sciendo

DOI 10.2478/ama-2018-0047

acta mechanica et automatica, vol.12 no.4 (2018)

STATISTICAL ANALYSIS OF HUMAN HEART RHYTHM WITH INCREASED INFORMATIVENESS

Serhii LUPENKO[*], Nadiia LUTSYK[*], Oleh YASNIY[*], Łukasz SOBASZEK[**]

[*]Faculty of Computer Information Systems and Software Engineering, Ternopil Ivan Pul'uj National Technical University, 46001, Ruska str. 56, Ternopil, Ukraine
[**]Institute of Technological Systems of Information, Mechanical Engineering Faculty, Lublin University of Technology, ul. Nadbystrzycka 38 D, 20–618 Lublin

lupenko.san@gmail.com, lutsyk.nadiia@gmail.com, oleh.yasniy@gmail.com, l.sobaszek@pollub.pl

received 10 July 2018, revised 17 December 2018, accepted 20 December 2018

Abstract: The new methods of statistical analysis of heart rhythm were developed based on its generalized mathematical model in a form of random rhythm function, that allows to increase the informativeness and detailed analysis of heart rhythm in cardiovascular information systems. Three information criteria (BIC, AIC and AICc) were used to determine the cumulative distribution functions that best describe the sample and to assess the unknown parameters of distributions. The usage of the rhythm function to analyse heart rhythm allows to consider much better its time structure that is the basis to improve the accuracy of diagnosis of cardiac rhythm.

Key words: Mathematical Model, Heart Rhythm, Statistical Analysis, Diagnosis, Information System

1. INTRODUCTION

The study of heart rhythm is one of the most promising non-invasive methods of diagnosis of the cardiovascular system and the adaptive capacity of the human body, since it is a sensitive indicator of the degree of concordance and order in the functioning of the human body as an integral system (Akaike, 1974; Ciucurel et al., 2018; Evaristo et al., 2018; Fumagalli et al., 2018; Gadhoumi et al., 2018; Galeotti and Scully, 2018).

Analysis of heart rhythm allows to perform early diagnostics of abnormal fetal, to identify autonomic neuropathy in diabetic patients, to assess the risk of death after myocardial infarction, to determine the measure of tension of human body regulatory process state, etc (Hammad et al., 2018; Isler et al., 2019; Koichubekoc et al., 2018; Li et al., 2018; Mustaqeem et al., 2017; Napoli et al., 2018; Serrano and Figiola, 2009; Sharma and Sunkria, 2018; Shen et al., 2015; Wang et al., 2018).

In modern cardiovascular information systems, the research of heart rhythm is implemented by recording and automated processing of cardiointervalogram or rhythmogram (Brandão et al., 2014).

Cardiointervalogram is the sequence of values that are equal to the time intervals between peak values of R-wave of electrocardiogram in the sequential cardiac cycles. Rhythmogram is a discrete process that is defined on a finite or on a counTab. set of moments of time equal to the moment of time when the peak values of electrocardiogram R-wave are recorded and rhythm cardiogram values are equal to the time intervals between peak values of electrocardiogram R-wave. Rhythmogram is more informative empirical curve related with the analysis of heart rhythm as compared with cardiointervalogram because it contains information about the moments of time at which the peak values of R-wave of electrocardiogram are recorded (Bozhokin and Suslova, 2014).

However, this kind of curve does not allow to determine more detailed features of heart rhythm, since it reflects only changes in the duration of the cardiac cycle. Therefore, it does not take into account the entire set of time intervals between single-phase values of heart cyclic signal for every phase. The latter data enable to represent the rhythm in completely.

Cardiointervalogram and rhythmogram have insufficient precision and informativeness for the heart rhythm analysis (Kotel'nikov et al., 2002). This points out the relevance and prospects of developing a new approach to modeling and analysis of heart rhythm. This would allow increasing the level of informativeness, the growth of reliability of heart rhythm analysis that provide the early diagnosis of cardiac diseases and regulatory activity of the human body as a whole.

2. METHODOLOGY

2.1. New approach for the analysis of heart rhythm

The growth of reliability and informativeness of heart rhythm analysis can be achieved by statistical analysis of not only RR-intervals which define the duration of the cardiac cycle but also by identifying statistical patterns for the greater amount of time intervals that separate the single-phase intervals of electrocardiogram. The electrocardiogram as a registered electrocardiosignal $\xi(\omega, t)$ is convenient to be treated theoretically as a deterministic function $\xi_{\omega}(t)$ of real argument ($t \in R$). For example, the following single-phase counting of electrocardiogram may be the starts of P-wave, Q-wave, R-wave, S-wave, T-wave, U-wave. That is, the object of the statistical analysis of rhythm is a set of sequences of

DOI 10.2478/ama-2018-0047

time intervals between single-phase counting of electrocardiogram, namely, the sequences $TT(t_{T,n})$, $PP(t_{P,n})$, $QQ(t_{Q,n})$, $RR(t_{R,n})$, $SS(t_{S,n})$ determined by the formulae:

$$PP(t_{P,n}) = P(t_{P,n}) - P(t_{P,n-1})$$
$$QQ(t_{Q,n}) = Q(t_{Q,n}) - Q(t_{Q,n-1})$$
$$RR(t_{R,n}) = R(t_{R,n}) - R(t_{R,n-1})$$
$$SS(t_{S,n}) = S(t_{S,n}) - S(t_{S,n-1}) \quad (1)$$
$$TT(t_{T,n}) = T(t_{T,n}) - T(t_{T,n-1})$$
$$UU(t_{U,n}) = U(t_{U,n}) - U(t_{U,n-1}); n = \overline{2,N},$$

where: $P(t_{P,n})$, $Q(t_{Q,n})$, $R(t_{R,n})$, $S(t_{S,n})$, $T(t_{T,n})$, $U(t_{U,n})$ are the time moments of start of P-wave, Q-wave, R-wave, S-wave, T-wave, U-wave, respectively, n is a number of cardiac cycles.

Under this approach, the more informative discrete function is constructed. Theoretically, it can be as dense as necessary and can be transformed into a function of real argument $T(t)$, $t \in R$. This function is called the rhythm function of cyclic process [18]. In this case, the electrocardiosignal is such a cyclic process.

The formation of the above sequences and rhythm function from electrocardiogram is schematically presented in Fig. 1.

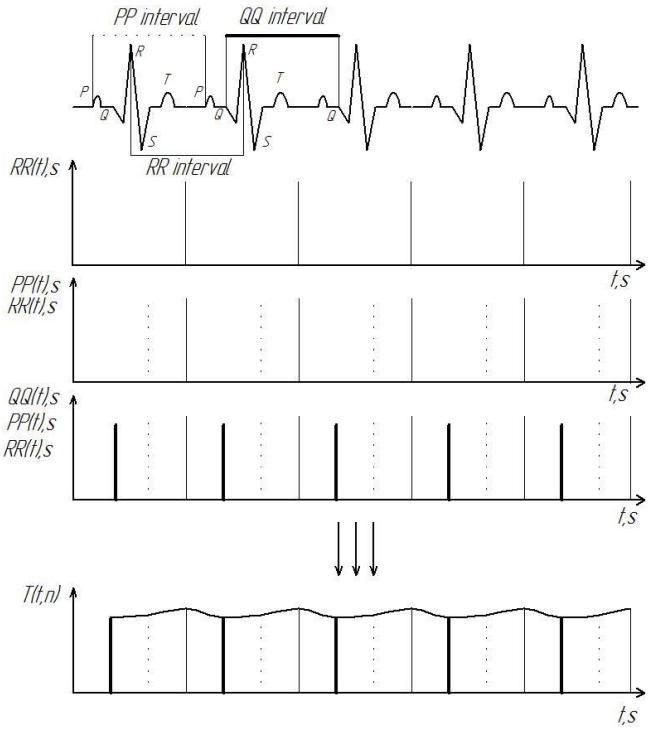

Fig. 1. Scheme of formation sequences of time intervals between single-phase counting of electrocardiogram

There were performed 20 experiments on the processing of electrocardiograms based on the presented above approach to the analysis of heart rhythm with a purpose of its verification. For example, let's present the results of one of these experiments. Namely, the electrocardiogram of male was registered. His age was 58 years. The graph of several cycles of registered electrocardiogram is shown in Fig. 2.

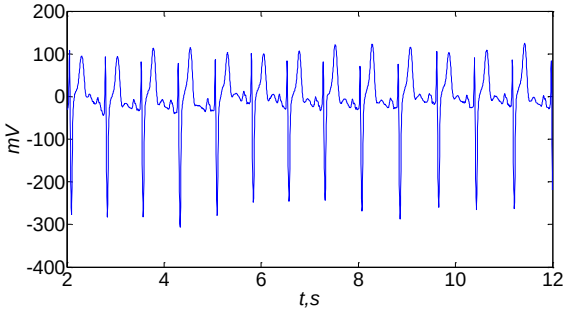

Fig. 2. The graph of registered electrocardiogram

The lengths of PP, TT, QQ, RR, SS intervals were estimated for every cardiac cycle of registered electrocardiogram by methods available in literature [19]. The plots of obtained sequences $TT(t_{T,n})$, $PP(t_{P,n})$, $QQ(t_{Q,n})$, $RR(t_{R,n})$, $SS(t_{S,n})$ are shown in Fig. 3.

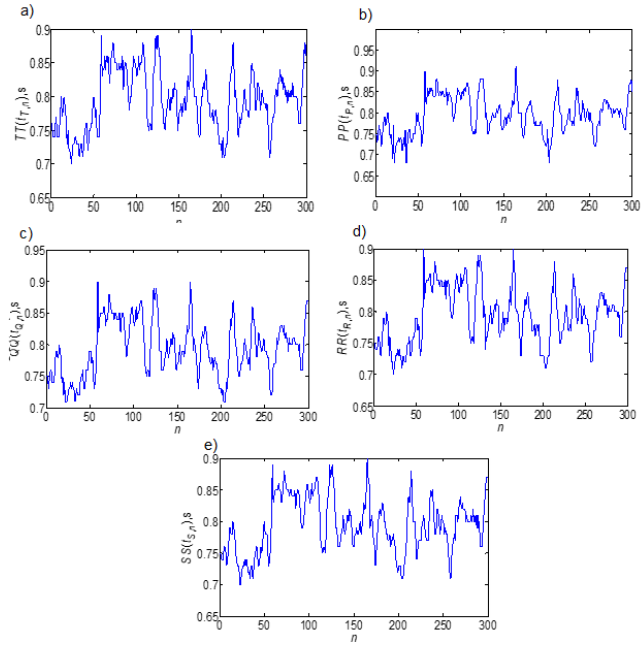

Fig. 3. Time intervals between a single-phase countings of ECG: a) $TT(t_{T,n})$, b) $PP(t_{P,n})$, c) $QQ(t_{Q,n})$, d) $RR(t_{R,n})$, e) $SS(t_{S,n})$

It is necessary to explain the mathematical model of the above mentioned sequences, in order to perform the statistical estimation of their stochastic characteristics. As in the case of mathematical models of rhythm cardiogram, these models of the above-indicated sequences can be both stationary and non-stationary random sequences. Therefore, the first step in explanation of the mathematical model as the foundation of statistical analysis of random sequences is the stationarity test, which can be done by applying the method of Foster-Stewart [8]. This method allows to check for trend components in the mathematical expectation (mean) and variance of the studied random sequences. Statistics of the criteria are as follows:

$$K = \sum_{i=2}^{n} K_i \quad (2)$$
$$d = \sum_{i=2}^{n} d_i \quad (3)$$

where: $d_i = u_i - l_i, K_i = u_i + l_i,$

$$u_i = \begin{cases} 1, \text{if the i}^{\text{th}} \text{ observation is an upper record} \\ 0, \text{otherwise.} \end{cases} \quad (4)$$

$$l_i = \begin{cases} 1, \text{if the i}^{\text{th}} \text{ observation is an lower record} \\ 0, \text{otherwise.} \end{cases} \quad (5)$$

The *K* statistics is used to check the trend in the variance of the random sequence, and the *d* statistics is employed to identify a trend in its mathematical expectation. It is obvious, that

$$0 \le K \le n-1 \text{ and } -(n-1) \le d \le n-1. \quad (6)$$

In the absence of trend, the random variables $t = \frac{d}{f}$ and $\tilde{t} = \frac{t-f^2}{l}$, where $l = \sqrt{2\ln n - 3.4253}$, $f = \sqrt{2\ln n - 0.8456}$ are distributed according to Student distribution with $v = n$ degrees of freedom.

If $|t|$, $|\tilde{t}| > t_{\frac{1+\alpha}{2}}$, then the null hypothesis of the absence of trend is rejected with the confidence level of α.

3. RESULTS

The values of corresponding statistics for different sequences $TT(t_{T,n})$, $PP(t_{P,n})$, $QQ(t_{Q,n})$, $RR(t_{R,n})$, $SS(t_{S,n})$ are presented in Tab. 1.

Tab. 1. The values of calculated statistics for sequences PP, TT, QQ, RR, SS of ECG (Fig. 3)

Intervals	$\lvert t\rvert$	$\lvert\tilde{t}\rvert$	$t_{0.975}$
TT	0	0.1966	1.9679
PP	0.6154	0.1989	1.9679
QQ	0.3076	1.2626	1.9679
RR	0	0.9068	1.9679
SS	0.3076	0.5550	1.9679

The verification of stationarity hypotheses confirmed that the mentioned above random sequences $TT(t_{T,n})$, $PP(t_{P,n})$, $QQ(t_{Q,n})$, $RR(t_{R,n})$, $SS(t_{S,n})$ do not contradict the hypothesis of stationarity.

The informative characteristics of heart rate analysis and methods of their study were determined. The significant characteristics of random sequences are their mathematical expectation, variance, probability density and cumulative distribution functions. Tab. 2 provides the values of mathematical expectation and variance of the given sequences.

Tab. 2. The mathematical expectation and variance of sequences TT, PP, QQ, RR, SS

Intervals	Expected value	Variance
TT	0.794	1.97e-3
PP	0.795	2e-3
QQ	0.795	1.9e-3
RR	0.795	1.95e-3
SS	0.795	1.94e-3

Three information criteria were used to determine the cumulative distribution functions that best describe the sample and the unknown parameters of distributions. These criteria are Akaike information criterion (AIC), Bayesian information criterion (BIC) and Akaike information criterion with a correction for finite sample sizes (AICc).

Tab. 3. The best fit distributions and their parameters obtained for sequences PP, TT, QQ, RR, SS

Interval	Inform.criter.	Name of distribution	Parameters of distribution
PP	AIC	GEV	$\xi = -0.299$ $\mu = 0.0448$ $\sigma = 0.7793$
	BIC	GEV	$\xi = -0.2802$ $\mu = 0.0435$ $\sigma = 0.7787$
	AICc	GEV	$\xi = -0.299$ $\mu = 0.0448$ $\sigma = 0.7793$
QQ	AIC	GEV	$\xi = -0.2819$ $\mu = 0.0433$ $\sigma = 0.779$
	BIC	GEV	$\xi = -0.2819$ $\mu = 0.0433$ $\sigma = 0.779$
	AICc	GEV	$\xi = -0.2819$ $\mu = 0.0433$ $\sigma = 0.779$
RR	AIC	GEV	$\xi = -0.283$ $\mu = 0.0435$ $\sigma = 0.779$
	BIC	GEV	$\xi = -0.283$ $\mu = 0.0435$ $\sigma = 0.779$
	AICc	GEV	$\xi = -0.283$ $\mu = 0.0435$ $\sigma = 0.779$
SS	AIC	GEV	$\xi = -0.2938$ $\mu = 0.0436$ $\sigma = 0.7792$
	BIC	GEV	$\xi = -0.2938$ $\mu = 0.0436$ $\sigma = 0.7792$
	AICc	GEV	$\xi = -0.2938$ $\mu = 0.0436$ $\sigma = 0.7792$
TT	AIC	GEV	$\xi = -0.2802$ $\mu = 0.0435$ $\sigma = 0.7787$
	BIC	Birnb.-Saund.	$\alpha = 0.7932$ $\beta = 0.0558$
	AICc	GEV	$\xi = -0.2802$ $\mu = 0.0435$ $\sigma = 0.7787$

The values of AIC information criterion can be found by the following formula (Akaike, 1974):

$$\text{AIC} = 2\ln L_{max} + 2k, \quad (7)$$

DOI 10.2478/ama-2018-0047

where: L_{max} is the maximum of likelihood function, k is the number of parameters of distribution. The best model is the one that minimizes the value of AIC. This criterion is obtained by the approximate minimization of Kullback-Leibler information entropy. It is a measure of the difference between the true distribution of data and the distribution of the model (Liddle, 2007).
The values of BIC information criterion can be determined by the following formula (Schwarz, 1978):

$$\mathrm{BIC} = -2\ln L_{max} + k\ln N, \tag{8}$$

where: N is the sample size.

While using BIC, the researcher suggests that the elements of the sample are independent and identically distributed. The application of AIC and BIC criteria usually shows good agreement between the conclusions about the best model (Liddle, 2007).

In the case when the sample size is insignificant, it is advisable to use the AICc. The value of this criterion is calculated using the formula (Sugiura, 1978):

$$\mathrm{AICc} = \mathrm{AIC} + 2\mathrm{k}(\mathrm{k}+1)/\mathrm{N}\text{-}\mathrm{k}\text{-}1. \tag{9}$$

As a result of the use of information criteria, it was found that these samples of different intervals (PP, TT, QQ, RR, SS) are best described by two distribution functions − generalized extreme values function (Fisher-Tippett distribution) and Birnbaum-Saunders distribution (see Tab. 3).

Recall that generalized extreme values (GEV) function has the following form:

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}. \tag{10}$$

where: ξ, μ, σ are the parameters of distribution (Coles, 2001).

The Birnbaum-Saunders (Birnb.-Saund.) distribution is described by the following expression:

$$F(x;\alpha,\beta) = \Phi\left(\frac{1}{\alpha}\left[\left(\frac{x}{\beta}\right)^{\frac{1}{2}} - \left(\frac{\beta}{x}\right)^{\frac{1}{2}}\right]\right). \tag{10}$$

where: $\Phi(x)$ is the cumulative function of normal distribution, α, β are the parameters of distribution (Coles, 2001).

The probability density function and cumulative distribution function of TT and PP intervals are shown in Fig. 4.

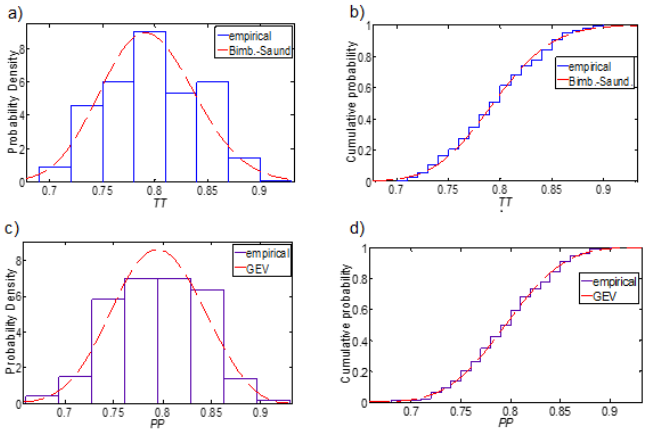

Fig. 4. The probability density function and cumulative distribution function of intervals a) TT interval, b) PP interval

It was found that the random sequences $TT(t_{T,n})$, $PP(t_{P,n})$, $QQ(t_{Q,n})$, $RR(t_{R,n})$, $SS(t_{S,n})$ are the stationary random processes. Their probabilistic characteristics, such as mathematical expectation, variance, probability density function and cumulative density function were estimated. The calculated mathematical expectation and variance are practically identical for every random sequence. This is also true for the estimated parameters of distributions. These sequences can be described well enough by Fisher–Tippett distribution.

4. CONCLUSIONS

Three information criteria (BIC, AIC and AICc) were used to determine the cumulative distribution functions that best describe the sample and to estimate the unknown parameters of distributions. It was found that the samples of different intervals (PP, TT, QQ, RR, SS) are well enough described by next distribution functions − generalized extreme values function (Fisher-Tippett distribution) and Birnbaum-Saunders distribution. Also, it was determined that the mentioned above random sequences are the stationary random processes. Their probabilistic characteristics, such as the mathematical expectation, variance, probability density function and cumulative density function were estimated. For this case, the calculated mathematical expectation and variance are practically identical for every random sequence. The mathematical expectation of samples was approximately equal to 0.795 s, and the variance is around 0.2 s^2. This is also true for the estimated parameters of distributions.

In general, the use of the rhythm function to analyze heart rhythm allows considering much better its time structure that is the basis to improve the accuracy of diagnosis of cardiac rhythm.

REFERENCES

1. **Akaike H.** (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(4), 716–723.
2. **Berkaya S.K., Uysal A.K., Gunal E.S, Ergin S., Gunal S., Gulmezoglu M.B.** (2018), A survey on ECG analysis, *Biomedical Signal Processing and Control*, 43, 216-235.
3. **Bozhokin S.V., Suslova I.B.** (2014), Wavelet Analysis of Non-stationary Signals in Medical Cyber-Physical Systems (MCPS). B S. Balandin, S. Andreev, & Y. Koucheryavy (Eds), *Internet of Things, Smart Spaces and Next Generation Networks and Systems*, Springer International Publishing.
4. **Brandão G.S., Sampaio A.A.C., Brandão G.S., Urbano J.J., Fonsêca N.T., Apostólico N., Oliveira E.F., Perez E.A., Almeida R.G., Dias I.S., Santos I.R., Nacif S.R., Oliveira L.V.F.** (2014), Analysis of heart rate variability in the measurement of the activity of the autonomic nervous system: technical note, *Manual Therapy, Posturology & Rehabilitation Journal*, 12, 243-251.
5. **Ciucurel C., Georgescu L., Iconaru E.I.** (2018), ECG response to submaximal exercise from the perspective of Golden Ratio harmonic rhythm, *Biomedical Signal Processing and Control*, 40, 156-162.
6. **Coles S.** (2001), *Extreme values, regular variation and point processes,* Springer, London.
7. **Evaristo R.M., Batista A.M., Viana R.L., Iarosz K.C., Szezech J.D. Jr., Godoy M.F.** (2018). Mathematical model with autoregressive process for electrocardiogram signals, *Communications in Nonlinear Science and Numerical Simulation*, 57, 415-421.
8. **Foster F.G., Stuart A.** (1954), Distribution-Free Tests in Time-Series Based on the Breaking of Records, *Journal of the Royal Statistical Society. Series B* (*Methodological*), 16(1), 1-22.

10. **Fumagalli F., Silver A.E**., **Tan Q., Zaidi N., Ristagno G.** (2018), Cardiac rhythm analysis during ongoing cardiopulmonary resuscitation using the Analysis During Compressions with Fast Reconfirmation technology, *Heart Rhythm*, 15(2), 248-255.
11. **Gadhoumi K., Do D., Badilini F., Pelter M.M., Hu X.** (2018), Wavelet leader multifractal analysis of heart rate variability in atrial fibrillation, *Journal of Electrocardiology*, 51(6), S83-S87.
12. **Galeotti L., Scully C.G.** (2018), A method to extract realistic artifacts from electrocardiogram recordings for robust algorithm testing, *Journal of Electrocardiology*, 51(6), S56-S60.
13. **Hammad M., Maher A., Wang K., Jiang F., Amrani M.** (2018), Detection of abnormal heart conditions based on characteristics of ECG signals, *Measurement*, 125, 634-644.
14. **Isler Y., Narin A., Ozer M., Perc M.** (2019), Multi-stage classification of congestive heart failure based on short-term heart rate variability, *Chaos, Solitons & Fractals*, 118, 145-151.
15. **Koichubekov B.K., Sorokina M.A., Laryushina Y.M., Turgunova L.G., Korshukov I.V.** (2018), Nonlinear analyses of heart rate variability in hypertension, *Annales de Cardiologie et d'Angéiologie*, 67(3), 174-179.
16. **Kotel'nikov S.A., Nozdrachev A.D., Odinak M.M., Shustov E.B., Kovalenko I.Yu., Davydenko V.Yu.** (2002), Cardiac Rhythm Variability: Approaches to Mechanisms, *Human Physiology*, 28(1), 114-127.
17. **Li J., Chen Ch., Yao Q., Zhang P., Wang J., Hu J., Feng F.** (2018), The effect of circadian rhythm on the correlation and multifractality of heart rate signals during exercise, *Physica A: Statistical Mechanics and its Applications*, 509, 1207-1213.
18. **Liddle A.R.** (2007), Information criteria for astrophysical model selection, *Monthly Notices of the Royal Astronomical Society: Letters*, 377(1), 74-78.
19. **Lupenko S., N. Lutsyk, Y. Lapusta.** (2015), Cyclic linear random process as a mathematical model of cyclic signals, *Acta Mechanica et Automatica*, 9(4), 219-224.
20. **Lytvynenko I., Maruschak P., Lupenko S., Panin S.** (2015), Segmentation and Statistical Processing of Geometric and Spatial Data on Self-Organized Surface Relief of Statically Deformed Aluminum Alloy, *Applied Mechanics and Materials*, 770, 288-293.
21. **Mustaqeem A., Anwar SM, Khan AR., Majid M.** (2017), A statistical analysis based recommender model for heart disease patients, *International Journal of Medical Informatics*, 108, 134-145.
22. **Napoli N.J., Demas M.W., Mendu S., Stephens C.L., Kennedy K.D, Harrivel A.R, Bailey R.E., Barnes L.E.** (2018), Uncertainty in heart rate complexity metrics caused by R-peak perturbations, *Computers in Biology and Medicine*, 103, 198-207.
23. **Schwarz G.** (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.
24. **Serrano E., Figliola A.** (2009), Wavelet Leaders: A new method to estimate the multifractal singularity spectra, *Physica A: Statistical Mechanics and its Applications*, 388(14), 2793-2805.
25. **Sharma L.D., Sunkaria R.K.** (2018), Stationary wavelet transform based technique for automated external defibrillator using optimally selected classifiers, *Measurement*, 125, 29-36.
26. **Shen C., Yu Z., Liu Z.** (2015), The use of statistics in heart rhythm research: a review, *Heart Rhythm*, 12(6), 1376-1386.
27. **Sugiura N.** (1978). Further analysts of the data by akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7(1), 13-26.
28. **Wang Y., Wei S., Zhang S., Zhang Y., Zhao L., Liu C., Murray A.** (2018), Comparison of time-domain, frequency-domain and non-linear analysis for distinguishing congestive heart failure patients from normal sinus rhythm subjects, *Biomedical Signal Processing and Control*, 42, 30-36.