УДК 616.12-008:004.891.3

Robust entropy estimator for heart rate variability

A. Martynenko[1], G. Raimondi[2], N. Budreiko[1]
[1]V. N. Karazin Kharkiv National University, School of Medicine, Ukraine
[2]University of Roma «Sapienza», Italy

Abstract

Introduction. Non-linear methods of analysis have found widespread use in the heart rate variability (HRV) technology, when the long-term HRV records are available. Using one of the effective nonlinear methods of analysis of HRV entropy for the standard 5-min HRV records is suppressed by unsatisfactory accuracy of available methods in case of short records (usually, doctors have 300–600 RRs during standard 5-min HRV record), as well as complexity and ambiguity of choosing additional parameters for known methods of calculating entropy ApEn, SampEn, MSE and etc.

The purpose of the work. Building a robust formula for calculating entropy (*EnRE*) with high accuracy for limited series of RR-intervals observed in a standard 5-minute HRV record, i. e. with $N \approx 300 - 600$. As well as demonstrating the capabilities of the *EnRE* formula on a series of statistical distributions and in diagnosing the congestive heart failure.

Materials and Methods. We used MIT-BIH long-term HRV records for Normal Sinus Rhythm (NSR) and Congestive Heart Failure (CHF). In order to analyze the accuracy of *EnRE*, we used the known statistical distributions (Normal, Uniform, Exponential, Lognormal, Pareto) with their precise entropy values.

The results of the study. We have shown the effectiveness of the developed *EnRE* formula for time series of limited length ($N = 300 - 600$) using the example of various types of statistical distributions that simulate human HRV ranges, and also demonstrated high accuracy of classification of cases of NSR and CHF for standard 5 min segments from MIT-BIH database of HRV records, using *EnRE* and *EnRE*(sort).

Conclusion. Proposed in the article is generalized form for Robust Entropy Estimator, which allows, for time series of limited length ($N = 300 - 600$), to calculate entropy value that differs from a precise one by 0,5–2,9%, as demonstrated for various random distributions. On standard 5-min segments from MIT-BIH database of HRV records, we have shown the usage of EnRE and EnRE (sort) for classification of cases of Normal Sinus Rhythm and Congestive Heart Failure with indicators of quality of division into groups: *Se* = 0,76; *Sp* = 0,98; *Acc* = 0,90.

Key words: Hearth rate variability; Entropy; Congestive heart failure.

ISSN 1812-7231 Klin. inform. telemed., 2019, vol. 14, iss. 15, pp. 67–73. https://doi.org/10.31071/kit2019.15.06
http://kit-journal.com.ua/en/index_en.html

References (34)

Introduction

The heart rate variability (HRV) is based on measuring (time) intervals between R-peaks (of RR-intervals) of an electrocardiogram (ECG) and plotting a rhythmogram on their basis with its subsequent analysis by various mathematical methods that are classified as Time-Domain, Frequency-Domain and Nonlinear [1, 2]. Using nonlinear methods in HRV and ECG analysis has proven to be very advantageous, and they are reviewed in detail, for example, in [3, 4]. Being constrained by a journal article's volume, we shall limit ourselves to the using of entropy in diagnostics of congestive heart failure (CHF). Congestive heart failure is a typical degeneration of the heart function characterized by the reduced ability of the heart to pump blood efficiently [5]. It is a difficult condition to manage, as are the CHF patients in clinical practice, having high mortality rate [6–10]. HRV analysis has given an insight into understanding the abnormalities of congestive heart failure, and can be used to identify the higher-risk CHF patients [11–15].

Nowadays, entropy-based measures, such as the typical approximate entropy (ApEn) [16] and sample entropy (SampEn) [17], are widely used in HRV analysis. SampEn is regarded as a modified version of ApEn, intended to solve such shortcomings as bias and relative inconsistency [17]. However, the traditional SampEn method is single-scale based and, therefore, fails to account for the multiple time scales inherent in cardiovascular systems [18–20]. Multiscale entropy (MSE) method was proposed in [20] and received much attention in the biomedical and mechanical fields [21–23]. Further MSE developing was transformed to multiscale multivariate entropy analysis [21, 23–26]. Existing entropy-based CHF studies are referred to in [27] and demonstrate high diagnostic effectiveness of entropy for HRV records with duration of more than 1000 RRs, i.e. greatly exceeding standard 5-min HRV records.

However, prevalence of the effective methodology of entropy analysis for standard 5-min HRV records is suppressed by unsatisfactory accuracy of available methods in case of short records: usually, doctors have 300–600 RRs during a standard 5-min HRV record; as well as complexity and ambiguity of choosing additional parameters for ApEn, SampEn, MSE and etc. Therefore, it appears there is a necessity for building a robust formula for calculating entropy with required accuracy for a limited series of RR-intervals observed in a standard 5-minute HRV record, i. e. with $N \approx 300 - 600$.

Materials and Methods

We used long-term HRV records by Massachusetts Institute of Technology – Boston's Beth Israel Hospital (MIT-BIH) from [28] (http://www.physionet.org), a free-access, on-line archive of physiological signals. Normal Sinus Rhythm (NSR) RR Interval Database includes beat annotation files for 54 long-term ECG recordings of subjects in normal sinus rhythm (30 men, aged 28,5 to 76, and 24 women, aged 58 to 73). Congestive Heart Failure (CHF) RR Interval Database includes beat annotation files for 29 long-term ECG recordings of subjects aged 34 to 79, with congestive heart failure (NYHA classes I, II, and III). Subjects include 8 men and 2 women; gender of the remaining 21 subjects is not known. The original electrocardiography (ECG) signals for both NSR and CHF RR interval databases were digitized at 128 Hz, and the beat annotations were obtained by automated analysis with manual review and correction.

We calculated entropy of limited time series by original formula proposed by Claude Shennon in 1948 [29] and here called Empirical Entropy (EnEmp) due to time series limitations:

$$EnEmp = -\sum_{i=1}^{N} P(x_i) ln\left(P(x_i)\right) \quad (1)$$

The problem of using the formula (1) in practice is its low accuracy for small number of points in a time series (e. g. $N<1000$), as well as slow descending to an accurate value with the increasing length of sequence. Shown in tab. 2 is the dependency of accuracy of calculating entropy according to formula (1) on the length of a series for a case of normal distribution, and in tab. 3 – for certain other types of distribution of a random value. Parameters of distribution for numerical analysis have been chosen according to coverage by a modelled distribution of a range of values of RR-intervals observed in a human (tab. 3). Accurate entropy values for associated distributions are given in tab. 1.

Special attention is given to a case of normal distribution, because with a lack of pathological changes, sinus rhythm proves to be rather close to normal distribution (see tab. 4a and tab. 4b).

Other distributions are considered as a necessary example of using numerical formulas for analysis of a limited series, because it is impossible in every specific case to precisely match the observed RR sequence with some definite random distribution. We can acknowledge the impossibility of applying the formula (1) to a short times series $N<1000$. Therefore, it appears there is a necessity for building a formula for calculating entropy with required accuracy for a limited series of RR-intervals observed in a standard 5-minute HRV record, i. e. with $N\approx 300-600$.

Tab. 1. Various probability distributions and correspondent Entropy [30].

Distribution	Probability density function	Entropy (*En*, nats)
Normal Distribution	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$En=ln\left(\sqrt{2\pi\sigma^2}\right)$
Uniform Distribution	$f(x)=\frac{x}{b-a}$	$En=ln(b-a)$
Exponential Distribution	$f(x)=\lambda exp(-\lambda x)$	$En=1-ln(\lambda)$
Lognormal Distribution	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}exp\left(-\frac{(ln(x)-\mu)^2}{2\sigma^2}\right)$	$En=\mu+ln\left(\sqrt{2\pi\sigma^2}\right)$
Pareto Distribution	$f(x)=\frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$	$En=ln\left(\frac{x_m}{\alpha}\right)+1+\frac{1}{\alpha}$

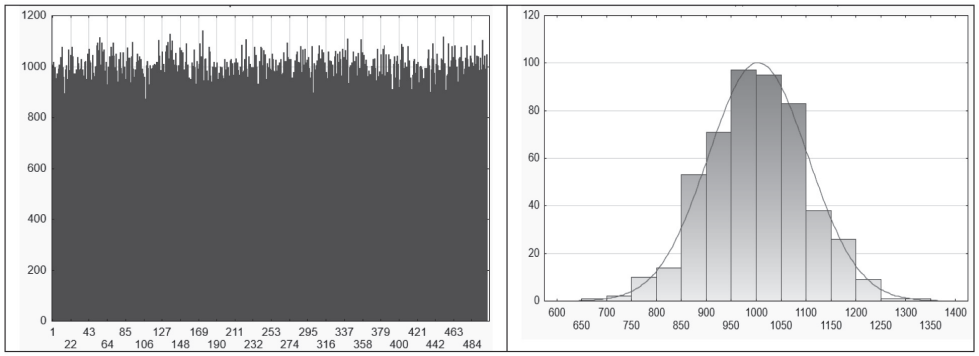
Tab. 2. Dependence of Entropy estimation accuracy from the length of time series for Normal Distribution ($\mu=1000$; $\sigma=100$).

Length of sample (N)	Accuracy of Entropy estimation (Relative Error, %)	
	Empirical Entropy (*EnEmp*)	Robust Entropy Estimator (*EnRE*)
N = 100	27,2	0,95
N = 300	17,3	0,45
N = 500	8,65	0,40
N = 1000	4,80	0,36
N = 5000	1,23	0,35
N = 10000	0,66	0,35
N = 20000	0,38	0,35

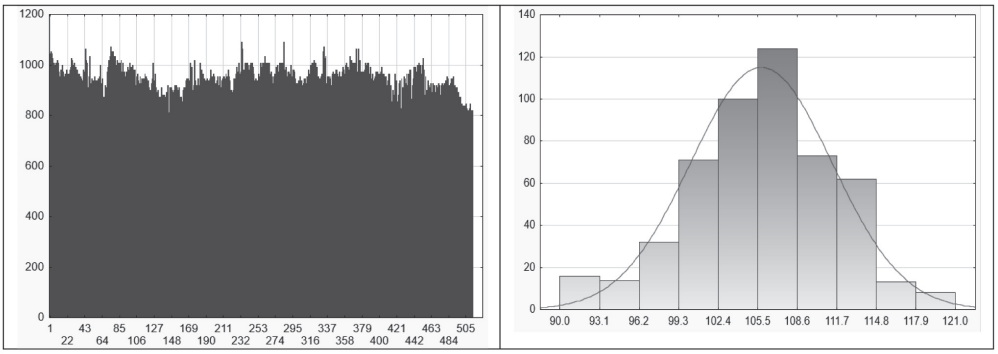
Tab. 3. Dependence from the length of time series of Entropy estimation accuracy and Correlation along simulated distribution parameters for various probability distributions.

Distribution	Length of sample	Empirical Entropy (*EnImp*)		Robust Entropy Estimator (*EnRE*)	
		Accuracy (Relative Error, %)	Correlation	Accuracy (Relative Error, %)	Correlation
Normal Distribution (M = 1000; σ = 100–200)	N = 100	29,99	0,46	1,58	0,95
	N = 500	11,51	0,945	0,51	0,987
	N = 1000	6,56	0,987	0,36	0,993
Uniform Distribution (a = 100–1000; b = 1500)	N = 100	32,8	0,78	2,9	0,99
	N = 500	13,40	0,973	2,89	0,998
	N = 1000	7,53	0,990	2,85	0,999
Exponential Distribution (λ = 0,0001–0,0011)	N = 100	46,2	0,45	2,9	0,98
	N = 500	28,38	0,903	2,83	0,997
	N = 1000	21,30	0,950	2,80	0,999
Lognormal Distribution (μ = 7; σ = 0,002–0,012)	N = 100	6,46	0,98	3,73	0,989
	N = 500	1,62	0,997	1,43	0,998
	N = 1000	0,985	0,999	1,04	0,999
Pareto Distribution (α = 2; s = 1000–2000)	N = 100	36,68	0,489	3,79	0,875
	N = 500	20,38	0,867	2,93	0,97
	N = 1000	15,65	0,946	2,75	0,987

Tab. 4a. RR simulated by Normal distribution (μ = 1000; σ = 50).



Tab. 4b. Real RR record (MIT Normal Sinus Rhythm RR Interval Database by Physionet.org).



Early in the last century, an Italian statistics professor Corrado Gini had proposed a way to measure the inequality among values of a frequency distribution (the Gini coefficient) [31]:

$$G = \frac{1}{2N^2 M} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\left| x_i - x_j \right| \right), \quad (2)$$

where M is a mean of x values. The Gini coefficient proved to be very popular in economics and sociology, and there are attempts to apply it to other areas as well, including HRV analyzing [32]. The Gini coefficient is an instance of generalized inequality index [33], and its alternative, as measure of deviation from balance – generalized entropy index – is derived from information theory as a measure of redundancy in data [34]. There are known limitations when using the Gini coefficient for data analysis: dependence on additive change of mean; a small selection substantially decreases the magnitude of the coefficient and etc.

Therefore, following the analysis of known definitions of measures of deviation from balance and degree of order, and following the discussion with colleagues, professor A. Martynenko has proposed generalized form for Robust Entropy Estimator (*EnRE*) for heart rate variability:

$$EnRE = ln\left(\frac{A}{N^{l/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\left(\left| RR_i - RR_j \right| \left| RR_j - MD \right| \right)^{1/k}}{\left(D_{ij} \right)^{m/2}} \right) \right), \quad (3)$$

where MD is median of RR values; D_{ij} – distance between RR_i and RR_j; A, l, m, k – estimated coefficients. Search conditions for coefficients A, l, m, k are the following:

1) accurate approximation for known distributions of a random value in ranges that represent models of RRs for heart rate variability;

2) independence of *EnRE* from *N* for initial time series and for series after sorting;

3) independence of *EnRE* from additive changes of mean.

After numerical researches, final results of which are presented in tab. 2 and tab. 3, the following coefficient values had been found: $A = \sqrt{\frac{2\pi e}{5}}$, $l = 3$, $m = 1$, $k = 2$. Let us note important characteristics of the found generalized form of *EnRE* and coefficients:

1) form of recording (3) and found coefficients *l, m, k* provide independence from additive change of mean series and from magnitude of selection N for basic series and for series after sorting;

2) value *EnRE* is sensitive to structural changes in series, such as, for example, sorting which increases the degree of order in series, decreasing the *EnRE*. This offers additional advantages in research, as shown below for a case of NSR and CHF groups classification;

3) readjusting coefficient *A* alone may be required to find the best *EnRE* value in another range of change in parameters of various random distributions, which can always be done using the method of least squares.

Mostly, we have used measures common for medicine, statistical measures of the performance of a binary classification test, – sensitivity (*Se*), specificity (*Sp*) and accuracy (*Acc*):

$$Se = TP/(TP + FN); \; Sp = TN/(TN + FP);$$
$$Acc = (TP + TN)/(TP + FN + FP + TN) \quad (4)$$

where TP is the number of the CHF patients correctly classified as the CHF group, TN is the number of the NSR subjects correctly classified as the NSR group, FP is the number of the NSR subjects falsely classified as the CHF group, and FN is the number of the CHF patients falsely classified as the NSR group.

The difference between means of the two independent selections (NSR and CHF) has been determined by a t-test for independent samples; the calculations for Discriminant Analysis have been done using statistical package Statistica 10.

Results and discussion

First of all, let us assess the accuracy of the proposed formula (3) for calculation of entropy:

1) given in tab. 2 are the values of *EnRE* for various lengths of series that represent a model of RR sequence by way of a normal distribution. It is evident that even for $N = 100$, relative error does not exceed 1%, while for $N = 300$ and more, the error is less than 0,5%;

2) given in tab. 3 are the results of numerical simulation of RR sequences of various length and by various types of random distributions, and given for these results are error values in calculating entropy and in correlation with precise values in case of changing the distribution parameters. Let us note, that in all cases for $N = 500$, relative error in calculating the precise value of entropy does not exceed 3%, while the magnitude of correlation is no worse than 0,97.

Given in tab. 5 are mean values of difference for RR entropy in the NSR group, calculated according to formula (1) and (3). Since in the NSR group RR distributions are close enough to a normal one (tab. 4), the difference between *EnEmp* and *EnRE* in case of changing the length of an RR segment behaves similarly to relative error of EnEmp in case of normal distribution (tab. 2). Consequently, in case of real HRV records, we can just as well rely on a conclusion derived from simulated distributions: it is impossible to use *EnEmp* for short RR segments ($N < 1000$), while the *EnRE* formula can be used for precise assessment of RR entropy, even in case of short records ($N \approx 300 - 600$).

Let us demonstrate the usage of *EnRE* for Detecting Congestive Heart Failure in short segments ($N = 500$) by MIT-BIH RR database. In [27], it has been shown that the minimal length of an RR-segment, for which it is possible to classify NSR and CHF groups by way of Multiscale Entropy Analysis, is $N = 1000$. The performance of such classification is: $Se = 0{,}70$; $Sp = 0{,}76$; $Acc = 0{,}74$. Given in tab. 6 are Mean and Standard deviation of *EnRE* for NSR and CHF for basic RR-intervals and series after sorting ($N = 500$). In both cases, the differences between groups are reliable to the degree of $p < 10^{-7}$.

Let us make an assessment: how greatly can entropy of series change after sorting in ascending order. For random series with equally probable outcomes, when every following element can be greater than, equal to, or less than the previous one, we have the following maximal value of structural entropy

$$EnStr = -3 \cdot \frac{1}{3} ln\left(\frac{1}{3} \right) = 1{,}099$$

After sorting without excluding the repetitions, when there are two equally probable outcomes, every following element of series is greater than or equal to the previous one, the maximal assessment of entropy will be

$$EnStr = -2 \cdot \frac{1}{2} ln\left(\frac{1}{2} \right) = 0{,}693$$

Finally, after sorting with excluding the repeating values, only one outcome is possible – every following element will be strictly greater than the previous one, and entropy will be equal to zero $EnStr = 0$.

Obviously, with real series, maximal assessments are not carried out, and in our case we deal with mean values (tab. 6)

Tab. 5. Dependence from the length of RR segment for mean difference between *EnEmp* and *EnRE* (NSR group).

Length of RR Segment (N)	Mean difference between *EnEmp* and *EnRE*
N = 300	16%
N = 500	14%
N = 600	11,5%
N = 800	9%
N = 1000	5%
N = 2000	0,4%

Tab. 6. Mean and Standard deviation of EnRE for NSR and CHF (original and sorted data).

HRV Record ($N = 500$)	Original RR		Sorted Time Series	
	EnRE	***p*–value**	***EnRE*(sort)**	***p*–value**
NSR	1,72 ± 0,47	10^{-7}	1,42 ± 0,51	10^{-7}
CHF	0,65 ± 0,76		0,55 ± 0,81	

NSR: *EnRE* − *EnRE*(sort) = 0,3,
CHF: *EnRE* − *EnRE*(sort) = 0,1,

which is expectedly less than maximal assessment of transition from basic series to sorted one (without excluding the repetitions): *EnRE* − *EnRE*(sort) = 0,403. Let us note a smaller change in entropy with sorting in case of a CHF group, than in case of an NSR, which points to a significant difference of CHF outcomes from an equally probable one (either greater than or equal to) and, as a consequence, to a greater aggregation of data between each other.

Using discriminant analysis, let us assess the quality of division into NSR and CHF groups and build classification functions:

1. When using one parameter of classification EnRE, we receive the following indicators of quality of division into NSR and CHF groups

Se = 0,66; Sp = 0,93; Acc = 0,83

and classification function (1 variable EnRE; Wilks' Lambda: 0,56 approx. F (1,81) = 62,43; $p < 10^{-7}$):

NSR = 4,96 * *EnRE* − 4,71
CHF = 1,87 * *EnRE* − 1,66

2. When using two parameters of classification *EnRE* & *EnRE*(sort) for sorted series, we receive the following indicators of quality of division into NSR and CHF groups

Se = 0,76; Sp = 0,98; Acc = 0,90

and classification function (2 variables *EnRE* and *EnRE*(sort); Wilks' Lambda: 0,44; approx. F (2,80) = 50,54; $p < 10^{-7}$):

NSR = 19,28 * *EnRE* − 13,85 * *EnRE*(sort) − 7,22
CHF = 6,91 * *EnRE* − 4,87 * *EnRE*(sort) − 1,97

Therefore, proposed generalized form for Robust Entropy Estimator (3) allows, with high accuracy, to divide NSR and CHF groups in short records ($N = 500$), which had remained unachieved in [27] by the way of Multiscale Entropy Analysis, and presents additional advantages provided by EnRE in case of structural changes in series (such as sorting). Quality of classification achieved by using two variables *EnRE* and *EnRE*(sort) is superior to results received in [27] by way of Multiscale Entropy Analysis for RR segments with length of $N = 1000$, $N = 2000$ and $N = 5000$.

Conclusions

Proposed in the article is generalized form for Robust Entropy Estimator (3), which allows, for time series of a limited length ($N = 300 - 600$), to find entropy value with high accuracy, which has been demonstrated for various types of random distributions (tab. 1–tab. 3). Parameters in generalized form for Robust Entropy Estimator (3) have been derived from the following criteria:

1) accurate approximation for known distributions of a random value in ranges that represent models of RRs for heart rate variability;

2) independence of *EnRE* from *N* for initial time series and for series after sorting;

3) independence of *EnRE* from additive changes of mean.

Important characteristics of the found generalized form of *EnRE* and coefficients are:

1) form of recording (3) and found coefficients l, m, k provide independence from additive change of mean series and from magnitude of selection N for basic series and for series after sorting;

2) value *EnRE* is sensitive to structural changes in series, such as, for example, sorting which increases the degree of order in series, decreasing the *EnRE*. This offers additional advantages in research, as shown for a case of NSR and CHF groups classification;

3) readjusting coefficient A alone may be required to find the best *EnRE* value in another range of change in parameters of various random distributions, which can always be done using the method of least squares.

Using the proposed generalized form for Robust Entropy Estimator (3) for MIT-BIH database of HRV records, we show in short records ($N = 500$) the usage of *EnRE* and *EnRE*(sort) for classification of Normal Sinus Rhythm (NSR) and Congestive

Heart Failure (CHF) cases. It is demonstrated, that, as opposed to [27] where by way of Multiscale Entropy Analysis it was possible to achieve division into NSR and CHF groups only from $N = 1000$ onwards, we have managed to perform said division for $N = 500$. Accuracy of classification achieved by using two variables *EnRE* and *EnRE*(sort) is superior to results received in [27] by way of Multiscale Entropy Analysis for RR segments with length of $N = 1000$, $N = 2000$ and $N = 5000$.

The studies were carried out in compliance with international bioethical standards and the provisions of the Helsinki Declaration (as amended in 2013). The authors of the article, A. Martynenko, G. Raimondi, N. Budreiko, confirm that they have no conflict of interest.

References

1. Task force of the European society of cardiology and the North American society of pacing and electrophysiology. Heart rate variability — standards of measurement, physiological interpretation, and clinical use. *Circulation*, 1996. vol. 93, iss. 5, pp.1043–1065.
2. Yabluchansky N., Martynenko A. [Heart Rate Variability for clinical practice]. 2010. Kharkiv, *Univer. Press*, 131 p. (In Russ.) depositary: http://dspace.univer.kharkov.ua/handle/ 123456789/1462.
3. Nayak S. K. at all. A Review on the Nonlinear Dynamical System Analysis of Electrocardiogram *Signal. J. Healthc. Eng*. 2018. Article ID 6920420, 19 p. doi: 10.1155/2018/6920420.
4. De Godoy M. F. Nonlinear Analysis of Heart Rate Variability: A Comprehensive Review. *J.Cardiology and Therapy,* 2016, vol. 3, no. 3, pp.528–533.
5. İşler Y., Kuntalp M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Comput. Biol. Med.*, 2007. iss. 37, pp. 1502–1510.
6. Wood A. J., Cohn J. N. The management of chronic heart failure. *N. Engl. J. Med.* 1996, iss. 335, pp. 490–498.
7. Nolan J. at all. Prospective study of heart rate variability and mortality in chronic heart. *Circulation* 1998, vol. 98, pp. 1510–1516.
8. Rector T. S., Cohn J. N. Prognosis in congestive heart failure. *Annu. Rev. Med.* 1994, vol. 45, pp. 341–350.
9. Arbolishvili G. N. at all. Heart rate variability in chronic heart failure and its role in prognosis of the disease. *Kardiologiia* [Cardiology]. 2005, vol. 46, pp. 4–11.
10. Smilde T. D. J., van Veldhuisen, D. J. and van den Berg, M. P. Prognostic value of heart rate variability and ventricular arrhythmias during 13-year follow-up in patients with mild to moderate heart failure. *Clin. Res. Cardiol.* 2009, vol. 98, pp. 233–239.
11. Acharya U. R. at all. Application of empirical mode decomposition (EMD) for automated identification of congestive heart failure using heart rate signals. *Neural Comput. Appl.* 2016, pp. 1–22, doi:10.1007/s00521-016-2612-1.
12. Yu, S. N., Lee M. Y. Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. *Comput. Biol. Med.* 2012, vol. 42, pp. 816–825.
13. Narin A., Isler Y., Ozer M. Investigating the performance improvement of hrv indices in chf using feature selection methods based on backward elimination and statistical significance. *Comput. Biol. Med.* 2013, vol. 45, pp. 72–79.
14. Jong T. L., Chang B., Kuo C. D. Optimal timing in screening patients with congestive heart failure and healthy subjects during circadian observation. *Ann. Biomed. Eng.* 2011, vol. 39, pp. 835–849.
15. Kumar M., Pachori R. B., Acharya U. R. Use of accumulated entropies for automated detection of congestive heart failure in flexible analytic wavelet transform framework based on short-term HRV signals. *Entropy* 2017, vol. 19, p. 92, doi:10.3390/e19030092.
16. Pincus S. M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA.* 1991, vol. 88, pp. 2297–2301.
17. Richman J. S., Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 2000, iss. 278, pp. H2039–H2049.
18. Humeau-Heurtier A. The multiscale entropy algorithm and its variants: A review. *Entropy* 2015, vol. 17, pp. 3110–3123.
19. Costa M., Goldberger A. L., Peng C. K. Multiscale entropy analysis of biological signals. *Phys. Rev. E.* 2005, vol. 71, Article ID 021906.
20. Costa M., Goldberger A.L., Peng C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* 2002, vol. 89, Article ID 068102.
21. Gao Z. K., Fang P. C., Ding M. S., Jin N. D. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. *Exp. Therm. Fluid Sci.* 2015, vol. 60, pp. 157–164.
22. Labate D. at all. Entropic measures of EEG complexity in Alzheimer's disease through a multivariate multiscale approach. *IEEE Sens. J.* 2013, vol. 13, pp. 3284–3292.
23. Azami H., Escudero J. Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals. *Physica A.* 2017, iss. 465, pp. 261–276.
24. Zhao L. N., Wei S. S., Tong H., Liu C. Y. Multivariable fuzzy measure entropy analysis for heart rate variability and heart sound amplitude variability. *Entropy.* 2016, vol. 18, p. 430. doi:10.3390/e18120430.
25. Li, P. at all. Multiscale multivariate fuzzy entropy analysis. *Acta Phys. Sin.* 2013, vol. 62, 120512.
26. Ahmed M. U., Mandic D. P. Multivariate multiscale entropy analysis. *IEEE Signal Proc. Lett.* 2012, vol. 19, .pp.91–94.
27. Chengyu L., Rui G. Multiscale Entropy Analysis of the Differential RR Interval Time Series Signal and Its Application in Detecting Congestive Heart Failure. *Entropy.* 2017, vol. 19, p. 251. doi:10.3390/e19060251.
28. Goldberger A. L. at all. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation.* 2000, vol. 101, pp. 215–220.
29. Shannon C. E. «A Mathematical Theory of Communication». *Bell System Technical Journal.* 1948. vol. 27, no 3, pp. 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
30. Lazo A. and Rathie P. «On the entropy of continuous probability distributions». *IEEE Transactions on Information Theory.* 1978, vol. 24, no. 1. doi:10.1109/TIT.1978.1055832.
31. Gini C. Variabilità e mutabilità. Reprinted in Pizetti, E.; Salvemini, T., eds. Memorie di metodologica statistica. Rome, Libreria Eredi Virgilio Veschi. 1955. 156 p.
32. Sànchez-Hechavarrìa M. E. and etc. Introduction of Application of Gini Coefficient to Heart Rate Variability Spectrum for Mental Stress Evaluation. *Arq Bras Cardiol.* 2019; [online]. ahead print, pp. 0–0. doi: 10.5935/abc.20190185.
33. Firebaugh G. Empirics of World Income Inequality. *American Journal of Sociology.* 1999, vol. 104, no. 6, pp. 1597–1630. doi:10.1086/210218.
34. Shorrocks A. F. The Class of Additively Decomposable Inequality Measures. *Econometrica*. 1980, vol. 48, no. 3, pp. 613–625. doi:10.2307/1913126.

Corresponding author

Professor **Alexander V. Martynenko**, PhD, D.Sc.
V. N. Karazin Kharkiv National University, School of Medicine,
Svobodi Sq., 4, Kharkiv, 61022, Ukraine
e-mail: Alexander.v.martynenko@karazin.ua
tel.: +380 97 248 1436

Робастна формула ентропії для аналізу варіабельності серцевого ритму

A. Martynenko[1], G. Raimondi[2], N. Budreiko[1]
[1]V. N. Karazin Kharkiv National University
[2]University of Roma «Sapienza»

Резюме

Вступ. Нелінійні методи аналізу знайшли широке застосування в технології варіабельності серцевого ритму (ВСР) при наявності довгих записів ВСР. Використання одного з ефективних методів нелінійних аналізу ВСР, — ентропії для стандартних 5-хв. записів ВСР стримується незадовільною точністю наявних методів в разі коротких записів (зазвичай, стандартна 5 хв. запис ВСР містить 300-600 комплексів RR), а також складністю і неоднозначністю вибору додаткових параметрів для відомих методів обчислення ентропії ApEn, SampEn, MSE.

Мета роботи. Побудова робастної формули обчислення ентропії (*EnRE*) з високою точністю для обмеженого ряду RR-інтервалів, які спостерігаються в стандартній 5-хвилинного запису HRV, тобто з $N \approx 300-600$. А також демонстрація можливостей формули *EnRE* на ряді статистичних розподілів і в діагностиці серцевої недостатності.

Матеріали та методи. У статті була використана база даних MIT-BIH з тривалими записами ВСР в разі нормального синусового ритму (NSR) і при серцевій недостатності (CHF). Для аналізу точності *EnRE* використовувалися відомі статистичні розподіли (нормальне, рівномірне, експоненціальне, логнормальне, Парето) з їх точними значеннями ентропії.

Результати. Була показана ефективність розробленої формули *EnRE* для тимчасового ряду обмеженої довжини ($N = 300-600$) на прикладі різних видів статистичних розподілів, що симулюють діапазони ВСР людини, а також продемонстрована висока точність класифікації випадків нормального синусового ритму і серцевою недостататочності для стандартних 5 хв. сегментів з бази даних MIT-BIH записів ВСР з використанням *EnRE* і *EnRE* (sort).

Висновки. У статті запропонована робастна формула обчислення ентропії (*EnRE*), яка дозволяє для часового ряду обмеженої довжини ($N = 300-600$) знайти значення ентропії, які відрізняються від точного на 0,5–2,9%, що було продемонстровано для різних випадкових розподілів. На стандартних 5 хв. сегментах з бази даних MIT-BIH записів ВСР було показано використання *EnRE* і *EnRE* (sort) для класифікації випадків нормального синусового ритму і серцевої недостататності з показниками якості поділу на групи: *Se* = 0,76; *Sp* = 0,98; *Acc* = 0,90.

Ключові слова: варіабельність серцевого ритму; ентропія; серцева недостатність.

Робастная формула энтропии для анализа вариабельности сердечного ритма

A. Martynenko[1], G. Raimondi[2], N. Budreiko[1]
[1]V. N. Karazin Kharkiv National University
[2]University of Roma «Sapienza»

Резюме

Введение. Нелинейные методы анализа нашли широкое применение в технологии вариабельности сердечного ритма (ВСР) при наличии длинных записей ВСР. Использование одного из эффективных методов нелинейных анализа ВСР, — энтропии для стандартных 5-мин. записей ВСР сдерживается неудовлетворительной точностью имеющихся методов в случае коротких записей (обычно, стандартная 5 мин. запись ВСР содержит 300–600 комплексов RR), а также сложностью и неоднозначностью выбора дополнительных параметров для известных методов вычисления энтропии ApEn, SampEn, MSE.

Цель работы. Построение робастной формулы вычисления энтропии (*EnRE*) с высокой точностью для ограниченного ряда RR-интервалов, наблюдаемых в стандартной 5-минутной записи HRV, т. е. с $N \approx 300-600$. А также демонстрация возможностей формулы *EnRE* на ряде статистических распределений и в диагностике сердечной недостаточности.

Материалы и методы. В статье была использована база данных MIT-BIH с продолжительными записями ВСР в случае нормального синусового ритма (NSR) и при сердечной недостаточности (CHF). Для анализа точности *EnRE* использовались известные статистические распределения (нормальное, равномерное, экспоненциальное, логнормальное, Парето) с их точными значениями энтропии.

Результаты. Была показана эффективность разработанной формулы *EnRE* для временного ряда ограниченной длины ($N = 300-600$) на примере различных видов статистических распределений, симулирующих диапазоны ВСР человека, а также продемонстрирована высокая точность классификации случаев нормального синусового ритма и сердечной недостататочности для стандартных 5 мин. сегментов из базы данных MIT-BIH записей ВСР с использованием *EnRE* и *EnRE* (sort).

Выводы. В статье предложена робастная формула вычисления энтропии (*EnRE*), которая позволяет для временного ряда ограниченной длины ($N = 300-600$) найти значение энтропии, отличающиеся от точного на 0,5–2,9%, что было продемонстрировано для различных случайных распределений. На стандартных 5 мин. сегментах из базы данных MIT-BIH записей ВСР было показано использование *EnRE* и *EnRE*(sort) для классификации случаев нормального синусового ритма и сердечной недостататочности с показателями качества разделения на группы: *Se* = 0,76; *Sp* = 0,98; *Acc* = 0,90.

Ключевые слова: вариабельность сердечного ритма; энтропия; сердечная недостаточность.