

# Nonlinear Quality Indices Based on a Novel Lempel-Ziv Complexity for Assessing Quality of Multi-Lead ECGs Collected in Real Time

Yatao Zhang<sup>\*,\*\*</sup>, Zhenguo Ma<sup>\*\*\*</sup>, and Wentao Dong<sup>\*\*\*\*</sup>

## Abstract

We compared a novel encoding Lempel-Ziv complexity (ELZC) with three common complexity algorithms i.e., approximate entropy (ApEn), sample entropy (SampEn), and classic Lempel-Ziv complexity (CLZC) so as to determine a satisfied complexity and its corresponding quality indices for assessing quality of multi-lead electrocardiogram (ECG). First, we calculated the aforementioned algorithms on six artificial time series in order to compare their performance in terms of discerning randomness and the inherent irregularity within time series. Then, for analyzing sensitivity of the algorithms to content level of different noises within the ECG, we investigated their change trend in five artificial synthetic noisy ECGs containing different noises at several signal noise ratios. Finally, three quality indices based on the ELZC of the multi-lead ECG were proposed to assess the quality of 862 real 12-lead ECGs from the MIT databases. The results showed the ELZC could discern randomness and the inherent irregularity within six artificial time series, and also reflect content level of different noises within five artificial synthetic ECGs. The results indicated the AUCs of three quality indices of the ELZC had statistical significance ( $>0.500$ ). The ELZC and its corresponding three indices were more suitable for multi-lead ECG quality assessment than the other three algorithms.

## Keywords

Complexity, ECG Quality Assessment, Encoding LZ Complexity, Entropy

## 1. Introduction

In practice, considerable amount of electrocardiogram (ECG) recordings collected via wearable devices and smart phones cannot be used in clinic because their quality is so poor that their main waveform cannot be identified, so the quality of ECGs needs to be classified before diagnosis [1-3].

At present, most assessing quality methods used quality indices of waveform and frequency characters of ECG. Waveform characters of ECG are firstly used to generate quality indices because the characters are easy to be identified and calculated. A simple and effective method just used quality indices of waveform characters, i.e., flat lines, saturation, baseline drift, high and low amplitude and steep slope, to classify quality of ECG [4], and its assessment accuracy, namely the accuracy of quality classification

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received November 14, 2018; January 8, 2019; accepted February 15, 2019.

Corresponding Author: Yatao Zhang (zytboy@sdu.edu.cn)

\* School of Control Science and Engineering, Shandong University, Jinan, China (zytboy@sdu.edu.cn)

\*\* School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, China (zytboy@sdu.edu.cn)

\*\*\* School of Computer Science and Technology, University of Science and Technology of China, Hefei, China (rj\_mazhenguo@163.com)

\*\*\*\* Dept. of Cardiovascular Surgery, Weihai Municipal Hospital, Weihai, China (147626597@qq.com)

achieved 85.7% on a test dataset from a specially collected database of the PhysioNet/Computing in Cardiology Challenge 2011 (CinC2011). Similarly, Kuzilek et al. [5] summarized six quality indices of waveform characters to assess ECG quality, and its accuracy achieved 83.6% on the same test dataset. Frequency characters, i.e., power spectrum density (PSD) can reflect energy distribution within time series, so they can help to obtain some detailed information within physiological series. So quality indices of PSD were used for assessing quality of ECG [6-8]. In [6], 35 quality indices generated from PSD were used for assessing ECG quality, and the corresponding assessment result achieved 90.4% on the aforementioned test dataset and was higher than that of the other methods that just employed quality indices of waveform characters. Actually, some ECG recordings can be correctly and fast classified as unacceptable ones by some obvious quality indices of waveform characters (i.e., flat line and high amplitude) without that of frequency characters. Therefore, some assessment methods combining waveform quality indices with frequency indices not only improve assessment accuracy but also help for calculation efficiency. Clifford et al. [7] and Zhang et al. [8] proposed an assessing method based on combination of quality indices of time and frequency characters respectively. Comparing with the previous methods, multi quality indices including waveform and frequency indices can provide relatively more comprehensive information within ECG so that higher assessment accuracy can be achieved. However, in quality assessment, the ECG recordings to be assessed are raw ones without any pretreatment so that both of the acceptable and unacceptable ECG recordings contain a lot of noise. So, waveforms of the acceptable signals are also seemingly disorderly so that they are similar to that of unacceptable in many cases. Thus, waveform of ECG is not favor to improve assessment accuracy and generalization ability of the assessment methods, or even worse. Furthermore, frequency bandwidths of random noise within ECG often overlap with that of normal ECG so that it is difficult to discern noise and ECG on frequency domain. Therefore, it is necessary to seek new characters within ECG and the corresponding quality indices to more accurately reflect the inherent irregularity within ECG instead of random components so that they can help for improving classification accuracy of quality assessment.

In fact, normal and abnormal ECGs should exhibit the inherent nonlinear irregularity instead of randomness when content of noise and unexpected components within ECG are relatively lower, otherwise the ECG exhibits randomness. So we can evaluate quality of ECG by measuring complexity of the ECG. At present, some common complexity algorithms, i.e., classical Lempel-Ziv complexity (CLZC), approximate entropy (ApEn), and sample entropy (SampEn) have been extensively used for biomedical signal analysis [9-12]. Zhang et al. [13] evaluated performance of CLZC on ECG quality assessment and concluded that the CLZC was only sensitive to content of high frequency noise. In practice the ECG are usually contaminated by the mixed noise composed from the high frequency, low frequency and power line noise instead of a high frequency noise, so CLZC shows an unsatisfactory classification performance when it is used for quality assessment. In addition, the ApEn and SampEn have been not yet reported to be used for assessing ECG quality. Zhang et al. [14] proposed an encoding Lempel-Ziv complexity (ELZC) to analyze the irregularity of physiological signal, and the algorithm could better discern randomness and the inherent irregularity within time series than other Lempel-Ziv complexity algorithms.

In this study, we performed the ApEn, SampEn, CLZC, and ELZC algorithms on three experimental schemes respectively, aiming to find a satisfactory complexity algorithm and its corresponding quality indices for ECG quality assessment. Firstly, we compared performance of the four complexity algorithms

on discerning the inherent irregularity within time series and randomness caused by noise and unexpected components. Furthermore we evaluated performance of the complexity algorithms on reflecting different noises and their content level within the ECG. Finally, we proposed three quality indices based on complexity values of the multi-lead ECG, and used the indices to assess quality of 862 real 12-lead ECG recordings selected from the CinC2011 dataset of the MIT database. Receiver operating characteristic curves (ROC) and the corresponding areas under the curve (AUC) of the quality indices were calculated respectively. The rest of this study is organized as follows: Section 2 describes the four complexity methods and the dataset including artificial time series and real ECG recordings. The experimental results and discusses are described in Section 3 and Section 4, respectively. Section 5 concludes the current study.

## 2. Materials and Methods

### 2.1 Artificial Synthetic and Real ECGs

This study employed three datasets including six typical artificial time series, the artificial synthetic noisy ECG and the real 12-lead ECG from the CinC2011 database.

First, in this study, six typical artificial time series were generated for evaluating performance of the aforementioned four complexity algorithms on discerning the inherent irregularity within time series and randomness caused by random component. Gaussian noise represents pure noise. The first mixed noise series MIX(0.4) mean that 40% of a period series of length  $N$  are randomly chosen and replaced with independent identically distributed random noise. The other mixed noise series MIX(0.2) is generated by the same way. Logistic (Logi) mapping series represent nonlinear irregularity series and are defined as

$$x_{n+1} = \mu \times x_n \times (1 - x_n) \quad (1)$$

where  $1 < \mu \leq 4$ . Logi(4.0) series represent the Logistic series that  $\mu$  is set as 4.0. Similarly, Logi(3.8) and Logi(3.5) are also generated. Actually the Logi(3.5) represent the periodic series.

Second, in this study, the artificial synthetic noisy ECG was constructed by two signals, a signal was the clear artificial ECG generated by the open source software ECGSYN [15], and the other one was the common real noise within ECG from the MIT Noise Stress Test Database (NSTDB) [16,17]. The NSTDB provides three common noises that can be typically found in real clinical application—i.e., baseline wander (BW), electrode motion (EM), and muscle artefacts (MA)—and each of three noises includes 48 real noise series. In addition, the other typical common noise, i.e., power line (PL) is also used for synthesizing the noisy ECG. In fact, the ECG is usually contaminated by a mixed noise. So we built a hybrid noise with each of the aforementioned noises on 25% proportion, and added to the clean artificial ECG. Here, we generated 48 clear artificial ECGs duration of 10 seconds with 360 Hz sampling rate and heart rate from 50 to 100 beats per minute. The clear artificial ECG was set to a 360-Hz sampling rate because sample rate of three common noises from the NSTDB is 360 Hz so that we can expediently construct the artificial synthetic noisy ECG.

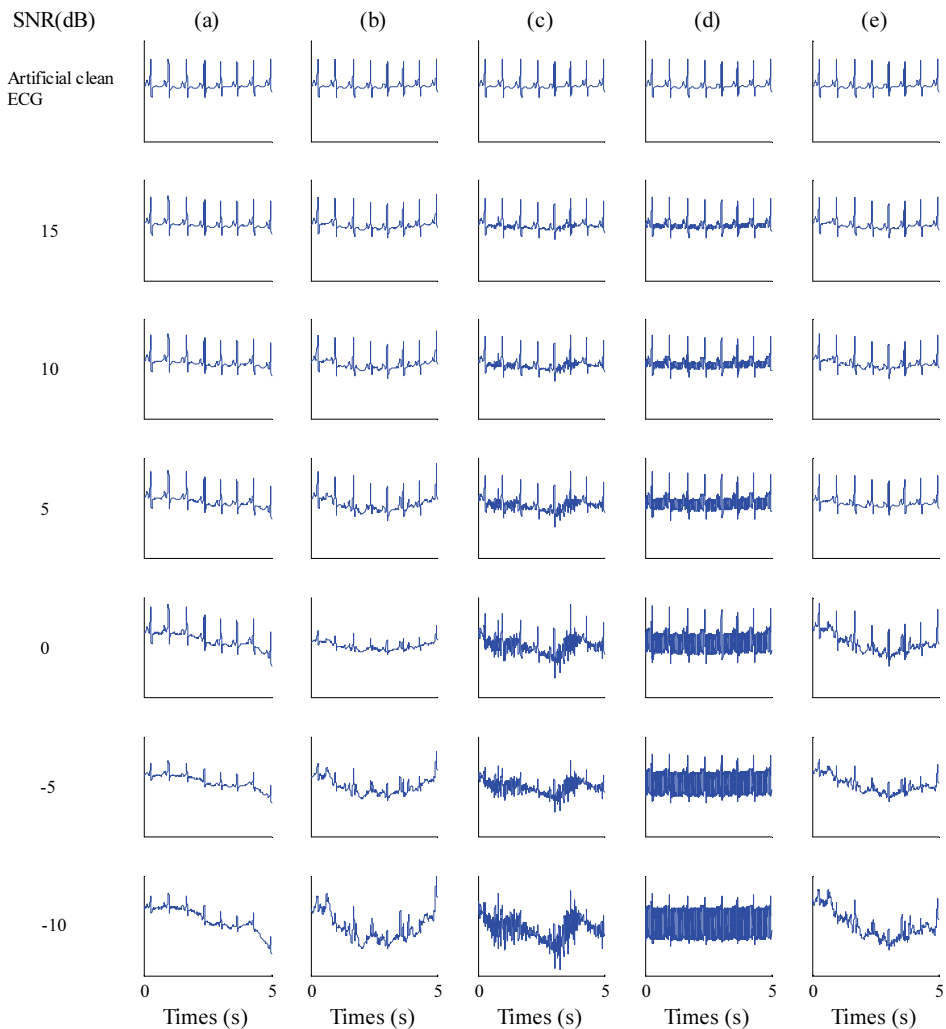
In this study, we generated five kinds of artificial synthetic noisy ECGs, i.e., the clean ECG plus BW, EM, MA, PL and hybrid noise. In practice, ECG usually cannot be used for clinical purpose when signal-

to-noise rate (SNR) is lower than -10 dB because the main waveform of ECG cannot be recognized due to a lot of noise. So SNRs of the five synthetic ECGs are from -10 dB to 15 dB with steps of 5 dB. For each SNR level, 48 repeats were yielded in this study. Finally each of five synthetic noisy ECGs with 360 Hz sampling rate was transformed to 500 Hz sampling rate using the resample function of MATLAB in this study. Fig. 1 shows the aforementioned five artificial synthetic noisy ECGs with SNR from -10 dB to 15 dB with steps of 5 dB.

SNR reflects the content level of noise within time series, and it is defined as

$$SNR = 10 \times \log_{10} \left( \frac{P_{signal}}{P_{noise}} \right) \quad (2)$$

where  $P_{signal}$  and  $P_{noise}$  denote the power of the clean ECG and the power of the noise, respectively.



**Fig. 1.** The artificial clean ECG and the artificial synthetic noisy ECG with SNR from -10 dB to 15 dB: (a) the clean and the clean plus BW noise, (b) the clean and the clean plus EM, (c) the clean and the clean plus MA, (d) the clean and the clean plus PL, and (e) the clean and the clean plus mixed noise.

Third, the real data was selected from training dataset of the CinC2011 [2] in this study. The training dataset includes 1000 12-lead ECGs with sampling rate 500 Hz, and it is available to all researchers and given quality label as “acceptable” or “unacceptable”. In this dataset, each recording lasts for 10 seconds. It is unnecessary to calculate the complexity of ECG recordings with lead-fall because their waveforms look like a straight line so that the signals are easy to be identified, furthermore the signals with lead-fall affect accuracy of the final results. So this study removed the ECG recordings with lead-fall from the training set, and then the remained ECGs included 767 acceptable and 95 unacceptable recordings.

## 2.2 The CLZC and ELZC

The CLZC consists of two processes, and the first process is performed to transform an original time series into a new binary symbolic sequence by comparing series  $X$  with a threshold, and the series is larger than the threshold, one maps the series to 1, otherwise, to 0. In this study, the mean of the series is usually selected as the threshold for the CLZC. Then the LZ value of the new symbolic sequence will be calculated on the traditional method [9]. In fact, the LZ complexity counter  $c(n)$  is closed related to the new subsequence of consecutive characters within the symbolic sequence. For calculating the counter  $c(n)$ , the symbolic sequence  $B$  is scanned from left to right and the  $c(n)$  is increased one unit when a new subsequence is encountered. First,  $S$  and  $Q$  are represented two subsequences of  $B$  respectively, and  $SQ$  is the concatenation of  $S$  and  $Q$ . The subsequence  $SQ\pi$  yielded from  $SQ$  after its last character is deleted ( $\pi$  is the operation deleting the last character in a sequence).  $v(SQ\pi)$  is the vocabulary of all subsequences of  $SQ\pi$ . Initially,  $c(n)=1$ ,  $S=s(1)$  and  $Q=s(2)$ , then  $SQ\pi=s(1)$ . Generally,  $S=s(1), s(2), s(3), \dots, s(r)$  and  $Q=s(r+1)$ , and so  $SQ\pi=S$ .  $Q$  is a subsequence of  $SQ\pi$  instead of a new sequence when it belongs to  $v(SQ\pi)$ . Then  $Q$  is replaced with  $s(r+1), s(r+2)$  and used to judge if it belongs to  $v(SQ\pi)$  or not. The aforementioned processes are repeat until  $Q= s(r+1), s(r+2), \dots, s(r+i)$  and it is a new sequence instead of a subsequence of  $SQ\pi$ , then  $c(n)=c(n)+1$ . Thereafter  $S=s(1), s(2), \dots, s(r+i)$  and  $Q=s(r+i+1)$ . This above procedure is repeated until  $Q$  is the last character of  $B$ . in fact the LZ complexity counter  $c(n)$  may be normalized as the  $C(n)$

$$C(n) = c(n) \frac{\log_{\alpha}(n)}{n} \tag{3}$$

where  $n$  is the length of signal  $X$ ,  $\alpha$  is the number of possible symbols contained in the new sequence. Generally, the normalized complexity  $C(n)$  is adopted instead of  $c(n)$  in practice.

For the ELZC, we leverage a novel symbolic process to generate an 8-state symbolic (3-bit binary) sequence, then the LZ value is calculated by the aforementioned calculation process. The novel symbolical process consists of the following steps [14].

Each  $x_i$  within the original signal  $X=x_1, x_2, \dots, x_n$  is transformed into a 3-bit binary symbol  $b_1(i)b_2(i)b_3(i)$ .

Step 1: the  $b_1(i)$  is determined by comparing  $x_i$  with the mean of signal  $X$ , and  $b_1(i)$  is set 0 when the  $x_i$  is less than the mean, otherwise the  $b_1(i)$  is 1.

Step 2: the  $b_2(i)$  is 0 when the difference between  $x_i$  and  $x_{i-1}$  is less than 0, otherwise the  $b_2(i)$  is set to 1. Initially,  $b_2(1)$  is set to 0.

Step 3: calculated process of the third digit  $b_3(i)$  is relatively complex, a variable *Flag* is first denoted as follows:

$$Flag(i) = \begin{cases} 0 & \text{if } |x_i - x_{i-1}| < dm \\ 1 & \text{if } |x_i - x_{i-1}| \geq dm \end{cases}, i = 2, 3, \dots, n \quad (4)$$

where  $dm$  is the mean distance between adjacent points within signal  $X$ . Subsequently,  $b_3(i)$  is calculated as follows:

$$b_3(i) = NOT(b_2(i) XOR Flag(i)), i = 2, 3, \dots, n \quad (5)$$

where  $b_3(1)$  is 0. The detailed symbolic process is described in [14].

### 2.3 The ApEn and SampEn

Pincus [18] proposed ApEn as a metric to quantified regularity of a time series, and meaning the probability of new pattern within time series when the dimension increases from  $m$  to  $m+1$ . The process of calculating ApEn is described as follows: Let  $S$  be a time series of length  $N$  and  $S = s_1, \dots, s_N$ , and reconstruct a vector  $x_i$  of the embedded dimension  $m$  and  $x_i = s_i, s_{i+1}, s_{i+2}, \dots, s_{i+m-1}$  for  $1 \leq i \leq N-m+1$  where  $m$  indicates the embedding dimension. The distance  $d_{ij}$  between the two vectors  $x_i$  and  $x_j$  is calculated where  $1 \leq i, j \leq N-m+1$ .

$$d_{ij} = \max |s(i+k) - s(j+k)|, k = 0, 1, \dots, m-1 \quad (6)$$

For each vector  $x_i$ , the number of the distance  $d_{ij}$  within  $r \times Std$  is found where  $Std$  is the standard deviation of the time series  $S$ , and the ratio of the number to the total number of vector  $N-m+1$  is calculated as  $C_i^m(r)$ .

$$C_i^m = \frac{1}{N-m+1} num\{d_{ij} < r\}, i = 1, 2, \dots, N-m+1 \quad (7)$$

Then the average degree of similarity for all of  $i$  is defined as

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (8)$$

Similarly, when the embedded dimension is  $m+1$ , the corresponding  $C_i^{m+1}(r)$  and  $\Phi^{m+1}(r)$  can be obtained.

$$C_i^{m+1} = \frac{1}{N-m} num\{d_{ij} < r\}, i = 1, 2, \dots, N-m \quad (9)$$

$$\Phi^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r) \quad (10)$$

Then, the ApEn is described as follows:

$$ApEn = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \quad (11)$$

The length  $N$  is not infinite, so the ApEn can be calculated by Eq. (12) when  $N$  is a finite number.

$$ApEn = \Phi^m(r) - \Phi^{m+1}(r) \quad (12)$$

In fact, the SampEn is a modified complexity method based on ApEn. Comparing with the ApEn, the SampEn does not include self-similar patterns, and it also does not depend on data size. Finally the SampEn can be given as

$$SmpEn = -\ln[B^{m+1}(r) / B^m(r)] \quad (13)$$

where  $B^m(r)$  is the probability that two  $m$ -dimension vectors will match, and similarly  $B^{m+1}(r)$  is the probability that two  $m+1$  dimension vectors will match, and  $m$  is the embedding dimension [19].

In this study, the ApEn and SampEn were calculated with the typical  $m$  setting of 2 and  $r$  setting of  $0.20 \times SD$  where  $SD$  was the standard deviation of the data series [20,21].

## 2.4 Testing Performance on Identifying the Irregularity within Time Series

The inherent irregularity within ECG especially within abnormal signals is different from randomness caused by random noise and unexpected components, although they look disorderly. In fact a quality index has poor performance on quality assessment when it confuses the inherent irregularity within signals and randomness. So, we need to firstly analyze performance of the ApEn, SampEn, CLZC and ELZC algorithms on discerning the inherent irregularity and randomness before they were used as quality index. In this study, this test was designed to calculate the aforementioned four complexities on six typical artificial time series mentioned in Section 2.1. In this test, 20 samples for each type of series were employed, and length of each sample was 100, 500, 2000, and 5000 points.

## 2.5 Analyzing Sensitivity to Different Noises and Different SNRs

In ECG quality assessment, all ECG recordings including acceptable and unacceptable ones contain a lot of noise because they are raw signals without any pretreatment. So a satisfied quality index should be able to reflect content level of noise within ECG, furthermore it is also necessary to reflect different types of noise. In order to evaluate sensitivity of the four complexity algorithms to content level of several common noises contained in ECG, this test was designed to compared the ApEn, SampEn, CLZC, and ELZC values of five aforementioned synthetic noisy ECGs in Section 2.1 (i.e., the clear artificial ECG plus BW, EM, MA, PL, and hybrid noises) at different SNRs.

## 2.6 Verifying on the Real Data

Generally, we cannot judge the quality of a multi-lead ECG recording based on complexity value of a lead. So the statistics (i.e. the mean, maximum and standard deviation) of complexity values of a multi-lead ECG recording are proposed as quality indices to assess quality of the recording in this study, and we defined three quality indices of the ELZC as follows: Let  $ELZC_i$  denote the ELZC value of the  $i$ th lead of a 12-lead ECG recording,  $i = 1, 2, \dots, 12$ , and the three quality indices are the mean, maximum and standard deviation of the  $ELZC_i$  respectively. For comparing performance of quality indices of the ELZC, we also similarly calculated the corresponding quality indices of the ApEn, SampEn, and CLZC, respectively.

In this test, for evaluating performance of the three quality indices of the ELZC, we calculated their ROCs and the corresponding AUCs on the real 12-lead ECG recordings that were collected from the training set of the CinC 2011 and described in Section 2.1. Similarly, we calculated ROCs of the ApEn, SampEn and CLZC respectively. Then, for calculating sensitivity ( $Se$ ) and specificity ( $Sp$ ) of each of the three quality indices, a 12-lead ECG recording was marked as an acceptable signal when its quality index was less than preset threshold. Otherwise the recording was recognized as an unacceptable one. In this test, the range of the threshold was from 0 to 1 in steps of 0.005. The two indices  $Se$  and  $Sp$  are described as follows:

$$\begin{aligned} Se &= \text{true positive} / \text{total number of unacceptable ECG recordings} \\ Sp &= \text{true negative} / \text{total number of acceptable ECG recordings} \end{aligned} \quad (14)$$

where true positive is number of identified unacceptable recordings among the unacceptable ECG recordings, and true negative is number of identified acceptable recordings among the acceptable ECG recordings. Actually, normalization of data is necessary because the value ranges of the ApEn, SampEn, CLZC and ELZC are different. Therefore, the Min-Max normalization approach was employed to normalize the values of the four complexity approaches before calculating  $Se$  and  $Sp$  in this study. The normalization method is described as follows:

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (15)$$

where  $x$  and  $x^*$  represent original time series and the new series after normalizing, respectively.

## 3. Results

### 3.1 Results of Discerning Nonlinear Properties within Time Series

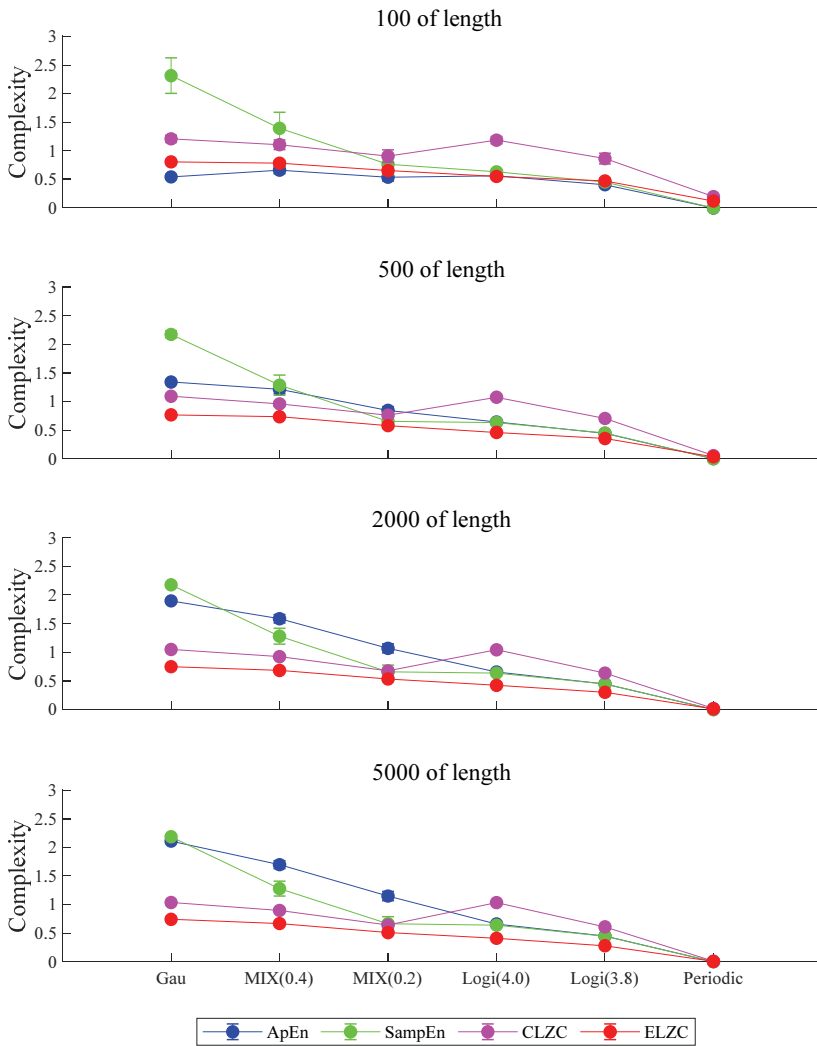
Fig. 2 shows ApEn, SampEn, CLZC and ELZC values of six aforementioned typical artificial time series including Gaussian noise, MIX(0.4), MIX(0.2), Logi(4.0), Logi(3.8) and periodic series on four time lengths 100, 500, 2000, and 5000, respectively. The ELZC and SampEn values exhibit monotonically decrease in order of Gaussian noise, MIX(0.4), MIX(0.2), Logi(4.0), Logi(3.8) and Logi(3.5) on all type time lengths. Similarly, the ApEn values monotonically decrease also in the aforementioned order when time lengths are 500, 2000, and 5000, respectively however the values exhibit fluctuations on MIX(0.4) and Logi(4.0) on 100 length. The CLZC values exhibit fluctuations between MIX(4.0) and Logi(4.0) on all time lengths.

### 3.2 Results of Sensitivity Analysis on SNR

Table 1 shows the ApEn, SampEn, CLZC, and ELZC values for five artificial synthetic noisy ECGs (i.e., the clean artificial ECG plus BW noise, the clean plus EM, the clean plus MA, the clean plus PL, and the clean plus hybrid noise) on several SNRs from -10 dB to 15 dB, in steps of 5 dB. The ELZC values exhibit monotonically decrease with increase of SNR on all of five artificial synthetic ECGs except



the clean ECG plus PL noise. In fact the ELZC values increase with increase of SNR on the clean ECG plus PL noise. The ApEn and CLZC values exhibit monotonically decrease with increase of SNR on the clean ECG plus MA noise. The SampEn values show the fluctuations on all of five artificial synthetic ECGs.



**Fig. 2.** The ApEn, SampEn, CLZC, and ELZC values of six typical artificial time series on time lengths 100, 500, 2000, and 5000, respectively.

### 3.3 ROC of the ELZC on Real Data

Fig. 3 shows ROC curves and their corresponding AUCs of three quality indices defined in Section 2.6 of the ApEn, SampEn, CLZC, and ELZC for classifying quality of the real 12-lead ECG recordings. The ROCs of the ELZC are the highest than that of the ApEn, SampEn, and CLZC on all of the quality indices. The corresponding AUCs show that the AUCs of the indices of ELZC are 0.505, 0.536, and 0.596, respectively. AUCs of the indices of the ApEn, SampEn, and CLZC are lower than 0.500.

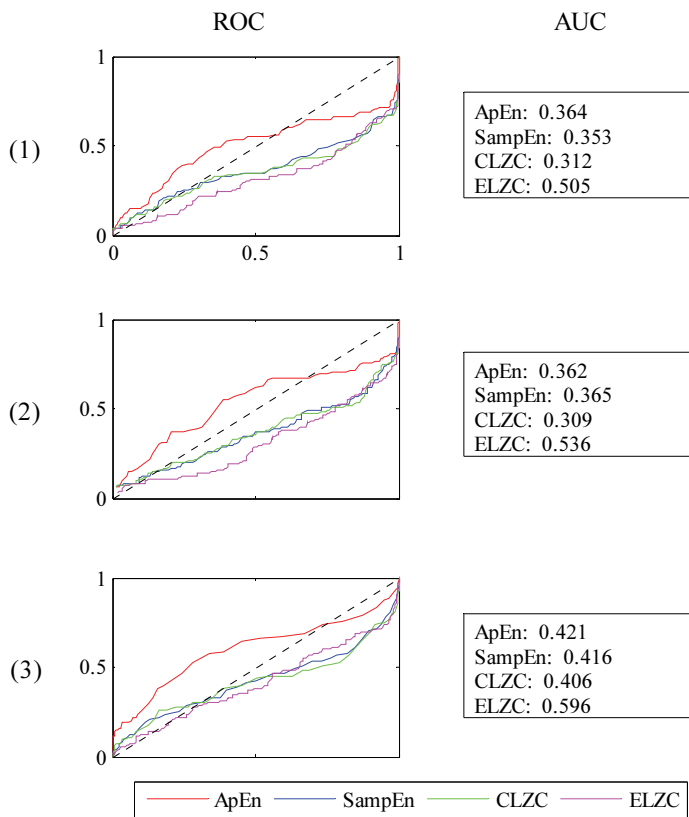
**Table 1.** The ApEn, SampEn, CLZC, and ELZC values for clean ECG plus BW, EM, MA, PL and mixed noise on several SNR from -10 dB to 15 dB with steps of 5 dB

Method	SNR (dB)	The clean plus BW noise	The clean plus EM noise	The clean plus MA noise	The clean plus PL noise	The clean plus hybrid noise
ApEn	-10	0.0977±0.0110	0.1602±0.0083	0.5524±0.0017	0.0806±0.0223	0.1883±0.0079
	-5	0.1324±0.0210	0.1866±0.0162	0.5523±0.0058	0.1554±0.0370	0.2127±0.0147
	0	0.1729±0.0177	0.2165±0.0158	0.5379±0.0208	0.2462±0.0626	0.2369±0.0159
	5	0.1718±0.0296	0.2067±0.0321	0.4874±0.0441	0.3762±0.1060	0.2219±0.0350
	10	0.1444±0.0317	0.1662±0.0341	0.3935±0.0653	0.4839±0.0647	0.1746±0.0357
	15	0.1175±0.0294	0.1304±0.0303	0.2737±0.0654	0.4799±0.0279	0.1339±0.0302
SampEn	-10	0.0919±0.0052	0.1280±0.0043	0.4201±0.0039	0.7877±0.0359	0.1510±0.0037
	-5	0.1048±0.0067	0.1384±0.0055	0.4222±0.0090	0.8631±0.0684	0.1598±0.0048
	0	0.1159±0.0098	0.1457±0.0107	0.4051±0.0258	0.9629±0.0901	0.1635±0.0112
	5	0.1138±0.0174	0.1382±0.0207	0.3500±0.0464	1.0069±0.0585	0.1509±0.0223
	10	0.0940±0.0193	0.1120±0.0226	0.2627±0.0529	0.8954±0.0937	0.1198±0.0241
	15	0.0708±0.0142	0.0833±0.0170	0.1772±0.0418	0.6627±0.1610	0.0872±0.0178
CLZC	-10	0.0716±0.0056	0.0855±0.0050	0.2804±0.0076	0.0877±0.0108	0.1129±0.0053
	-5	0.0778±0.0064	0.0922±0.0050	0.2727±0.0102	0.1001±0.0120	0.1160±0.0058
	0	0.0824±0.0075	0.0975±0.0069	0.2593±0.0144	0.1194±0.0107	0.1159±0.0066
	5	0.0836±0.0104	0.0984±0.0067	0.2404±0.0225	0.1331±0.0225	0.1155±0.0082
	10	0.0806±0.0111	0.0946±0.0086	0.2122±0.0320	0.1495±0.0260	0.1062±0.0111
	15	0.0774±0.0098	0.0881±0.0110	0.1760±0.0355	0.1598±0.0220	0.0948±0.0109
ELZC	-10	0.5666±0.0141	0.4967±0.0086	0.5067±0.0025	0.0877±0.0052	0.5460±0.0076
	-5	0.5290±0.0201	0.4741±0.0135	0.5036±0.0054	0.0941±0.0044	0.5265±0.0125
	0	0.4721±0.0282	0.4388±0.0183	0.4920±0.0099	0.1018±0.0065	0.4922±0.0183
	5	0.4009±0.0318	0.3911±0.0239	0.4757±0.0140	0.1111±0.0111	0.4416±0.0267
	10	0.3326±0.0280	0.3349±0.0278	0.4485±0.0187	0.1262±0.0126	0.3735±0.0340
	15	0.2855±0.0269	0.2801±0.0297	0.4101±0.0230	0.1387±0.0111	0.3076±0.0302

## 4. Discussions

Fig. 2 shows that the ELZC and SampEn have a satisfied performance for discerning the inherent irregularity within time series and randomness caused by random noise and unexpected components on all time lengths 100, 500, 2000, and 5000. The ApEn can achieve a satisfied performance on short signals i.e. time length 100 however it exhibits a poor performance on medium and long signals, i.e., time lengths 500, 2000, and 5000. The CLZC confuses the inherent irregularity within time series and randomness on all time lengths. The reason why the ELZC achieves the satisfied performance is that the ELZC is performed on a new sequence generated from the original time series for calculating complexity values. For the ELZC, the new sequence can reflect properly inherent properties within time series instead of randomness because the transformation process preserves the inherent information of the original series, losing random components, i.e., noise and unexpected components [14]. Fig. 2 means that the ELZC and

SampEn meet a fundamental requirement for assessing ECG quality because they can identify the inherent irregularity and randomness from seemingly disorderly signals, but the ApEn and CLZC cannot. However the SampEn relies on two parameters, i.e.,  $m$  and  $r$ , so that performance of the SampEn maybe poor when  $m$  and  $r$  are set to other values. In contrast, the ELZC presents a more reliable and satisfied performance on discerning the inherent irregularity and randomness within signals than other complexity approaches, i.e., the ApEn, SampEn, and CLZC.



**Fig. 3.** ROCs and their corresponding AUCs of three quality indices for classifying the real 12-lead ECG recordings, respectively: (1) the mean, (2) the maximum, and (3) the standard deviation.

In test 2, randomness caused by noise and unexpected components can be negligible in the clear artificial ECG because few random components are contained in the artificial clear ECG that are generated by the open source software ECGSYN. So the SNRs of the synthetic noisy ECG (i.e., the clean ECG mixed BW, the clean ECG mixed EM, the clean ECG mixed MA, and the clean ECG plus the hybrid noise) can be considered to reflect content level of the added noises (i.e., BW, EM, MA, PL, and mixed noises), then the randomness within the synthetic noisy ECG increases with decreasing of SNR. Pure PL noise is periodic signal, and its irregularity is nearly zero, so irregularity within the clean artificial ECG plus PL increases with increase of SNR. Actually, Table 1 shows that the ELZC exhibits the proper change trend on all of five synthetic noisy ECGs however the ApEn and CLZC exhibit incorrect change trend on all of the synthetic ECGs except the clean mixed MA and the clean mixed PL, and the SampEn

exhibits an improper trend on five synthetic ECGs. The results mean that the ELZC can be used to measure content level of random noise contained in ECG, but the ApEn, SampEn and CLZC cannot.

Fig. 3 shows the three quality indices, i.e., the mean, maximum and standard deviation of the ELZC of the real 12-lead ECG recordings for classifying quality of ECG have a relatively satisfied classification performance, and their corresponding AUCs, i.e. 0.505, 0.536, and 0.596 exhibit statistically significant ( $>0.500$ ). So the indices of ELZC can be employed for classifying quality of multi-lead ECG. Conversely, the AUCs of the three indices of the ApEn, SampEn, and CLZC are lower than 0.500 so that they cannot be used to assess quality of multi-lead ECG. Actually the three quality indices of the ELZC are employed to assess ECG quality in conjunction with other time domain and frequency domain quality indices.

## 5. Conclusions

In this study, we compared a novel ELZC with ApEn, SampEn, and CLZC for multi-lead ECG quality assessment. The results indicate that the ELZC achieves satisfied performance for distinguishing randomness and inherent irregularity within time series, and it can also efficiently reflect content level of noise contained in ECG because it exhibits proper change trend with increase of SNR on all of five synthetic noisy ECGs. Finally, we proposed three quality indices derived from complexity of the multi-lead ECG for multi-lead ECG, and validated the indices on the real 12-lead ECG signals from the CinC2011 of the MIT database. The results indicate the three indices (i.e., the mean, maximum and standard deviation) of the ELZC have statistical significance for multi-lead ECG quality assessment. In practice, three quality indices of the ELZC need to be combined with the other indices of time and frequency domain for ECG quality assessment.

## Acknowledgement

Project supported by the China Postdoctoral Science Foundation (No. 2017M612280), the National Natural Science Foundation of China (No. 61473174), and the Natural Science Foundation of Shandong Province, China (No. ZR2015AM015 and No. ZR201702190055).

## References

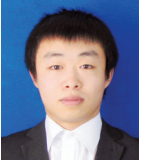
- [1] G. D. Clifford and G. B. Moody, "Signal quality in cardiorespiratory monitoring," *Physiological Measurement*, vol. 33, no. 9, article no. E01, 2012.
- [2] The PhysioNet Computing in Cardiology Challenge 2011 - Improving the Quality of ECGs Collected using Mobile Phones [Online]. Available from: <https://physionet.org/content/challenge-2011/1.0.0/>.
- [3] Q. Li and G. D. Clifford, "Signal quality and data fusion for false alarm reduction in the intensive care unit," *Journal of Electrocardiology*, vol. 45, no. 6, pp. 596-603, 2012.
- [4] P. Langley, L. Y. Di Marco, S. King, D. Duncan, C. Di Maria, W. Duan, M. Bojarnejad, D. Zheng, J. Allen, and A. Murray, "An algorithm for assessment of quality of ECGs acquired via mobile telephones," in *Proceedings of the 38th Computing in Cardiology*, Hangzhou, China, 2011, pp. 281-284.

- [5] J. Kuzilek, M. Huptych, V. Chudacek, J. Spilka, and L. Lhotska, "Data driven approach to ECG signal quality assessment using multistep SVM classification," in *Proceedings of the 38th Computing in Cardiology*, Hangzhou, China, 2011, pp. 453-455.
- [6] S. Zaunseder, R. Huhle and H. Malberg, "CinC challenge - Assessing the usability of ECG by ensemble decision trees," in *Proceedings of the 38th Computing in Cardiology*, Hangzhou, China, 2011, pp. 277-280.
- [7] G. D. Clifford, J. Behar, Q. Li, and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," *Physiological Measurement*, vol. 33, no. 9, pp. 1419-1433, 2012.
- [8] Y. T. Zhang, C. Y. Liu, S. S. Wei, C. Z. Wei, and F. F. Liu, "ECG quality assessment using a kernel support vector machine and genetic algorithm with a feature matrix," *Journal of Zhejiang University-SCIENCE C (Compute & Electron)*, vol. 15, no. 7, pp. 564-573, 2014.
- [9] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75-81, 1976.
- [10] K. Balasubramanian, S. S. Nair, and N. Nagaraj, "Classification of periodic, chaotic and random sequences using approximate entropy and Lempel-Ziv complexity measures," *Pramana*, vol. 84, no. 3, pp. 365-372, 2015.
- [11] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short datasets," *Annals of Biomedical Engineering*, vol. 41, no. 2, pp. 349-365, 2013.
- [12] A. G. Ravelo-Garcia, J. L. Navarro-Mesa, U. Casanova-Blancas, S. Martin-Gonzalez, P. Quintana-Morales, I. Guerra-Moreno, Jose M. Canino-Rodriguez, and E. Hernandez-Perez, "Application of the permutation entropy over the heart rate variability for the improvement of electrocardiogram-based sleep breathing pause detection," *Entropy*, vol. 17, no. 3, pp. 914-927, 2015.
- [13] Y. Zhang, S. Wei, C. Di Maria, and C. Liu, "Using Lempel-Ziv complexity to assess ECG signal quality," *Journal of Medical and Biological Engineering*, vol. 36, no. 5, pp. 625-634, 2016.
- [14] Y. Zhang, S. Wei, H. Liu, L. Zhao, and C. Liu, "A novel encoding Lempel-Ziv complexity algorithm for quantifying the irregularity of physiological time series," *Computer Methods and Programs in Biomedicine*, vol. 133, pp. 7-15, 2016.
- [15] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289-294, 2003.
- [16] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, 2001.
- [17] G. B. Moody, W. E. Muldrow, and R. G. Mark, "Noise stress test for arrhythmia detectors," in *Proceedings of the 11th Computing in Cardiology*, Park City, UT, 1984, pp. 381-384.
- [18] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297-2301, 1991.
- [19] Y. Chen, M. Oyama-Higa, and T. D. Pham, "Identification of mental disorders by hidden Markov modeling of photoplethysmograms," in *Biomedical Informatics and Technology*. Heidelberg: Springer, 2014, pp. 29-39.
- [20] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, article no. 021906, 2005
- [21] S. Zurek, P. Castiglioni, M. Kosmider, G. Parati, P. Guzik, and J. Piskorski, "A characteristic ridge-ledge in entropy surfaces of cardiovascular time series estimated by the Norm Component Matrix algorithm," in *Proceedings of the 39th Computing in Cardiology*, Krakow, Poland, 2012, pp. 281-284.



**Yatao Zhang** <https://orcid.org/0000-0002-6152-0806>

He received Ph.D. degree in Biomedical Engineering from Shandong University, China, in 2015. He is currently a senior Lecturer at School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, China. He has published 9 original journal/conference papers, and his research interests mainly focus on ECG signal processing, machine learning, data mining and big data processing for physiological signals.



**Zhenguo Ma** <https://orcid.org/0000-0001-7660-735X>

He received B.S. degree in School of Mechanical, Electrical & Information Engineering from Shandong University, Weihai, China, in Jun. 2018. Now he is a graduate student in School of Computer Science and Technology, University of Science and Technology of China. His current research interests include ECG signal classification and machine learning.



**Wentao Dong** <https://orcid.org/0000-0001-8632-3473>

He received M.S. degree in Department of Clinical Medicine from Weifang Medical University, Weifang, China, in 2007. Now he is a doctor in Department of Cardiovascular Surgery, Weihai municipal hospital, Weihai, China. His current research interests include Classification of arrhythmia and annotation of ECG.