DOI 10.15622/ia.21.2.4

M. ABRAMOV, E. TSUKANOVA, A. TULUPYEV, A. KOREPANOVA,
S. ALEKSANIN

IDENTIFICATION OF DETERIORATION CAUSED BY AHF, MADS OR CE BY RR AND QT DATA CLASSIFICATION

Abramov M., Tsukanova E., Tulupyev A., Korepanova A., Aleksanin S. **Identification of Deterioration caused by AHF, MADS or CE by RR and QT Data Classification.**

Abstract. A sharp deterioration of the patient's condition against the backdrop of the development of life-threatening arrhythmias with symptoms of acute heart failure (AHF), multiple organ dysfunction syndrome (MODS) or cerebral edema (CE) can lead to the death of the patient. Since the known methods of automated diagnostics currently cannot accurately and promptly determine that the patient is in a life-threatening condition leading to the fatal outcome caused by AHF, MODS or CE, there is a need to develop appropriate methods. One of the ways to identify predictors of such a state is to apply machine learning methods to the collected datasets. In this article, we consider using data analysis methods to test the hypothesis that there is a predictor of death risk assessment, which can be derived from the previously obtained values of the ECG intervals, which gives a statistically significant difference for the ECG of the two groups of patients: those who suffered deterioration leading to the fatal outcome caused be MODS, AHF or CE, and those with favorable outcome. A method for unifying ECG data was proposed, which allow, based on the sequence of RR and QT intervals, to the construct of a number that is a characteristic of the patient's heart condition. Based on this characteristic, the patients are classified into groups: the main (patients with fatal outcome) and control (patients with favorable outcome). The resulting classification method lays the potential for the development of methods for identifying the patient's health condition, which will automate the detection of its deterioration. The novelty of the result lies in the confirmation of the hypothesis stated above, as well as the proposed classification criteria that allow solving the urgent problem of an automatic detection of the deterioration of the patient's condition.

Keywords: ECG-based patient classification, deterioration identification, medical prediction, ECG analysis, machine learning, artificial intelligence, data science, logistic regression, mortality risk stratification, RR interval, QT interval.

1. Introduction. Currently, in medical practice, the generally accepted method for analyzing an electrocardiogram (ECG) is a visual assessment of changes in ECG curves. Despite the great diagnostic efficiency of such an analysis, it obviously has a subjective component, and in the area of borderline states of health, its diagnostic ability is extremely unsatisfactory [1, 2]. Usually, 6 waves can be distinguished on the ECG record: P, Q, R, S, T, U. The intervals between the waves, the duration, shape, amplitude of the waves carry information that allows doctors to draw conclusions whether the patient suffers from diseases associated with the cardiovascular system. The RR interval is an indicator that characterizes the duration of the cardiac cycle and is measured in any ECG lead. The availability of research on this indicator is undeniable and has long been used in assessing heart rate variability [3, 4, 5]. For example, a decrease in

SDNN (standard deviation of RR intervals) of less than 50 milliseconds is a highly specific (88%) sign in predicting fatal outcomes in patients with myocardial infarction [6].

The QT interval, which reflects the time between the beginning of the depolarization process (the beginning of the Q wave) and the completion of the process of repolarization of the ventricular myocardium (the end of the T wave), is one of the most significant ECG parameters in cardiology. Upward or downward deviation of its value from the norm is associated with a high risk of malignant ventricular arrhythmias. Any genesis of changes in the QT interval (congenital or acquired) is equally dangerous as a risk factor for the development of ventricular tachyarrhythmias [7]. It should be noted that cardiac fluctuations are not strictly periodic and are characterized by rhythm variability [33].

There are good results in research on the spectral characteristics of heart activity [34–36], which uses mathematical modeling of the human cardiovascular system and shows its connection with different patient states. There are also results in the field of data analysis and artificial intelligence, allowing a posteriori in an automatic mode to form a class of patients with an increased risk of stroke [8] or to predict the state of the pilot, using the results of the ECG as an indicator of the pilot’s physical condition [9]. Thus, the study of the ECG has shown itself to be promising in predicting the various conditions of patients; however, until now, the question of the relationship between the state of deterioration of health in patients, manifested by symptoms of multiple organ dysfunction syndrome (MODS), acute heart failure (AHF) or cerebral edema (CE), and their ECG data, has not yet been studied. Research in this area would potentially make it possible to predict changes in this condition on the basis of ECG signal. Timely identification of such conditions and subsequent prediction of their changes are urgent tasks. Moreover, with the development of such conditions, deterioration occurs quickly, without leaving sufficient time to prepare for treatment, and standard visual analysis does not allow a sufficiently accurate forecast of the patient's condition.

Thus, on the one hand, the problem of automating the assessment of the patient's condition leading to a fatal outcome from the development of MODS, AHF or CE is urgent, its solutions are in demand; on the other hand, there are no existing studies aimed at the automation of this assessment. The aim of this study is to automate the assessment of the patient's condition based on ECG data using machine learning methods. This article presents the result of the first stage of the study, namely, the verification using data analysis methods of the hypothesis that there is a predictor of death risk assessment, which can be derived from the previously obtained values of the ECG intervals, which gives a statistically significant difference for the ECG of the two groups of patients: those who suffered deterioration leading to the fatal outcome caused be MODS, AHF or CE, and those with a favorable outcome.

The practical significance of the study lies in the formation of the potential for solving the problem of automated assessment of the patient's condition based on ECG data and predicting outcomes in case of deterioration in health using machine learning models. Such a diagnosis could allow earlier recognition and faster response to potentially life-threatening conditions of patients. Doctors could take the necessary measures before the moment of critical changes. The theoretical significance of the study lies in the proposition of the estimation method for classifying patients into the main (with MODS, AHF or CE, which led to a fatal outcome) and control (recovered) groups based on the ECG records, as well as the formation of a basis for the development of a classification methodology that allows achieving higher accuracy metrics, and predicting deterioration. The novelty of the study is attributable to the lack of methods, models and algorithms for identifying the state of deterioration and predicting its changes as a result of the development of the syndrome of acute cardiovascular, multiple organ failure or cerebral edema. The new feature based on the sequence of ECG records allowing patient’s classification into main and control group is proposed, its level of significance in a logistic regression model is 3.64×10^{-6}, a new method is proposed for solving the urgent problem of identifying the deterioration of the patient's condition on the basis of ECG analyzes, which differs from the existing ones in the formulation of the problem being solved, and in the accuracy of the results obtained.

Machine learning and data analysis methods are increasingly used today in applications to problems from the field of medicine and healthcare [10–24]. These are generally the tasks of diagnostics, predicting the course of the disease development, assistance in prescribing treatment, etc. [10–24]. An important role among these tasks is those related to monitoring and diagnostics of the most probable scenarios of the course of the disease. For example, in [10], on the basis of data collected using a smartphone (voice data on calls, data on the use of smartphones, reflecting social and physical activity), authors draw the conclusions about the user's mood, and the prospects for bipolar disorder. In [11], machine learning methods are used in diabetes monitoring to determine the risk of diabetes based on the personal medical history of patients. In addition, machine learning methods are also used to study the brain, as, for example, in [21] — for the motor imagery EEG classification.

There are other examples of the successful use of artificial intelligence and data analysis methods in tasks from the field of medicine, in particular, a wide range of works using image analysis methods to search for anomalies on X-rays and other images [16].

One of the applications of machine learning in medicine is also the automation of the recognition of critical cases for various diseases of patients. [19–21, 22, 23]. For example, [19] proposed a model for automating the analysis of breast cancer from images. In [20], the authors applied unsupervised machine learning methods to cluster ICU patients to identify patients at increased risk of mortality based on laboratory data. The authors of [21] proposed a new method for analyzing the relationship between the results of electroencephalography and human cognitive functions. The work [23] proposed methods for identifying patients with osteoporosis who have an increased risk of hip fracture.

Another common area of application of machine learning and data analysis methods in medicine is the processing of information related to the human cardiovascular system [12, 13, 17, 18, 22, 24]. Authors of [24] proposed a model for heart disease prediction; they trained three classifiers on the Cleveland dataset of 303 data instances that contains the following patient features: gender, type of chest pain, cholesterol level, maximum heart rate, etc. The authors used the following ECG features: the slope of the peak exercise ST segment V1, ST depression induced by exercise relative to rest, maximum heart rate achieved, the characteristics of resting ECG. The results showed that LR and SVM have high effectiveness (87% and 85% accuracy, respectively). The work [12] is focused on the automation of several waveforms detection on the ECG signal with the use of machine learning methods. In [13], these methods are used to develop a classification of heart diseases based on patient data from the Cleveland dataset. The author used models different from those in [24]. 6 machine learning models were used to classify patients into healthy and suffering from cardiovascular diseases subjects; CNN showed the best performance with the highest accuracy in 85.86%. [17] is focused on the ECG anomaly leading to different types of heart diseases automatic detection with MATLAB. The following features were considered: interval and amplitude values of P wave, Q wave, T wave, QRS complex, and ST segment, FIR filters showed good results. Results of [18] showed a great performance of machine learning techniques in automatic coronary artery disease detection based on ECG time series. The authors used 50 ECG features, including average and standard deviation of RR interval values, corrected QT interval values, etc. Authors of [22] proposed a method for heart disease identification with missing data handling: they replaced missing values with the mean values during pre-processing. Three ML models were trained on the data from the Cleveland dataset.

There are also works on heart disease detection which don't use ECG information. In [14], authors tried to predict cardiovascular and cerebrovascular events in patients with hypertension based on data on the status of use of medical resources, detailed information about diseases, their types, treatment methods, prescriptions, and the status of the clinic in which the treatment was carried out. In [15], machine learning methods are used to assess the development of coronary heart disease in breast cancer survivors.

Thus, the application of machine learning to ECG analysis has shown good results in identifying conditions associated with the cardiovascular system. However, the task of the analysis of ECG measurements to identify the subsequent death of the patient as a result of the development of MODS, AHF or CE has not yet been addressed.

The purpose of the general study is to automate the assessment of the patient's condition based on ECG data using machine learning methods. As the first stage of the study, this work solves the problem of testing the hypothesis using data analysis methods that there is a predictor of death risk assessment, which can be derived from the basis of previously obtained values of the ECG intervals, which gives a statistically significant difference for the ECG of the two groups of patients: those who suffered deterioration leading to the fatal outcome caused be MODS, AHF or CE, and those with a favorable outcome. The dataset provided by the Federal State Public Enterprise Nikiforov's All-Russian Center for Emergency and Radiation Medicine of the Emergencies Ministry of Russia (the Nikiforov's ARCERM) is characterized by a wide variation in the number of ECG tests taken (from 6 to 407) in different patients. This variation is explained by the different number of hospitalization days of patients, different reasons for admission to the hospital, and an increase in the frequency of testing when the condition worsens. Since the analysis of such data by statistical methods is difficult, it is necessary to unify them, that is, from the sets of values of the RR and QT intervals for each patient, obtain one value. In other words, in this work, the following problem was solved: on the basis of all the obtained values of the intervals, it is necessary to form one, which will be a characteristic of the patient's ECG records, and the differences in which for two groups (patients who died as a result of deterioration with symptoms of AHF, MODS and CE, and patients with a favorable outcome) will be statistically significant.

2. Materials and methods. The dataset was collected, the criterion derived from ECG data was proposed. The formulated hypothesis was tested using logistic regression.

2.1. Dataset. The first stage of the study was carried out on the basis of anonymized data on 120 patients of both sexes who were consecutively hospitalized in the Federal State Public Enterprise Nikiforov's All-Russian Center for Emergency and Radiation Medicine of the Emergencies Ministry of Russia (the Nikiforov's ARCERM) in the period from 01.01.2015 to 31.12.2017. The main group (MG) consisted of 60 patients with a lethal outcome as a result of deterioration; the control group (CG) consisted of 60 patients with a favorable outcome. The dataset provided by ARCERM at the first stage consisted of tables containing the following information: patient diagnoses, data on all ECG records, general data about the patient, data on events significant for the patient's condition that occurred during his stay in the hospital (surgery, life-threatening change in the rhythm of heart contractions, etc.). Below we will take a closer look at each of the tables:

Patient diagnosis table was formed on the basis of epicrises; each patient in the dataset had from 1 to 37 diagnoses.

ECG diaries contained records of all available amenable to analysis of ECG records in the form of diaries. All patients underwent a standard 12-lead ECG: 4 electrodes were placed on the arms and legs distal from the shoulder and thigh; 6 electrodes were placed on the chest surface at the following points (V1 – fourth intercostal space along the right sternal line, V2 – fourth intercostal space along the left sternal line, V3 – in the middle of the distance between V2 and V4, V4 – fifth intercostal space along the midclavicular line, V5 – in horizontal plane V4 along the anterior axillary line, V6 – in the horizontal plane V4 along the mid-axillary line). The duration of the QT interval on the resting ECG using the interval editing module was measured by the classical method of E. Lepeshkin and B. Surawicz [25], which is based on drawing a straight line tangent along the line of maximum slope of the descending part of the T-wave until it intersects with the isoline (slope method). Measurements were performed in standard lead II and, in the case of a pronounced U-wave, in chest lead V5. For ECG measurements, MAC 1600 ECG Analysis System and MUSE NX were used. The table contains the following information:

- Date and time of measurement;
- Data on the rhythm of heart contractions. In the presented data set, 4 main types of rhythm were distinguished: 1) sinus; 2) atrial fibrillation and atrial flutter; 3) ectopic; and 4) with a permanent pacemaker;
- Whether the rhythm has changed since the last diary. The exact time of the change is unknown; the rhythm can change from one of the first three types described above to another of the first three types;
- Whether the patient was in the intensive care unit at the time of the ECG recording;

- Characteristics of the general condition of the patient (satisfactory, moderate, severe, etc.);
- Data on the patient's hemodynamics (stable, unstable), on the patient's breathing and consciousness, which together determine whether the patient is in deterioration. This information was extracted from complete diaries kept by doctors in a free-form text while observing patients;
- Whether the patient experienced deterioration during the ECG recording;
- Was the operation performed on the patient close to the time of the ECG recording;
- Whether sympathomimetic drugs have been administered to the patient.

General patient data contained information on the following patient characteristics: age, gender, date of admission, date of a discharge / fatal outcome, number of days the patient was hospitalized, cause of fatal outcome, and hospitalization department.

RR and QT intervals contained data on the RR and QT intervals corresponding to the measurement diaries. Each diary corresponds to a different number of pairs of RR and QT intervals; for each pair of intervals, their lengths are known, as well as the presence of supraventricular and ventricular extrasystoles.

Next, let's consider the primary statistics of the dataset. Patients were treated in the departments of cardiology, cardiovascular surgery, intensive care, neurology, neurosurgery, emergency surgery, therapy, traumatology, hematology, and rehabilitation. In the main group, 26 (43.3%) patients had multiple organ dysfunction syndrome as the cause of fatal outcome, 28 (46.7%) patients had an acute cardiovascular failure, and 6 (10%) had cerebral edema. The gender distribution in both groups was the same and amounted to 50% of men and 50% of women. The duration of hospitalization ranged from 1 to 172 bed-days. Among the patients of the main group, diseases of the circulatory system prevailed as the main pathology and amounted to about 50% of cases, taking into account the concomitant pathology, the number of those suffering from diseases of the circulatory system is greater – 83%. In the control group, diseases of the circulatory system also prevailed, amounting to more than 60% (Table 1), also taking into account the concomitant pathology, the number of those suffering from diseases of the circulatory system was greater – 82%.

During hospitalization, all patients underwent electrocardiogram testing. Thus, each patient received from 1 to 38 ECGs. Due to the variability of heart rate between ECG tests, and also due to the fact that the quality of the records did not always allow using the entire ECG record, each processed record contained a different number of pairs of RR and QT intervals: from 3 to 28.

Table 1. Clinical characteristics of patients

Pathology (ICD code)	MG *n* (%)	CG *n* (%)
Main pathology		
Diseases of the circulatory system (I00–I99)	29 (48,3)	41 (68,3)
Neoplasms (C00–D48)	16 (26,7)	5 (8,3)
Endocrine system diseases (E00–E90)	2 (3,3)	5 (8,3)
Infectious and parasitic diseases (A00–B99)	5 (8,3)	0,0
Trauma (S00–T98)	6 (10,0)	0,0
Diseases of the digestive system (K00–K93)	2 (3,3)	9 (15,0)
Concomitant pathology		
Diseases of the circulatory system (I00–I99)	23 (38.3)	9 (15.0)
Neoplasms (C00–D48)	3(5.0)	2 (3.3)
Endocrine system diseases (E00–E90)	18 (30.0)	15 (25.0)
Diseases of the Genitourinary System (N00–N99)	26 (43.3)	12 (20.0)
Diseases of the nervous system (G00–G99)	7 (11.7)	7 (11.7)
Diseases of the digestive system (K00–K93)	22 (36.7)	18 (30.0)
Diseases of the respiratory system (J00–J99)	23 (38.3)	7 (11.7)

The distribution by the type of the rhythm for all ECG records was comparable in both groups and amounted to about 60% for sinus rhythm, about 25-30% for atrial fibrillation (Table 2). The original dataset contained 3 patients who received a permanent pacemaker. Since the pacemaker affects the heart rhythm, the changes in the ECG of these patients may be of a different nature, so these patients were excluded from the study at the moment. The final dataset contained 117 patients: 58 from the main group and 59 from the control group.

Table 2. Electrocardiographic characteristics of patients

Rhythm	Main group	Control group
	n (%)	*n* (%)
Sinus rhythm	166 (61,7)	123 (65,0)
Atrial fibrillation	63 (23,3)	63 (33,3)
Permanent pacemaker	9 (3,3)	3 (1,7)
Ectopic rhythm	31 (11,7)	0 (0,0)

At the second stage of the study, the model was tested on new data of 38 patients of both sexes who were consecutively admitted to the Federal

State Public Enterprise Nikiforov's All-Russian Center for Emergency and Radiation Medicine of the Emergencies Ministry of Russia (the Nikiforov's ARCERM) during the period from 01.01.2018 to 31.12.2019. The main group (MG) consisted of 19 patients with a lethal outcome, the control group (CG) consisted of 19 patients with a favorable outcome. Sex distribution in both groups (main and control): 9 men and 10 women. The duration of the hospitalization ranged from 1 to 105 bed-days.

2.2. Data preprocessing. The dataset is characterized by a high degree of heterogeneity, for example, the table with diagnoses contains 1986 diagnoses, 1621 of them are unique within the dataset. Also, each patient has a variable number of diagnoses, days of hospitalization, ECG records, RR and QT intervals, etc. An additional difficulty when working with this data is the small size of the dataset. Thus, most of the data is either inapplicable for analysis, or requires a significant degree of preprocessing, so a subset of features was extracted from the general patient data set.

Based on the information about patients collected at the first stage, a database was formed, which contained information about the gender and age of patients, the number of days of hospitalization, pairs of RR and QT intervals from all ECG records.

Based on the available data, two tables were generated. In the first table, each row contains the patient identification number; the group to which the patient belongs (MG or CG); the patient’s gender; age; the number of days of hospitalization, and information about whether the type of the heart rate has changed or not. Table 3 shows the general view of the row in the table.

Table 3. General view of a table row containing information about patients

Patient number	Patient group	Patient sex	Patient age	Number of days of hospitaliza tion	Was there a change in rhythm?
$n : n \in$ 1..117	$g : g \in$ {0;1} (0 for MG; 1 for CG)	$p : p \in \{0,1\}$ (0 for female; 1 for male)	$v : v \in [34, 94]$	$k : k \in$ 0..172	$r : r \in$ {0;1} (0 if not; 1 if it was)

The second dataset table contains information about all measured intervals (RR and QT interval sets). Since each patient underwent several ECGs, and each ECG contains several RR and QT intervals, there are several rows in the table for each patient, each of which records the patient's number, as well as the values of the RR and QT intervals. For each patient, there are as many rows as there are RR and QT intervals (from 6 to 407). Table 4 shows a general format of the row of the analyzed dataset.

Table 4. General format of the table row containing information about the RR and QT intervals

Patient ID	RR interval value	QT interval value
$n : n \in 1..117$	$\mathrm{rr} : \mathrm{rr} \in 244..3042$	$\mathrm{qt{:}qt} \in 141..692$

2.3. The proposed method for the unification of ECG data. To simplify the further presentation of the research results, we introduce the following notation:

Notation. Each patient number *i* has n_i pairs of RR and QT intervals.

Notation. For the patient number *i* the value of the *j*-th RR interval is denoted by RR_{ij}.

Notation. For the patient number *i* the value of the *j*-th QT interval is denoted by QT_{ij}.

Notation. For the patient number *i* the set of all the values of the patient's RR intervals is denoted by RR_i.

Notation. For the patient number *i* the set of all the values of the patient's QT is denoted by QT_i.

In addition, we will use the notation for some numerical characteristics, used for unifying the data on the recorded ECG.

Notation. For patient number *i* the mean value of cubes of the values of RR intervals is denoted by $\overline{\mathrm{RR}_i}^{(3)}$. To put it another way, $\sum_{j=1}^{n_i} \frac{\mathrm{RR}_{ij}^{\ 3}}{n_i} = \overline{\mathrm{RR}_i}^{(3)}$.

Notation. For patient number *i* the mean value of cubes of the values of QT intervals is denoted by $\overline{\mathrm{QT}_i}^{(3)}$. To put it another way, $\sum_{j=1}^{n_i} \frac{\mathrm{QT}_{ij}^{\ 3}}{n_i} = \overline{\mathrm{QT}_i}^{(3)}$.

Notation. For patient number *i* as the **criterion value** will be referred the special dimensionless constant K_i, which is denoted by the following: $K_i = \log_{\overline{\mathrm{QT}_i}^{(3)}} \overline{\mathrm{RR}_i}^{(3)}$.

This criterion was chosen because it reflects the relationship between the RR and QT intervals in all of the patient's ECGs. This criterion was chosen among a set of functions with similar properties, as having the highest classifying ability.

It is proposed to consider the values of the K_i criterion or the analyzed dataset as a marker for classifying patients according to the values of the RR and QT intervals into two classes: the main group and the control group. Figures 1, 4 show the distribution of test values in both groups. Figures 2, 3 show the distribution of the criterion value in the MG and CG groups, respectively. According to Shapiro-Wilk normality test, criterion value isn't distributed normally, p-value is equal 0.2833 for the distribution of the criterion value in both groups, 0.4829 for MG, and 0.7372 for CG.

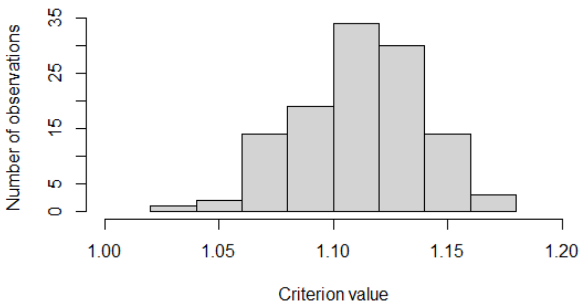

Fig. 1. Distribution of criterion values in both group

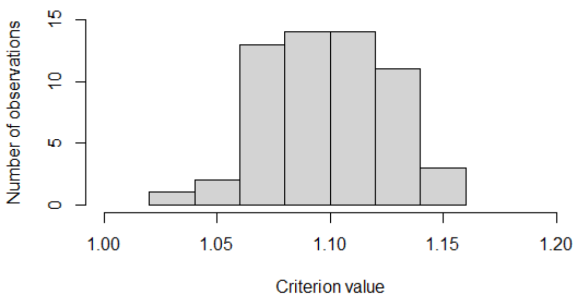

Fig. 2. Distribution of criterion values in the MG group

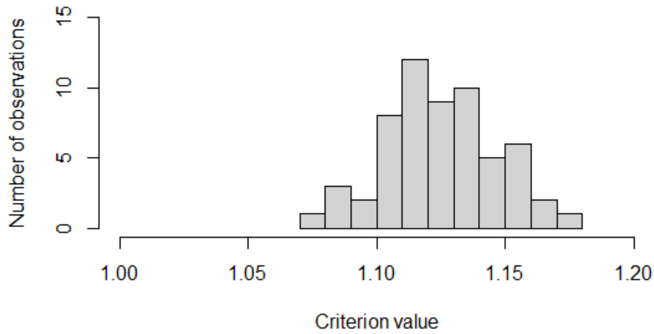

Fig. 3. Distribution of criterion values in the CG group

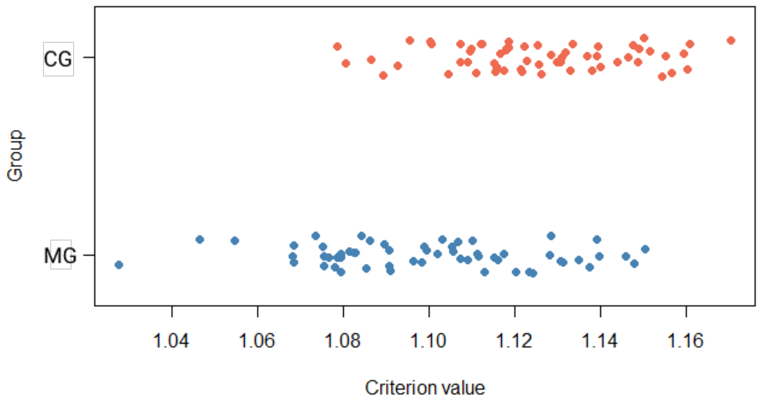

Fig. 4. Scatterplot of criterion value distribution in two groups (1 – MG, 0 – CG)

Hypothesis. There is a predictor of death risk assessment, which can be derived from the previously obtained values of the ECG intervals, which gives a statistically significant difference for the ECG of the two groups of patients: those who suffered deterioration leading to the fatal outcome caused be MODS, AHF or CE, and those with a favorable outcome. The criterion value can be used for patient classification into one of two groups.

To test the hypothesis, a logistic regression model was applied, this model doesn’t require specific distribution of predictors. Logistic regression is one of the machine learning methods [25], a statistical model used to predict the likelihood of an object belonging to one of two classes using a logistic function. The following assumptions are made about the probability of an object belonging to classes 1 and 0:

$$P(y = 1 \mid x) = h_w(x),$$
$$P(y = 0 \mid x) = 1 - h_w(x),$$

$$h_w(x) = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + \ldots + w_n x_n)}},$$

where $x_j,\ j \in 1..n$ denotes attributes of the object x, $w_j : j \in 0..n$ — regression coefficients. In our case, the patients are the objects, the criterion value is the attribute of the object, the patient groups are the classes (1 for MG, and 0 for CG).

3. Results. This section describes the results of testing the proposed criterion during the first and the second stages of the study.

3.1. Testing the criterion. Table 5 shows the characteristics of the logistic regression model between the patient group and the criterion value. Figure 5 shows a plot of sensitivity and specificity (ROC-curve) for the initial data, figure 6 – graph of the logistic function.

Since the *p-value* of the test of the association between criterion value and the patient group is 3.64×10^{-6}, which is less than 0.05, then we can talk about the presence of a statistically significant association. As the size of the dataset was bigger than 30 subjects, and the maximum likelihood estimates are asymptotically normal, for the p-value calculation z-score was applied that uses normal distribution.

Table 5. Results of logistic regression between the group and the value of the criterion for the initial data

	Estimate	Std. Error	*z* value	*p*-value
(Intercept)	-49.791	10.762	-4.626	3.72×10^{-6}
slope	44.753	9.664	4.631	3.64×10^{-6}

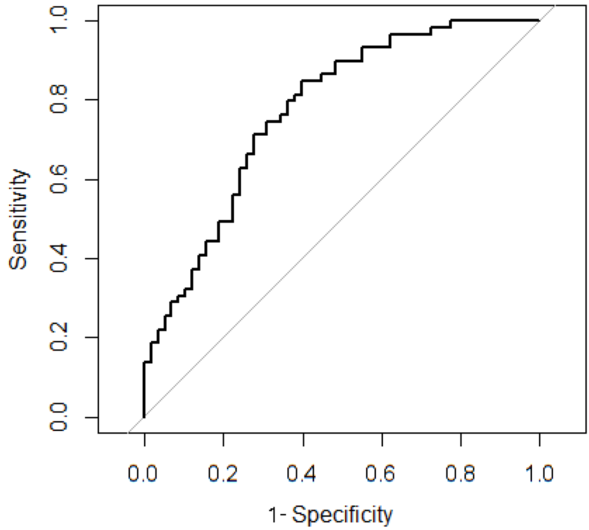

Fig. 5. A plot of sensitivity and specificity for the initial data (ROC-curve)

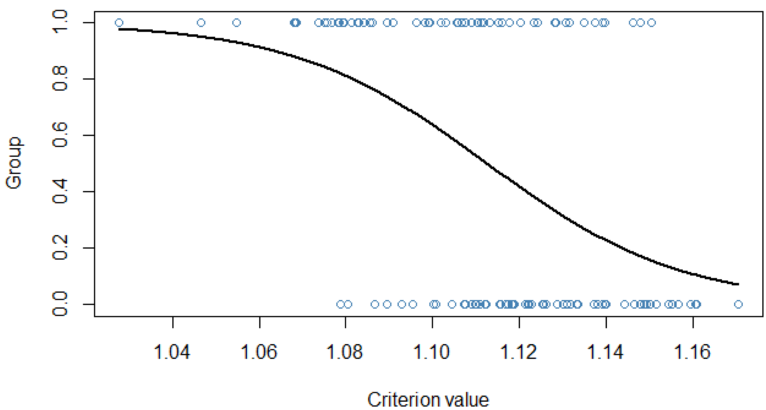

Fig. 6. Logistic function graph

The area under the ROC-curve was 0.772, which testifies to the good quality of the classification. Thus, it is possible to accept the hypothesis formed for the initial data. An increase in the criterion value by 1 entails an increase in the logarithm of the odds ratio of dying by 44.7.

Now, knowing that the value of the patient's criterion is a statistically significant indicator in determining the patient's group, it is possible to test how well this model is able to classify patients into the groups indicated above (MG and CG). In the role of the metric for assessing the quality of classification, we used sensitivity, that is, the proportion of all correctly classified patients relative to all patients in this group in the sample (Table 6).

Table 6. Results of classification using logistic regression.

Patient group	Number of correctly classified	Number of all patients	Classification sensitivity in percentage
All patients	82	117	75%
MG patients	40	58	78%
CG patients	42	59	73%

3.2. Testing the criterion on new data. The proposed model was tested on new data obtained at the second stage of data collection. The new dataset contained information on 38 patients, 19 of whom belonged to the MG group and 19 to the CG group. Table 7 shows the results of applying the resulting logistic regression model to new data.

Table 7. Testing the model on new data

Patient group	Number of correctly classified	Number of all patients	Classification sensitivity in percentage
All patients	25	38	65%
MG patients	14	19	73%
CG patients	11	19	53%

3.3. Testing the criterion for patients with different causes of a fatal outcome. As a result of the work done, a model was built to classify patients into two groups: MG (with a fatal outcome as a result of deterioration) and CG (with a favorable outcome), based on data from the sequence of ECG tests. In the resulting dataset, deterioration manifested itself as symptoms of the following three conditions: acute heart failure, multiple organ dysfunction syndrome, and cerebral edema. Knowing the patient's risk of fatal outcome for whatever reason could help in treating the patient in a life-threatening condition. Also, deterioration leading to fatal outcome for various reasons can have a different clinical picture. From these points of view, it is of interest to evaluate the ability of criterion to classify patients who had deterioration leading to the different causes of fatal outcome. Figure 7 shows a scatter diagram of the criterion values for patients with different causes of fatal outcome, Table 8 shows the results of the classification of patients with different causes of fatal outcome on a combined dataset consisting of data obtained at the first and second stages. For better visualization, groups of patients with different causes of death are clearly separated along the OX axis in the diagram.

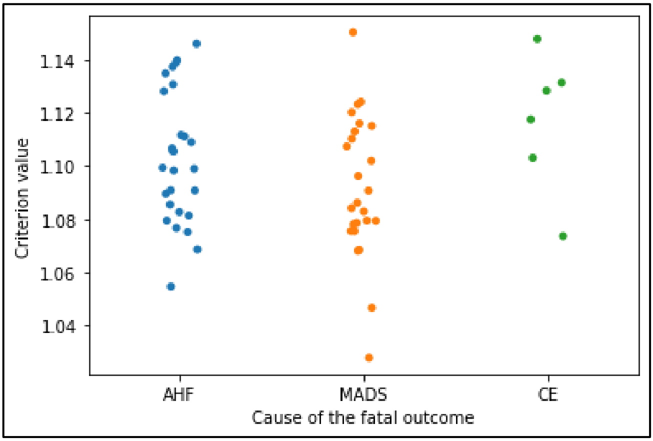

Fig. 7. Scatter diagram of the criterion values for patients with different causes of fatal outcome

Table 8. Testing the model for patients with different causes of fatal outcome

Patient group	Number of correctly classified	Number of all patients	Classification sensitivity in percentage
All patients	107	155	69%
Patients with MODS	22	30	73%
Patients with AHF	30	41	73%
Patients with CE	2	6	33%
CG patients	53	78	68%

Thus, according to the results of testing the model for patients with different causes of fatal outcome, this model best categorizes patients with a life-threatening state of deterioration that subsequently led to multiple organ dysfunction syndrome, and the worst categorizes those whose cause of fatal outcome was cerebral edema.

3.4. Software implementation. The developed model was applied in the implementation of a prototype web service that allows using ECG data to determine whether a patient suffered from a life-threatening deterioration. The web service was developed using PHP for the server-side, HTML, CSS, JavaScript for the client-side and is hosted on the page at https://ecg.dscs.pro. To use it, you need to upload a file in excel format with data on the values of the pairs of RR and QT intervals of the patient in the following format, presented in Table 9:

Table 9. File format with data on the values of the patient's RR and QT intervals for uploading to the web service form

RR value	QT value
$rr : rr \in 244..3042$	$qt : qt \in 141..692$

Figure 8 shows the main page and interface of the web service prototype, Figure 9 shows the screen with the output of the result.

Fig. 8. Web service prototype interface. Translation: «Analysis of the patient's condition according to ECG data», «Upload a document», «Calculate»

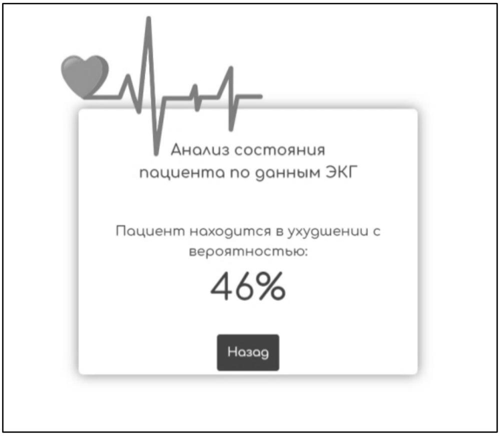

Fig. 9. A screen displaying the result. Translation: «Analysis of the patient's condition according to ECG data», «The patient is in deterioration with a probability of 46%»

4. Discussion. The difference in the distribution of the criterion value in two groups (MG, CG) is statistically significant, so the proposed method for unifying the RR and QT interval values can be used as an independent criterion for classifying patients into MG (with a fatal outcome as a result of worsening with symptoms of MODS, AHF and CE) and CG (with a favorable outcome). This follows from the fact that for the initial data, the *p-value* in the logistic regression was close to 0, and the area under the ROC curve was 0.772. Thus, the hypothesis that there is a predictor of death risk assessment, which can be derived from the previously obtained values of the ECG intervals, which gives a statistically significant difference for the ECG of the two groups of patients: those who suffered deterioration leading to the fatal outcome caused be MODS, AHF or CE, and those with favorable outcome was confirmed. Each ECG analysis results in a set of RR and QT interval values. Thus, based on a sequence of ECG records, the value of the K_i criterion can be calculated to identify whether the patient had a state of deterioration leading to fatal outcome using a logistic regression model. The results of classifying patients into risk groups on the new dataset are worse than the results on the original dataset; nevertheless, the likelihood of correct identification of deterioration in patients is quite high. The model was tested on patients with fatal outcomes for various causes. The model showed itself in the best way when identifying deterioration in patients whose cause of fatal outcome was the syndrome of multiple organ failure: the sensitivity of determining such patients was 78%.

On the one hand, the approach proposed in the article is limited by the fact that it does not take into account the individual characteristics of the patient, on the other hand, taking into account the high-quality characteristics of the proposed criterion, it is very likely that the proposed indicator sufficiently aggregates information about the state of the patient's cardiovascular system. The main distinctive feature of this study is that it aims to find common features of patients with different causes of fatal outcome and derive the predictor of deterioration from their ECG data. The results obtained show that the proposed estimation method succeeds in finding similar features in ECGs of patients with MODS and AHF, but works poorly for patients with CE. Given the small number of the patients with this condition, we probably will further exclude them from the study.

The methodological choices were constrained by the size of the dataset as well as the characteristics of data received from ARCERM. In the future, the dataset will be expanded, and the number of considered features will also be increased, which will allow the use of a wider toolkit, which may positively affect the results of the study.

5. Conclusion. The logistic regression model used in the article demonstrated the possibility of accepting the hypothesis that the proposed method for unifying the values of the RR and QT intervals for one patient makes a statistically significant feature for determining the patient group (the main one – with a fatal outcome and the control one – with a favorable outcome). The constructed estimation method showed good results on new data as well. Thus, the hypothesis is accepted that there is a predictor of death risk assessment, which can be derived from the previously obtained values of the ECG intervals, which gives a statistically significant difference for the ECG of the two groups of patients: those who suffered deterioration leading to the fatal outcome caused be MODS, AHF or CE, and those with a favorable outcome.

The obtained criterion allows, with a sensitivity of at least 65%, to identify patients who had a state of deterioration as a result of the development of acute heart failure syndrome, multiple organ failure or cerebral edema, based on the results of ECG analysis. The model showed the best sensitivity (83%) when identifying the condition of patients whose cause of death was multiple organ dysfunction syndrome. As a result, the constructed model was used to develop a prototype web service for assessing the patient's condition and automatically determining life-threatening deterioration of condition based on ECG data. The practical significance of the result is based on the fact that such a diagnosis can allow taking the necessary measures before the moment of critical changes. Patients may show signs of deterioration long before they become unstable. The ability to recognize these signs can lead to a decrease in mortality. The theoretical significance of the result lies in the construction of a mathematical model for classifying patients into those who had a favorable outcome, and those who have a fatal outcome as a result of the development of MODS, AHF or CE, as well as the basic formation for the development of a methodology for predicting the development of life-threatening conditions. Several new questions emerge in light of the discoveries presented here. One of the possible further research directions lies in the inclusion of the obtained results into a more complex model that takes into account other information about the patient, such as gender, age, the presence of a change in rhythms, etc. Bayesian trust networks or algebraic Bayesian networks can serve as such a model [26–29]. The advantage of the latter is that they can work with incomplete and inaccurate data, which can be an important advantage in the further work with medical data [28, 30, 31]. This will allow more information about the patient to be used, which can increase the chance of correct classification of the patient's risk group. The second direction can be the study of the dynamics of changes in ECG, in other words, the study of changes in RR and QT interval values over time, and building a model based on time series, which describes the dynamics of changes in the studied interval values. It will also provide new information about the patient that can help predict disease outcomes. The third possible direction is the construction of a model for classifying deterioration in terms of probable causes of fatal outcome, that is, a model that allows predicting the risk of the development of acute heart failure, multiple organ dysfunction syndrome, or cerebral edema.

References

1. [Preventive cardiology. Under. ed. G.M. Kositsky.]. Meditsina [Medicine]; 1987. 512 p. (InRuss.).
2. Mjasnikov A.L. Jeksperimental'nye nekrozy miokarda [Experimental myocardial necrosis]. M. Medicina. 1963. 204 p. (InRuss.).
3. Brodsky M., Wu D., Penes P., Kanakis Ch., Rosen K. Arrythmias documented by 24 hour continiusnelectrocardiographic monitoring in 50 male medical students without apparent heart diseases. The American journal of cardiology. 1977. vol. 39. No. 3. pp. 390–395. doi: 10.1097/00132586-197710000-00002
4. Umetini K., Singer D., McCarty R., Atkinson M. 24 Hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. Journal of the American College of Cardiology. 1998. vol. 31. No. 3. pp. 593–601. doi: 10.1016/s0735-1097(97)00554-8
5. Makarov L.M. Holterovskoe monitorirovanie. [Holter monitoring.] M: ID "Medpraktika – M". 2008. 504 p. (InRuss.).
6. Kleiger R.E., Miller J.P., Bigger Jr J.T., Moss A.J. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. The

American journal of cardiology. 1987. vol. 59. №. 4. pp. 256–262. doi: 10.1016/0002-9149(87)90795-8

7. Anischenko V.S. Znakomstvo s nelinejnoj dinamikoj [Acquaintance with nonlinear dynamics.]. Izdatelstvo LKI. 2018. 224 p. (InRuss.).
8. Rathakrishnan K., Min S.N., Park S.J. Evaluation of ECG Features for the Classification of Post-Stroke Survivors with a Diagnostic Approach. Applied Sciences. 2021. vol. 11. No. 1: 192.
9. Wang X., Gong G., Li N., Ding L., Ma Y. Decoding pilot behavior consciousness of EEG, ECG, eye movements via an SVM machine learning model. International Journal of Modeling, Simulation, and Scientific Computing. 2020. vol. 11. №. 04: 2050028.
10. Antosik-Wojcinska A.Z., Dominiak M., Chojnacka M. Kaczmarek-Majer K., Opara K.R., Radziszewska. W., Olwert A., Swiecicki L. Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling. International Journal of Medical Informatics. 2020. vol. 138: 104131.
11. Srivastava A.K., Kumar Y., Singh P.K. A Rule-Based Monitoring System for Accurate Prediction of Diabetes: Monitoring System for Diabetes. International Journal of E-Health and Medical Communications (IJEHMC). 2020. vol. 11. №. 3. pp. 32–53.
12. Aziz, S., Ahmed, S. & Alouini, M.S. ECG-based machine-learning algorithms for heartbeat classification. Sci Rep 11. 2021. vol. 18738. https://doi.org/10.1038/s41598-021-97118-5.
13. Tougui I., Jilbab A., El Mhamdi J. Heart disease classification using data mining tools and machine learning techniques. Health and Technology. 2020. vol. 10. pp. 1137–1144.
14. Park J., Kim J.W., Ryu B., Heo E., Jung S.Y., Yoo S. Patient-Level Prediction of Cardio-Cerebrovascular Events in Hypertension Using Nationwide Claims Data. Journal of medical Internet research. 2019. vol. 21. №. 2: e11757.
15. Guo A., Zhang K.W., Reynolds K., Foraker R.E. Coronary heart disease and mortality following a breast cancer diagnosis. BMC medical informatics and decision making. 2020. vol. 20: 88.
16. Kovalev M.S., Utkin L.V., Kasimov E.M. SurvLIME: A method for explaining machine learning survival models. Knowledge-Based Systems. 2020. vol. 203. pp. 106–164.
17. Sohal H., Jain S. Interpretation of cardio vascular diseases using electrocardiogram: A study. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, 2018. pp. 159–164.
18. Yao L.K., Liu C.C., Li P., Wang J.K., Liu Y.Y., Li W., Wang X.P., Li H., Zhang H. Enhanced Automated Diagnosis of Coronary Artery Disease Using Features Extracted From QT Interval Time Series and ST–T Waveform. IEEE Access. 2020. vol. 8. pp. 129510–129524.
19. George K., Sankaran P., Joseph K.P. Computer assisted recognition of breast cancer in biopsy images via fusion of nucleus-guided deep convolutional features. Computer Methods and Programs in Biomedicine. 2020. vol. 194: 105531.
20. Hyun S., Kaewprag P., Cooper C., Hixon B., Moffatt-Bruce S. Exploration of Critical Care Data by Using Unsupervised Machine Learning. Computer Methods and Programs in Biomedicine. 2020. vol. 194: 105507.
21. Hang W.L., Feng W., Liang S., Wang Q., Liu X.J., Choi K.S. Deep Stacked Support Matrix Machine Based Representation Learning for Motor Imagery EEG Classification. Computer Methods and Programs in Biomedicine. 2020. vol. 193: 105466.
22. Louridi N., Amar M., El Ouahidi B. Identification of Cardiovascular Diseases Using Machine Learning. Proceedings of the 7th Mediterranean Congress of

Telecommunications 2019, CMT 2019, Fez, Morocco, 24–25 October 2019. 2019. pp. 1–6

23. Villamor E., Monserrat C., Del Rio L., Romero-Martin J.A., Ruperez M.J. Prediction Of Osteoporotic Hip Fracture In Postmenopausal Women Through Patient-Specific FE Analyses And Machine Learning. Computer Methods and Programs in Biomedicine. 2020. vol. 193: 105484.
24. Perumal R., Kaladevi A.C. Early Prediction of Coronary Heart Disease from Cleveland Dataset using Machine Learning Techniques. Int. J. Adv. Sci. Technol. 2020. vol. 29. pp. 4225–4234.
25. Lepeshkin E., Surawicz B. The measurement of the QT interval of the electrocardiogram. Circulation, 1952;6:378-388 https://doi.org/10.1161/01.cir.6.3.378
26. Rymarczyk T., Kozłowski E., Kłosowski G., Niderla K. Logistic Regression for Machine Learning in Process Tomography. Sensors. 2019. vol. 19. No. 15: 3400. doi: 10.3390/s19153400
27. Bishop C. Pattern Recognition and Machine Learning. Springer-Verlag New York, 2006. 738 p.
28. Delen D., Topuz K., Eryarsoy E. Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. Operations research and management sciences. 2020. vol. 281. pp. 575–587.
29. Kharitonov N.A., Maximov A.G., Tulupyev A.L. Algebraic Bayesian networks: the use of parallel computing while maintaining various degrees of consistency. International Conference on Information Technologies. Springer, Cham. 2019. pp. 446–456.
30. Nielsen T.D., Jensen F.V. Bayesian networks and decision graphs. Springer Science Business Media. 2009. 464 p.
31. Tulupyev A.L., Stolyarov D.M., Mentyukov M.V. Representation of the local and global structure of an algebraic Bayesian network in Java applications. Proceedings of SPIIRAS. 2017. vol. 5. No. 0. pp. 71–99.
32. Kharitonov N., Malchevskaia E., Zolotin A., Abramov M. External consistency maintenance algorithm for chain and stellate structures of algebraic bayesian networks: statistical experiments for running time analysis. International Conference on Intelligent Information Technologies for Industry. Springer, Cham. 2018. pp. 23–30.
33. Baevskij R.M., Ivanov G.G., Gavrilushkin A.P., Dovgalevskij P.Ja., Kukushkin Ju.A., Mironova T.F., Priluckij D.A., Semenov A.V., Fedorov V.F., Flejshman A.N., Medvedev M.M., Chirejkin L.V. [Analysis of Heart Rate Variability When Using Different Electrocardiographic Systems (Part 1)]. Vestnik aritmologii [Bulletin of Arithmology]. 2002. No. 24. pp 65–86.
34. Ponomarenko V.I., Karavaev A.S., Borovkova E.I., Hramkov A.N., Kiselev, A.R., Prokhorov M.D., Penzel T. Decrease of coherence between the respiration and parasympathetic control of the heart rate with aging. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2021. vol. 31. no. 7: 073105.
35. Karavaev A.S., Ishbulatov Yu Y.M., Prokhorov M.D., Ponomarenko V.I., Kiselev A.R., Runnova A.E., Hramkov A.N., Semyachkina-Glushkovskaya O.V., Kurths J., Penzel T. Simulating dynamics of circulation in the awake state and different stages of sleep using non-autonomous mathematical model with time delay. Frontiers in Physiology. 2020. vol. 11: 1656.
36. Karavaev A.S., Borovkova E.I., Runnova A.E., Kiselev A.R., Zhuravlev M.O., Ponomarenko V.I.,.Prokhorov M.D., Koronovskii A.A., Hramov A.E. Experimental Observation of the Self-Oscillatory Dynamics of the Regulation Contours of the Cardiovascular System. Radiophysics and Quantum Electronics. 2019. vol. 61. no. 8. pp. 681-688.

Abramov Maxim — Ph.D., Head of laboratory, senior researcher, Laboratory of theoretical and interdisciplinary computer science, St Petersburg Federal Research Center of the Russian Academy of Sciences; Assistant professor, Computer science department, faculty of mathematics and mechanics, SPbSU. Research interests: information security, social engineering attacks, analysis of the security of users of information systems from social engineering attacks of intruders; analysis of the dissemination of information in social networks based on the models used in the analysis of the security of users of information systems from social engineering attacks, analysis and modeling of social networks, client-server technologies, researching the relationship between content posted by users on social networks and offline behavior, business analytics, social computing, business intelligence. The number of publications — 100. mva@dscs.pro; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-3337.

Tsukanova Ekaterina — Chief specialist, The Federal State Budgetary Institute "The Nikiforov Russian Center of Emergency and Radiation Medicine" The Ministry of Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters (EMERCOM of Russia). Research interests: statistical research methods in medicine, integrative approach in functional diagnostics, decision support systems for early detection, prevention, selection and assessment of the effectiveness of therapy for life-threatening arrhythmias, rational organization of the diagnostic service in a modern hospital. The number of publications — 8. Eka_77@bk.ru; 4/2, Academician Lebedev St., 194044, St. Petersburg, Russia; office phone: +7(812)607-5900.

Tulupyev Alexander — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of theoretical and interdisciplinary computer science, St Petersburg Federal Research Center of the Russian Academy of Sciences; Professor, Computer science department, faculty of mathematics and mechanics, SPbSU. Research interests: uncertain knowledge and data representation and processing, application of mathematics and computer science in sociocultural and epidemiological studies, software technologies and development of information systems with databases. The number of publications — 400. alt@dscs.pro; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-3337.

Korepanova Anastasia — Junior researcher, laboratory of theoretical and interdisciplinary computer science, St Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: information security, social engineering attacks, analysis of the security of users of information systems from social engineering attacks of intruders, analysis and modeling of social networks, machine learning, data analysis. The number of publications — 13. aak@dscs.pro; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-3337.

Aleksanin Sergei — Ph.D., Dr.Sci., Professor, Corresponding member, Director, The Federal State Budgetary Institute "The Nikiforov Russian Center of Emergency and Radiation Medicine" The Ministry of Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters (EMERCOM of Russia). Research interests: disaster medicine, problems of a person's functional state in extreme and unfavorable environmental conditions, occupational pathology, radiation medicine, preventive and restorative medicine. The number of publications — 356. medicine@nrcerm.ru; 4/2, Academician Lebedev St., 194044, St. Petersburg, Russia; office phone: +7(812)607-5900.

Acknowledgements. The work was carried out within the framework of the project under the state order of the St. Petersburg Federal Research Center of the Russian Academy of Sciences SPIIRAN No. FFZF-2022-0003.

УДК 004.89 **DOI** 10.15622/ia.21.2.4

М.В. Абрамов, Е.И. Цуканова, А.Л. Тулупьев, А.А. Корепанова, С.С. Алексанин

ИДЕНТИФИКАЦИЯ КЛИНИЧЕСКОГО УХУДШЕНИЯ В РЕЗУЛЬТАТЕ РАЗВИТИЯ ОСН, СПОН ИЛИ ОГМ ПОСРЕДСТВОМ КЛАССИФИКАЦИИ НА ОСНОВЕ ДАННЫХ ОБ ИНТЕРВАЛАХ RR И QT

Абрамов М.В., Цуканова Е.И., Тулупьев А.Л., Корепанова А.А., Алексанин С.С. **Идентификация клинического ухудшения в результате развития ОСН, СПОН или ОГМ посредством классификации на основе данных об интервалах RR и QT.**

Аннотация. Резкое ухудшение состояния на фоне развития жизнеугрожающих аритмий с симптомами острой сердечной недостаточности (ОСН), синдрома полиорганной недостаточности (СПОН) или отёка головного мозга (ОГМ) может привести к гибели пациента. Поскольку известные методы автоматизированной диагностики в настоящий момент не могут достаточно точно и своевременно определить, что пациент находится в жизнеугрожающем состоянии, ведущем к летальному исходу от ОСН, СПОН или ОГМ, существует необходимость в разработке соответствующих методов. Одним из способов выявить предикторы такого состояния является применение методов машинного обучения к накопленным наборам данных. В данной статье решалась задача проверки с помощью методов анализа данных гипотезы о наличии зависимости между результатами измерения ЭКГ и последующим летальным исходом пациента в результате развития СПОН, ОСН или ОГМ. Был предложен метод комбинирования данных, сводящейся к тому, чтобы на основе характеристик ЭКГ для каждого пациента предложить алгоритм, на вход которого подаются пары интервалов RR и QT, а на выходе получается число, которое является характеристикой состояния пациента. На основе полученной характеристики производится классификация пациентов на группы: основную (пациенты с летальным исходом) и контрольную (выжившие пациенты). Полученная модель классификации закладывает потенциал для разработки методов идентификации клинического состояния пациента, что позволит автоматизировать получение сигнала о его ухудшении. Новизна результата заключается в подтверждении гипотезы о наличии зависимости между результатами измерения ЭКГ и последующим летальным исходом пациента в результате развития СПОН, ОСН или ОГМ, а также предложенном критерии и модели классификации, которые позволяют решать актуальную задачу автоматической фиксации ухудшения состояния пациентов.

Ключевые слова: классификация пациентов на основе ЭКГ, идентификация ухудшения клинического состояния, прогнозирование по медицинским данным, анализ ЭКГ, машинное обучение, искусственный интеллект, наука о данных, логистическая регрессия, стратификация риска смертности, интервал RR, интервал QT.

Абрамов Максим Викторович — канд. техн. наук, руководитель лаборатории, старший научный сотрудник, лаборатория теоретических и междисциплинарных проблем информатики, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук"; доцент, кафедра компьютерных наук, математико-механический факультет, СПбГУ. Область научных интересов: информационная безопасность, социоинженерные атаки, анализ защищенности пользователей информационных систем от социоинженерных атак злоумышленников, анализ распространения информации в социальных сетях на основе моделей, используемых при анализе защищенности пользователей информационных систем от атак социальной инженерии, анализ и моделирование социальных сетей, клиент-серверные технологии, исследование взаимосвязи между контентом, размещаемым пользователями в социальных сетях, и поведением в оффлайне; бизнес-аналитика, социальные вычисления, бизнес-аналитика. Число научных публикаций — 100. mva@dscs.pro; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3337.

Цуканова Екатерина Игоревна — главный специалист, ФГБУ Всероссийский центр экстренной и радиационной медицины имени А.М. Никифорова МЧС России. Область научных интересов: статистические методы исследования в медицине, интегративный подход в функциональной диагностике, системы поддержки принятия решений для раннего выявления, профилактики, выбора и оценки эффективности терапии жизнеугрожающих аритмий, рациональная организация диагностической службы в современном стационаре. Число научных публикаций — 8. Eka_77@bk.ru; Академика Лебедева, 4/2, 194044, Санкт-Петербург, Россия; р.т.: +7(812)607-5900.

Тулупьев Александр Львович — д-р физ.-мат. наук, профессор, главный научный сотрудник, лаборатория теоретических и междисциплинарных проблем информатики, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук"; профессор, кафедра компьютерных наук, математико-механический факультет, СПбГУ. Область научных интересов: представление и обработка неопределенных знаний и данных, применение математики и информатики в социокультурных и эпидемиологических исследованиях, программные технологии и разработка информационных систем с базами данных. Число научных публикаций — 400. alt@dscs.pro; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3337.

Корепанова Анастасия Андреевна — младший научный сотрудник, лаборатория теоретических и междисциплинарных проблем информатики, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук". Область научных интересов: информационная безопасность, социоинженерные атаки, анализ защищенности пользователей информационных систем от социоинженерных атак злоумышленников, анализ и моделирование социальных сетей, машинное обучение, анализ данных. Число научных публикаций — 13. aak@dscs.pro; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3337.

Алексанин Сергей Сергеевич — д-р мед. наук, профессор, член-корреспондент, директор, ФГБУ Всероссийский центр экстренной и радиационной медицины имени А.М. Никифорова МЧС России. Область научных интересов: медицина катастроф, проблемы функционального состояния человека в экстремальных и неблагоприятных условиях окружающей среды, профессиональная патология, радиационная медицина, профилактическая и восстановительная медицина. Число научных публикаций — 356. medicine@nrcerm.ru; Академика Лебедева, 4/2, 194044, Санкт-Петербург, Россия; р.т.: +7(812)607-5900.

Поддержка исследований. Работа выполнена в рамках проекта по государственному заказу Санкт-Петербургского Федерального исследовательского центра РАН (СПб ФИЦ РАН) - СПИИРАН № FFZF-2022-0003.

Литература

1. [Preventive cardiology. Under. ed. G.M. Kositsky.]. Meditsina [Medicine]; 1987. 512 p. (InRuss.).
2. Mjasnikov A.L. Jeksperimental'nye nekrozy miokarda [Experimental myocardial necrosis]. M. Medicina. 1963. 204 p. (InRuss.).
3. Brodsky M., Wu D., Penes P., Kanakis Ch., Rosen K. Arrythmias documented by 24 hour continius electrocardiographic monitoring in 50 male medical students without apparent heart diseases. The American journal of cardiology. 1977. vol. 39. No. 3. pp. 390–395. doi: 10.1097/00132586-197710000-00002
4. Umetini K., Singer D., McCarty R., Atkinson M. 24 Hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. Journal of the American College of Cardiology. 1998. vol. 31. No. 3. pp. 593–601. doi: 10.1016/s0735-1097(97)00554-8
5. Makarov L.M. Holterovskoe monitorirovanie. [Holter monitoring.] M: ID ”Medpraktika – M”. 2008. 504 p. (InRuss.).
6. Kleiger R.E., Miller J.P., Bigger Jr J.T., Moss A.J. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. The American journal of cardiology. 1987. vol. 59. №. 4. pp. 256–262. doi: 10.1016/0002-9149(87)90795-8
7. Anischenko V.S. Znakomstvo s nelinejnoj dinamikoj [Acquaintance with nonlinear dynamics.]. Izdatelstvo LKI. 2018. 224 p. (InRuss.).
8. Rathakrishnan K., Min S.N., Park S.J. Evaluation of ECG Features for the Classification of Post-Stroke Survivors with a Diagnostic Approach. Applied Sciences. 2021. vol. 11. No. 1: 192.
9. Wang X., Gong G., Li N., Ding L., Ma Y. Decoding pilot behavior consciousness of EEG, ECG, eye movements via an SVM machine learning model. International Journal of Modeling, Simulation, and Scientific Computing. 2020. vol. 11. №. 04: 2050028.
10. Antosik-Wojcinska A.Z., Dominiak M., Chojnacka M. Kaczmarek-Majer K., Opara K.R., Radziszewska. W., Olwert A., Swiecicki L. Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling. International Journal of Medical Informatics. 2020. vol. 138: 104131.
11. Srivastava A.K., Kumar Y., Singh P.K. A Rule-Based Monitoring System for Accurate Prediction of Diabetes: Monitoring System for Diabetes. International Journal of E-Health and Medical Communications (IJEHMC). 2020. vol. 11. №. 3. pp. 32–53.
12. Aziz, S., Ahmed, S. & Alouini, M.S. ECG-based machine-learning algorithms for heartbeat classification. Sci Rep 11. 2021. vol. 18738. https://doi.org/10.1038/s41598-021-97118-5.
13. Tougui I., Jilbab A., El Mhamdi J. Heart disease classification using data mining tools and machine learning techniques. Health and Technology. 2020. vol. 10. pp. 1137–1144.
14. Park J., Kim J.W., Ryu B., Heo E., Jung S.Y., Yoo S. Patient-Level Prediction of Cardio-Cerebrovascular Events in Hypertension Using Nationwide Claims Data. Journal of medical Internet research. 2019. vol. 21. №. 2: e11757.
15. Guo A., Zhang K.W., Reynolds K., Foraker R.E. Coronary heart disease and mortality following a breast cancer diagnosis. BMC medical informatics and decision making. 2020. vol. 20: 88.
16. Kovalev M.S., Utkin L.V., Kasimov E.M. SurvLIME: A method for explaining machine learning survival models. Knowledge-Based Systems. 2020. vol. 203. pp. 106–164.

17. Sohal H., Jain S. Interpretation of cardio vascular diseases using electrocardiogram: A study. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, 2018. pp. 159–164.
18. Yao L.K., Liu C.C., Li P., Wang J.K., Liu Y.Y., Li W., Wang X.P., Li H., Zhang H. Enhanced Automated Diagnosis of Coronary Artery Disease Using Features Extracted From QT Interval Time Series and ST–T Waveform. IEEE Access. 2020. vol. 8. pp. 129510–129524.
19. George K., Sankaran P., Joseph K.P. Computer assisted recognition of breast cancer in biopsy images via fusion of nucleus-guided deep convolutional features. Computer Methods and Programs in Biomedicine. 2020. vol. 194: 105531.
20. Hyun S., Kaewprag P., Cooper C., Hixon B., Moffatt-Bruce S. Exploration of Critical Care Data by Using Unsupervised Machine Learning. Computer Methods and Programs in Biomedicine. 2020. vol. 194: 105507.
21. Hang W.L., Feng W., Liang S., Wang Q., Liu X.J., Choi K.S. Deep Stacked Support Matrix Machine Based Representation Learning for Motor Imagery EEG Classification. Computer Methods and Programs in Biomedicine. 2020. vol. 193: 105466.
22. Louridi N., Amar M., El Ouahidi B. Identification of Cardiovascular Diseases Using Machine Learning. Proceedings of the 7th Mediterranean Congress of Telecommunications 2019, CMT 2019, Fez, Morocco, 24–25 October 2019. 2019. pp. 1–6
23. Villamor E., Monserrat C., Del Rio L., Romero-Martin J.A., Ruperez M.J. Prediction Of Osteoporotic Hip Fracture In Postmenopausal Women Through Patient-Specific FE Analyses And Machine Learning. Computer Methods and Programs in Biomedicine. 2020. vol. 193: 105484.
24. Perumal R., Kaladevi A.C. Early Prediction of Coronary Heart Disease from Cleveland Dataset using Machine Learning Techniques. Int. J. Adv. Sci. Technol. 2020. vol. 29. pp. 4225–4234.
25. Lepeshkin E., Surawicz B. The measurement of the QT interval of the electrocardiogram. Circulation, 1952;6:378-388 https://doi.org/10.1161/01.cir.6.3.378
26. Rymarczyk T., Kozłowski E., Kłosowski G., Niderla K. Logistic Regression for Machine Learning in Process Tomography. Sensors. 2019. vol. 19. No. 15: 3400. doi: 10.3390/s19153400
27. Bishop C. Pattern Recognition and Machine Learning. Springer-Verlag New York, 2006. 738 p.
28. Delen D., Topuz K., Eryarsoy E. Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. Operations research and management sciences. 2020. vol. 281. pp. 575–587.
29. Kharitonov N.A., Maximov A.G., Tulupyev A.L. Algebraic Bayesian networks: the use of parallel computing while maintaining various degrees of consistency. International Conference on Information Technologies. Springer, Cham. 2019. pp. 446–456.
30. Nielsen T.D., Jensen F.V. Bayesian networks and decision graphs. Springer Science Business Media. 2009. 464 p.
31. Tulupyev A.L., Stolyarov D.M., Mentyukov M.V. Representation of the local and global structure of an algebraic Bayesian network in Java applications. Proceedings of SPIIRAS. 2017. vol. 5. No. 0. pp. 71–99.
32. Kharitonov N., Malchevskaia E., Zolotin A., Abramov M. External consistency maintenance algorithm for chain and stellate structures of algebraic bayesian networks: statistical experiments for running time analysis. International Conference on Intelligent Information Technologies for Industry. Springer, Cham. 2018. pp. 23–30.

33. Baevskij R.M., Ivanov G.G., Gavrilushkin A.P., Dovgalevskij P.Ja., Kukushkin Ju.A., Mironova T.F., Priluckij D.A., Semenov A.V., Fedorov V.F., Flejshman A.N., Medvedev M.M., Chirejkin L.V. [Analysis of Heart Rate Variability When Using Different Electrocardiographic Systems (Part 1)]. Vestnik aritmologii [Bulletin of Arithmology]. 2002. No. 24. pp 65–86.
34. Ponomarenko V.I., Karavaev A.S., Borovkova E.I., Hramkov A.N., Kiselev, A.R., Prokhorov M.D., Penzel T. Decrease of coherence between the respiration and parasympathetic control of the heart rate with aging. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2021. vol. 31. no. 7: 073105.
35. Karavaev A.S., Ishbulatov Yu Y.M., Prokhorov M.D., Ponomarenko V.I., Kiselev A.R., Runnova A.E., Hramkov A.N., Semyachkina-Glushkovskaya O.V., Kurths J., Penzel T. Simulating dynamics of circulation in the awake state and different stages of sleep using non-autonomous mathematical model with time delay. Frontiers in Physiology. 2020. vol. 11: 1656.
36. Karavaev A.S., Borovkova E.I., Runnova A.E., Kiselev A.R., Zhuravlev M.O., Ponomarenko V.I., Prokhorov M.D., Koronovskii A.A., Hramov A.E. Experimental Observation of the Self-Oscillatory Dynamics of the Regulation Contours of the Cardiovascular System. Radiophysics and Quantum Electronics. 2019. vol. 61. no. 8. pp. 681-688.