DOI: https:// 10.15407/kvt196.02.003

UDC 004.021

FAINZILBERG L.S.[1]**,** DSc. (Engineering), Professor,
Chief Researcher of the Department
of Intelligent Automatic Systems
e-mail: fainzilberg@gmail.com
DYKACH Ju.R.[2]**,** Student of Biomedical Engineering Faculty,
e-mail: jul.dykach@gmail.com
[1] International Research and Training Center
for Information Technologies and Systems
of the National Academy of Sciences of Ukraine
and Ministry of Education and Science of Ukraine,
40, Acad. Glushkova av., Kyiv, 03187, Ukraine
[2] The National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»
37, Peremohy av., Kiev, 03056, Ukraine

LINGUISTIC APPROACH FOR ESTIMATION OF ELECTROCARDIOGRAMS'S SUBTLE CHANGES BASED ON THE LEVENSTEIN DISTANCE

Introduction. *The linguistic approach, based on the transition from electrocardiogram (ECG) to codogram, gained fame for the analysis of heart rhythm. To expand the functionality of the method, it is advisable to study the possibility of simultaneously monitoring the dynamics of changes in the duration of cardiac cycles and the indicator of simmetry T-wave that carries information about ischemic changes in the myocardium.*

The purpose of the article *is to develop algorithmic and software components to solve this problem and conduct experimental studies on model and real data.*

Methods. *ECG of certain groups was automatically encoded, Levenshtein distance was calculated between ECG pairs for group and the reference codogram of the group was constructed. The evaluation of the results of experimental studies was carried out on the basis of traditional methods of statistical analysis.*

Results. *It is shown that based on the Levenshtein distance between all pairs of codograms of the test group, the reference codogram of the group of patients with coronary artery disease (CAD) and the group of healthy volunteers can be determined. It was established that making decisions based on the comparison of the ECG codogram of the person with the reference codogram allows for the separation of representatives of the indicated groups with sensitivity S_E = 72% and specificity C_P = 79% even in those cases when the traditional analysis of the ECG in 12 leads is not informative.*

Conclusions. *The proposed approach allows to obtain additional diagnostic information when solving actual problems of practical medicine.*

Keywords: *linguistic approach, diagnostic sign of ECG, Levenshtein distance.*

INTRODUCTION

The physiological processes occurring in biological systems are often repetitive in time. Such processes generate specific signals, which are commonly called cyclical in the scientific literature [1, 2]. A typical example of a cyclical signal is an electrocardiogram (ECG), reflecting the cyclical nature of the circulatory system and respiratory organs of a living organism.

Despite the fact that electrocardiography for over a hundred years has been the most common method of functional diagnostics in cardiology, the sensitivity and specificity of traditional ECG examinations are not high enough. Thus, in work [3] it was shown that resting ECG, assessed by generally accepted criteria, remains normal in approximately half of patients with chronic coronary artery disease (CAD), including during episodes of chest discomfort.

Modern digital electrocardiographs, which implement traditional approaches to the analysis and interpretation of ECG, also do not provide the required reliability of diagnostic results. Moreover, experienced clinicians still prefer visual interpretation of ECG, not completely trusting computer algorithms, which, because of the complexity of real signals, often lead to errors at the stage of recognition of informative fragments [4].

Therefore, scientists are constantly looking for alternative approaches to computer processing of cyclic signals, in particular, ECG. One of these approaches is based on the transformation of the original signal into a word (sequence of characters), for the analysis of which the concepts of formal languages are used. Such an approach in various publications is called linguistic [5, 6], structural [7] or syntactic [8].

In cardiology practice, the linguistic approach has become famous for analyzing heart rhythm [9, 10]. It has long been known that the heart rhythm is a universal reaction of the organism to any influence from the external and internal environment [11]. Mathematical analysis of the heart rhythm allows you to obtain important information about the functional state of all parts of the regulation of human life, both in normal conditions and in various pathologies [12]. Computer technologies of mathematical analysis of heart rate variability of the heart rate (HRV) are still actively used to assess the state of the autonomic nervous system and adaptive reserves of the body [13].

However, only on the basis of control over the sequence of lengths of the $R-R$ -intervals it is impossible to judge the functional state of the heart itself as the main system-forming organ, in particular, the ischemia of the heart muscle. To increase the credibility of the conclusion about the functional state of the body, it is reasonable to supplement the analysis of the dynamics of the $R-R$ -intervals with an analysis of the dynamics of other ECG indicators.

In [14, 15], a method for analyzing an ECG signal was proposed, which provides for encoding ECG symbols of a given alphabet that carry information about the increment signs of both the $R-R$ -interval lengths and the amplitudes of R -wave of adjacent cycles. As a result, the observed ECG generates words (codograms), the processing of which by machine learning methods expands the capabilities of the mathematical analysis of heart rhythm.

The International Scientific and Training Center for Information Technologies and Systems of the National Academy of Sciences of Ukraine and the Ministry of Education and Science of Ukraine has developed an innovative method for processing electrocardiograms (ECG), which is called fasegraphy [16, 17]. A distinctive feature of the method is the use of intelligent IT for processing the observed time signal $z(t)$ on the phase plane $z(t),\dot{z}(t)$, where $\dot{z}(t)$ is the rate of change of the electrical activity of the heart [18]. This difference made it possible for the *first time* to implement a procedure for reliably determining the novel ECG feature (indicator β_T) characterizing the symmetry of the T-wave of the cardiac cycles [19].

Large-scale clinical studies conducted with the help of the FASEGRAPH® software and hardware complex, which implements fasegraphy method, confirmed that measuring the indicator β_T and evaluating its standard deviation RMS β_T makes it possible to increase the accuracy of detecting the initial signs of myocardial ischemia, even in cases where the analysis of traditional ECG features in 12 leads is not informative [20].

It follows that, remaining within the framework of linguistic analysis, it is advisable to study the possibility of simultaneously analysis the dynamics of changes $R-R$- intervals and dynamics of changes β_T-indicator on the sequence of cardiac cycles.

The purpose of the article is to develop algorithmic and software components for solving this problem.

BASIC COMPONENTS OF LINGUISTIC ECG ANALYSIS

Recall that the general scheme of linguistic analysis of the time signal $z(t)$ suggests segmentation $z(t)$ into a sequence of separate fragments, reflecting the alternation of elementary events during the development of the process under study [6]. Thus, a transition is made from the k-implementation of the signal $z_k(t)$ observed on a limited time interval $t\in[0,T]$ to the word $S_k=\alpha_1\alpha_2\cdots\alpha_K$ as *finite* chain of *characters* $\alpha_j\in \mathrm{A}$, $j=1,...,K$ from the alphabet of the "names" of the fragments. A set of all-possible words (not necessarily finite) forms a formal language for which the grammar is built [21]:

$$G=\langle\Omega_0,\Omega_T,P_G,\omega_0\rangle, \qquad (1)$$

where Ω_0 is a set of *non-terminal* symbols (variables); Ω_T — a set of *terminal* symbols (constants), $\Omega_T\cup\Omega_0=\mathrm{A}$, $\Omega_T\cap\Omega_0=\varnothing$; P_G — a set of *grammatical rules* (substitution rules); $\omega_0\in\Omega_0$ — *initial* (root) non-terminal character.

In the majority of works devoted to linguistic analysis of signals, it is assumed that the alphabet of reference fragments is known in advance [22], and the construction of grammar makers adequate to the set of observed signals is carried out by man on the basis of informal knowledge of an expert in the subject area.

Despite the fact that in a number of papers, particular, in [23], the main theoretical concepts of the general formulation of the problem of restoring grammars on a training set of observations were revealed, it should be stated that there are still no universal methods for practical solution of this difficult problem.

Therefore, following the works [14, 15], we will make the transition from the k-th observed signal $z_k(t)$ to the word $S_k = \alpha_1\alpha_2\cdots\alpha_K$, analyzing the sequence of sign differences in the values of ECG indices on adjacent cycles.

Let we have N ECG cycles and values of several indicators are determinedon on each of cycles , for example, the duration of RR intervals and the index β_T of T wave symmetry.

Denote each of these sequences as:

$$x_1, x_2, \ldots x_N . \tag{2}$$

Let us determine the values of the indicator variable V_i, $i = 2,\ldots,N$, by the signs of the increments of the quantities on each cycle in relation to the previous cycle:

$$V_i = \begin{cases} +1, \text{ if } & x_i - x_{i-1} > 0, \\ -1, \text{ if } & x_{i-1} - x_i > 0. \end{cases} \tag{3}$$

As a result, for the indicators RR and β_T , we will get two sequences of indicator variables $V_i^{(RR)}$ and $V_i^{(\beta)}$ respectively, and each ECG cycle will be encoded with one of the four symbols of the alphabet $\mathrm{A} = \{a, b, c, d\}$ as follows (Tabl. 1).

The sequence of characters received in accordance with Table 1 forms a N-bit word S_k (codogram) that uniquely encodes the k-th processed ECG.

For illustration, a flowchart of the codogram's generation S, where N is the total number of registered cardiac cycles (Fig. 1).

The proposed method of linguistic analysis and interpretation of ECG is based on the Levenshtein distance $L(S_1, S_2)$ between two words S_1, S_2 of N and M symbols, respectively. Recall that the Levenshtein distance is equal to the minimum number of editing operations such as insertion, deletion, and replacement of a character for converting a word S_1 into a word S_2 [24, 25].

The algorithm for calculating the Levenshtein distance is as follows.

Table 1. **Principle of ECG cycle coding**

The value of the indicator variable $V_i^{(RR)}$	+1	+1	-1	-1
The value of the indicator variable $V_i^{(\beta_T)}$	+1	-1	+1	-1
Symbol	***a***	***b***	***c***	***d***

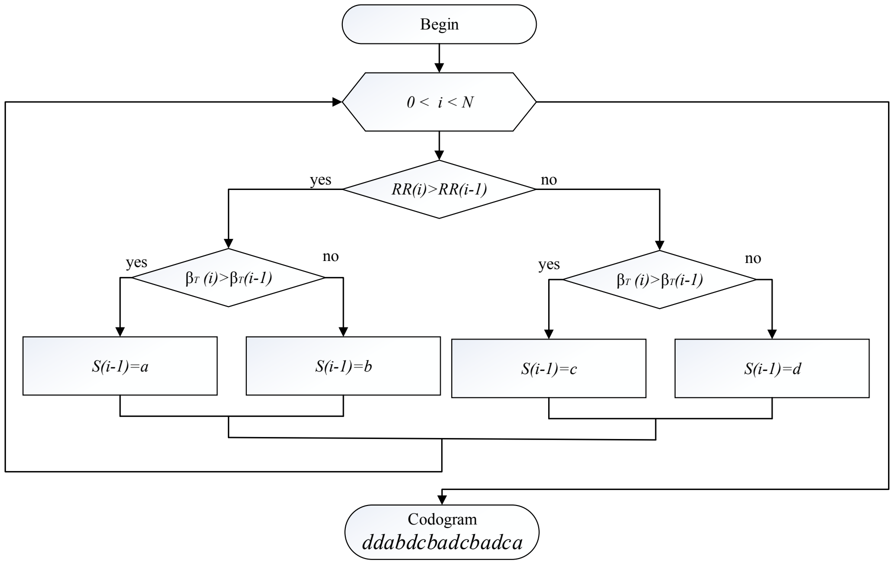

Fig. 1. The algorithm's block diagram for the formation of codograms

We form a matrix D of dimension $N+1,\ M+1$ and fill the first row and first column of the matrix D as follows:

$$D(i,0) = i, \quad \forall i = 0 \ldots N,$$
$$D(0,j) = j, \ \forall j = 0 \ldots M. \tag{4}$$

The remaining elements of the matrix D $(i > 0,\ j > 0)$ is filled in accordance with the rule:

$$D(i,j) = \min\{D(i,j-1)+1,\ D(i-1,j)+1,\ D(i-1,j-1)+m(S_1(i),S_2(j)\}, \tag{5}$$

where

$$m(S_1(i),S_2(j)) = \begin{cases} 0, \text{ if } & S_1(i) = S_2(j), \\ 1, \text{ if } & S_1(i) \neq S_2(j). \end{cases} \tag{6}$$

As a result the value $D(N,M)$ determines the Levenshtein distance.

In general, there may be several optimal paths from cell $D(1,\ 1)$ to cell $D(N,M)$ that provide the minimum number of editing operations. Figure 2 illustrates this fact and shows the "optimal" sequence of transition from the word

$$S_1 = \text{ddabdcbadcbadca} \tag{7}$$

to the word

$$S_2 = \text{bacdaaacdadccbb}. \tag{8}$$

		d	d	a	b	d	c	b	a	d	c	b	a	d	c	a
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
b	1	1	2	3	3	4	5	6	7	8	9	10	11	12	13	14
a	2	2	2	2	3	4	5	6	6	7	8	9	10	11	12	13
c	3	3	3	3	3	4	4	5	6	7	7	8	9	10	11	12
d	4	3	3	4	4	3	4	5	6	6	7	8	9	9	10	11
a	5	4	4	3	4	4	4	5	5	6	7	8	8	9	10	10
a	6	5	5	4	4	5	5	5	5	6	7	8	8	9	10	10
a	7	6	6	5	5	5	6	6	5	6	7	8	8	9	10	10
c	8	7	7	6	6	6	5	6	6	6	6	7	8	9	9	10
d	9	8	7	7	7	6	6	6	7	6	7	7	8	8	9	10
a	10	9	8	7	8	7	7	7	6	7	7	8	7	8	9	9
d	11	10	9	8	8	8	8	8	7	6	7	8	8	7	8	9
c	12	11	10	9	9	9	8	9	8	7	6	7	8	8	7	8
c	13	12	11	10	10	10	9	9	9	8	7	7	8	9	8	8
b	14	13	12	11	10	11	10	9	10	9	8	7	8	9	9	9
b	15	14	13	12	11	11	11	10	10	10	9	8	8	9	10	10

Fig. 2. Possible ways to ensure the optimal transition from the word S_1 to the word S_2

Table 2. **The optimal transition from word S_1 to word S_2**

Step	**Source word**	**Operation**	**Result of editing**
1	S_1 = ddabdcbadcbadca	Replacement d → b	S_1 = bdabdcbadcbadca
2	S_1 = bdabdcbadcbadca	Deletion *d*	S_1 = babdcbadcbadca
3	S_1 = babdcbadcbadca	Replacement b → c	S_1 = bacdcbadcbadca
4	S_1 = bacdcbadcbadca	Deletion *c*	S_1 = bacdbadcbadca
5	S_1 = bacdbadcbadca	Replacement b → a	S_1 = bacdaadcbadca
6	S_1 = bacdaadcbadca	Replacement d → a	S_1 = bacdaaacbadca
7	S_1 = bacdaaacbadca	Replacement b → d	S_1 = bacdaaacdadca
8	S_1 = bacdaaacdadca	Replacement a → c	S_1 = bacdaaacdadcc
9	S_1 = bacdaaacdadcc	Insertion *b*	S_1 = bacdaaacdadccb
10	S_1 = bacdaaacdadccb	Insertion *b*	S_1 = bacdaaacdadccbb

The step on i symbolizes the removal of the next symbol from the word S_1, the step on j — inserting the next word S_1 symbol into the word S_2, and the step on both indices symbolizes the replacement of the S_1 symbol with the next S_2 symbol or the absence of changes.

Table 2 demonstrates one of the optimal options shown in Fig. 2.

In addition to the classic Levenshtein distance, a number of its modifications are known, for example, the Damerau-Levenshtein distance, in which an additional editing operation in the form of transposition - permutation of two adjacent symbols, or the Levenshtein extension where different prices for elementary editing operations is introdused.

INFORMATION TECHNOLOGY OF LINGUISTIC ECG ANALYSIS

A number of important tasks can be formulated based on the calculation of the Levenshtein distance between the ECG codograms, including:

Task 1. Study of the properties of the Levenshtein distance based on the analysis of model signals generating an ECG of a realistic shape

Task 2. Decision making on the affiliation of the examined person one of two groups, for example, a group of patients or a group of conditionally healthy people, a group of trained or untrained, etc.

Task 3. Evaluation of the intra-individual characteristics of the codograms of one person over a sufficiently large segment of observations.

Task 4. Study of the possibility of using Levenshtein distance to assess changes in the functional state of the body under the influence of physical and emotional stress, medication, surgery, etc.

To perform the necessary research, an information technology (IT) has been developed, the structure of which is shown in Fig. 3.

To study the properties of the Levenshtein distances a software simulator that implements a stochastic model of generating artificial electrocardiograms of a realistic form [26] was uncluded in the IT. The generative model generates a sequence of ECG cycles, built on the basis of the sum of asymmetric Gaussian functions.

$$z_0(t) = \sum_k A_i \exp\left[-\frac{(t-\mu_k)^2}{2[b_k(t)]^2} \right] \tag{9}$$

with a given level of distortion of parameters that characterize the shape of the k-th fragment $k \in \{P, Q, R, S, ST, T\}$

Studies of the possibilities of the proposed approach for solving other problems were carried out on the basis of real ECGs accumulated in the database of the FASEGRAPH® complex. The complex consists of an ECG recorder with finger electrodes and a computer program that provides automatic ECG processing and determination of traditional and original diagnostic features, in particular, the duration of RR- intervals and parameters β_T beat to beat.

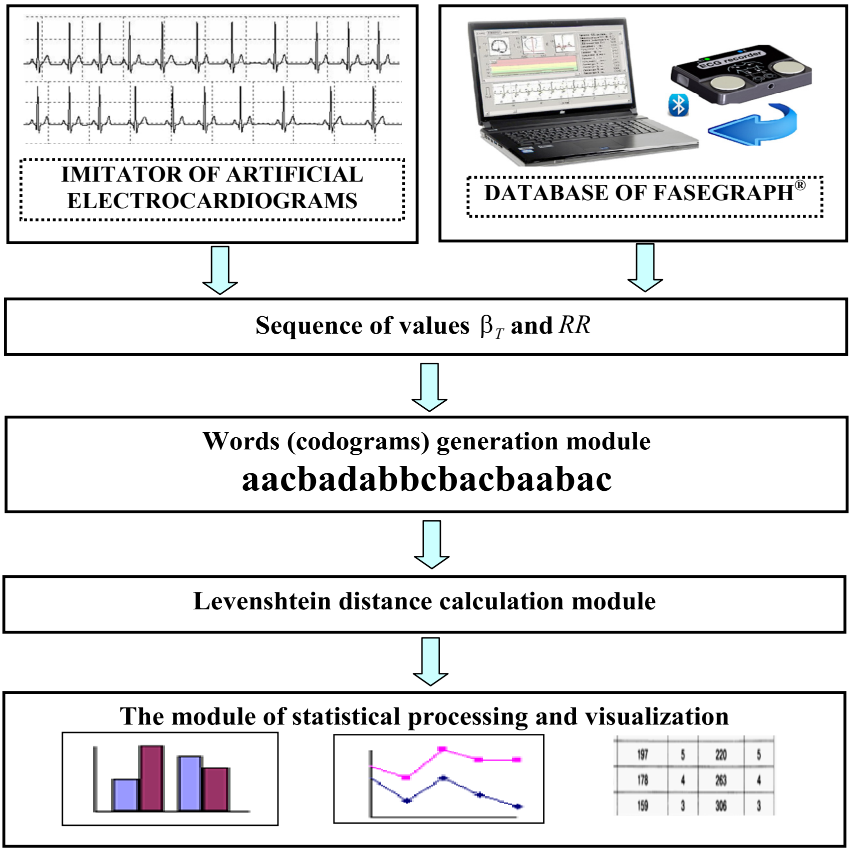

Fig. 3. Structure of information technology

According to the accumulated data, corresponding codograms were formed and *reference codograms* of individual groups of subjects were determined, for example, reference codograms of verified patients and conditionally healthy volunteers, athletes and people who are not actively involved in sports, etc.

The algorithm for constructing reference ECG codograms is as follows.

Let Q_1 and Q_2 ECG which represent one of the two studied groups was recorded as a result of the experiments. Each of Q_1 ECG of the first group (for example, a group of patients) is encoded in words $S_q^{(1)}$, $q=1,...,Q_1$, in accordance with Table 1. According to formulas (4) - (6), the Levenshtein distance $L_{\mu\nu}(S_\mu^{(1)}, S_\nu^{(1)})$ between each pair $S_\mu^{(1)}, S_\nu^{(1)}$, $\mu=1,...,Q_1$, $\nu=1,...,M_1$ encoded ECG is determined.

Next, we form a square $Q_1 \times Q_1$-matrix of distances $L_{\mu\nu}(S_\mu^{(1)}, S_\nu^{(1)})$, $\mu=1,...,Q_1$, $\nu=1,...,Q_1$ between all pairs of words corresponding to the ECG of the first group of subjects

$$\Lambda = \begin{pmatrix} L_{11}, & L_{12}, \ldots, L_{1Q_1} \\ L_{21}, & L_{22}, \ldots, L_{2Q_1} \\ & \ldots \\ L_{Q_1 1}, & L_{Q_1 2}, \ldots, L_{Q_1 Q_1} \end{pmatrix}.$$

The row of this matrix, the sum of the elements of which is minimal, defines the reference word $S_0^{(1)}$ of the first group, i.e.

$$S_0^{(1)} = \arg\min_{1 \le \nu \le Q_1} \sum_{\mu=1}^{Q_1} L_{\mu\nu} . \tag{10}$$

The reference word $S_0^{(2)}$ of the second group is determined in a similar way by analyzing the sum of the elements of the Levenshtein distance matrix $L_{\mu\nu}(S_\mu^{(2)}, S_\nu^{(2)})$, $\mu = 1,\ldots,Q_2$, $\nu = 1,\ldots,Q_2$ constructed for all pairs of codograms of the second group, i.e.

$$S_0^{(2)} = \arg\min_{1 \le \nu \le Q_2} \sum_{\mu=1}^{Q_2} L_{\mu\nu} . \tag{11}$$

Reference codograms (10), (11) allow solving various tasks, for example, relating the analyzed ECG to the first or second groups based on a comparison of the Levenshtein distance between the code word S_t of the analyzed ECG and the reference words $S_0^{(1)}$ and $S_0^{(2)}$:

$$\text{ECG belongs to the first group, if } L(S_t, S_0^{(1)}) \le L(S_t, S_0^{(2)}), \tag{12}$$

$$\text{ECG belongs to the second group, if } L(S_t, S_0^{(1)}) > L(S_t, S_0^{(2)}). \tag{13}$$

A similar rule can be used to make a decision about the level of fitness of a person on the basis of Levenshtein's distance between his codogram S_t and reference words $S_0^{(1)}$ and $S_0^{(2)}$, constructed separately for groups of athletes and people who are not actively involved in sports.

To assess the intraindividual features of the codograms (task 3), the data of the concrete person was extracted from the database, which were registered for a sufficiently large period of time. We analyzed only those persons who had no serious organic cardiac abnormalities during the observation period.

On the basis of processing the Q codograms of a concrete person we determined a reference word S_0 for this person and calculated the Levenshtein distance $L(S_1, S_0)$ $L(S_2, S_0)$, … , $L(S_Q, S_0)$ relative to S_0. The values $L(S_1, S_0)$ $L(S_2, S_0)$, … , $L(S_Q, S_0)$ were considered as the implementation of a random value for which an estimate of the distribution was constructed and its statistical characteristics were determined.

For the study of changes in Levenshtein distances in the course of physical and emotional loads, drip injections of drugs and other experimental studies (task 4), we used the electrocardiogram accumulated in the database, recorded during the corresponding observation period T_0, and plotted the changes of the distances $L(S_t, S_0)$ during the experiment relatively discrete time $t = 1,2,..., T_0$.

RESULTS OF EXPERIMENTAL RESEARCHES

Let us briefly review the first practical results obtained using the developed IT.

Figure 4 shows graphs of test signals generated by an artificial ECG simulator. The first signal (the signal A) is a periodic function - a sequence of ECG cycles of a given shape without distortion. Obviously, the codogram of such a signal contain only one letter:

$$S_A = \boldsymbol{dddd\ldots dddd.}$$

The codogram S_A was used as a reference for calculating the Levenshtein distances to the codograms of two other test signals: a signal B that simulated a sequence of reference cycles with distortions of the shape of the wave T only, and a signal C that additionally distorted the durations RR - intervals. The standard deviation of β_T for both signals was RMS β_T = 0,3 units, and the standard deviation of the RR -duration (the standard indicator of the mathematical analysis of heart rate variability) was $SDNN = 140$ ms.

The codograms constructed for the distorted signals had the form:

$$S_B = \boldsymbol{aacddccbaabdccbacabbcaaddcddadbbcaabbdaddadbabcc\ldots}$$

$$\boldsymbol{\ldots babccdbabbdccbabcbababdcdacbabadcbadcdcbaddcd}$$

$$S_C = \boldsymbol{bbabcbcabccaabddcbdcabdcaaabddcabdcbabbcbccabdccabdc\ldots}$$

$$\boldsymbol{\ldots\ bccbabdcbdcbcabdcbdcbdcabddcbcabcaabddca.}$$

Table 3 summarizes the results of the comparison of distorted signals B and C with the reference signal A.

As can be seen from Table 3, the Levenshtein distance increases as the level of ECG cycle distortion increases. However, it was established that the statistical dependence of the distance $L(S_t, S_0)$ of the analyzed codogram S_t relative to the reference one S_0 and index $SDNN$ is substantially nonlinear: the coefficient of determination of the linear dependence $R^2 < 0,5$.

Table 3. **The results of the comparison with the test signal** ***A***

SIGNAL	Increment of RMS β_T, un.	Increment of $SDNN$, ms	Levenshtein distance, un.
B	0,03	0	$L(S_B, S_A) = 46$
C	0,03	140	$L(S_C, S_A) = 71$

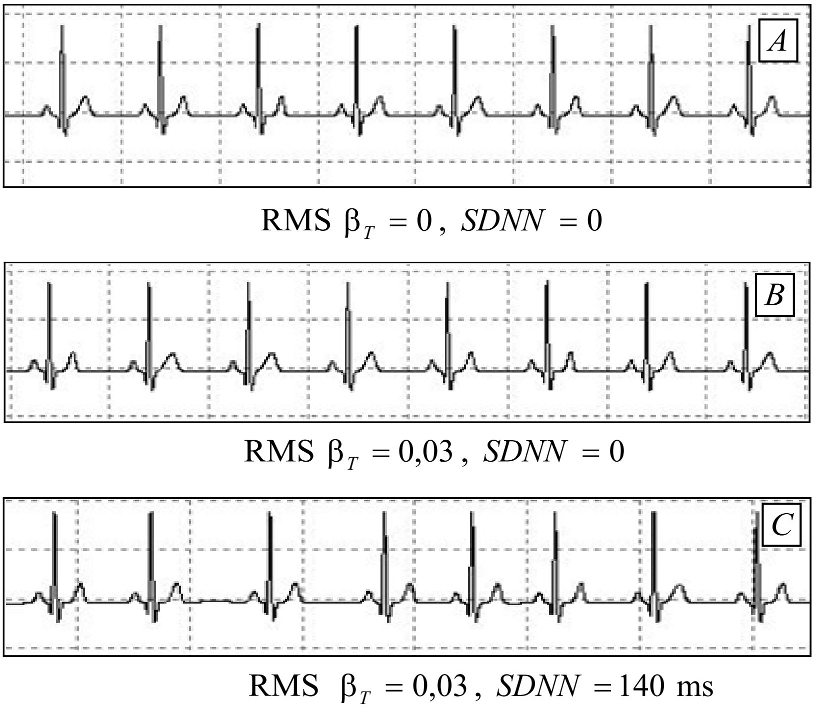

Fig. 4. Test signals: A — without distortion, B — distortion of β_T only, C — both distortion β_T and RR-intervals

Interesting results were obtained for athletes and people who are not actively involved in sports. The studies were conducted on the basis of real ECG, obtained when testing young volunteers of both sexes aged 18–24, which were divided into two groups:

• **Group 1:** $Q_1 = 47$ highly qualified athletes who are engaged in boxing, various types of wrestling and triathlon;

• **Group 2**: $Q_2 = 113$ people not involved in sports.

Investigations were carried out while performing the Martin test: 20 squats in 30 seconds. We recorded ECG all person from each group before the load, at the height of the load, and after a 3-minute rest. The values of the average heart rate (HR) and the average index β_T characterizing the symmetry of the T wave were determined in these three states.

Following the previously obtained results [26], we will assume that an adequate response of the organism to the load is that with the load both indicators increase, and after resting they return to their original state, i.e.

$$(HR^{(2)} - HR^{(1)}) > (HR^{(3)} - HR^{(1)}), \tag{14}$$

$$(\beta_T^{(2)} - \beta_T^{(1)}) > (\beta_T^{(3)} - \beta_T^{(1)}), \tag{15}$$

where $\eta^{(1)}, \eta^{(2)}, \eta^{(3)}$ — the value of the corresponding indicator in the initial state, at the height of the load and after a 3-minute rest.

We studied the differences in the body's response to exercise in athletes and people who are not actively involved in sports, based on conditions (14), (15), and also based on an estimate of the Levenshtein distance $L(S^{(2)}, S^{(1)})$ and $L(S^{(3)}, S^{(1)})$ between the codogram $S^{(2)}$ at load height and codogram $S^{(3)}$ after resting with respect to to the original codogram $S^{(1)}$.

As a working hypothesis, it was assumed that the condition

$$L(S^{(2)}, S^{(1)}) > L(S^{(3)}, S^{(1)}) \tag{16}$$

also confirms an adequate response of the organism to the load.

The results of the research are summarized in Tables 4 and 5, in the cells of which the symbol "+" and "-" denote fulfillment or non-fulfillment of conditions (14) – (16).

As can be seen from the data given in Tables 4 and 5, the athletes and non-athletes observed small differences in the frequency of occurrence of events W_i, $i = 1,\ldots,8$, which form a complete group. To confirm the reliability of the detected differences, additional data processing was carried out on the basis of the calculation of confidence intervals.

From probability theory it is known [29] that the frequency P^* of a random event, calculated from a sample of observations by volume Q, with reliability of inference γ determines the confidence interval $\mathbf{I} = [P^{(1)}, P^{(2)}]$ of probability P, the boundaries of which are determined by formulas:

Table 4. **Athletes reaction to load**

Random event	**Indicator's recovery after the load**			**Event frequency, %**
	HR	β_T	***L* (*S*** $(S^{(2)}, S^{(2)})$	
	Condition (14)	**Condition (15)**	**Condition (16)**	
W_1	+	+	+	23,40
W_2	-	+	+	2,13
W_3	+	-	+	2,13
W_4	-	-	+	2,13
W_5	+	+	-	36,17
W_6	-	+	-	10,64
W_7	+	-	-	14,89
W_8	-	-	-	8,51

Table 5. **The response of non-athletes to the load**

Random event	Indicator's recovery after the load			Event frequency, %
	HR	β_T	$L\,(S^{(2)}, S^{(2)})$	
	Condition (14)	Condition (15)	Condition (16)	
W_1	+	+	+	26,55
W_2	-	+	+	6,19
W_3	+	-	+	8,85
W_4	-	-	+	2,65
W_5	+	+	-	34,51
W_6	-	+	-	7,08
W_7	+	-	-	11,50
W_8	-	-	-	2,65

$$P^{(1)} = \frac{P^* + \frac{1}{2}\frac{t_\gamma^2}{Q} - t_\gamma\sqrt{\frac{P^*(1-p^*)}{Q} + \frac{1}{4}\frac{t_\gamma^2}{Q^2}}}{1 + \frac{t_\gamma^2}{Q}} \tag{17}$$

$$P^{(2)} = \frac{P^* + \frac{1}{2}\frac{t_\gamma^2}{Q} + t_\gamma\sqrt{\frac{P^*(1-P^*)}{Q} + \frac{1}{4}\frac{t_\gamma^2}{Q^2}}}{1 + \frac{t_\gamma^2}{Q}}, \tag{18}$$

where $t_\gamma = \arg\Phi^*\left(\frac{1+\gamma}{2}\right) > 0$, and $\Phi^*(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{\tau^2}{2}} d\tau$.

On the basis of expressions (17), (18), according to the data of Tables 4 and 5, taking into account the volumes $Q_1 = 47$ and $Q_2 = 113$ of the respective samples, the confidence intervals of the probability of random events $W_i, i = 1,...,8$ are determined. It has been established that with reliability of inference $\gamma = 99\,\%$ the reliability intervals of the probability for events W_2, W_3, W_8 do not overlap in the group of athletes and people who are not actively involved in sports (Table 6). And this with high credibility indicates that the probabilities of these events are different. For example, in athletes $P(W_2) < 0{,}397$, while $P(W_2) > 0{,}498$ in people who are not actively engaged in sports.

Table 6. **Significant differences in confidence intervals of probabilities**

Random event	Confidence intervals of probabilities		Reliability of inference γ
	Athletes	Non-athletes	
W_2	[0,100 ; 0,397]	[0,498 ; 0,727]	99 %
W_3	[0,100 ; 0,397]	[0,786 ; 0,941]	99 %
W_8	[0,675 ; 0,940]	[0,173 ; 0,383]	99 %

Note that the indicators HR and β_T represent the averaged values of the parameters of individual ECG cycles, and the condition (16) characterizes the dynamics of the RR-intervals and the index β_T beat to beat. Therefore, condition (16), to a greater extent than conditions (14), (15), characterizes transient processes when performing load tests.

The found statistically significant differences in the response of the body of athletes and non-athletes to the load test may be related both to the athlete's overloads at previous trainings and also to indicate that in the process of regular workouts the athlete's body learns to more economically adapt to the load test. Both are reflected in the features of the dynamics of change HR, β_T and $L(S^{(2)}, S^{(1)})$. Of course, the study of the found fact requires further deeper investigations on representative samples of observations.

Recently, scientists have paid attention to intraindividual ECG changes of a healthy person at rest [29], which is not a precursor of any pathology. In this regard, it is interesting to study the intra-individual changes in the codograms of the same person over a fairly long observation period.

The basis of such studies was based on the analysis of the ECG series of two subjects, registered for six years. $Q_1 = 26$ codograms of the first test subject (male) and $Q_2 = 25$ codograms of the second test subject (woman) are analyzed.

Based on the processing of the available codograms, the reference codogram of the first person

$$S_0^{(1)} = \boldsymbol{bccabcbadabdcadcabcabccbdcbbcabccaddcbbdaba\ldots}$$

$$\boldsymbol{\ldots cbccbdabadadccdbcadccbaccbdcabdabbcabcabdcbacabbcadccbda}$$

and the second person

$$S_0^{(2)} = \boldsymbol{addaddacdadadcbdabcaddacbdcadabdadcadcadacd\ldots}$$

$$\boldsymbol{\ldots adbccadcbbcadbcdadcadcdcdbcdaddacbacbcdacdadaacbabbcada}$$

was constructed using the method described above.

Levenshtein distance between reference codograms is

$$L(S_0^{(1)}, S_0^{(2)}) = 52 .$$

Estimates of distributions of random variables $L(S_t^{(1)}, S_0^{(1)})$ and $L(S_t^{(2)}, S_0^{(2)})$, $t = 1,\ldots,Q_2$ (histograms) corresponding to ECG's codograms recorded during the observation period are presented in the (Fig. 5).

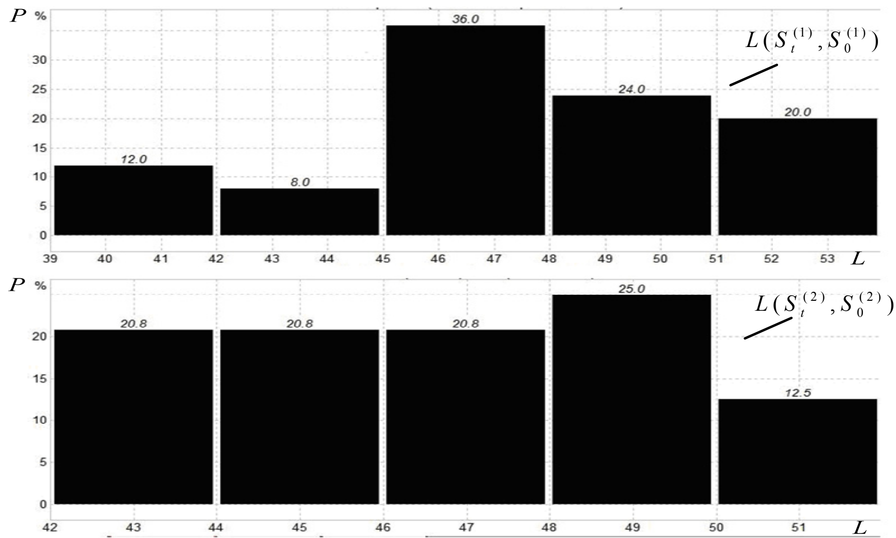

Fig. 5. Histograms of distance distributions $L(S_t^{(1)}, S_0^{(1)})$ and $L(S_t^{(2)}, S_0^{(2)})$

As can be seen from Fig. 5, Levenshtein distances $L(S_t^{(1)}, S_0^{(1)})$ $L(S_t^{(2)}, S_0^{(2)})$ varied over a fairly wide range of values for both individuals during the observation period. For comparison, we note that during this period there were also significant fluctuations in traditional indicators β_T and standard deviation of RR -intervals which amounted

$$\beta_T^{(1)} = 0{,}771 \pm 0{,}079\,,\ \beta_T^{(2)} = 0{,}752 \pm 0{,}125\,,\ SDNN^{(1)} = 37{,}9 \pm 16{,}1\,,$$
$$SDNN^{(2)} = 32{,}5 \pm 10{,}5\,.$$

The study of the diagnostic capabilities of the proposed method was carried out using a database of real ECG recorded at the Ischemic Heart Diseases Department of the V.D. Strazhesko's Research Institute of Cardiology Academy of Medical Sciences of Ukraine (Kyiv) and four clinics in Germany — Essen University Hospital (Essen), Katholical Hospital "Phillpusstift" (Essen), German Heart Center (Berlin).

The clinical material consisted of 100 ECG records of patients with chronic ischemic heart disease (CAD), whose diagnosis was previously established by coronary angiography (class V_1), and 100 ECG records of healthy volunteers included in the control group (class V_2)

By the formulas (10), (11) two reference codograms for the specified classes are defined: the reference codogram of patients with ischemic heart disease

$$S_0^{(1)} = \boldsymbol{adcbdadcadabdabcadabdadcbdab}$$

and reference codogram of a healthy volunteers

$$S_0^{(2)} = \boldsymbol{cbcdcabdcabddcaadcaa}.$$

The Levenshtein distance between these codograms was

$$L(S_0^{(1)}, S_0^{(2)}) = 15.$$

Based on the processing of the available data, it has been established that decision making according to the rules (12), (13) provides sensitivity $S_E = 72\%$ and specificity $C_P = 79\%$.

For illustration, we present the results of the ECG evaluation for two patients — a verified patient (male, 69 years old), whose codogram was

$$S_t^{(1)} = \boldsymbol{adcabdadcadabdaddabdaadabdbdda}$$

and a representative of the control group — a male, 54 years old, whose codogram was

$$S_t^{(2)} = \boldsymbol{bdcbbcdcabcdcabcdcbaa}.$$

It is easy to verify that $L(S_t^{(1)}, S_0^{(1)}) = 13$ and $L(S_t^{(1)}, S_0^{(2)}) = 15$, i.e.

$$L(S_t^{(1)}, S_0^{(1)}) < L(S_t^{(1)}, S_0^{(2)})$$

and in accordance with the rule (12), the survey must be reduced to the CAD-group.

Similarly, for the second person we have $L(S_t^{(2)}, S_0^{(1)}) = 14$ and $L(S_t^{(2)}, S_0^{(2)}) = 8$, i.e.

$$L(S_t^{(2)}, S_0^{(1)}) > L(S_t^{(2)}, S_0^{(2)})$$

and in accordance with the rule (13) the person should be reduced to a healthy group.

It is important to note that traditional signs of myocardial ischemia (elevation or depression of the ST segment) not observed on all patient's ECG. And this means that conventional electrocardiography would classify all processed ECG into healthy group.

At the same time experiments show that it is possible to classify representatives of the classes even on such "complex" clinical material on the basis comparison of the Levenshtein distances.

Fig. 6 presents estimates of the conditional distributions $P(L(S_t, S_0^{(1)}))$ and $P(L(S_t, S_0^{(2)}))$ of Levenshtein distances with respect to the reference codograms of the sick and the healthy person.

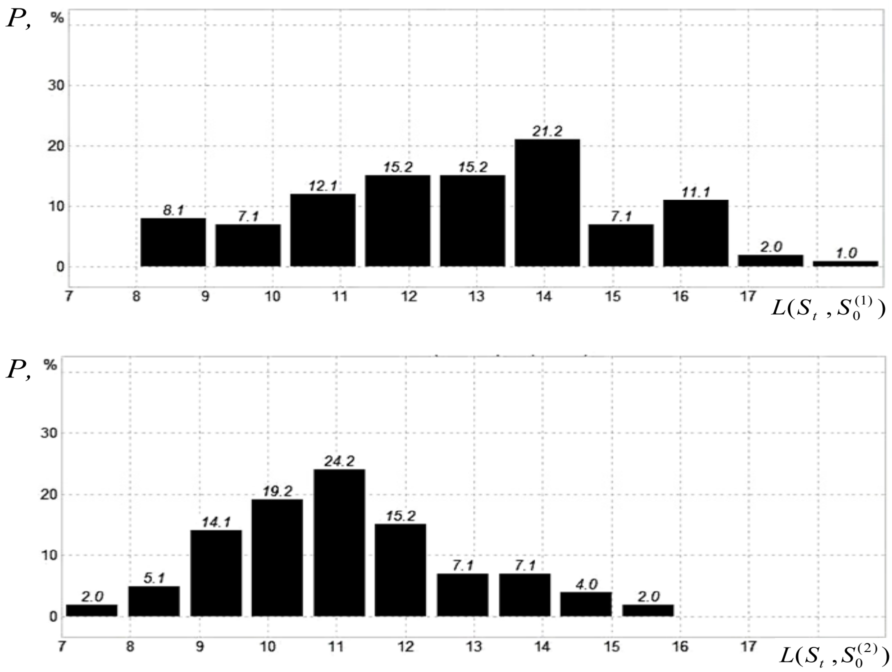

Fig. 6. Conditional distributions of Levenshtein distances to patient (top) and healthy (bottom) standards

Testing the hypothesis of homogeneity of conditional distributions $P(L(S_t, S_0^{(1)}))$ and $P(L(S_t, S_0^{(2)}))$ according to the Kolmogorov-Smirnov criterion showed that the hypothesis of equality of distributions should be rejected with high statistical significance $(p < 0{,}001)$. The similar fact was confirmed by Mann-Whitney U test.

Consequently

$$P(L(\cdot) \mid V_1) \not\equiv P(L(\cdot) \mid V_2). \tag{19}$$

If also the diagnostic sign $L(\cdot)$ is conditionally independent of indicators $SDNN$ and RMS, i.e.

$$P(L, SDNN, RMS\ \beta_T \mid V_k) \equiv P(L \mid V_k) P(SDNN, RMS\ \beta_T \mid V_k), \quad k = 1{,}2, \tag{20}$$

then, according to a theorem proved in [30], the Levenshtein distance is guaranteed a useful diagnostic sign in conjunction with $SDNN$ and RMS in terms of reducing the probability of erroneous decisions.

The analysis of the Levenshtein distances was also useful for illustrating the dynamics of ECG changes during the drug treatment of cardiovascular pathologies.

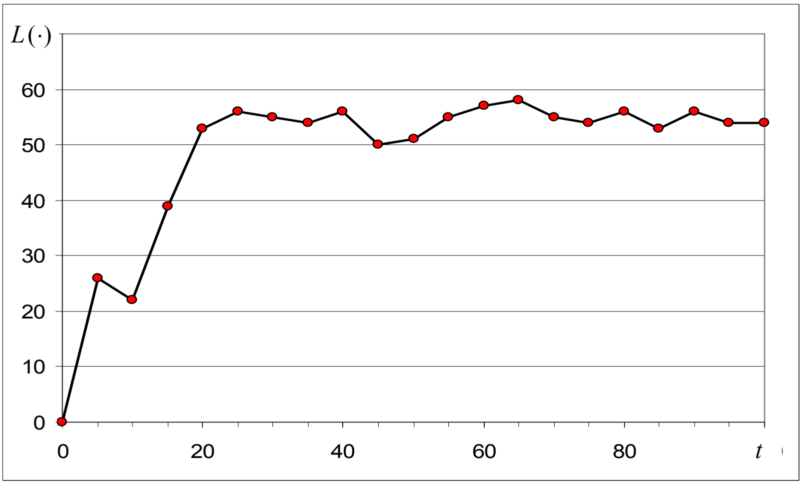

Fig. 7. Schedule of changes in the distance L (S_t, S_0) in the process of intravenous drug infusions

Fig. 7 is a graph of the change in the distance $L(S_t, S_0)$ between ECG codograms, which were recorded during intravenous infusion of drugs to patient R. (male, 72 years old), diagnosed with atrial fibrillation.

Patient R. was hospitalized at the Scientific and Practical Center for Prevention and Clinical Medicine (Kyiv). ECGs were recorded every 5 minutes for 1 hour and 20 minutes in the process of intravenous infusion of Tivomax preparations followed by the addition of the drug Armadin.

Before the administration of the drugs, there was a pronounced extrasysitivity on the patient's ECG, which decreased significantly to 20-th minute. With the further administration of drugs, the heart rhythm gradually normalizes. By the end of 100 minutes, the standard deviations of the RR -interval were within the functional norm ($SDNN = 30$ ms). As can be seen from Fig. 7, the positive dynamics of intravenous infusion of drugs is clearly illustrated by the graph of Levenshtein distances $L(S_t, S_0)$ changing.

The analysis of the Levenshtein distance was also used to monitor the condition of patient S. (male, 61) diagnosed with atrial fibrillation combined with arterial hypertension, who was hospitalized for 47 days in a hospital for scientists of the National Academy of Sciences of Ukraine. Monitoring was carried twice a day (Fig. 8).

The patient was treated with antiarrhythmic drugs (Digoxin et al.) and blood pressure lowering drugs (Captopres, etc.). Periodically intravenous infusions of a number of drugs were conducted.

During the entire period of treatment, the patient's condition was stable, which was reflected only by insignificant fluctuations of the Levenshtein distance between codograms except for some moments of time (Fig. 8). One of such measurements, marked by an arrow in Fig. 8, was associated with the treadmill study, when the heart rate was 139 beats / min, and the T -wave symmetry index assumed an excessively high value $\beta_T = 2{,}05$ unit.

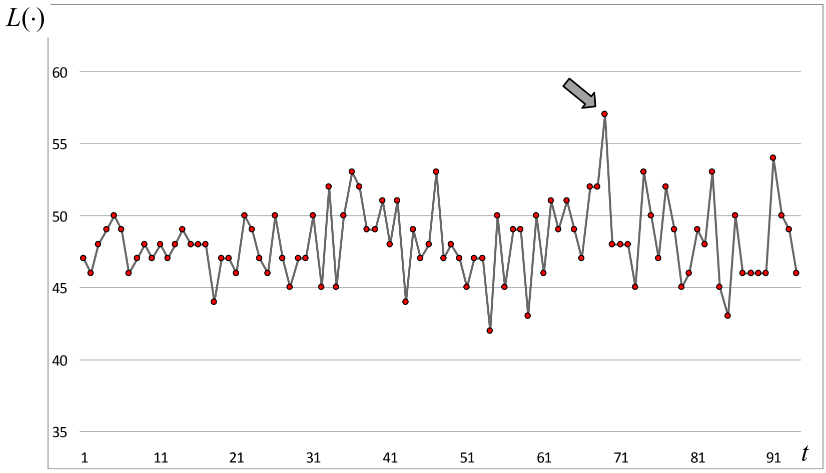

Fig. 8. Monitoring of state patient S

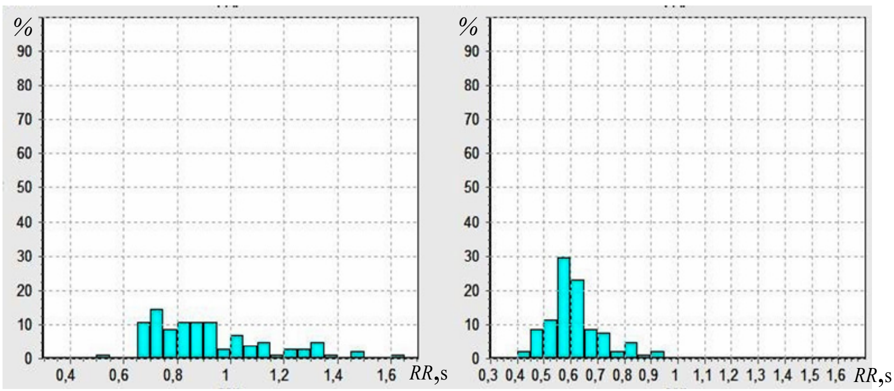

Fig. 9. Histograms *RR*- intervals before (left) and after (right) treatment of patient S

Daily monitoring of ECG indices using the method fasegraphy combined with monitoring of Levenshtein distances, confirmed the effectiveness of the medical treatment of the patient B.: the index decreased from the initial value of ms to the value of ms. heart rate variability decreased by 73%, what clearly illustrates the comparison of the histograms of the durations of the RR -intervals (Fig. 9).

In conclusion we note that the considered approach of ECG coding with subsequent analysis of the Levenshtein distance between codograms can naturally be generalized to the cases when not only the intervals $R-R$ and indicator β_T are used for ECG coding, but also other informative indicators, in particular, amplitudes of R and T waves.

CONCLUSIONS

The proposed approach is based on converting sequences of parameters that characterize the form of individual electrocardiogram cycles into a word — a finite string of characters of a given alphabet. The words (codograms) obtained in this way are analyzed on the basis of the Levenshtein distance, which determines the minimum number of editing operations (inserting, deleting, and replacing a character) to convert one word into another.

The calculation of the Levenshtein distances between all pairs of codograms of the group of subjects allows one to determine the reference codogram of this group. Comparison of the Levenshtein distances, including the analysis of the editorial distance between the current codogram and the reference codogram, makes it possible to make diagnostic decisions about patient ownership of the corresponding group.

An information technology based on the implementation of the components of the proposed approach has been developed. Experimental studies conducted using model and real data confirmed the potential diagnostic effectiveness of the proposed approach for obtaining additional information in solving actual applied problems. Of course, from the point of view of evidence-based medicine, such research should be continued on a significantly larger amount of clinical material.

REFERENCES

1. Kanjilal P. P., Bhattacharya J., Saga G. Robust Method for Periodicity Detection and Characterisation of Irregular Cyclical Series in Terms of Embedded Periodic Components. *Phys. Rev.* 1999. Vol. 59. P. 4013–4025.
2. Fainzilberg L.S. Generalized Method of Processing Cyclic Signals of Complex Form in Multidimension Space of Patameters. *Journal of Automation and Information Sciences*. 2015. Vol. 47. Issue 3. P. 24–39. DOI: 10.1615/JAutomatInfScien.v47.i3.30
3. Connolly D.C., Elveback L.R., Oxman H.A. Coronary heart disease in residents of Rochester, Minnesota: Prognostic value of the resting electrocardiogram at the time of initial diagnosis of angina pectoris. *Mayo.Clin.Proc*. 1984. Vol. 59. P. 247–250.
4. Solopov V.N., Sadykova A.R., Fedoseeva T.S. Limitations of automatic computerized analysis of an electrocardiogram. *Kazan Medical Journal*. 2012. Vol. 93. No. 4. P. 687–691. (In Russian).
5. Pavlidis T. Linguistic analysis of waveforms. *Software Eng.* 1971. Vol. 2. No. 4. P. 203–225. – https://doi.org/10.1016/B978-0-12-696202-4.50019-X.
6. Mottl N.V., Muchnik I.B., Jakovled V.G. Optimal segmentation of experimental curves. *Automation and remote control.* 1983. Issue 8. P. 84–95. (In Russian).
7. Haralick R.M. Structural pattern recognition homomorphisms and arrangements. *Pattern recognition*. 1978. Vol. 10. No. 3. P.223–236.
8. Goldberger A.L. Fractal mechanisms in the electrophysiology of the heart. *IEEE Eng. Med. Biol*. 1992. No 11. P. 47–52.
9. Cherkay A.D., Vlasov Ju.A. Linguistic analysis of the sequence of time intervals between heartbeats. *Theory and practice of automation of electrocardiographic and clinical research*. Kaunas: Kaunas Medical Institute. 1977. P. 128–131. (In Russian)
10. Maksimov A.V. ECG signals analyzing using translating grammars. *News of the Taganrog University of Radio Engineering. Thematic issue "Computer technology in engineering and management practice". 2*001. P. 210–216. (In Russian).
11. Tsidypov Ch.S. Pulse diagnosis of Tibetan medicine. Novosibirsk: Science, 1988. 134 p. (In Russian).
12. Bajevsky R.M., Ivanov G.G. Heart rate variability: theoretical aspects and possibilities of clinical use. *Ultrasound and functional* diagnostics. 2001. No 3. P C. 108–127. (In Russian).
13. McCrafy R. New Frontiers in Heart Rate Variability and Social Coherence Research: Techniques, Technologies, and Implications for Improving Group Dynamics and Outcomes. *Front Public Health*. 2017. No. 5. P. 2–13. https:// doi:10.3389/fpubh.2017.00267.
14. Uspenskiy V.M. Diagnostic System Based on the Information Analysis of Electrocardiogram. *Proceedings of MECO 2012. Advances and Challenges in Embedded Computing (Montenegro, June 19–21)*. 2012. P. 74–76.

15. Kolesnikova O.V., Krivenko S.S. Information analysis of electrocardiosignals: justification and possibilities. *Papers of scientific works of the First International Scientific and Practical Conference "Information Systems and Technologies in Medicine" (ISM-2018).* Kharkiv: KNURE. 2018. P. 161–163. (In Ukrainian)
16. Fainzilberg L.S. The basic of fasegraphy. Kyiv: Osvita Ukraini, 2017. 264 p. (In Russian).
17. Fainzilberg L.S. Computer diagnostics based on the electrocardiogram phase portrait. Kyiv: Osvita Ukraini, 2013. 191 p. (In Russian).
18. Fainzilberg L.S. ECG Averaging based on Hausdorff Metric. *International Journal of Biomagnetism.* 2003. Vol. 5. № 1. P. 236–237.
19. Fainzilberg L.S. Intellectual possibilities and prospects for the development of fasegraph information technology of processing signals of complex shape. *Kibernetika i vyčislitelnaâ tehnika*. 2016. Issue 186. P. 56–77. (In Russian). https://doi.org/10.15407/usim.2018.03.003
20. Dyachuk, D.D., Gritsenko, V.I., Fainzillberg, L.S. 2017. Application of a method of a phasagraphy during the screening of ischemic heart disease. Methodical recommendations of the Ministry of Health of Ukraine No. 163.16 / 13.17, K.: Ukrainian Center for Scientific Medical Information and Patent and Licensing, 32 p. (In Ukrainian).
21. Gonzalez R.C. Syntactic pattern recognition. Introduction and survey. Proceedings of the National Electronic Conference. 1972. Vol. 27. No. 1. P. 27–32.
22. Braverman E.M., Muchnic I.B. Structural methods for processing empirical data. Moscow: Nauka. 1983. 492 p. (In Russian).
23. Biermann A.W. A Grammatical inference program for linear languages. *International Conference of System Science (Hawall, 1971).* 1971. Vol. S1. P. 117–141.
24. Levenshtein V.I. Binary codes with corrections for fallouts, insertions and substitutions of characters. *Reports of the USSR Academy of Sciences*. 1965. Vol. 163. No. 4. P. 845–848. (In Russian).
25. Wagner R.A., Fischer M.J. The String-to-String Correction Problem. *Journal of the ACM.* 1971. Vol. 21. Issue 1. P. 168–173. https:// doi: 10.1145/321796.321811
26. Fainzilberg L.S. Simulation models of artificial electrocardiogram generation under internal and external disturbances conditions. *Journal of Qafgaz University. Matheмatics and Computer Scienc*e. 2012. No. 34. P. 92–104. (In Russian).
27. Fainzilberg L.S., Orikhovska K.B. Information technology of the organism adaptation reserves assessment in field conditions. *Kibernetika i vyčislitelnaâ tehnika*. 2015. Issue 181. P. 4–22. (In Russian).
28. Gornan I.I. Probability theory and mathematical statistics for researchers and engineers. Kyiv: IPMMS NAS of Ukraine, 2003. 244 p. (In Ukrainian).
29. Schijvenaars B.J.A, Van Herpen G., Kors J.A. Intraindividual variability in electrocardiograms. *Journal of Electrocardiology*. 2008. Vol. 41. Issue 3. P. 190–196. doi: 10.1016/j.jelectrocard.2008.01.012
30. Fainzilberg L..S. Guaranteed estimate of the features utility while statistical recognition of two classes. *Journal of Automation and Information Sciences*. 2010. Vol. 42. Issue 10. P. 32–48.

Received 01.04.2019

ЛІТЕРАТУРА

1. Kanjilal P. P., Bhattacharya J., Saga G. Robust Method for Periodicity Detection and Characterisation of Irregular Cyclical Series in Terms of Embedded Periodic Components. *Phys. Rev.* 1999. Vol. 59. P. 4013–4025.
2. Fainzilberg L.S. Generalized Method of Processing Cyclic Signals of Complex Form in Multidimension Space of Patameters. *Journal of Automation and Information Sciences*. 2015. Vol. 47. Issue 3. P. 24–39. DOI: 10.1615/JAutomatInfScien.v47.i3.30
3. Connolly D.C., Elveback L.R., Oxman H.A. Coronary heart disease in residents of Rochester, Minnesota: Prognostic value of the resting electrocardiogram at the time of initial diagnosis of angina pectoris. *Mayo.Clin.Proc.* 1984. Vol. 59. P. 247–250.
4. Солопов В.Н., Садыкова А.Р., Федосеева Т.С. Ограничения автоматического компьютерного анализа электрокардиограммы. *Казанский медицинский журнал.* 2012. Т. 93. № 4. С. 687–691.

5. Pavlidis T. Linguistic analysis of waveforms. *Software Eng.* 1971. Vol. 2. No. 4. P. 203–225. — https://doi.org/10.1016/B978-0-12-696202-4.50019-X
6. Моттль В.В, Мучник И.Б., Яковлев В.Г Оптимальная сегментация экспериментальных кривых. *Автоматика и телемеханика.* 1983. Вып. 8. С. 84–95.
7. Haralick R.M. Structural pattern recognition homomorphisms and arrangements. *Pattern recognition.* 1978. Vol. 10. No. 3. P.223–236.
8. Goldberger A.L. Fractal mechanisms in the electrophysiology of the heart. *IEEE Eng. Med. Biol.* 1992. No 11. P. 47–52.
9. Черкай А.Д., Власов Ю.А. Лингвистический анализ последовательности временных интервалов между ударами сердца. *Теория и практика автоматизации электрокардиографических и клинических исследований*. Каунас: Каунаский медицинский институт, 1977. С. 128–131.
10. Максимов А.В. Анализ ЭКГ-сигналов с использованием транслируюших грамматик. *Известия Т*аганрогского радиотехнический университета. Тематический выпуск «Компьютерные технологии в инженерной и управленческой практике». 2001. С. 210–216.
11. Цыдыпов Ч.С. Пульсовая диагностика тибетской медицины. Новосибирск: Наука, 1988. 134 с.
12. Баевский Р.М., Иванов Г. Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения. *Ультразвуковая и функциональная диагностика.* 2001. № 3. С. 108–127.
13. McCrafy R. New Frontiers in Heart Rate Variability and Social Coherence Research: Techniques, Technologies, and Implications for Improving Group Dynamics and Outcomes. *Front Public Health.* 2017. No. 5. P. 2–13. — https:// doi:10.3389/fpubh.2017.00267.
14. Uspenskiy V.M. Diagnostic System Based on the Information Analysis of Electrocardiogram. *Proceedings of MECO 2012.* Advances and Challenges in Embedded Computing (Montenegro, June 19–21). 2012. P. 74–76.
15. Колеснікова О.В., Крівенко С.С. Інформаційний аналіз електрокардіосигналів: обгрунтування і можливості. *Збірник наукових праць Першої Міжнародної науково-практичної конференції «Інформаційні системи та технології в медицині»* (ISM–2018). Харків: ХНУРЕ, 2018. С. 161–163.
16. Файнзильберг Л.С. Основы фазаграфии. Киев: Освита Украины, 2017. 264 с.
17. Файнзильберг Л.С. Компьютерная диагностика по фазовому портрету электрокардиограммы. Киев: Освита Украины, 2013. 191 с.
18. Fainzilberg L.S. ECG Averaging based on Hausdorff Metric. *International Journal of Biomagnetism.* 2003. Vol. 5. № 1. P. 236–237.
19. Файнзильберг Л.С. Интеллектуальные возможности и перспективы развития фазаграфии — информационной технологии обработки сигналов сложной формы. *Кибернетика и вычислительная техника.* 2016. Вып. 186. С. 56–77.
20. Дячук Д.Д., Гриценко В.І., Файнзільберг Л.С. и др. Застосування методу фазаграфії при проведенні скринінгу ішемічної хвороби серця. Методичні рекомендації МОЗ України № 163.16/13.17. К.: Український центр наукової медичної інформації і патентно-ліцензійної роботи, 2017. 32 с.
21. Gonzalez R.C. Syntactic Pattern Recognition – Introduction and Survey. *Proceedings of the National Electronic Conference.* 1972. Vol. 27. No. 1. P. 27–32.
22. Браверман Э.М. , Мучник И.Б. Структурные методы обработки эмпирических данных. М.: Наука. 1983. – 492 с.
23. Biermann A.W. A Grammatical Inference Program for Linear Languages *Intern. Conf. of System Science* (Hawall, 1971). 1971. Vol. S1. P. 117–141.
24. Левенштейн В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов. *Докл. АН СССР.* 1965. Т. 163. № 4. С. 845–848.
25. Wagner R.A., Fischer M.J. The String-to-String Correction Problem. *Journal of the ACM.* 1971. Vol. 21. Issue 1. P. 168–173. — https:// doi:10.1145/321796.321811.
26. Файнзильберг Л.С. Имитационные модели порождения искусственных электрокардиограмм в условиях внутренних и внешних возмущений. *Journal of Qafgaz University — Mathematics and Computer Science.* 2012. № 34. С. 92–104.

27. Файнзильберг Л.С., Ориховская К.Б. Информационная технология оценки адаптационных резервов организма в полевых условиях. *Кибернетика и вычислительная техника.* 2015. Вып. 181. С. 4–22.
28. Горбань І.І. Теорія ймовірності і математична статистика для наукових працівників та інженерів. К., ІПММС НАН України, 2003. 244 с.
29. Schijvenaars B.J.A, Van Herpen G., Kors J.A. Intraindividual variability in electrocardiograms. *Journal of Electrocardiology.* 2008. Vol. 41. Issue 3. P. 190–196.
30. Fainzilberg L.S. Guaranteed Estimate of the Features Utility while Statistical Recognition of Two Classes. *Journal of Automation and Information Sciences*. 2010. Vol. 42. Issue 10. P. 32–48.

Отримано 01.04.2019

Файнзільберг Л.С.[1], д-р техн. наук, професор,
голов. наук. співроб. відд.
інтелектуальних автоматичних систем
e-mail: fainzilberg@gmail.com
Дикач Ю.Р.[2], студент
факультет біомедичної інженерії
e-mail: jul.dykach@gmail.com
[1] Міжнародний науково-навчальний центр
інформаційних технологій та систем
НАН України та МОН України,
пр.-т. Академіка Глушкова, 40, м. Київ, 03187, Україна,
[2] Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
пр.-т. Перемоги, 37, м. Київ, 03056, Україна

ЛІНГВІСТИЧНИЙ ПІДХІД ДЛЯ ОЦІНЮВАННЯ ТОНКИХ ЗМІН ЕЛЕКТРОКАРДІОГРАМИ НА ОСНОВІ ВІДСТАНІ ЛЕВЕНШТЕЙНА

Вступ. Лінгвістичний підхід, оснований на переході від ЕКГ до кодограми, здобув популярність для аналізу серцевого ритму. Для розширення функнальних можливостей методу доцільно вивчити можливості одночасного контролю динаміки тривалості серцевих циклів і оригінального показника, який несе інформацію про ішемічні зміни міокарда.

Мета статті — розробити алгоритмічні і програмні компоненти для розв'язання цього завдання і провести експериментальні дослідження за модельними і реальними даними.

Методи. Забезпено автоматичне кодування ЕКГ, обчислення відстаней Левенштейна між парами ЕКГ певної групи випробовуваних і побудова референтної кодограми групи. Оцінювання результатів експериментальних досліджень проводилося на основі традиційних методів статистичного аналізу.

Результати. Показано, що на основі відстаней Левенштейна між усіма парами кодограм групи випробовуваних можна визначити референтну кодограму групи хворих на ішемічну хворобу серця (ІХС) і групи здорових добровольців. Встановлено, що прийняття рішень на основі порівняння кодограми ЕКГ випробуваного з еталонною кодограмою забезпечує поділ представників зазначених груп з чутливістю S_E = 72% і специфічністю C_P = 79% навіть в тих випадках, коли традиційний аналіз ЕКГ у 12 відведеннях виявляється неінформативним.

Висновки. Запропонований підхід дає змогу отримати додаткову діагностичну інформацію для вирішення актуальних завдань практичної медицини.

Ключові слова: *лінгвістичний підхід, діагностична ознака ЕКГ, відстань Левенштейна.*

Файнзильберг Л.С.[1], д-р техн. наук, профессор,
глав. науч. сотр. отд.
интеллектуальных автоматических систем
e-mail: fainzilberg@gmail.com
Дыкач Ю.Р.[2], студент,
факультет биомедицинской инженерии
e-mail: jul.dykach@gmail.com
[1] Международный научно-учебный центр
информационных технологий и систем
НАН Украины и МОН Украины,
пр-т. Академика Глушкова, 40, г. Киев, 03187, Украина
[2] Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»,
пр-т. Победы, 37, г. Киев, 03056, Украина

ЛИНГВИСТИЧЕСКИЙ ПОДХОД ДЛЯ ОЦЕНИВАНИЯ ТОНКИХ ИЗМЕНЕНИЙ ЭЛЕКТРОКАРДИОГРАММЫ НА ОСНОВЕ РАССТОЯНИЯ ЛЕВЕНШТЕЙНА

Введение. Лингвистический подход, основанный на переходе от ЭКГ к кодограмме, получил известность для анализа сердечного ритма. Для расширения функциональных возможностей метода целесообразно изучить возможности одновременного контроля динамики изменения продолжительности сердечных циклов и оригинального показателя, несущего информацию об ишемических изменениях миокарда.

Цель статьи — разработать алгоритмические и программные компоненты для решения этой задачи и провести экспериментальные исследования на модельных и реальных данных.

Методы. Обеспечивалось автоматическое кодирование ЭКГ, вычисление расстояний Левенштейна между парами ЭКГ определенной группы испытуемых и построение референтной кодограммы группы. Оценка результатов экспериментальных исследований проводилась на основе традиционных методов статистического анализа.

Результаты. Показано, что на основе расстояний Левенштейна между всеми парами кодограмм группы испытуемых можно определить референтную кодограмму группы больных ишемической болезнью сердца (ИБС) и группы здоровых добровольцев. Установлено, что принятие решений на основе сравнения кодограммы ЭКГ испытуемого с эталонной кодограммой обеспечивает разделение представителей указанных групп с чувствительностью $S_E = 72\%$ и специфичностью $C_P = 79\%$ даже в тех случаях, когда традиционный анализ ЭКГ в 12 отведениях оказывается неинформативным.

Выводы. Предложенный подход позволяет получить дополнительную диагностическую информацию при решении актуальных задач практической медицины.

Ключевые слова: *лингвистический подход, диагностический признак ЭКГ, расстояние Левенштейна.*