DOI: https://doi.org/10.15407/kvt203.01.077

УДК 004.75+004.932.2:616

KRYVOVA O.A., Researcher,
Medical Information Systems Department
ORCID: 0000-0002-4407-5990
e-mail: ol.kryvova@gmail.com
KOZAK L.M., DSc (Biology), Senior Researcher,
Leading Researcher of the Medical Information Systems Department
ORCID: 0000-0002-7412-3041
e-mail: lmkozak52@gmail.com
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. Glushkov av., Kyiv, 03187, Ukraine

INFORMATION TECHNOLOGY FOR CLASSIFICATION OF DONOSOLOGICAL AND PATHOLOGICAL STATES USING THE ENSEMBLE OF DATA MINING METHODS

Introduction. *The digital technologies implementation provides registration of large amounts of bio-medical data (ECG, EEG, electronic medical records) as a basis for assessing and predicting the patients` condition. Data Mining methods allow to identify the most informative indicators and typological groups, to classify the person` functional state and the patients` disease stages to predict their changes.*

The purpose *of the paper is to develop information technology for the classification of human health states using an set of Data Mining methods and to carry out its validation on examples of a operators` functional state and patient's disease severity.*

Results. *The developed IT unites several stages: I — data pre-processing; II — clustering, selecting the homogeneous groups (data segmentation); III — predictors` identification; IV — classifying the studied states, development of predictive models using machine learning algorithms (Decision trees, Support vector machines, neural networks) and the method cross-validation. The proposed IT was used to classify the operators` functional state and the patients` severity in case of disease progression.*

Conclusions. *The IT use to assess the operators` activity successes made it possible to identify the most informative HRV indicators, changes in which can predict the operators` reliability, taking into account the type of vegetative regulation. Assessing the disease activity of children with dysplasia with IT use made it possible to identify diagnostic markers of CCC and develop diagnostic rules for determining the stages of the disease by ECG parameters (T wave symmetry, an integral indicator of the ST_T segment shape).*

Keywords: *information technology, Data Mining, machine learning models, severity of the patient.*

INTRODUCTION

At the current stage of digital medicine development, accompanied by the use of multifunctional monitoring systems, individual mobile health monitoring devices, there is a problem of interpretation of untreated primary arrays of heterogeneous medical data. One approach to solving it is to develop and apply information technology using Data Mining methods. The basic definition of Data Mining is the process of identifying patterns in data arrays (previously unknown) and using them to predict health states and make decisions [1].

In recent years, more and more researches have been done to improve patients' health. Multilevel schemes are developed, which use different types and methods of adaptive learning and combine various sources of clinical information (EHR, laboratory data, monitors, medical images). In recent decades, researchers have noted that the direction of Data Mining application and machine learning methods, namely the patients’ classification into risk groups to predict treatment outcomes, mortality, disease stages etc., was formed [2–10]. Analysis of the literature data for 2008–2019 leads to the conclusion that in terms of accuracy and clarity of the results the intellectual analysis methods, which integrate hybrid methods and previous models of clinical risk stratification, should be prefered [11].

PROBLEM STATEMENT

In the early 2000s, examples of successful use of Data Mining for biomedical data analysis appeared [2]. The research was mainly aimed at improving the diagnostic accuracy of the diseases identification by medical databases [3] and at developing the decision support systems [4] and diferent studies by medical and biological information [8].

Different types of machine learning methods are used to develop classification diagnostic models: logistic regression, decision tree methods, random forest (RF), support vector machine (SVM) or ensembles of classification models, genetic algorithms, artificial neural or deep learning networks [4–7].

Among the growing number of works on the application of Data Mining and machine learning technologies, the trend of the clinical direction of predictive models, which use new multi-sensory, multi-resource and multiprocessor information merging schemes, stands out. The architecture of such systems consists of hybrid multilevel schemes, combines uncontrolled and controlled teaching methods and methods of features selection. This approach makes it possible to identify clinically significant patterns using data of monitoring, clinical measures, tools and treatment outcomes [9–11].

For almost 30 years, more than twenty classical tools (systems) for assessing and forecasting the patients’ condition have been developed and updated [12–16]. Among them are severity scores, which quantitatively or qualitatively determine the severity of the patient's condition and classify him into specific risk group based on the analysis of deviations of anatomical, physiological, biochemical parameters. Determining the severity of the condition the decision to hospitalize the patient in the intensive care unit can be made. For example, in intensive care units in the United States and the EC use scoring systems to assess the patients’ condition. These are several scales: Simplified Acute Physiology Score (SAPS II) [14], Acute

Physiology and Chronic Health Evaluations (APACHE II and III) [15], Mortality Probability Models (MPM II-24) [16]. In addition to standardized scales designed for the general population, a number of specialized scales have been developed to assess the activity (stages, severity) of individual diseases.

A number of studies on the stratification of patients into risk groups according to clinical data and treatment outcomes have demonstrated the superiority of models developed by Data Mining methods over classical scales [17–19].

Much attention is paid to the choice of informative characteristics for the analysis of prenosological and pathological human conditions. Many researchers have determined that the cardiovascular system (CVC) is one of the main indicators of adaptive capacity and responses of the whole organism [20, 21].

One of the most common methods of studying the mechanisms of regulation of the cardiovascular system is the analysis of heart rate variability, which has become a reliable and powerful tool for research in cardiology, assessment of human functional status (FS) [21]. It is proved that the adaptive reactions of the heart to constantly changing physiological conditions are reflected in changes in heart rate variability, which provides information about the systemic reactions of the body during deteriorating health and under the influence of external stress. It is the CVS functioning level that can determine the boundary between the prenosological state (health) and the disease, as well as affect the disease severity.

Methods of HRV analysis are being actively developed [21–29], the technology of analysis is being improved, mathematical approaches to the analysis of nonlinear dynamics of heart rhythm are involved, which has expanded the list of informative indicators for assessing the human condition. Currently, studies of this condition (norm and pathology) are carried out using estimates of irregularity and chaotic rhythm, such as fractal dimension, entropy parameters [27]. The following approaches are proposed for use: wavelet transform [28], the method of multispectral analysis of CVC [29], analysis in the phase plane [23, 24]. Thus, one of the common and effective approaches to detecting changes in human health is to assess the relationship of this condition with the CVS state, which allows to determine functional changes in physiological systems, identify the boundary between prenosological state (health) and disease, as well as affect the disease severity.

The purpouse of the paper is to develop information technology for the classification of human health using sets of Data Mining methods by objective and expert characteristics.

DEVELOPMENT OF INFORMATION TECHNOLOGY FOR CLASSIFICATION OF FUNCTIONAL CONDITION AND HEALTH STATE

Large amounts of information, the need for it adequate analysis with the possibility of further forecasting and planning of appropriate activities necessitate the development and application of new technologies for assessing the current state of both individual health and population health of Ukraine on objective and expert indicators.

Note the effectiveness of the use of Data Mining methods to determine the risk groups according to clinical data, to assign patients to the appropriate group by health markers with further prediction of its changes and evaluation of the effectiveness of treatment. We have developed a method for detecting markers of the cardiovascular system state, which is based on Data Mining models, which are based on the analysis of heart rate variability (HRV) [30]. The development of the method takes into account the experience of using the constructed Data Mining models to determine population health clusters that are homogeneous in terms of medical and demographic indicators [31].

We have formed an ensemble of Data Mining methods, developed a research scheme that uses a combination of filtering methods, cluster analysis algorithms (*k*-means, EM) and classification (Decision Trees, Neural networks, SVM) using informative ECG features. The application of these methods makes it possible to combine the possibilities of solving specific tasks at the stages of analysis: reducing the sample size, selecting criteria / markers of the appropriate health level and classification of a particular subject / patient to the appropriate group according to his health.

Let consider in more detail an ensemble of used Data Mining methods.

Selection of informative parameters (filtering). The choice of variables follows from two tasks: 1) to find informative variables strongly connected with the target feature, 2) to define a small parameters subset, keeping enough information on initial indicators.

A peculiarity of the initial data in our study was a large number of indicators — ECG parameters. The multilevel system of indicators, calculated by automated ECG analysis, had a total of 240 features. Such a large amount of primary data is characteristic of many tasks in various fields of medical research. The correlation matrix was calculated among the predictors to avoid the problem of multicollinearity. The correlation coefficient (R > 0,7) is used as criteria for deciding whether variable may be excluded from the analysis because another input variable contains the same information.

As you know, there are several reasons for the negative impact of a large number of non-informative parameters on the learning algorithm quality, three of which are considered basic [32]. One important reason is that as the parameters number increases, more learning objects are needed for reliable classification. In addition, with increasing parameters number decreases the statistical reliability of the algorithm on the control data. The advantage of selecting informative features is the increase in the accuracy of the classification algorithm, generalization ability, achieving the possibility for the best interpretation of data.

Usually a preliminary selection of parameters is carried out before the start of machine learning algorithms. Statistical criteria for correlation of each of the primary features with the target feature and ordering (for example, by the size and significance of Chi-Square Pearson, F — Fisher) are used. Further selection of a set (combinations) of informative parameters is performed using classification algorithms for greater accuracy [33].

Clustering. One of the effective methods of data processing is their segmentation using cluster analysis methods (unsupervised learning). The clustering process divides the data set into cluster groups or subclasses [3]. Clustering (subgroups) allows you to use all available information to build multiple models, and then make more accurate predictions for the model.

We used two most popular algorithms, namely *k*-means, EM, which are implemented in the module Data Miner STATISTICA 10 [35, 36].

K–means method. The patients were divided into groups using the generalized *k*–means method. This method makes it possible to distribute observations (from space *Xn*) into *k* clusters according to the following criteria.

The first criterion for recalculating cluster centers is the minimization of the objective function (F_1) by the sum of the squares of the distances between each object x_i and the center of the cluster μ_n, to which it belonged at each iteration:

$$F_1 = \sum_{n=1}^{k} \sum_{x \in X_n} \| x_i - \mu_i \|^2 \to \min, \qquad (1)$$

where x_i is the set of n observations, n is the number of objects to be divided into k groups (clusters), μ_n — cluster centers.

And the second criterion (F2) determines that the distances sum between the clusters should be as large as possible:

$$F_2 = \sum_{n,i=1}^{k} \| \mu_i - \mu_n \|^2 \to \max. \qquad (2)$$

Additionally, in contrast to the classical method *k*–means, a cross-check is performed on *N* random samples, which allows to minimize the error and to select the optimal number of clusters. If the error function (average distance between cluster centers) for a solution *k*+1 clusters is not 5% better than the solution for k clusters, then the solution with *k* clusters will be final (optimal).

Fuzzy clustering algorithm (EM). The expectation-maximization (EM) algorithm assumes that the data correspond to a linear combination of distributions (normal, lognormal, binomial):

$$P(x) = \sum_{i=1}^{k} w_i \cdot p_i(x), \sum_{i=1}^{k} w_i = 1, w_i \geq 0, \qquad (3)$$

where k is a number of components in a mixture of distributions $P(x)$, w_i — is weights of components, $p_i(x)$ — distribution density of components.

At each step of the iterative process, the expectation parameters are estimated and the likelihood function is calculated until the maximum of logarithmic likelihood is reached. The *k*-fold cross-validation use with the error function evaluation (loglikelihood) helps to determine the final number of clusters.

One of the accepted methods of estimating the required number of clusters is the Cluster Validity Indices method [37].

Classification and Regression Trees (CART). Decision trees have become the most common approach to solving the problem of assessing the patient's condition [17], to detect ischemia of the heart [38], to classify the stages of heart disease [39], as well as to identify changes in human functional states [30].

The advantage of the decision tree method is that there are no requirements for data distribution, their type. This approach facilitates the interpretation of the results, the model is displayed as a tree, the structure of which is determined by logical rules (IF — THAT). Its purpose is to predict the target variable based on other features known as predictors, which makes it possible to detect complex interactions.

We used the CART algorithm, a recursive method that allows us to develop classification and regression models. According to the CART algorithm, the data set is distributed across all variables sequentially into segments. The purpose of sequential segmentation is to obtain uniformity of data on the selected attribute, reducing uncertainty in the partition node.

In the CART algorithm for predictor selection and division into two nodes, the index as a measure of uncertainty (*Gini*) is used:

$$Gini(d) = 1 - \sum_{i \neq j}^{k} (P_i^d)(P_j^d), \tag{4}$$

where P_i is the probability of classification in node d as i or j.

In each node, the reduction of the impurity is maximized.

To summarize the result, the optimal size tree is selected by cutting branches in combination with the method of estimating the error of cross-checking (algorithms minimal cost-complexity tree pruning, V-fold cross-validation).

The method of **Support Vector Machine** is based on vector space model, which aims to find such a surface distribution between classes, which is the most remote from all points of the learning set any of the classes. If we denote the learning data set $D = \{X_i, y_i\}$, where X is the vector of the i-point and y_i is the corresponding class label, then the linear classifier has the form:

$$f(x) = sign(W^T X_i + b), \tag{5}$$

where W^T is weight vector and b is constant.

The optimization problem is solved, namely, the task of achieving the maximum gap between the reference points:

$$\frac{1}{2} W^T W \to \min. \tag{6}$$

For all (X_i, y_i) € D is satisfied when

$$y_i(W^T X_i + b) \geq 1. \tag{7}$$

This method is implemented in STATISTICA Data Mining module, there is a possibility of transition to a nonlinear model using other core functions.

ANN Neural network is a mathematical apparatus that simulates the work of a network of brain neurons. The components of the neural network consist of inputs x_i, which are fed to the neurons synapses that are connected by axons in several hidden layers and the final outputs y_i. The neuron state is described by a function

$$S = \sum_{i}^{n} x_i w_i, \tag{8}$$

where n is a number of inputs, w_i — weights i — synapse. The output value of the axon is

$$Y = f(S), \tag{9}$$

where $f(S)$ is the activation function.

When learning the network, the task is to minimize the objective error function by the method of least squares:

$$E(w)=\frac{1}{2}\sum_{j=1}^{k}(y_j-d_j)^2, \tag{10}$$

where y_j is the value of the j^{th} output of the neural network, d_j — target value of the j^{th} output, k — is a number of neurons in the output layer.

Classification quality indicators. To evaluate the performance of the proposed model the sensitivity, specificity, accuracy, and F-score are calculated.

The sensitivity is the proportion of positive instances that are correctly classified as positive. The specificity is the proportion of negative instances that are correctly classified as negative. The accuracy is the proportion of instances that are correctly classified.

$$Sensitivity\,(Recall)=\frac{TP}{TP+FN}, \tag{11}$$

$$Precision=\frac{TP}{TP+FP}, \tag{12}$$

$$Specificity=\frac{TN}{FP+TN}, \tag{13}$$

$$predicive\ Accuracy=\frac{TP+TN}{TP+TN+FP+FN}, \tag{14}$$

$$F\,score=\frac{2\times Recall\times Precision}{Recall+Precision}, \tag{15}$$

where — TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives respectively.

For multiclass case, these measures can be obtained from the confusion matrix by comparing numbers of instances for each class in the matrix against instances of all the other classes. F-score, since it combines precision and recall into a single number evaluating the whole system performance [40].

To solve the tasks for a particular subject of analysis, the formation of an appropriate ensemble of the considered set of methods was done. The initial data were indicators of heart rate variability, objective indicators of the studied physiological systems and expert assessments of the human health state.

The proposed information technology for the classification of functional states and human health consists of four main stages (Fig. 1).

Stage 1. Data pre-processing. At stage 1, the input data is pre-processed, checked for completeness, the presence of emissions, type compliance, reformatting. The target feature is determined by the specific task of the analysis: information about the response of body systems to external influences, expert data on the severity of the condition (disease activity) of patients and so on. At this stage, the number of primary HRV indicators was reduced and the most informative ones were selected regarding the target feature using filtration methods.

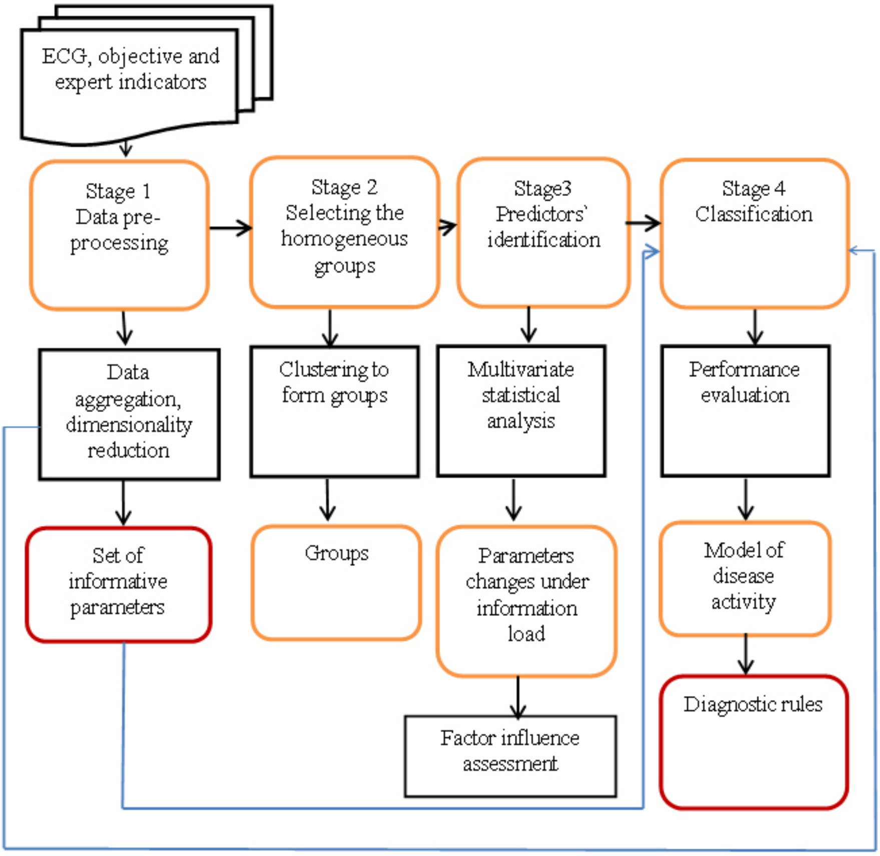

ECG, objective and expert indicators
Stage 1 Data pre-processing
Stage 2 Selecting the homogeneous groups
Stage3 Predictors` identification
Stage 4 Classification
Data aggregation, dimensionality reduction
Clustering to form groups
Multivariate statistical analysis
Performance evaluation
Set of informative parameters
Groups
Parameters changes under information load
Model of disease activity
Factor influence assessment
Diagnostic rules

Fig. 1. Stages of information technology for the classification of functional states and human health

At this stage, the primary selection methods of informative parameters are used with statistical correlation criteria (Pearson's Chi-Square, *F*), that results in a reduction in the study volume to further determine the classification groups for the gradation of studied state changes.

In *stage 2 — clustering to form groups*, the number and composition of typological groups were determined by sets of informative indicators. At this stage, cluster analysis methods (*k* –means, EM) are used.

Stage 3. Predictors` identification. The transition to step 3 is carried out if there is a need for analysis of the repeated measurements. That is, when the purpose of the study is to identify changes in the informative indicators associated with changes in the factor (e.g., response to exercise). Methods of repeated analysis of variance (RepANOVA) are used, which allows to determine differences in informative indicators changes in certain subgroups, as well as to provide a statistical assessment of the factor influence. The result obtained at this stage will be a set of informative features that are statistically significantly related to the factor.

Stage 4. Classification of the human condition severity. In this stage, informative indicators set were tested, which are predictors of the CVS state as attributes of the state classification model. This step is performed if the initial data contains the target attribute (class label) provided by the experts.

Algorithms CART, ANN, and SVM were used. Comparisons of classification features sets selected by different models and general classification accuracy of different models were performed. The efficiency indicators of the models for each class were calculated (sensitivity, specificity, accuracy).

For samples of small number, cross-validation (10 -fold) was used to optimize the complexity of the model. The model was chosen according to two criteria: high enough accuracy and optimal complexity.

Calculated according to the CART algorithm, the decision tree with the optimal size allows to formulate classification rules for health of each severity (logical conditions for the values of a small set of ECG parameters).

If the classification quality is unsatisfactory, it is possible to return to the previous stages 1, 2, 3 using other selection methods of features subsets (or model parameters). In the presence of a test sample, the quality of classification models is checked on it.

The end result is the classification rules according to the informative set of ECG parameters, which determine the patient's condition severity.

During the development, the procedures of the Data Miner module of the STATISTICA 10 package were used. Note that the Data Mining algorithms are implanted in the Weka, RapidMiner, SAS Enterprise Miner software and in the modern design tool Python, R.

STUDY OF FUNCTIONAL STATE AND HUMAN HEALTH WITH THE USE OF DEVELOPED INFORMATION TECHNOLOGY

The proposed IT is used to solve problems aimed at studying the operators\` functional state (prenozological state) and to classify the patients\` severity in case of disease progression.

Determination of specific changes in operators\` HRV indicators (prenosological state). Verification of the developed IT was carried out according to the experimental study of the reliability of operator activity under information load, which was performed by employees of the Research Institute of Military Medicine of the Armed Forces of Ukraine [41].

The condition of CVS regulatory mechanisms was studied by ECG recording (for 2 min) using Cardio Sens AIC (KHAI Medica, Kharkiv). The analysis was performed on the main HRV indicators, which belong to the generally accepted informative characteristic set of human functional state (statistical characteristics, spectral analysis, spectral components in the ranges ULF, VLF, LF, HF).

The professionally important qualities of military operators and their reliability of activities were assessed by tests consisted of information-intensive tasks: the dynamic memorization quality test (DMQ); the test of determining the speed and accuracy of the reaction to a moving object (RMO); the attention concentration and short-term memory test (ACSM). Factors influencing the operators\` FS were determined according to the training process stages: 1 — rest state; 2 — QDM; 3 — RMO; 4 — ACSM; 5 — recovery state [42].

At the preparatory stage of each test, the individual optimal load level (τ_{lim}), which the operator can still perform without errors, was determined. At the training stage, the tasks complexity increased by 10 % of the determined individual optimal level. The percentage of errors made was used as an indicator of the operator activity reliability at different levels of test task complexity. The technique of the training cycle is described in detail in the works [41, 42].

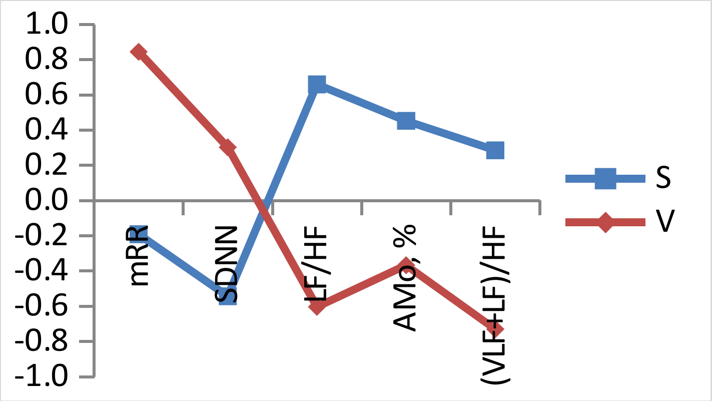


Fig. 2. Graf of means of indicators in two groups (S — sympathonics, V — vagotonics)

At the IT first stage the complex of the HRV main indicators on groups is defined:

— statistical characteristics: mode of RR-intervals (M_0RR), amplitude mode (AMo, %), standard deviation of RR-intervals (SDNN), stress index (SI);

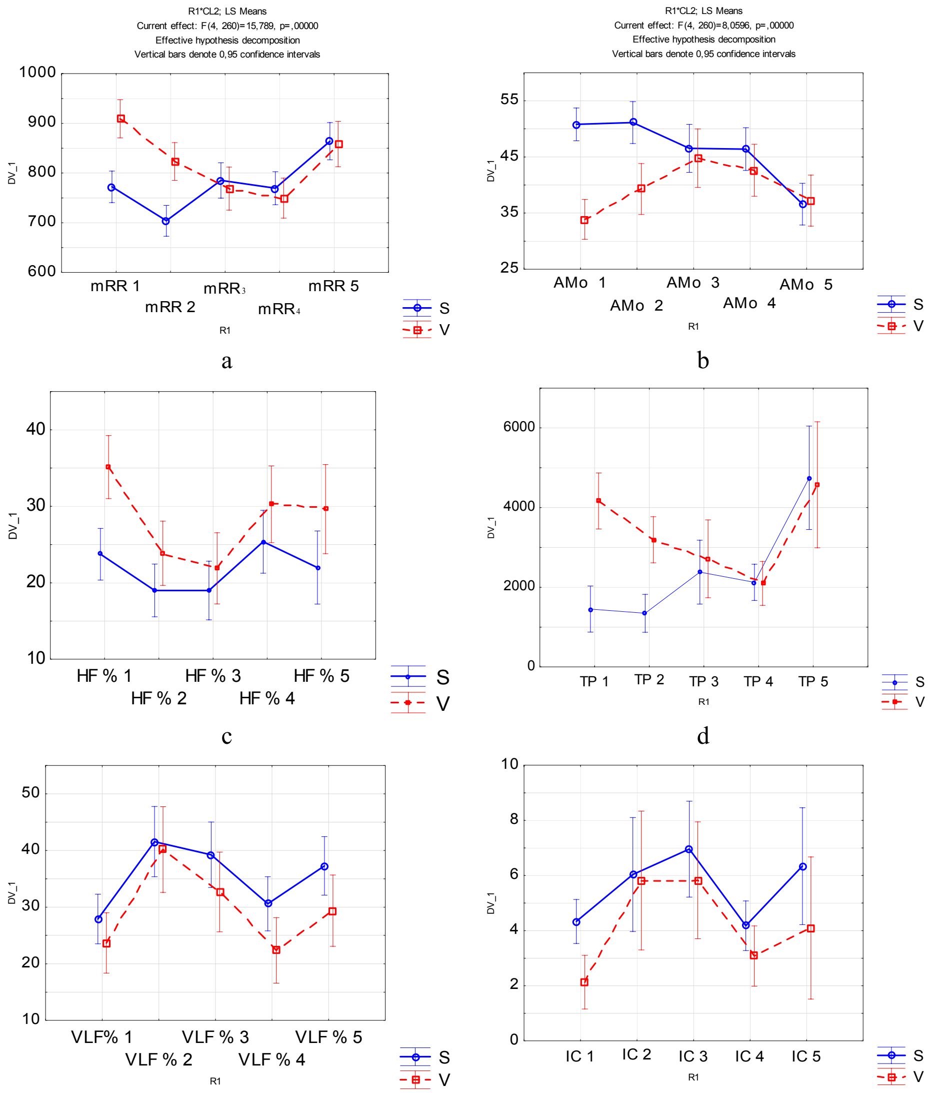
— spectral indicators: total spectral power of the TP spectrum (0.003–0.4 Hz), spectral components in the bands ULF (< 0.015 Hz), VLF (0.015–0.04 Hz), LF (0.04–0,15 Hz), HF (0.15–0.4 Hz), the activation indices of subcortical centers VLF/HF, index centralization IC = (VLF + LF)/HF.

At the second stage of IT with the help of cluster analysis the group at rest state (1) was determined by the vegetative regulation type: 1) predominance of sympathetic division (S, sympathonics), 2) predominance of parasympathetic division (V, vagotonics). In fig. 2 standardized average values of indicators for which clustering was performed are provided. The group of sympathonics (S) includes 42, vagotonics (V) 28 operators.

Heart rate parameters in groups with different types of vegatative regulation, which differed significantly at rest stage (1), undergo significant changes during the training cycle, and at the stage of recovery (5) there is no significant difference between the two typological groups (Fig. 3).

If at the initial stage (1) the spectrum was dominated by components of the activity of the autonomous control loop (HF, LF), then after performing tests in both groups there is a redistribution of power spectrum against the background of decreasing mode of RR-intervals. The power of the high-frequency component (HF) decreases, the low-frequency component of the spectrum (VLF) increases, and the LF (first-order slow-wave power) increases, which reflects the activity of the vasomotor center.

At the same time, it was determined that the test loads of dynamic memory (DMQ) and rapid response (RMO) cause greater changes in HRV than the activation of attention concentration and memory (ACSM). Characteristically, after performing all test loads (step 5), the components of the heart rate spectrum return to values at rest state (1), except for the spectrum total power (TP) due to an increase in the LF component in the sympathonics` group.



a

b

c

d

e

f

Fig. 3. Changes in spectral components at the stages of training (1, 2, 3, 4, 5) in sympathonics and vagotonics: a) mode; b) the RR mode amplitude; c) HF - high frequency component; d) TP — total power; e) VLF — power in the region of very low frequencies; f) (VLF + LF) / HF — centralization index

At the same time, it was determined that the test loads of dynamic memory (DMQ) and rapid response (RMO) cause greater changes in HRV than the activation of attention concentration and memory (ACSM). Characteristically, after performing all test loads (step 5), the components of the heart rate spectrum return to values at rest state (1), except for the spectrum total power (TP) due to an increase in the LF component in the sympathonics` group.

According to the literature, it is known that in the bases of mechanisms of formation of the low-frequency component (VLF) are stressors that activate the renin-angiotensin-aldosterone system and increase the catecholamines concentration in plasma. The VLF component power is associated with the activity of suprasegmental (hypothalamic) centers of vegetative regulation, which are transmitted through the sympathetic part of the VNS [21].

Thus, after performing NPS and PPO tests changes in heart rate regulation occur, namely: acceleration of heart rate (decrease in MoRR), increase in low-frequency (VLF) and high-frequency (HF) heart rate fluctuations, increase in the centralization index (influence of the central control loop), indicating the stressful nature of these loads and significant psycho-emotional stress of the operators.

Development of a classification model of disease activity in children with dysplasia. Connective tissue dysplasia (CTD) is a systemic disease that arises at an early age, has many manifestations in the cardiovascular system, musculoskeletal system and other organs. To predict the disease development, it is important to know the diagnostic criteria that characterize the stages of disease activity. The purpose of the study is to determine these criteria according to the system of ECG indicators.

The classification of the severity of the condition of children with CTD was developed according to the arrays of ECG indicators, as well as indicators of the severity of the condition of patients determined by expert physicians. The final indicator of CVS state is the final assessment (FA), which is formed from complex assessments of lower level: health rate regulation, myocardial status and additional features (quantitative and qualitative assessments of different coding systems, arrhythmias, risk of sudden cardiac events etc.). Complex indicators are calculated in points (0–100).

The study was based on data from laboratory and clinical examination of 25 children with CTD manifestations. Disease activity was measured by the Juvenile Arthritis Activity Scale (JADAS) [43]. 6-channel ECG recording was performed for 5–20 minutes using a Cardio Plus P device.

Cluster analysis methods (*k*-means with 10-fold cross-validation) allowed identifying two typological groups for comprehensive assessments of the CVS regulation, myocardium condition and its reserves:

- group 1 (16 children) had a low level of complex assessment: $FA_1 = 58.3 \pm 8.1$;
- group 2 (9 children) - significantly differed by higher complex assessment: $FA_2 = 68.3 \pm 5{,}2$. ($I = 68{,}3 \pm 5{,}2$).

The optimal set of CVS state predictors is determined. The set of predictors consists of the following primary ECG parameters: cardiac arrhythmia (Heart rhythm disorders), T-wave amplitude (lead II), integrated indicator of the STT form (lead II), QRS — alpha angle, T-wave symmetry ratio. The error of the regression model (by CART algorithm) for a set of 5 parameters ECG is 18.8 %, the correlation coefficient R = 0.88.

Classification models of disease activity stages were developed. CVS FA predictors were tested as attributes on CART, Neural network, SVM models. The target variable — disease activity was determined by 3 gradations provided by experts (1 — the initial stage of activity, 2, 3 — subsequent stages of inflammation increasing). The quality comparison of 3 models in the training sample gave such training errors: CART — 0 %, Neural network — 24 %, SVM — 44%. That is, for a small sample, the best result was for the C&RT decision tree model — 100% classification accuracy.

After 10-fold cross-validation of the CART models revealed 4 indicators, which determine the disease activity with an overall classification accuracy of 88 %. The most significant attributes of the disease activity model and their contribution to the CART model (by rank) is shown (Tab. 1).

The optimal classification tree, which was determined after 10-fold cross-validation is given (Fig.4). It should be noted that the distribution of the training sample into groups with different activity was unbalanced.

The quality measures of the classification of these disease stages are shown (Tab.2). The average F-score = 0,94.

The quality measures of the classification of these disease stages. Thus according to the optimal model, the CTD stages are classified with high accuracy.

In accordance with the tree splitting conditions, logical rules for the severity classification of the condition are formulated, in particular the basic classification rules are:

- Low disease activity level D_1:
  if Ind STT > 49,5 and Ind STT ≤ 84,0 and Amp T(II) ≤ -160,5 then $D_1 = 1$
  if Ind STT > 49,5 and Ind STT ≤ 84,0 and Amp T(II) > -160,5 and α-QRS ≤ 75,5 and SimmT(I) ≤ -0,57 then $D_1 = 1$
- Middle disease activity level $D_2 = 2$:
  if Ind STT > 49,5 and ≤ 84,0 and Amp T(II) > -160,5 and α-QRS > 75,5 then D2 = 2
  if Ind STT > 49,5 and Ind STT ≤ 84,0 and Amp T(II) > -160,5 and α-QRS ≤ 75,5 and SimmT(I) > 0,57 then $D_2 = 2$
- High disease activity level D3 =3:
  if Ind STT ≤ 49,5 then $D_3 = 3$
  if Ind STT > 49,5 and Ind STT > 84,0 then $D_3 = 3$.

Table 1. **Predictor importance for the classification of CTD activity stages**

The best predictors	Variable importance rank
T - wave symmet ratio (I)	100
Ampl. T (II)	86
Ind. form STT (II)	76
α QRS	68

Table 2. **Classification results for CTD activity detection**

Measures (%)	Disease activity		
	1	2	3
Sensitivity	100	81,8	100
Specificity	85,7	100	100
Predictive Accuracy	92	92	100

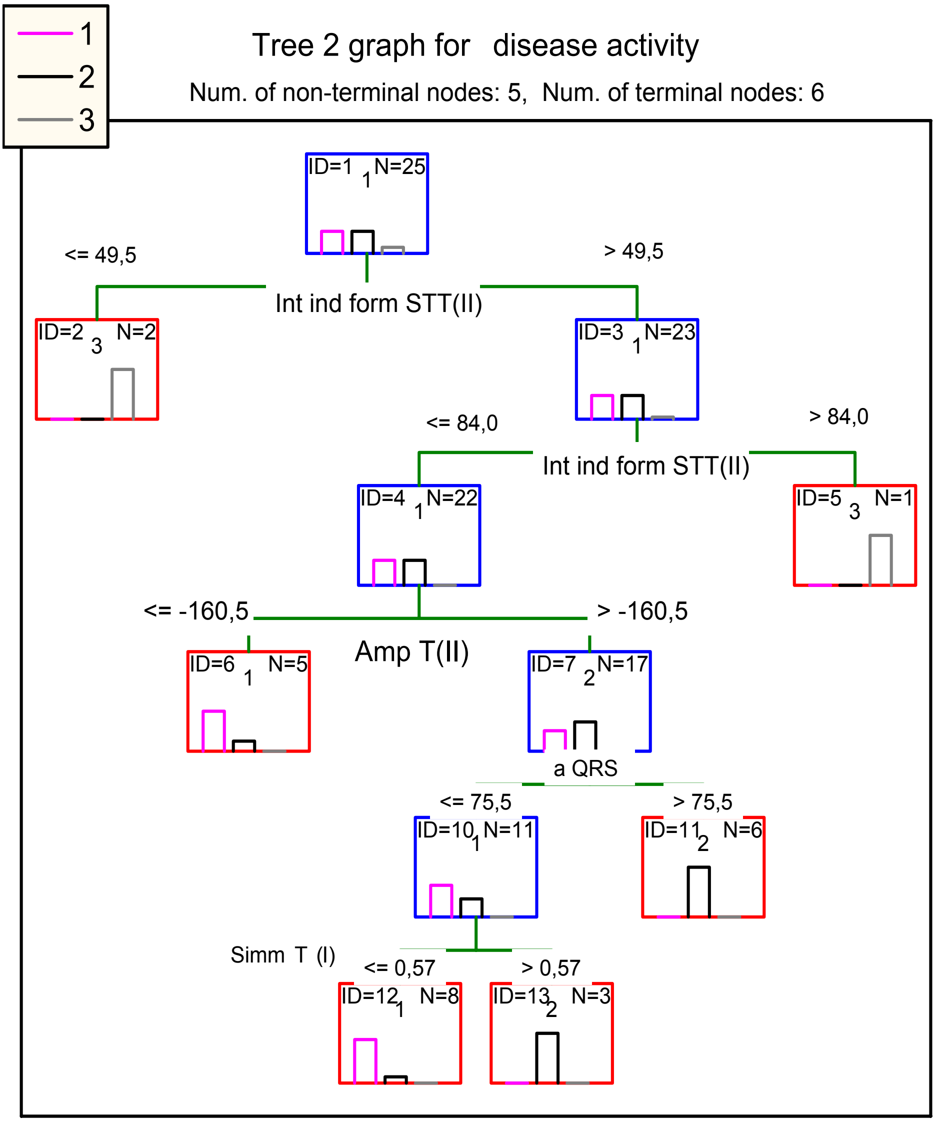


Fig. 4. Classification tree of CTD activity

Thus, with the help of the developed information technology the ECG indicators are determined, the changes of which can be markers of CVS disorders in the case of inflammatory processes in children diagnosed with juvenile arthritis, rheumatic disease. Markers of the initial stages of activity were determined by the following ECG parameters: α-QRS angle, chahge of the the T–wave (I) symmetry ratio.

Changes in the STT form (less than 49,5 or more than 84 points) indicates increased disease activity.

According to experts, the use of the proposed information technology to determine the CTD activity according the ECG parameters will allow the physician to identify the initial stages of the process in an outpatient setting. The advantage of this approach is the possibility of simultaneous assessment of CVS functional changes and the disease activity level before the clinical manifestations of the inflammatory process. The STT shape indicator gives an opportunity to select a group of children with apropriate changes in the STT segment. Such changes reflect dysmetabolic, hypoxic changes of the myocardium that accompany the manifestations of inflammatory processes [44]. At the same time, it is also necessary to take into account changes in the T wave amplitude, changes in the angle alpha angle QRS, the T wave symmetry index.

Prospects for solving the problem of physician` information support. Further development of a clinical desition support system in disease severity determining will be aimed at analyzing large arrays of clinical, laboratory and instrumental data in order to improve the classification accuracy for an extended range of tasks.

CONCLUSIONS

The created information technology, which combines the generalized stages: data pre-processing to reduce the studied data set, clustering (data segmentation, likelihood function biulding), predictors` identification by analysis of Data Mining models and classification of human condition with formation of final characteristics allows to determine pecularities of human functional state change under external factors influence and severity patients by analysis of heart rate variability and expert characteristics.

The combination of Data Mining methods used at different stages of IT allows solving consistently the necessary tasks: by filtering indicators, the relevant features are determined; the use of clustering provides the homogeneous groups detection; the decision tree method (CART algorithm) makes it possible to build a classification rules and high classification accuracy.

Using the developed IT, specific changes in HRV indicators in operators, which occur under the influence of various types of information loads, are determined taking into account the type of vegetative regulation. Loads of dynamic memorization and rapid response cause greater changes in HRV than activation of attention concentration and short-term memory. Thus, the following shifts in HRV regulation occur during the performance of these tasks: acceleration of heart rate (decrease in MoRR), increase of low-frequency (VLF) and high-frequency (HF) heart rate fluctuations, increase of centralization index (influence of central regulation loop), which indicates stress loads and significant psycho-emotional strain in operators. In the recovery state after all test loads, only in sympathonics, the spectrum total power (due to an increase in the LF component) does not return to the initial values.

The use of developed models and technologies to classify the patients` severity allowed to assess the CVS state of children with dysplasia, identify markers of stages of different disease activity and build diagnostic rules, the use of which make it possible to predict the disease severity and to adjust treatment tactics.

REFERENCES

1. Ian H. Data Mining Practical Machine Learning Tools and Techniques Witten, Eibe Frank and Mark A. Hall Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition. Morgan Kaufmann, 2011, 665 p.
2. Yoo I., Alafaireet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J. F. Data Mining in Healthcare and Biomedicine: A Survey of the Literature. *Journal of medical systems*. 2012, no 36(4), pp. 2431–2448.
3. Chen M., Hao Y. , Hwang K., Wang L., Wang L. Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. *IEEE Access 2017*;5:8869-8879.
4. Safdar S., Zafar S., Zafar N., Khan N.F. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. *Artificial Intelligence Review.* 2018, 50 (4), pp. 597–623.
5. Roopa C. K., Harish B. S. Survey on various Machine Learning Approaches for ECG Analysis. *International Journal of Computer Applications.* 2017, no 9, vol. 163, pp.25–33.
6. Mohan S., Thirumalai C., Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 2019. 7:81542–81554.
7. Goldstein B.A., Navar A.M., Pencina M.J., Ioannidis J.P. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2017, Jan; 24(1):198-208.
8. Antomonov M.Yu. Algorithmization of the choice of adequate mathematical methods in the analysis of medical and biological data. *Kibernetika i vyčislitel`naâ tehnika*. 2007, Iss. 153, pp. 12–23. (In Russian)
9. Georga E.I., Tachos N.S., Sakellarios A.I., Kigka V.I., Exarchos T.P., Pelosi G. Artificial intelligence and data mining methods for cardiovascular risk prediction Cardiovascular Computing. *Methodologies and Clinical Applications*. 2019, pp. 279–301
10. Amin M., Chiam Y. Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics.* 2019, Vol. 36, pp. 82–93.
11. Kaieski N., da Costa C.A., da Rosa Righi R., Lora P.S. Application of artificial intelligence methods in vital signs analysis of hospitalized patients: A systematic literature review. *Applied Soft Computing*. 2020, Vol. 96,
12. Owens W.D., Felts J.A., et al. A physical status classification: A study of consistency of ratings. *Anesthesiology*. 1978, Vol. 49, pp. 239–243.
13. Lemeshow S., Le Gall J.R: Modeling the severity of illness of ICU patients. *JAMA*. 1994, Vol 272, pp.1049–1055.
14. Le Gall J.R., Lemeshow S., Saulnier F: A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993, 270 (24), pp. 2957–2963.
15. Knaus W.A., Draper E.A., Wagner D.P., Zimmerman J.E: APACHE II: A severity of disease classification system. *Cri.t Care Med* .1985, 13:818-829.
16. Lemeshow S., Teres D., Klar J., Avrunin J.S., Gehlbach S.H., Rapoport J. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993, 270, pp. 2478-86
17. Trujillano J., Badia M, Serviá L. Stratification of the severity of critically ill patients with classification trees. *BMC medical research methodology*. 2009, V 9, no 7, pp. 83–95.
18. Kim S., Kim W., Park R.W. A Comparison of intensive care unit mortality prediction models through the use of Data Mining Techniques. *Health Inform Res* 2011,17, pp. 232–43.
19. Allyn J. et all. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS one* 2017, 12(1), pp. 1–12.
20. Amosov N.M. Thinking about health. Moskow: 1978, 178 p. (In Russian)
21. Baevsky R.M., Berseneva A.P. Introduction to prenosological diagnostics. Moskow: Slovo, 2008, 174 p. (In Russian)
22. HRV analysis software URL: http://www.nevrokard.eu/maini/hrv.html (last access 20.10.2020)
23. Fainzilberg L.S. Computer diagnostics based on the phase portrait of an electrocardiogram. Kyiv: Osvita Ukrainy. 2013, 191 p. (In Russian)

24. Gritsenko V.I., Fainzilberg L.S. Intelligent information technologies in digital medicine on the example of phasagraphy. Kyiv: Naukova Dumka. 2019, 423 p. (In Russian)
25. Fainzilberg L.S., Dykach Ju.R. Linguistic approach for estimation of electrocardiograms's subtle changes based on the Levenstein distance. *Cybernetics and Computer Engineering*. 2019, no. 2 (196), pp. 3–26.
26. Gritsenko V.I., Fainzilberg L.S. Current state and prospects for the development of digital medicine. *Cybernetics and Computer Engineering.* 2020, no. 1 (199), pp. 59–84.
27. Richman J.S. Randall M.J. Physiological time–series analysis using approximate entropy and sample entropy. *Am J. Physiol. Heart Circ. Physiol.* 2000, Vol. 278, № 6, pp. H22039–H2049.
28. İşler Y., Kuntalp M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine*. 2007, Vol. 37, no. 10, pp. 1502–1510.
29. Valupadasu R., Chunduri B. R., Chanagoni V. Identification of Cardiac Ischemia using bispectral analysis of ECG. *Biomedical Engineering and Sciences* (IECBES). 2012: IEEE EMBS Conference on, Langkawi. 2012, pp. 999–1003.
30. Romanyuk O.A., Kozak L.M., Kovalenko A.S., Kryvova O.A. Digital transformation in medicine: from formalized medical documents to information technologies of digital medicine. *Cybernetics and Computer Engineering.* 2018, no. 4(194), pp. 61–78.
31. Krivova O.A., Kozak L.M. Comprehensive assessment of regional demographic development. *Kibernetika i vyčislitel`naâ tehnika*. 2015, Iss 182, pp. 70–84 (In Russian)
32. Wolf L., Shashua A. Features Selection for Unsupervised and Supervised Inference: The Emergence of Sparsity in a Weight-Based Approach. *J. Machine Learning Res*. 2005, V. 6, pp. 1855–1887.
33. Guyon I., Elisseeff A. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*. 2003, V 3, pp. 1157–1182.
34. Mandel I.D. Cluster analysis. Moscow: Finance and Statistics. 1988. 128 p. (In Russian)
35. Tzortzis G., Likas A. The MinMax k-Means clustering algorithm. *Pattern Recognition*. 2014, no 47 (7), pp. 2505–2516.
36. McLachlan G. Krishnan T. *The EM algorithm and extensions*. New York, United States: Wiley. 1997, 274 p.
37. Wang K., Wang B., Peng L. CVAP: Validation for cluster analyses. *Data Science Journal.* 2009, no 8, pp. 88–93.
38. Fayn J. A classification tree approach for cardiac ischemia detection using spatiotemporal information from three standard ECG leads. *IEEE Trans. Biomed. Eng*. 2011, V. 58, no 1, pp. 95–102.
39. Pecchia L., Melillo P. Bracale M. Remote health monitoring of heart failure with data mining via CART method on HRV features. *IEEE Transactions Biomedical Engineering*. 2011, V. 58(3), pp. 800–804.
40. Sokolova M., Lapalme G. A systematic analysis of performance measures for classification tasks. *Information processing & management.* 2009, V. 45, N 4, pp. 427–437.
41. Kalnish V.V., Shvets A.V. Information technology for psychophysiological support of high reliability of operator activities. *Kibernetika i vyčislitel`naâ tehnika*. 2014, Iss. 177, pp. 54–67. (In Russian)
42. Shvets A.V., Kalnysh V.V. Features of influence of various psychophysiological states on reliability of operator` activity. *Military medicine of Ukraine*. 2009, no 1, pp. 84–91. (In Ukrainian)
43. Consolaro A., Ruperto N, Bazso A. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. *Arthritis & Rheumatism*, 2009, vol. 61, pp. 658–666.
44. Ansari S., Farzaneh N, Duda M, Horan K. A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records. *IEEE reviews in biomedical engineering*. 2017, Vol. 10, pp. 264–298.

Received 31.11.2020

ЛІТЕРАТУРА

1. Ian H. Data Mining Practical Machine Learning Tools and Techniques Witten, Eibe Frank and Mark A. Hall Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition. Morgan Kaufmann, 2011. 665 p.
2. Yoo I., Alafaireet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J. F. Data Mining in Healthcare and Biomedicine: A Survey of the Literature. *Journal of medical systems*. 2012. No 36(4). P. 2431–2448.
3. Chen M., Hao Y. , Hwang K., Wang L., Wang L. Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. *IEEE Access.* 2017;5:8869-8879.
4. Safdar S., Zafar S., Zafar N., Khan N.F. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. *Artificial Intelligence Review.* 2018, 50 (4), 597-623.
5. Roopa C. K., Harish B. S. Survey on various Machine Learning Approaches for ECG Analysis. *International Journal of Computer Applications.* 2017. no 9. Vol. 163. pp.25–33.
6. Mohan S., Thirumalai C., Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 2019. 7:81542–81554.
7. Goldstein B.A., Navar A.M., Pencina M.J., Ioannidis J.P. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2017, Jan; 24(1):198-208.
8. Антомонов М.Ю. Алгоритмизация выбора адекватных математических методов при анализе медико-биологических данных. *Кибернетика и вычислительная техника*. 2007. Вып. 153. С. 12–23.
9. Georga E.I., Tachos N.S., Sakellarios A.I., Kigka V.I., Exarchos T.P., Pelosi G. Artificial intelligence and data mining methods for cardiovascular risk prediction Cardiovascular Computing. *Methodologies and Clinical Applications*. 2019. P. 279–301
10. Amin M., Chiam Y. Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics.* 2019. Vol. 36. P. 82–93.
11. Kaieski N., da Costa C.A., da Rosa Righi R., Lora P.S. Application of artificial intelligence methods in vital signs analysis of hospitalized patients: A systematic literature review. *Applied Soft Computing*. 2020. Vol. 96.
12. Owens W.D., Felts J.A., et al. A physical status classification: A study of consistency of ratings. *Anesthesiology*. 1978. Vol. 49. P. 239–243.
13. Lemeshow S., Le Gall J.R: Modeling the severity of illness of ICU patients. *JAMA*. 1994, Vol 272. P.1049–1055.
14. Le Gall J.R., Lemeshow S., Saulnier F: A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993, 270 (24). P.: 2957–2963.
15. Knaus W.A., Draper E.A., Wagner D.P., Zimmerman J.E: APACHE II: A severity of disease classification system. *Cri.t Care Med* .1985. 13:818-829.
16. Lemeshow S., Teres D., Klar J., Avrunin J.S., Gehlbach S.H., Rapoport J. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA.*1993. 270:2478-86
17. Trujillano J., Badia M, Serviá L. Stratification of the severity of critically ill patients with classification trees. *BMC medical research methodology*. 2009. V 9. no 7. P. 83–95.
18. Kim S., Kim W., Park R.W. A Comparison of intensive care unit mortality prediction models through the use of Data Mining Techniques. *Health Inform Res* 2011;17:232-43.
19. Allyn J. et all. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PloS one* 2017. 12(1). P. 1–12.
20. Амосов Н.М. Раздумья о здоровье — М: 1978. 178 с.
21. Баевский Р.М., Берсенева А.П. Введение в донозологическую диагностику. М.: Слово, 2008. 174 с.
22. HRV analysis software. URL: http://www.nevrokard.eu/maini/hrv.html (дата звернення: 20.10.2020)
23. Файнзильберг Л.С. Компьютерная диагностика по фазовому портрету электрокардиограммы. К.: Освита Украины, 2013. 191 с.

24. Гриценко В.И., Файнзильберг Л.С. Интеллектуальные информационные технологии в цифровой медицине на примере фазаграфии. Киев: Наукова Думка, 2019. 423 с.
25. Fainzilberg L.S., Dykach Ju.R. Linguistic approach for estimation of electrocardiograms's subtle changes based on the Levenstein distance. *Cybernetics and Computer Engineering*. 2019. No. 2 (196). P. 3–26.
26. Gritsenko V.I., Fainzilberg L.S. Current state and prospects for the development of digital medicine. *Cybernetics and Computer Engineering*. 2020. No. 1 (199). P. 59–84.
27. Richman J.S. , Randall M.J.Physiological time–series analysis using approximate entropy and sample entropy. *Am J. Physiol. Heart Circ. Physiol.* 2000. Vol. 278. № 6. P. H22039–H2049.
28. İşler Y., Kuntalp M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine*. 2007. Vol. 37. No. 10. P. 1502–1510.
29. Valupadasu R., Chunduri B. R., Chanagoni V. Identification of Cardiac Ischemia using bispectral analysis of ECG. *Biomedical Engineering and Sciences* (IECBES). 2012: IEEE EMBS Conference on, Langkawi. 2012. P. 999–1003.
30. Romanyuk O.A., Kozak L.M., Kovalenko A.S., Kryvova O.A. Digital transformation in medicine: from formalized medical documents to information technologies of digital medicine. *Cybernetics and Computer Engineering.* 2018. No. 4(194). P. 61–78.
31. Кривова О.А., Козак Л.М. Комплексная оценка регионального демографического развития. *Киб.и выч.техн.*, 2015. Вып 182. С. 70–84.
32. Wolf L., Shashua A. Features Selection for Unsupervised and Supervised Inference: The Emergence of Sparsity in a Weight-Based Approach. *J. Machine Learning Res*. 2005. V. 6. P. 1855–1887.
33. Guyon I., Elisseeff A. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*. 2003. V 3. P. 1157–1182.
34. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика. 1988. 128 с.
35. Tzortzis G., Likas A. The MinMax k-Means clustering algorithm. *Pattern Recognition*. 2014. No 47 (7). Pp 2505–2516.
36. McLachlan G. Krishnan T. *The EM algorithm and extensions*. New York, United States: Wiley. (1997) 274 p.
37. Wang K., Wang B., Peng L. CVAP: Validation for cluster analyses. *Data Science Journal*. 2009. No 8. Pp. 88–93.
38. Fayn J. A classification tree approach for cardiac ischemia detection using spatiotemporal information from three standard ECG leads. *IEEE Trans. Biomed. Eng*. 2011. V. 58. no 1. P. 95–102.
39. Pecchia L., Melillo P. Bracale M. Remote health monitoring of heart failure with data mining via CART method on HRV features. *IEEE Transactions Biomedical Engineering*, 2011 V. 58(3). P. 800–804.
40. Sokolova M., Lapalme G. A systematic analysis of performance measures for classification tasks. *Information processing & management*. 2009 (45). N 4. V. 45. P. 427–437.
41. Кальниш В.В., Швец А.В. Информационная технология психофизиологического обеспечения высокой надежности операторской деятельности. *Киберн. И выч. техн.* 2014. Вып. 177. С. 54–67.
42. Швець А.В., Кальниш В.В. Особливості впливу різних психофізіологічних станів на надійність операторської діяльності. *Військова медицина України*. 2009. № 1. С. 84–91.
43. Consolaro A., Ruperto N, Bazso A. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. *Arthritis & Rheumatism*. 2009. Vol. 61. P. 658–666.
44. Ansari S., Farzaneh N, Duda M, Horan K. A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records. *IEEE reviews in biomedical engineering*. 2017. Vol. 10. P. 264–298.

Отримано 31.11.2020

Кривова О.А., наук. співроб.
відд. медичних інформаційних систем
ORCID: 0000-0002-4407-5990
e-mail: ol.kryvova@gmail.com
Козак Л.М., д-р біол. наук, старш. наук. співроб.,
провід. наук. співроб. відд. медичних інформаційних систем
ORCID: 0000-0002-7412-3041
e-mail: lmkozak52@gmail.com
Міжнародний науково-навчальний центр інформаційних
технологій та систем НАН України та МОН України,
пр. Акад. Глушкова, 40, м. Київ, 03187, Україна

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КЛАСИФІКАЦІЇ ДОНОЗОЛОГІЧНИХ ТА ПАТОЛОГІЧНИХ СТАНІВ ЗДОРОВ'Я З ВИКОРИСТАННЯМ АНСАМБЛЮ МЕТОДІВ DATA MINING

Вступ. Впровадження цифрових технологій забезпечує реєстрацію великих обсягів біомедичних даних (ЕКГ, ЕЕГ, електронних медичних записів) як основи для оцінювання і прогнозування стану пацієнтів. Методи Data Mining дають змогу виявити найбільш інформативні показники, типологічні групи, класифікувати функційний стан людини і стадії захворювання для прогнозування їхніх змін.

Метою роботи є розроблення інформаційної технології класифікації стану здоров'я людини за допомогою комплексу методів Data Mining за об'єктивними та експертними характеристиками.

Результати. Розроблена інформаційна технологія об'єднує кілька етапів: I — попереднє оброблення даних; II — кластеризація, вибір однорідних груп (сегментація даних); III — ідентифікація предикторів; IV — класифікація досліджуваних станів, розроблення прогнозних моделей за допомогою алгоритмів машинного навчання (дерев рішень (Decision trees, опорних векторних машин Support vector machine, нейронних мереж) та методу перевірки навчальної вибірки (cross-validation). Запропоновану ІТ використано для дослідження функційного стану операторів та класифікації тяжкості стану пацієнтів у разі прогресування захворювання.

Висновки. Використання інформаційної технології для оцінювання успішності діяльності операторів дало можливість виділити найінформативніші показники ВРС, за змінами яких можна прогнозувати надійність діяльності операторів з урахуванням типу вегетативної регуляції. Оцінювання активності захворювання дітей з дисплазією з використанням ІТ дало змогу ідентифікувати діагностичні маркери ССС та розробити діагностичні правила для визначення стадій захворювання за параметрами ЕКГ (симетрія зубця Т, інтегральний показник форми сегмента STT).

Ключові слова: *інформаційна технологія, Data Mining, моделі машинного навчання, тяжкість стану пацієнта.*