. For sake of comparison and validation, the NSR database was divided into two groups, first with 9 ECG records (Normal-1) and second, also, with 9 ECG records (Normal-2). Likewise, the VT/VF database was divided into two groups, first with 15 ECG records (VT/VF-1) and second also with 15 ECG records (VT/VF-2). From each record the modified limb lead II was only considered for analysis. The resolution is 200 samples per mV for nsrdb and 400 samples per mV for cudb. The sampling frequency of normal sinus rhythm signal from NSR is 128 Hz and that of VT/VF signal from cudb is 250 Hz. Since the sampling frequency does influence upon the calculated indices it is necessary to have the same sampling frequency for all the records. For this reason ECG signals from NSR database are first re-sampled at 250 Hz. Then each record is divided into segments of equal time duration (14 sec), with 3500 samples/ segment in both normal sinus rhythm and VT/VF database. A total of 1000 segments from normal sinus rhythm and from VT/VF data base, each, are analyzed. All the records are normalized before analysis. Also all the signals from both database are filtered using an 8-point moving average filter to remove high-frequency noise.

2.2 Symbolic Dynamics and Sequential spectrum

The application of symbolic analysis requires coarse grained representation of the signal, which is explained below.

Static and dynamic transformations. Symbolic dynamics, as an approach to investigate complex systems, facilitates the analysis of dynamic aspects of the signal of interest. The concept of symbolic dynamics is based on a coarse-graining of the dynamics [5]. That is the range of original observations or the range of some transform of the original observations such as the first difference between the consecutive values, is partitioned into a finite number of regions and each region is associated with a specific symbolic value so that each observation or the difference between successive values is uniquely mapped to a particular symbol depending on the region into which it falls. The former mapping is called static transformation and the latter dynamic transformation. As a simple example of static transformation a binary partition will lead to two symbols, '0' and '1'. If an observation x_i is below a specific threshold, x_{th}, then we associate it with a '0' and otherwise we symbolize it with a '1', as shown in the eqn. below.

$$S_i = \begin{cases} 0 & if\ x_i < x_{th} \\ 1 & if\ x_i \geq x_{th} \end{cases} \quad (1)$$

The threshold x_{th} is usually the mean or the median of the time series. Details of the dynamic transformation are given in the beginning of Sec. 2.2.2 below.

Thus the original observations are transformed into a series of same length but the elements are only a few different symbols (letters from the same alphabet), the transformation being termed symbolization. A general rule of thumb is the partitions must be such that the individual occurrence of each symbol is equiprobable with all other symbols or the measurement range covered by each region is equal. This is done to bring out ready differences between random and nonrandom symbol sequences. The transformations into symbols have to be chosen context dependent. For this reason, we use complexity measures on the basis of such context-dependent transformations, which have a close connection to physiological phenomena and are relatively easy to interpret. This way the study of dynamics simplifies to the description of symbol sequences. Some detailed information is lost in the process but the coarse and robust properties of the dynamic behavior is preserved and can be analyzed [5].

Mono-sequences, Binary Occupancy and Sequential spectrum of ECG signals. In this study, we use dynamic transformation approach for the symbolic dynamics [6]. Such a differenced symbolization scheme is relatively insensitive to extreme noise spikes in the data. In this approach arithmetic differences between adjacent data points of the ECG signal define the symbolic values. We symbolize the positive difference as a '*1*' and the negative difference as a '*0*' as shown in the eqn. below.

$$S_i = \begin{cases} 1 & if\ x_i - x_{i-1} \geq 0 \\ 0 & if\ x_i - x_{i-1} < 0 \end{cases} \qquad (2)$$

After symbolization the next step is to partition the symbol sequence into windows of width *W* symbols each. It is sliding window technique and if the shift is smaller than *W*, then the consecutive windows overlap. The next step is the construction of temporal patterns in each window, by the identification of short ordered sequences of only one type of symbol (*0* or *1*), termed tuples or mono-sequences, from the symbol series by gathering groups of symbols in the temporal order. The mono-sequence *Nx0* (or *Nx1*), is a homogeneous sub-sequence containing *N* (*N=1, 2, …, W*), only one type of symbol *S=0* (or *1)*. The lengths of the mono-sequences correspond to their respective frequencies. Next the tuples [*NxS*], i.e. mono-sequences of length *N* are counted in the *j*-th window. Thus we evaluate the cardinality, L_j [*NxS*], for all possible values of *N*, and arrive at the distribution of cardinalities in the *j*-th window. From the cardinalities we compute the respective binary occupancies, O_j [*NxS*], for the mono-sequences [*NxS*] in the window *j* as given by the eqn. (3) below.

$$O_j[NxS] = L_j[NxS].\frac{N}{W} \qquad (N = 1, 2, \dots, W) \qquad (3)$$

Each plot of binary occupancies, O_j [*NxS*] with (*S=0* or *S=1*), Vs the length of the mono-sequences, *N*, constitutes Sequential-spectrum for the window *j*. Thus, the binary occupancies characterize the distribution of monotonic intervals of length *N*, in the analyzed time series: decreasing intervals for *S=0* and increasing intervals for *S=1*. Though this distribution is referred to as spectrum, this is neither a transformation to frequency domain nor the inverse transform does exist. The shape, distribution and width of the sequential spectrum contain information about the analyzed time series.

Approximate Entropy (ApEn). Approximate entropy (ApEn) is a measure of irregularity in the data without any *a priori* knowledge of the system generating them [14]. ApEn is scale invariant and model independent, evaluates both dominant and

subordinant patterns in the data, and discriminates series for which clear feature recognition is difficult. It is immune to low level noise and robust to meaningful information with a reasonable number of data points. Large values of ApEn indicate more complexity or irregularity in the data and vice versa. In the following a short description of the formal implementation of the ApEn is given, for further details see [15]. Given a time series with M data points, x(1), x(2),…, x(M). To compute ApEn m-dimensional vector sequences, $p_m(i)$, are constructed from data series like [$p_m(1)$, $p_m(2)$, …, $p_m(M-m+1)$], where 1≤ ف≤ چ -m+1. If the distance between two vectors $p_m(i)$ and $p_m(j)$ is defined as $| p_m(i) - p_m(j)|$, then

$C_i^m(d)$ = [Number of vectors such that $| p_m(i) - p_m(j)| < d$]/(M-m+1), where m specifies the pattern length and d defines the criterion of similarity. $C_i^m(d)$ is considered as the mean of the fraction of patterns of length m that resemble the pattern of the same length that begins at index i. ApEn is computed using the eqn. (4) below.

Let $\Phi^m(d) = \sum_{i=1}^{M-m+1} \ln(C_i^m(d))/(M-m+1))$

and $\Phi^{m+1}(d) = \sum_{i=1}^{M-m} \ln(C_i^{m+1}(d)/(M-m))$

then $ApEn(m,d,M) = \Phi^m(d) - \Phi^{m+1}(d) \quad for\ m \geq 1$

where $ApEn(0,d,M) = -\Phi^1(d)$ (4)

Pincus and Goldberger [15] have shown that with m=2, d= 10–۲۵٪ علاعءکم لغغگ deviation of the M data points/ samples (100-900 data points) will lead to statistically reliable and reproducible results. In this study, we use m=2, d=0.0025 and M=3500. This is the key to a measure of irregularity: small values of ApEn indicate regularity and large values imply substantial fluctuations or irregularity in the time series.

Approximate Entropy of Binary Occupancies and *t*-tests. We define the approximate entropies, $ApEn_0$ and $ApEn_1$, of the binary occupancies, *O*[*Nx0*] and *O*[*Nx1*], respectively, in the sequential spectra. $ApEn_0$ and $ApEn_1$ are computed for each of the sequential-spectrum for the window or ECG segment in the normal and VT/VF groups. In this work we use approximate entropies, $ApEn_0$ and $ApEn_1$, and show that each one is capable of distinguishing between normal and VT/VF subjects. To assess the use of these parameters individual and pair-wise significance tests (Student's *t*-tests) are performed. To compare the regularity of the fluctuation functions between the normal and VT/VF groups we also compute the mean and standard deviation of the difference between the corresponding approximate entropies, $ApEn_0$ and $ApEn_1$, of the two groups. Parameters are regarded as statistically significant if $p < 0.05$.

Receiver Operating Characteristic (ROC) Analysis and C-statistic. As mentioned above, individual and pair-wise significance tests (Student's *t*-tests) are used to evaluate the statistical differences between the approximate entropy values, $ApEn_0$ and $ApEn_1$, of the binary occupancies, *O*[*Nx0*] and *O*[*Nx1*] respectively, for normal and VT/VF groups. If significant differences between groups are found, then the ability of the non-linear analysis method to discriminate normal from VT/VF subjects is evaluated using receiver operating characteristic (ROC) plots in terms of C-statistics. ROC curves are obtained by plotting sensitivity values (which represent the proportion of the patients with diagnosis of VT/VF who test positive) along the y axis against the corresponding (1-specificity) values (which represent the proportion of the correctly identified normal subjects) for all the available cutoff points along the x axis. Accuracy is a related parameter that quantifies the total number of subjects (both normal and VT/VF) precisely classified. The area under ROC curve (AUC), also called C-statistic, measures this discrimination, that is, the ability of the test to correctly classify those

with and without the disease and is regarded as an index of diagnostic accuracy. The optimum threshold is the cut-off point in which the highest accuracy (minimal false negative and false positive results) is obtained. This can be determined from the ROC curve as the closet value to the left top point (corresponding to 100% sensitivity and 100% specificity). A C-statistic value of 0.5 indicates that the test results are better than those obtained by chance, where as a value of 1.0 indicates a perfectly sensitive and specific test.

3. Results and Discussion

To test for statistical significance of sequential spectrum approach, first we analyze the ECG data from normal and VT/VF subjects of Group-I and show that approximate entropy of the binary occupancies, $ApEn_0$, alone is sufficient to distinguish between normal and VT/VF subjects. Next, we validate our approach conducting another case study on normal and VT/VF subjects from Group-II. $ApEn_0$ is analyzed from segments of 3500 samples and averaged to obtain mean values for the entire recording period.
The sampling frequency of NSR database is 128 Hz and that of VT/VF database is 250 Hz. Since the sampling frequency does influence upon the calculated indices it is necessary to have the same sampling frequency for all the records. For this reason ECG signals from NSR database are first re-sampled at 250 Hz. For sake of comparison and validation, as mentioned earlier, the normal sinus rhythm database (NSR) was divided into two groups, first with 9 ECG records (Normal-1) and second with 9 ECG records (Normal-2). Likewise, the VT/VF database was divided into two
groups, first with 10 ECG records (VT/VF-1) and second also with 10 ECG records (VT/VF-2). Then each record is divided into segments of equal time duration (14 sec), with 3500 samples/ segment in both normal sinus rhythm and VT/VF database. A total of 1000 segments from normal sinus rhythm and from VT/VF data base, each, are analyzed. Sequential spectrum analysis is applied to these segments from both the groups to decide whether a particular segment belongs to normal, or VT/VF group. Dynamic transformation as given in Eq. (1) is first applied on each segment to arrive at a symbol string with a range of two possible symbols {0, 1} (binary symbolization) and binary occupancies, O_j [*NxS*] with (*S=0* or *S=1*) are computed. This is repeated for all the segments (all values of *j*) and a plot of average binary occupancies Vs the length of the mono-sequences is made.

Fig.1(a) shows a comparison of averaged sequential spectra of the binary occupancies, *O*[*Nx0*] for normal and VT subjects from group-I and Fig.1(b) shows a comparison of sequence spectra of the binary occupancies, *O*[*Nx1*] for the same normal and VT subjects from group-I. It is found from Figs. 1(a) and 1(b) that the sequential spectra corresponding to *O*[*Nx0*] and *O*[*Nx1*] for VT subjects show a broader width compared to those of normal subjects. All binary occupancies for the range n= 1 to 9 contribute significantly in the normal case than those for VT. However, all other binary occupancies contribute significantly in the VT case than those for normal.
Fig.2(a) shows a comparison of averaged sequential spectra of the binary occupancies, *O*[*Nx0*] for normal and VF subjects from group-I and Fig.2(b) shows a comparison of sequence spectra of the binary occupancies, *O*[*Nx1*] for the same normal and VF subjects from group-I. It is found from Figs. 2(a) and 2(b) that the sequential spectra corresponding to *O*[*Nx0*] and *O*[*Nx1*] for VF subjects show a broader width compared to those of normal subjects. All binary occupancies for the range n= 1 to 11 contribute

significantly in the normal case than those for VF. However, all other binary occupancies contribute significantly in the VF case than those for normal.

The distribution of approximate entropies $ApEn_0$ and $ApEn_1$ for the normal, VT and VF groups (Group-I) are shown using Box-whiskers plots in Figs. 3(a) and 3(b), respectively. In Fig. 3(a) for $ApEn_0$, the boxes (inter-quartile range) of normal and VT/VF subjects are non-overlapping. In Fig. 3(b) for $ApEn_1$, the boxes (inter-quartile range) of normal and VT/VF subjects are non-overlapping. These plots show that either $ApEn_0$ or $ApEn_1$ is sufficient to distinguish between normal and VT/VF subjects. The results of statistical analysis of non-paired Student's *t*-test for normal, VT and VF groups of Group-I are depicted in Table 1. All values are expressed as mean ± Standard Deviation (median) [95% Confidence Interval]. For normal subjects, we find the following approximate entropies (mean ± S.D.): $ApEn_0 = 0.0580 \pm 0.0515$ and $ApEn_1 = 0.0921 \pm 0.0787$ respectively. For VT subjects we find the following approximate entropies (mean ± S.D.): $ApEn_0 = 0.1903 \pm 0.1265$ and $ApEn_1 = 0.2006 \pm 0.1211$, both different from normal. For VF subjects we find the following approximate entropies (mean ± S.D.): $ApEn_0 = 0.1819 \pm 0.1301$ and $ApEn_1 = 0.1859 \pm 0.1290$, both different from normal. These distributions show that either $ApEn_0$ or $ApEn_1$ alone is sufficient to distinguish between normal and VT/VF subjects. It can be observed that all the three classes show a larger value of $ApEn_1$ as compared to that of $ApEn_0$. This implies a reduced regularity in binary occupancies

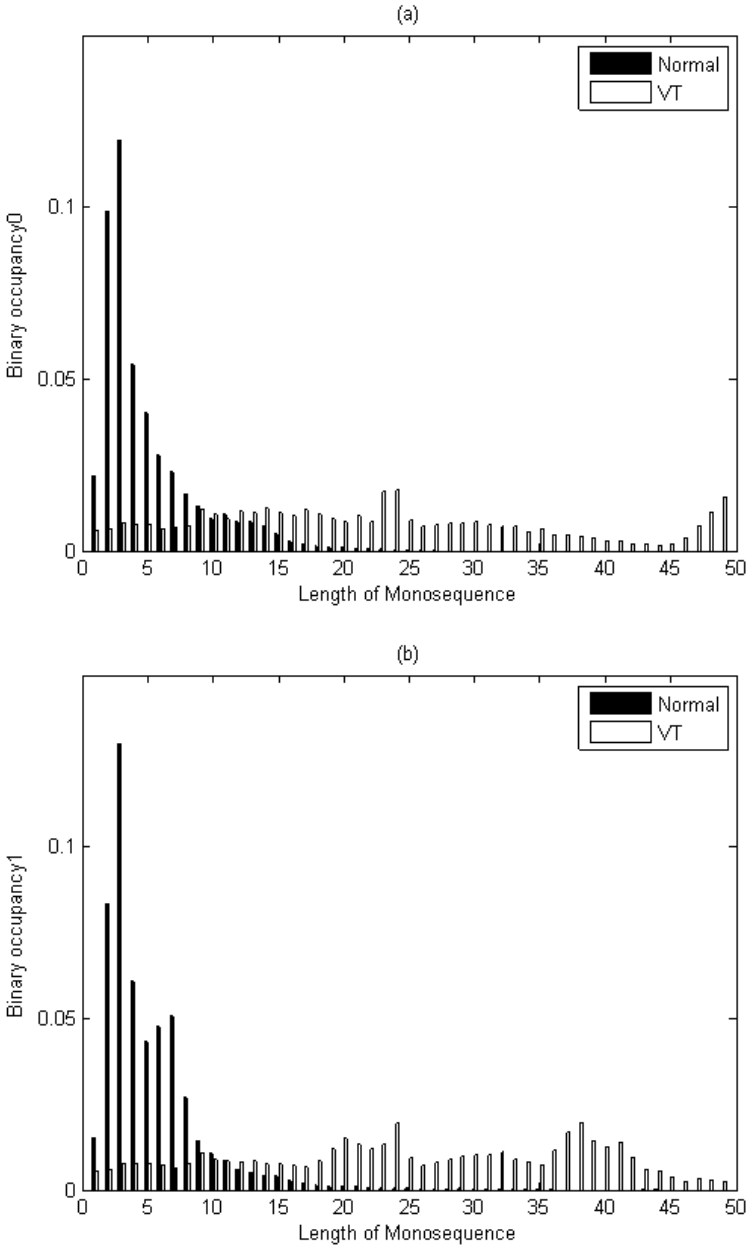

Figure 1. Comparison of sequential spectra of normal and VT subjects from Group-I. (a) for binary occupancies O[Nx0] (b) for binary occupancies O[Nx1].

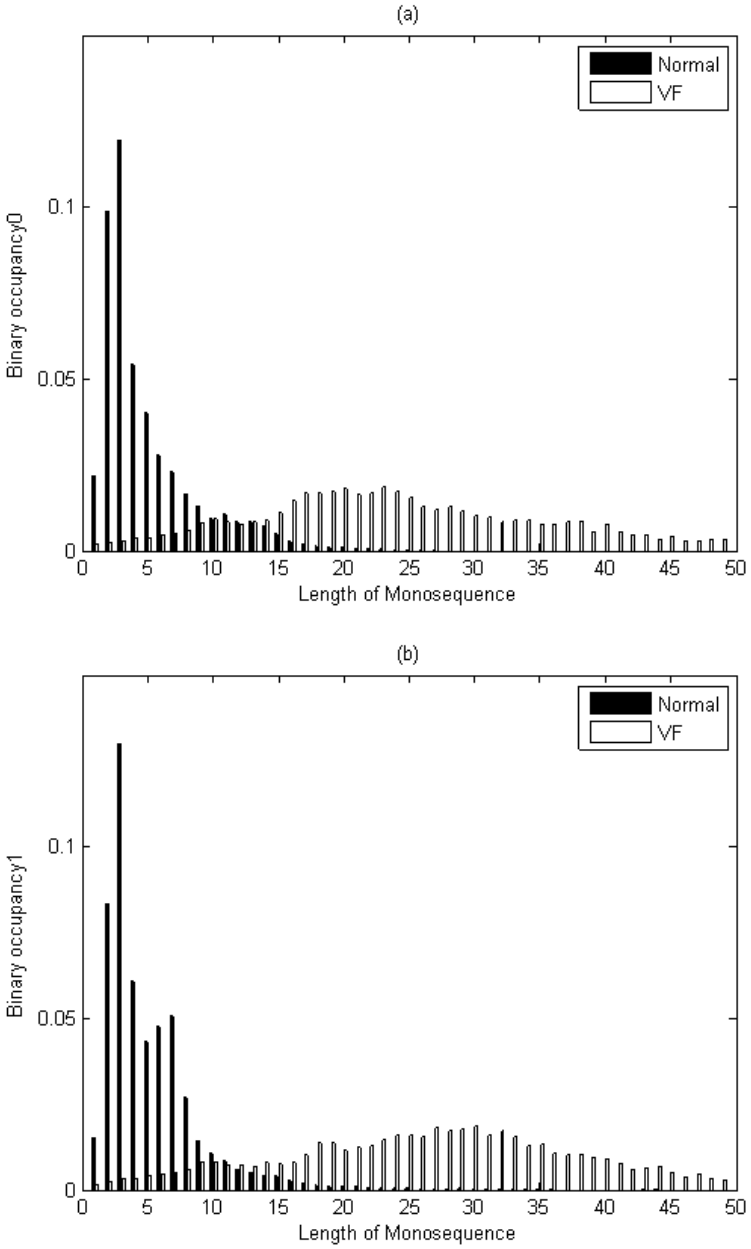

Figure 2. Comparison of sequential spectra of normal and VF subjects from Group-I. (a) for binary occupancies O[Nx0] (b) for binary occupancies O[Nx1].

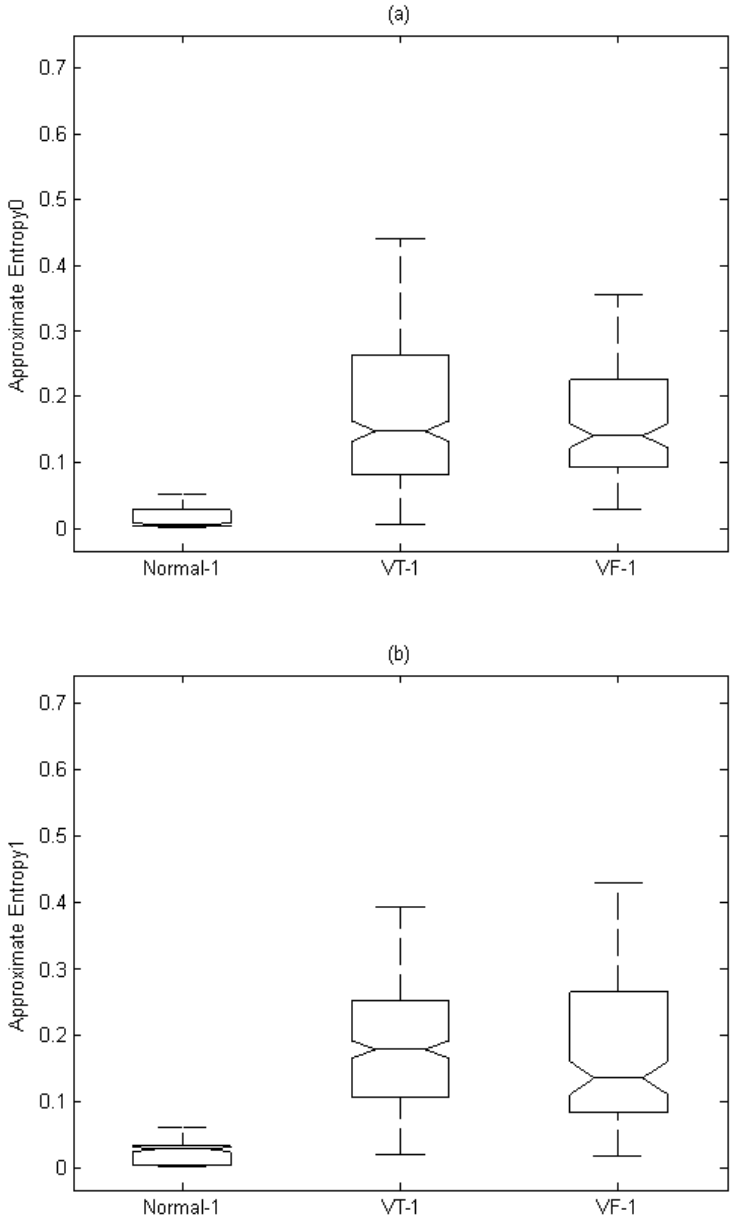

Figure 3. The distribution of approximate entropies (a) ApEn0 and (b) ApEn1 using Box-whiskers plots (without outliers) for normal, VT and VF subjects from Group-I.

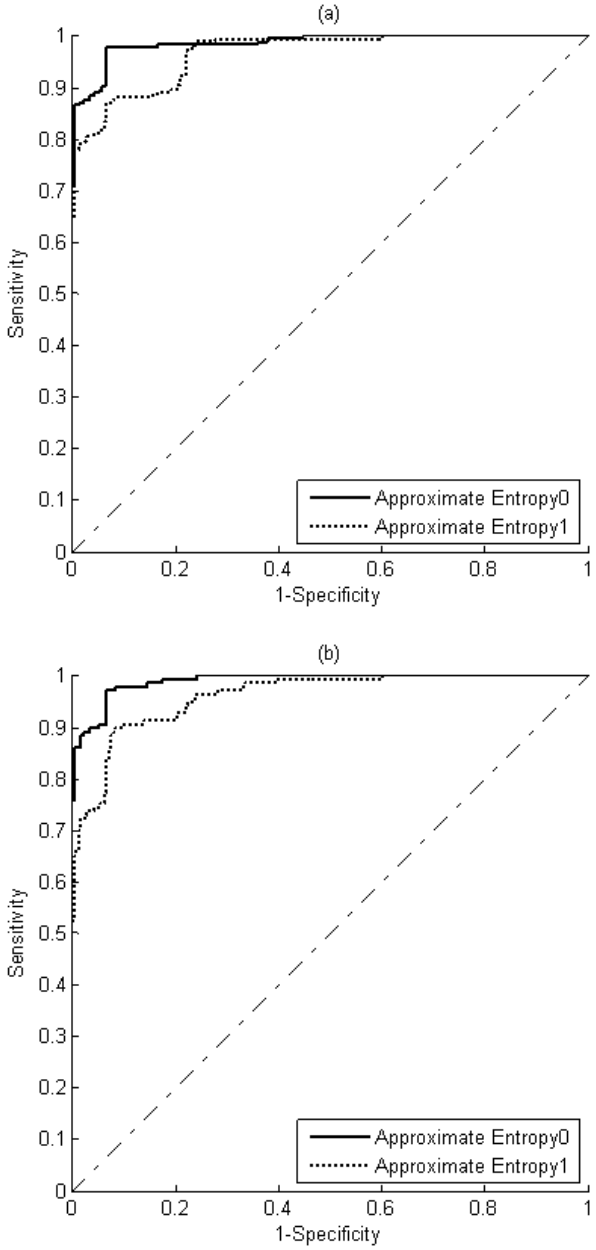

Figure 4. ROC curve for ApEn0 (solid line) and ApEn1 (dotted line) (a) between normal and VT and (b) between normal and VF. The diagonal line (dash-dot line) from 0,0 to 1,1 represents ROC curve that can not discriminate between normal and VT/VF from Group-I.

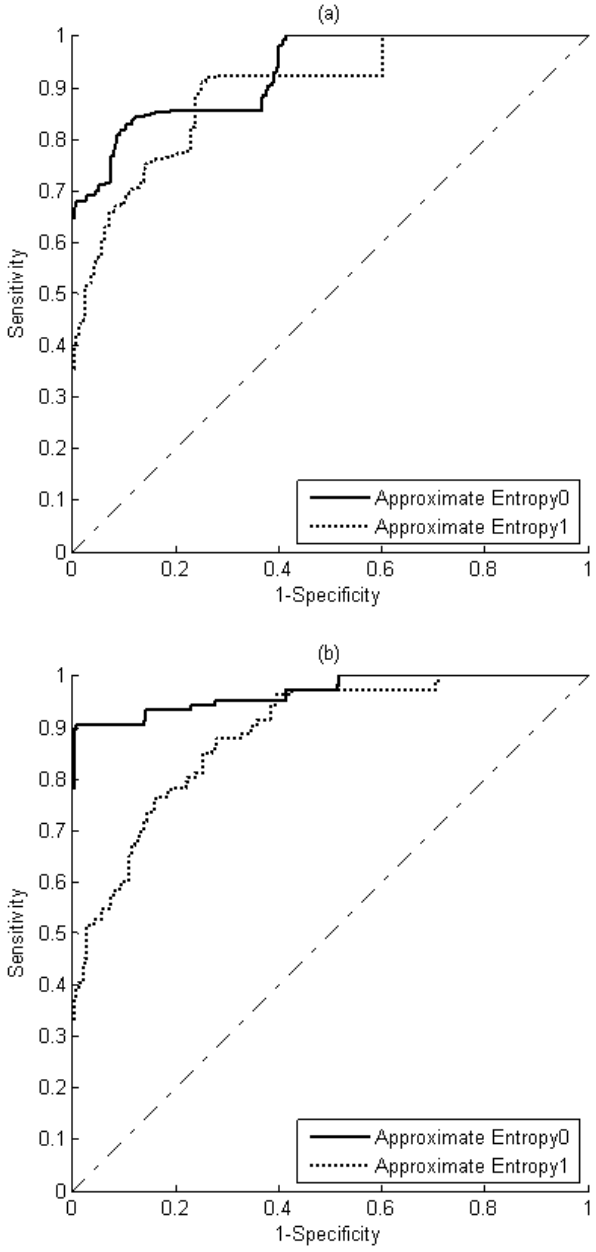

Figure 5. ROC curve for ApEn0 (solid line) and ApEn1 (dotted line) (a) between normal and VT and (b) between normal and VF. The diagonal line (dash-dot line) from 0,0 to 1,1 represents ROC curve that can not discriminate between normal and VT/VF from Group-II.

$O[Nx1]$ compared to $O[Nx0]$. It is also found that $ApEn_0$ and $ApEn_1$ for VT/VF group are always larger than the respective values of the normal group, the former being significantly larger. This implies an decrease in the regularity of binary occupancies $O[Nx0]$ and $O[Nx1]$ in the VT/VF group compared to normal group. Of course, experimental studies are necessary to confirm the mechanisms behind the decrease in the regularity of binary occupancies of VT/VF subjects.

Table 1. Descriptive results of sequence spectrum analysis for Group-I. All values are expressed as mean ± SD (median) [95% CI]. (non-paired Student's t-test; p < 0.0001)

Subject	$ApEn_0$	$ApEn_1$
Normal	0.0160± 0.0165 (0.005375) [0.0140 0.0181]	0.0277±0.0250 (0.02617) [0.0245 0.0309]
VT	0.1856± 0.1261 (0.1485) [0.1711 0.2001]	0.1927±0.1209 (0.1790) [0.1788 0.2066]
VF	0.1789±0.1303 (0.1416) [0.1569 0.2009]	0.1811±0.1307 (0.1363) [0.1590 0.2031]

We also perform the Student's *t*-test for paired data. The results are tabulated in Tables 2 and 3. Parameters are regarded as statistically significant if $p < 0.05$. The values of test statistic and *p*-value reveal that both the approximate entropies $ApEn_0$ and $ApEn_1$ for both the groups are statistically significant. Next, the ability of the $ApEn_0$ and $ApEn_1$ to discriminate normal from VT/VF subjects (Group-I) is evaluated using receiver operating characteristic (ROC) plots, which are shown in Fig. 4(a) for normal and VT and in Fig. 4(b) for normal and VF respectively, with $ApEn_0$ (shown

Table 2. p-values and tstat (test statistic) values of paired t-test for ApEn0 and ApEn1 of normal and VT subjects from Group-I.

Parameter	ApEn0-VT	ApEn1-VT
ApEn0-normal	p= 0; tstat = -20.6824	
ApEn1-normal		p= 0; tstat = -20.7837

Table 3. p-values and tstat (test statistic) values of paired t-test for ApEn0 and ApEn1 of normal and VF subjects from Group-I.

Parameter	ApEn0-VF	ApEn1-VF
ApEn0-normal	p= 0; tstat = -19.1237	
ApEn1-normal		p= 0; tstat = -17.6403

by solid line) and $ApEn_1$ (shown by dotted line). It is found, from both the figures that, $ApEn_0$ performs better than $ApEn_1$. For the case of $ApEn_0$, in Fig. 4(a), it is found that the area under the curve (AUC) is 0.98566 with sensitivity = 98.0%, specificity = 93.3%, positive predictivity = 94.7%, and accuracy = 95.9%. For the case of $ApEn_1$, in Fig. 4(a), it is found that the area under the curve (AUC) is 0.96587 with sensitivity = 88.1%, specificity = 92.1%, positive predictivity = 86.3%, and accuracy = 89.9%. For the case of $ApEn_0$, in Fig. 4(b), it is found that the area under the curve (AUC) is found to be 0.98936 with sensitivity = 97.1%, specificity = 93.3%, positive predictivity = 89.3%, and accuracy = 94.7%. For the case of $ApEn_1$, in Fig. 4(b), it is found that the area under the curve (AUC) is 0.96016 with sensitivity = 89.8%, specificity = 91.7%, positive predictivity = 86.0%, and accuracy = 91.0%. The above results substantiate our finding that $ApEn_0$ outperforms $ApEn_1$ and that $ApEn_0$ alone is sufficient to distinguish between normal and VT/VF subjects.

Finally, we validate our approach conducting another case study on normal and VT/VF subjects from Group-II. The results of statistical analysis of non-paired Student's *t*-test for normal and VT/VF groups of Group-II are depicted in Table. 4. All values are expressed as mean ± Standard Deviation (median) [95% Confidence Interval]. For normal subjects, we find the following approximate entropies (mean ± S.D.): $ApEn_0$ = 0.1187±0.0153 and $ApEn_1$ = 0.1756±0.0289 respectively. For VT/VF subjects we find the following approximate entropies (mean ± S.D.): $ApEn_0$ = 0.0824±0.0121 and $ApEn_1$ = 0.1042±0.0167, both different from normal. These distributions show that either $ApEn_0$ or $ApEn_1$ alone is sufficient to distinguish between normal and VT/VF subjects. We also perform the Student's *t*-test for paired data. The results are tabulated in Tables 5 and 6. Parameters are regarded as statistically significant if $p < 0.05$. The values of test statistic and *p*-value reveal that both the approximate entropies $ApEn_0$ and $ApEn_1$ for both the groups are statistically significant. Next, the ability of the $ApEn_0$ and $ApEn_1$ to discriminate normal from VT/VF subjects (Group-II) is evaluated using receiver operating characteristic (ROC) plots, which are shown in Fig. 5(a) for normal and VT and in Fig. 5(b) for normal and VF respectively, with $ApEn_0$ (shown by solid line) and $ApEn_1$ (shown by dotted line). It is found, from both the figures that, $ApEn_0$ performs better than $ApEn_1$. For the case of $ApEn_0$, in Fig. 5(a), it is found that the area under the curve (AUC) is 0.92899 with sensitivity = 83.7%, specificity = 88.3%, positive predictivity = 90.7%, and accuracy = 85.7%. For the case of $ApEn_1$, in Fig. 5(a), it is found that the area under the curve (AUC) is 0.89303 with sensitivity = 88.3%, specificity = 76.3%, positive predictivity = 83.5%, and accuracy = 83.2%. For the case of $ApEn_0$, in Fig. 5(b), it is found that the area under the curve (AUC) is found to be 0.96836 with sensitivity = 90.6%, specificity = 99.2%, positive predictivity = 98.0%, and accuracy = 96.5%. For the case of $ApEn_1$, in Fig. 5(b), it is found that the area under the curve (AUC) is 0.88683 with sensitivity = 76.4%, specificity = 83.8%, positive predictivity = 67.5%, and accuracy = 81.5%. The above results again substantiate our finding that $ApEn_0$ outperforms $ApEn_1$ and that $ApEn_0$ alone is sufficient to distinguish between normal and VT/VF subjects. The difference in accuracy and other measures of Group-II can be attributed to age differences, and differing male-to-female ratios between groups I and II.

Table 4. Descriptive results of sequence spectrum analysis for Group-II. All values are expressed as mean ± SD (median) [95% CI]. (non-paired Student's t-test; p < 0.0001)

Subject	ApEn0	ApEn1
Normal	0.0226±0.0198 (0.0267) [0.0201 0.0251]	0.0479±0.0452 (0.0389) [0.1681 0.1830]
VT	0.1383±0.1263 (0.1073) [0.1245 0.1521]	0.1698±0.1456 (0.1210) [0.1540 0.1857]
VF	0.1852±0.1323 (0.1497) [0.1597 0.2106]	0.2043±0.1590 (0.1538) [0.1737 0.2349]

Table 5. p-values and tstat (test statistic) values of paired t-test for ApEn0 and ApEn1 of normal and VT subjects from Group-II.

Parameter	ApEn0-VT	ApEn1-VT
ApEn0-normal	p= 0; tstat = -14.0562	
ApEn1-normal		p= 0; tstat = -12.5369

Table 6. p-values and tstat (test statistic) values of paired t-test for ApEn0 and ApEn1 of normal and VF subjects from Group-II.

Parameter	ApEn0-VF	ApEn1-VF
ApEn0-normal	p= 0; tstat = -18.6023	
ApEn1-normal		p= 0; tstat = -14.0283

4. Conclusion

We apply sequential spectrum analysis to the first difference of nonstationary raw ECG time series from normal and VT/VF subjects. We show that this approach can identify the monotonicity in the ECG signals and tell us how long such periods are, when the signal is increasing or decreasing. The quantified approximate entropies of the binary occupancies of the sequential spectrum are found to have potential in discriminating normal and VT/VF subjects and thus can significantly add to the prognostic value of traditional cardiac analysis. These approximate entropies can easily be analyzed from ambulatory ECG recordings without time consuming preprocessing and hence, may have practical implications for risk stratification. The nonlinear methods applied to time series usually demand more computations and a long ECG episode duration. Acquiring long records just for screening purpose is not amenable. Although the ECG data we use contains both 30 minutes and 20 hours duration records, our method uses short-term segments, of the order of 14 sec duration. Hence the method is suitable for screening large population in a short time.

5. References

[1] Szi-Wen Chen.: Complexity-Measure-Based Sequential Hypothesis Testing for Real-Time Detection of Lethal Cardiac Arrhythmias, EURASIP JASP. 20957, 1--8 (2007).

[2] Chen SW.: A Two-stage Discrimination of Cardiac Arrhythmias using a Total Least Squares-based Prony Modeling Algorithm. IEEE Trans. Biomed. Eng. 47(10), 1317--1327 (2000).

[3] Xu L., Zhang D., Wang K., Wang L.: Arrhythmic Pulses Detection using Lempel-Ziv Complexity Analysis. EURASIP JASP. 1--12(2006).

[4] Daw C.S., Finney C.E.A, Tracy E.R.: A Review of Symbolic Analysis of Experimental Data. Rev. Sci. Instrum. 74(2), 915--930 (2003).

[5] Xu J.H., Liu Z.R., Liu R.: The Measures of Sequence Complexity for EEG Studies. Chaos. 4 (11), 2111--2119 (1994).

[6] Daw C.S.: Observing and Modeling Nonlinear Dynamics in an Internal Combustion Engine. Phys. Rev. Lett. 57(3), 2811--2819 (1998).

[7] Finney C.E.A., Nguyen K., Daw C.S., Halow J.S.: Symbol-sequence Statistics for Monitoring Fluidization. Proc. ASME HTD. 405--411 (1998).

[8] Kurths J., Voss A., Saparin P., Witt A., Kleiner H.J., Wessel N.: Quantitative Analysis of Heart Rate Variability. Chaos. 5, 88--94 (1995).

[9] Porta A., D'Addio G., Pinna G.D., Maestri R., Gnecchi-Ruscone T., Furlan R., Montano N., Guzzetti S., Malliani A.: Symbolic Analysis of 24h Holter Heart Period Variability Series: Comparison between Normal And Heart Failure Patients. Computers in Cardiology. 32, 575--578 (2005).

[10] Eleonora Tobaldini, Alberto Porta, Shun-Guang Wei, Zhi-Hua Zhang, Joseph Francis, Karina Rabello Casali, Robert Weiss M., Robert Felder B., Nicola Montano.: Symbolic Analysis detects Alterations of Cardiac Autonomic Modulation in Ventriclar Tachycardia/ Ventricular Fibrillation Rats. Auton Neurosci. 150(1-2), 21--6 (2009).

[11] Wessel N., Schwarz U., Saparin P.I., Kurths J.: Symbolic Dynamics for Medical Data Analysis. Proc. EUROATTRACTOR2000. 46--62 (2000).

[12] Voss A., Kurths J., Kleiner H.J., Witt A., Wessel N., Saparin P., Osterziel K.J., Schurath R., Dietz, R.: The Application of Methods of Nonlinear Dynamics for the Improved and Predictive Recognition of Patients Threatened by Sudden Cardiac Death. Cardiovasc. Res. 31, 419--433 (1996).

[13] Robert Stepien A.: New method for analysis of nonstationary signals. Nonlinear Biomedical Physics. 5(3), 1--8 (2011).

[14] Cysarz D., Bettermann H., Van Leeuwen P.: Entropies of Short Binary Sequences in Heart Period Dynamics. Am. J. Physiol. Heart Circ. Physiol. 278: H2163--H2172 (2000).

[15] Pincus S.M., Goldberger A.L.: Physiological Time-series Analysis: What does Regularity Quantify? Am. J. Physiol. Heart Circ. Physiol. 266(4 Pt 2), H1643--H1656 (1994).

[16] Ary Goldberger L., Luis Amaral A.N., Leon Glass, Jeffrey Hausdorff M., Plamen Ivanov Ch., Roger Mark G., Joseph Mietus E., George Moody B., Chung-Kang Peng, Eugene Stanley H.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. 101, e215-e220 (2000), http://circ.ahajournals.org/cgi/content/full/101/23/e215.