Journal of Advances in Computer Research
Quarterly pISSN: 2345-606x eISSN: 2345-6078
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 6, No. 3, August 2015), Pages: 9-22
www.jacr.iausari.ac.ir

Complexity-Measure-Based Approach To Detect Life Threatening Cardiac Arrhythmias Using First-Order Difference Of Electrocardiogram Signals

Chandrakar Kamath
"ShanthaNilaya", 107,Ananthnagar, Manipal, India
Chandrakar.kamath@gmail.com

Received: 2014/08/08; Accepted: 2014/09/01

Abstract

The aim of this study is to evaluate how far Lempel-Ziv complexity (LZC) of the binary symbolic sequences resulting from static or dynamic transformation (partitioning first-order difference) of the short-term electrocardiogram (ECG) signals (only 2 seconds duration) has the potential in discriminating normal and ventricular tachycardia/ventricular fibrillation (VT/VF) subjects. The statistical analyses show that LZC from either transformation is sufficient to distinguish between normal and VT/VF subjects. Between the two, LZC of dynamic transformation is found to outperform LZC of static transformation. The receiver operating characteristic curve analysis confirms the robustness of this new approach which exhibits an average sensitivity of about 99.1% (100.0%), specificity of about 100.0% (100.0%), precision of around 98.9% (100.0%), and accuracy of about 99.5% (100.0%), with LZC to distinguish between normal and VT (VF) subjects. The presented method is simple, computationally efficient, and well suited for real time implementation in automatic external or implantable cardioverter-defibrillators.

Keywords: Electrocardiogram, Lempel-Ziv complexity, Symbolic sequences, Ventricular tachycardia, Ventricular fibrillation

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. It is crucial for the patient to receive immediate medical intervention when either VF or VT occurs. As a consequence the development of new non-invasive 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]. Sequential hypothesis testing of binary sequences has been employed to detect ventricular fibrillation [4]. Though the method shows an improvement over previous methods, the accuracy is not high enough for clinical applications. Gustavo Santos used regularized least squares technique with a radial basis kernel to predict ventricular arrhythmia [5]. Power spectrum of raw ECG signals and power spectrum of beat-to-beat intervals were tried as input vectors to a neural network.

Though the approach is said to lead to more accurate risk stratification, no details of accuracy and other measures are available. A number of methods are available to detect VT/VF in the literature. Some of them with good results are given below. Jun Wang et al. have tried symbolic dynamics and information entropy to discriminate VT/VF from NSR [6]. Zhang HX et al. use complexity measure and complexity rate information to detect VT/VF [7]. Complexity dispersity method has been tried in qualitative chaos analysis of VT/VF based on symbolic complexity [8]. The method showed accuracy of 100%, but with segments of 5 sec length. Inappropriate defibrillator discharge or anti-tachycardiac pacing remains an important clinical problem in implantable cardioverter-defibrillator therapy as they lead to unnecessary pain and sometimes proarrhythmic effects. As an implication in real time applications the value for specificity is more important than the value for sensitivity. However, in these studies no details pertinent to sensitivity and specificity are available.

Besides manual defibrillation by an emergency paramedic in the recent years, bystander defibrillation with automated external defibrillators (AEDs) has also been recommended for resuscitation. A reliable automated classification system combined with computationally efficient real time implementable algorithm can resolve this issue. This work is an attempt to develop such an automated computationally fast system to discriminate between normal and Ventricular tachycardia/ fibrillation subjects.

Physiological data more often show complex structures which cannot 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 [9]. 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 [10]. 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 [11]. 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 [12]. Under mechanical systems, symbolic methods were applied to combustion data from internal combustion engines to study the onset of combustion instabilities [13] and in multiphase flow data-symbolization were found to

be useful in characterizing and monitoring fluidized-bed measurement signals [14]. Symbolic dynamics, as an approach to investigate complex systems, has found profound use in the analysis of heart rate variability signals [15]-[19]. 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 dynamic transformation [18] is employed as the symbolic method and LZC [20]-[22] as a measure of complexity of the first-order difference of ECGs to classify ECG signals obtained from standard Holter recordings from PhysioNet database [23] into normal and VT/VF subjects. Dynamic transformation is preferred when the time series, like ECG, is either nonstationary or has very long-time-scale variations. Binary dynamic transformation amounts to partitioning first-order difference of the ECG signal and it is shown that it leads to better separation of the classes than binary static transformation. The rationale behind the application of LZC is that it is suitable for short-term segments of the ECG signal. Receiver operating characteristic (ROC) plots were used to evaluate the ability of the complexity measure to discriminate normal from VT/VF subjects. It is found that LZC yielded excellent results with an average sensitivity of about 99.1% (100.0%), specificity of about 100.0% (100.0%), precision of around 98.9% (100.0%), and accuracy of about 99.5% (100.0%), to distinguish between normal and VT (VF) subjects. In the testing phase, for LZC analysis, segments of only 2 seconds duration are used. As mentioned above, in real time applications the value for specificity is more important than the value for sensitivity. With this approach an average specificity of 100.0% and an average sensitivity about 99.0% is achieved. The usual nonlinear methods applied to time series (other than symbolic dynamics) usually demand long-term series, longer than 20 seconds length. Although the ECG data used contains both 30 minutes and 20 hours duration records, this method uses short-term segments, of the order of 2 sec duration in the final testing phase. Hence the method is suitable for screening large populations in a short time. The presented method is simple, computationally efficient, and hence, is well suited for real time implementation in automatic external or implantable cardioverter-defibrillators.

2. Methods and material

The following subsection 2.1 discusses the data used for analysis. The next three subsections 2.2, 2.3, and 2.4 discuss in depth the methods based on coarse-graining of the system dynamics. Subsection 2.5 depicts the LZC algorithm and subsection 2.6 describes the statistical tests and receiver operating characteristic (ROC) plots to evaluate the performance parameters.

2.1 ECG records

All the ECG records used are from the benchmark PhysioNet databases [23]. 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 (12 seconds), with 3000 samples/ segment in both normal sinus rhythm and VT/VF database. A total of 2000 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. In the testing phase, however, for LZC analysis segments of 2 seconds duration are used.

2.2 Symbolic Dynamics

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 [9]. The original time series is transformed into a series of dicretized symbols that are further processed to extract information about the generating process. Some detailed information is lost in the process but the coarse and robust properties of the dynamic behaviour is preserved and can be analyzed [9].

2.3 Static transformation and Dynamic transformation

During coarse-graining of the system dynamics, 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. 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 [14]. This is done to bring out ready differences between random and non-random symbol sequences. The transformations into symbols have to be chosen context dependent. For this reason, complexity measures on the basis of such context-dependent transformations, which have a close connection to physiological phenomena and relatively easy to interpret are employed. This way the study of dynamics simplifies to the description of symbol sequences.

2.4 Binary Symbol-sequence from Static transformation of ECG signals

This is the simplest possible partition involving the division of data range into two parts (binary partition). Those data which are above a cut-off value are assigned a symbolic value of '1', while those below the cut-off value are assigned a symbolic value of '0'. In this case the cut-off value is chosen to be equal to the median of the data, x_{median}. The time series x_i is transformed into the symbolic sequence S_i, where $S_i \in \{0, 1\}$, as given below.

$$\text{if } x_i \geq x_{median} \quad S_i = 1 \text{ and if } x_i < x_{median} \quad S_i = 0 \tag{1}$$

2.5 Binary Symbol-sequence from first-difference of ECG signals

In this study, dynamic transformation approach for the symbolic dynamics [14] is employed. Dynamic transformation amounts to partitioning first-order difference of the ECG signal. 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. The positive difference is symbolized as a '1' and the negative difference as a '0' as shown in the eqn. below.

$$\text{if } x_i - x_{i-1} \geq 0 \quad S_i = 1 \text{ and if } x_i - x_{i-1} < 0 \quad S_i = 0 \tag{2}$$

2.6 Lempel-Ziv Complexity Measure

The Lempel-Ziv complexity algorithm was proposed by Lempel and Ziv to evaluate the randomness of finite sequences. It is rather a simple-to-compute nonparametric measure of complexity suitable for finite length one-dimensional signals related to the number of distinct substrings and the rate of their recurrence. Larger values of LZC imply higher complexity data. Since LZC analyzes finite symbol-sequences it is essential that the given signal must first be coarse-grained. In this study, a binary conversion is used. This binary string is scanned from left to right and a complexity counter c(N) is incremented by one unit every time a new subsequence pattern is encountered in the scanning process, and the immediate next symbol is regarded as the beginning of the next subsequence pattern. The LZC can be estimated using the following algorithm [21].

1. Let P denote the original string sequence i.e. $P= \{s_1, s_2, s_3,\ldots\}$, with s_i defined as in Eq. (1). Let S and Q denote two sub-sequences of P and SQ be concatenation of S and Q. Also, let $SQ\pi$ be a sequence derived from SQ after its last character is deleted (π implying deletion of last character in the sequence) and $\upsilon(SQ\pi)$ denote the vocabulary of all different sub-sequences of $SQ\pi$.

2. At the beginning, the complexity counter c(N)=1, $S=s_1$, $Q=s_2$, $SQ=s_1, s_2$, and therefore, $SQ\pi=s_1$.

3. In general, with $S=s_1, s_2, s_3,\ldots,s_r$ and $Q=s_{r+1}$, $SQ\pi= s_1, s_2, s_3,\ldots,s_r$. If Q belongs to $\upsilon(SQ\pi)$ then Q is subsequence of $SQ\pi$ and not a new sequence.

4. With S intact, change Q to s_{r+1}, s_{r+2} and check if Q belongs to $\upsilon(SQ\pi)$ or not.

5. Keep repeating previous steps until Q does not belong to $v(SQ\pi)$. Now $Q=s_{r+1}$, $s_{r+2},\ldots,s_{r+i}$ is not a subsequence of $SQ\pi=s_1, s_2,..,s_{r+i-1}$. So increase c(N) by 1.

6. Thereafter, S is renewed to $S=s_1, s_2,\ldots, s_{r+i}$ and Q to $Q=s_{s+i+1}$.

7. Repeat the previous steps until Q is the last character. At this point in time, the number of sub-sequences in P is c(N), which corresponds to measure of complexity.

To arrive at a measure of complexity independent of sequence length, c(N) must be normalized. If the length of the sequence is n and the number of different symbols is α, it has been shown that the upper bound of c(N) is [21]

$$c(N) < N / ((1-\epsilon_N) \log_\alpha(N)) \quad (3)$$

where ε_N is a small quantity and $\varepsilon_N \rightarrow 0$ ($N\rightarrow\infty$). In general, $N/\log_\alpha(N)$ is the upper limit of c(N), i.e.,

$$\lim_{N\rightarrow\infty} c(N) = b(N) = N / \log_\alpha(N) \quad (4)$$

For a binary conversion $\alpha=2$, $b(N)= N/\log_2(N)$ and c(N) can be normalized by b(N) as

$$C(N) = c(N) / b(N) \quad (5)$$

C(N), the normalized LZC, reflects the arising rate of new patterns along with the sequence and thus captures the temporal structure of the sequence. A larger value of LZC means that the chance of generating a new pattern is greater, so the sequence is more complex, and vice versa.

2.7 t-Tests and Receiver Operating Characteristic (ROC) Analysis

Individual and pair-wise significance tests (Student's t-tests) are used to evaluate the statistical differences between the LZC values of normal and VT/VF groups. If significant differences between groups are found, then the ability of the nonlinear 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 cut-off 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 significance of LZC and efficacy of the first-order difference of ECG approach, first LZC of the ECG data from normal and VT/VF subjects of Group-I for

the following two cases (i) with binary static transformation and (ii) with binary dynamic transformation (partitioning first-order difference of ECG) are compared and then it is shown that dynamic transformation outperforms static transformation. Next, this approach is validated conducting another case study on normal and VT/VF subjects from Group-II.

The ECG records used are pre-processed, grouped, and segmented as mentioned in Sec. 2.1. LZC 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). LZC is computed and averaged to obtain mean values for the entire recording period. This is repeated for all the segments of all the three classes.

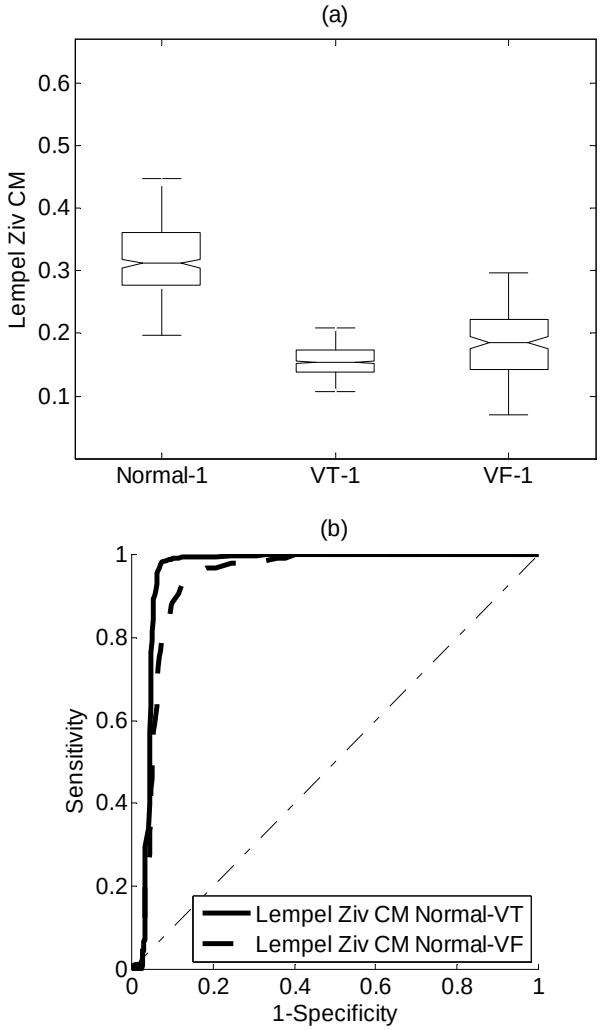

Figure 1. (a) The distribution of LZC values (with static transformation) using Box-whiskers plots (with outliers) for normal, VT and VF subjects from Group-I. (b) ROC curves for LZC with static transformation, for normal and VT (solid line), and normal and VF (dash-dot line). The diagonal line (dotted line) from 0,0 to 1,1 represents ROC curve that cannot discriminate between normal and VT/VF from Group-I.

The distribution of LZC values, with static transformation of ECGs, for the normal, VT and VF groups (Group-I) are shown using Box-whiskers plots in Figure 1(a). The boxes (inter-quartile range) of normal and VT/VF subjects are non-overlapping. This plot shows that LZC 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, the following LZC (mean ± SD) is observed: 0.3147±0.02744. For VT subjects the following LZC (mean ± SD) is found: 0.1547±0.01113, different from normal. For VF subjects the following LZC (mean ± SD) is observed: 0.1841±0.02637, again different from normal. These distributions show that LZC is sufficient to distinguish between normal and VT/VF subjects. It can be observed that LZC for VT/VF group are always smaller than that of the normal group. This implies a decrease in the complexity of VT/VF group compared to normal group.

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

Subject	LZC
Normal	0.3147±0.02744 (0.3119) [0.3074 0.3242]
VT	0.1547±0.01113 (0.154) [0.1524 0.1590]
VF	0.1841±0.02637 (0.1848) [0.1740 0.1907]

Table 2. p-values and tstat (test statistic) values of paired t-test for LZC analysis (with static transformation) of normal and VT/VF subjects from Group-I.

Subject	VT	VF
Normal	$p = 8.2742 \times 10^{-177}$; tstat = 37.5407	p = 0; tstat = 19.2343

Figure 2(a) depicts the distribution of LZC values, with dynamic transformation of ECGs, for the normal, VT and VF groups (Group-I) using Box-whiskers plots. The boxes (inter-quartile range) of normal and VT/VF subjects are found to be non-overlapping. From this plot it is seen that LZC can be used 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 3. All values are expressed as mean ± Standard Deviation (median) [95% Confidence Interval]. For normal subjects, the following LZC (mean ± SD) is seen: 0.8114±0.0292. For VT subjects the following LZC (mean ± SD) is seen: 0.3861±0.04696, different from normal. For VF subjects the following LZC (mean ± SD) is observed: 0.2821±0.02045, different from normal. These distributions show that LZC can be readily used to separate normal and VT/VF subjects. It is found that LZC for VT/VF group are always smaller than that of the normal group. This implies a decrease in the complexity of VT/VF group compared to normal group. Of course, experimental studies are necessary

to confirm the mechanisms behind the decrease in the complexity of signals in VT/VF subjects.

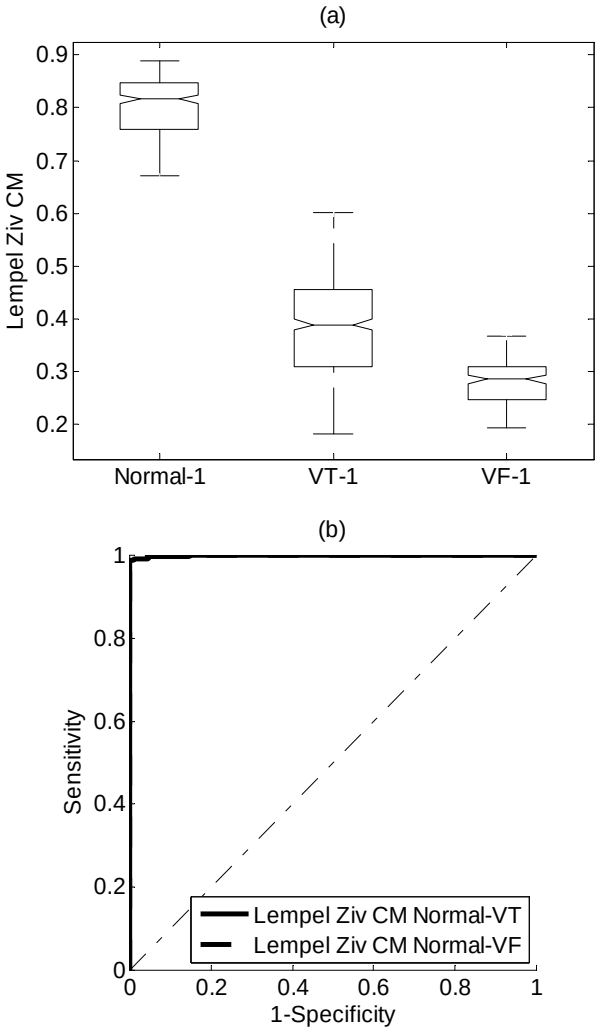

Figure 2. (a) The distribution of LZC values (with dynamic transformation) using Box-whiskers plots (with outliers) for normal, VT and VF subjects from Group-I. (b) ROC curves for LZC with dynamic transformation, for normal and VT (solid line), and normal and VF (dash-dot line). The diagonal line (dotted line) from 0,0 to 1,1 represents ROC curve that cannot discriminate between normal and VT/VF from Group-I.

Although both, static and dynamic transformations, perform well in separating normal from VT/VF groups, comparing paired t-test results (p-value and tstat) from Tables 2 and 4, it is found that dynamic transformation outperforms static transformation. This finding is substantiated using ROC plots, which are shown in Figure 1(b), for static transformation and in Figure 2(b), for dynamic transformation, respectively, with normal and VT (shown by solid line) and normal and VF (shown by dash-dot line). It is found, comparing both the figures that, dynamic transformation performs better than static transformation. For the case of static transformation, in Figure 1(b), it is found that (i) for normal and VT separation, the area under the curve

(AUC) is 0.95693 with sensitivity = 97.9%, specificity = 93.1%, precision = 94.6%, and accuracy = 95.8% and (ii) for normal and VF separation, the area under the curve (AUC) is 0.93597 with sensitivity = 93.1%, specificity = 87.7%, precision = 77.5%, and accuracy = 89.4%. For the case of dynamic transformation, in Figure 2(b), it is found that (i) for normal and VT separation, the area under the curve (AUC) is 0.99942 with sensitivity = 99.1%, specificity = 100.0%, precision = 98.9%, and accuracy = 99.5% and (ii) for normal and VF separation, the area under the curve (AUC) is 1.0000 with sensitivity = 100.0%, specificity = 100.0%, precision = 100.0%, and accuracy = 100.0%. Comparing these measures, it is obvious that using dynamic transformation for symbolization has an advantage over the usual static transformation.

Table 3. Descriptive results of LZC analysis (with dynamic transformation) for Group-I. All values are expressed as mean ± SD (median) [95% CI]. (non-paired Student's t-test; $p < 0.0001$)

Subject	LZC
Normal	0.8114±0.0292 (0.8163) [0.7952 0.8069]
VT	0.3861±0.04696 (0.3889) [0.3795 0.3985]
VF	0.2821±0.02045 (0.2849) [0.2788 0.2989]

Table 4. p-values and tstat (test statistic) values of paired t-test for LZC analysis (with dynamic transformation) of normal and VT/VF subjects from Group-I.

Subject	VT	VF
Normal	p = 0; tstat = 68.6044	$p = 5.0044 \times 10^{-317}$; tstat = 91.3541

Finally, our approach is validated 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. 5. All values are expressed as mean ± Standard Deviation (median) [95% Confidence Interval]. For normal subjects, the LZC (mean ± SD) is 0.8264±0.0125. For VT subjects, the LZC (mean ± SD) is 0.2956±0.05352, different from normal. For VF subjects the LZC (mean ± SD) is: 0.2853±0.03897, again different from normal. The paired t-test results (p-value and tstat) are shown in Table 6 and the ROC plots in Figure 3. It is found that (i) for normal and VT separation, the area under the curve (AUC) is 0.99861 with sensitivity = 99.0%, specificity = 99.0%, precision = 99.0%, and accuracy = 98.0% and (ii) for normal and VF separation, the area under the curve (AUC) is 1.0 with sensitivity = 100.0%, specificity = 100.0%, precision = 100.0%, and accuracy = 100.0%. The above results substantiate our finding that LZC with first-order difference of ECGs outperforms LZC with static transformation and that former is preferred to distinguish between normal and VT/VF subjects.

Table 5. Descriptive results of LZC analysis (with dynamic transformation) for Group-II. All values are expressed as mean ± SD (median) [95% CI]. (non-paired Student's t-test; p < 0.0001)

Subject	LZC
Normal	0.8264±0.0125 (0.8278) [0.8141 0.8220]
VT	0.2956±0.05352 (0.3042) [0.2990 0.3305]
VF	0.2853±0.03897 (0.2888) [0.2823 0.3147]

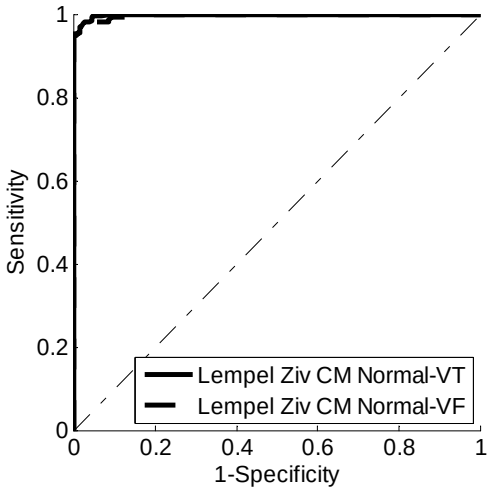

Figure 3. ROC curves for LZC with dynamic transformation, for normal and VT (solid line), and normal and VF (dash-dot line). The diagonal line (dotted line) from 0,0 to 1,1 represents ROC curve that cannot discriminate between normal and VT/VF from Group-II.

Table 6. p-values and tstat (test statistic) values of paired t-test for LZC analysis (with dynamic transformation) of normal and VT/VF subjects from Group-II.

Subject	VT	VF
Normal	$p = 7.4889 \times 10^{-283}$; tstat = 61.7038	$p = 5.5555 \times 10^{-296}$; tstat = 77.0900

Figure 4 shows a synthesized ECG signal comprising NSR, VT, and VF rhythms in sequence, each 3500 samples long, together with the corresponding LZC variation. For LZC analysis segments of 2 seconds duration are used. Two empirically found thresholds (at 0.53 and 0.21 marked by horizontal solid lines) are used for separating NSR and VT/ VF rhythms. The sudden drop in the LZC indicates that the patient entered crucial episode of VT.

The presented method is simple, computationally efficient, and well suited for real time implementation in automatic external or implantable defibrillators. One limitation of the current study is the small sample size. Although this study reports symbolic dynamics to yield very sensitive measures based on the p-value generated from the t-statistics, factors like high variance, age differences, and differing male-to-female ratios between groups will have an impact on the results when statistical analyses are carried out on small sample sizes. Nevertheless, the results of this study provide sufficient evidence to warrant the execution of larger studies that can provide more statistically robust confirmation of the application of symbolic dynamics as a reliable measure of cardiac health.

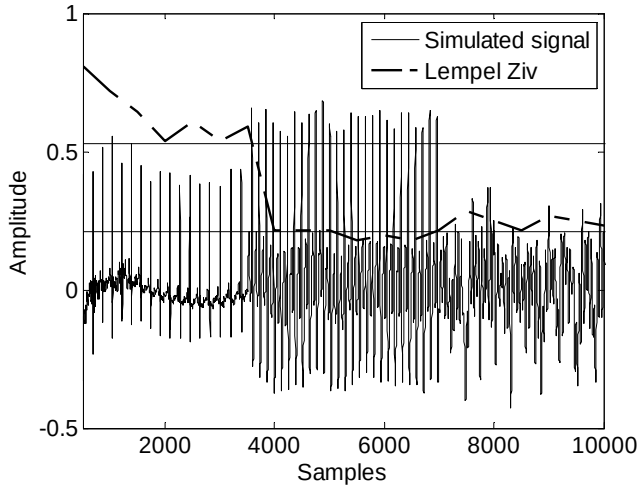

Figure 4. Variation of LZC (dotted line) for a simulated signal (solid line) comprising NSR, VT, and VF rhythms in sequence. The two horizontal solid lines mark the empirical thresholds to separate NSR from VT/VF.

4. Conclusion

LZC analysis is applied to the first-order difference of nonstationary raw ECG time series from normal and VT/VF subjects. The quantified complexity measure LZC of the binary symbolic sequences 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. This complexity measure can easily be analyzed from ambulatory ECG recordings without time consuming pre-processing and hence, may have practical implications for risk stratification. Inappropriate defibrillator discharge or anti-tachycardiac pacing remains an important clinical problem in implantable cardioverter-defibrillator therapy as they lead to unnecessary pain and sometimes proarrhythmic effects. As an implication in real time applications the value for specificity is more important than the value for sensitivity and with this approach the average specificity is 100.0% and average sensitivity is about 99.0%. The usual nonlinear methods applied to time series (other than symbolic dynamics) usually demand long-term series, longer than 20 seconds length. Although the ECG data used contains both 30 minutes and 20 hours duration records, this method uses short-term segments, of the order of 2 sec

duration in the final testing phase. Hence the method is well suited for real time implementation in automated external or implantable cardioverter-defibrillators.

References

[1] S.W. Chen, "Complexity-Measure-Based Sequential Hypothesis Testing for Real-Time Detection of Lethal Cardiac Arrhythmias," EURASIP Journal on Advances in Signal Processing, 2007.

[2] S.W. Chen, "A two-stage discrimination of cardiac arrhythmias using a total least squares-based Prony modeling algorithm," IEEE Transactions on Biomedical Engineering, 2000.

[3] L. Xu, D. Zhang, K. Wang, and L. Wang, "Arrhythmic pulses detection using Lempel-Ziv complexity analysis," EURASIP Journal on Applied Signal Processing, 2006.

[4] J. Pardey, "Detection of Ventricular Fibrillation by Sequential Hypothesis Testing of Binary Sequences," Computers in Cardiology, 2007.

[5] G. Santos, "Towards Ventricular Arrhythmia Prediction from ECG Signals," MIT Computer Science and Artificial Intelligence Laboratory, 2006.

[6] J. Wang and J. chen, "Symbolic dynamics of ventricular tachycardia and ventricular fibrillation," Physica A: Statistical Mechanics and its Applications 2010.

[7] H.X. Zhang, Y.S. Zhu, and Z.M. Wang, "Complexity measure and complexity rate information based detection of ventricular tachycardia and fibrillation," Medical and Biological Engineering and Computing, 2000.

[8] Z. Hongxuan and Z. Yishena, "Qualitative chaos analysis for ventricular tachycardia and fibrillation based on symbolic complexity," Medical Engineering and Physics, 2001.

[9] A. Voss, N. Wessel, V. Baier, K.J. Osterziel, J. Kurhts, R. Dietz, and A. Schirdewan, "Symbolic dynamics-a powerful tool in non-invasive biomedical signal processing," Report, University of Applied sciences Jena, Germany, 2007.

[10] J.G. Brida, "Symbolic time series analysis in economics," Tech. Rep. 10/00, Universidad de la Rep´ublica, Facultad de Ciencias Sociales, Departamento De Economia, 2000.

[11] C.S. Daw, C.E.A Finney, and E.R. Tracy, "A review of symbolic analysis of experimental data," Review of Scientific Instruments, 2003.

[12] J.H. Xu, Z.R. Liu, and R. Liu, "The measures of sequence complexity for EEG studies," Chaos, 1994.

[13] C.S. Daw, "Observing and modeling nonlinear dynamics in an internal combustion engine," Phys. Rev. Lett., 1998.

[14] C.E.A Finney, K. Nguyen, C.S. Daw, and J.S. Halow, "Symbol-sequence statistics for monitoring fluidization," Proceedings of the ASME Heat Transfer Division 1998.

[15] J. Kurths, A. Voss, P. Saparin, A. Witt, H.J. Kleiner, and N. Wessel, "Quantitative analysis of heart rate variability," Chaos, 1995.

[16] A. Porta, G. D'Addio, G.D. Pinna, R. Maestri, T. Gnecchi-Ruscone, R. Furlan, N. Montano, S. Guzzetti, and A. Malliani, "Symbolic analysis of 24h Holter heart

period variability series: Comparison between normal and heart failure patients," Computers in Cardiology, 2005.

[17] E. Tobaldini, A. Porta, S.G. Wei, Z.H. Zhang, J. Francis, K.C. Rabello, M.R. Weiss, R.B. Felder, and N. Montano, "Symbolic analysis detects alterations of cardiac autonomic modulation in Ventricular tachycardia/Ventricular fibrillation rats," Auton Neurosci., 2009.

[18] N. Wessel, U. Schwarz, P.I. Saparin, and J. Kurths, "Symbolic dynamics for medical data analysis," Proceedings of EUROATTRACTOR, 2000.

[19] A. Voss, J. Kurths, H.J. Kleiner, A. Witt, N. Wessel, P. Saparin, K.J. Osterziel, R. Schurath, and R. Dietz, "The application of methods of nonlinear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death," Cardiovasc. Res., 1996.

[20] A. Lempel and J. Ziv, "On the complexity of finite sequences," IEEE Transactions on Information Theory, 1976.

[21] C. Gómez and R. Hornero, "Entropy and Complexity Analyses in Alzheimer's disease: An MEG Study," The Open Biomedical Engineering Journal, 2010.

[22] Y. Li, S. Tong, D. Liua, Y. Gai, X. Wang, J. Wang, Y. Qiu, and Y. Zhu, "Abnormal EEG complexity in patients with schizophrenia and depression," Clinical Neurophysiology, 2008.

[23] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C-K Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation, 2000.
(Available at http://circ.ahajournals.org/cgi/content/full/101/23/e215).