COMPUTATIONAL INTELLIGENCE METHODS TO PATIENTS STRATIFICATION IN THE MEDICAL MONITORING SYSTEMS

Bakumenko N. S. – PhD, Associate Professor of Theoretical and Applied Systems Engineering Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.

Strilets V. Y. – PhD, Associate Professor of Theoretical and Applied Systems Engineering Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.

Ugryumov M. L. – Dr.Sc., Professor of Theoretical and Applied Systems Engineering Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.

Zelenskyi R. O. – Assistant of the Department of Urology and Pediatric Urology, Kharkiv Medical Academy of Postgraduate Education, Kharkiv, Ukraine.

Ugryumova K. M. – PhD, Associate Professor, Department of Mathematical Modeling and Artificial Intelligence, National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine.

Starenkiv V. P. – Dr.Sc., Professor, Head of the Radiation Oncology Department, State Organization “Grigoriev Institute for Medical Radiology and Oncology of the National Academy of Medical Sciences of Ukraine”, Kharkiv, Ukraine.

Artiukh S. V. – PhD, Senior Researcher of the Radiotherapy Group of the Radiology Department State Organization “Grigoriev Institute for Medical Radiology and Oncology of the National Academy of Medical Sciences of Ukraine”, Assistant Professor of Radiology and Radiation Medicine, Kharkiv National Medical University, Kharkiv, Ukraine.

Nasonova A. M. – PhD, Clinical Oncologist, Head of the Clinical Oncology and Hematology Department, State Organization “Grigoriev Institute for Medical Radiology and Oncology of the National Academy of Medical Sciences of Ukraine”, Kharkiv, Ukraine.

ABSTRACT

Context. In modern medical practice the automation and information technologies are increasingly being implemented for diagnosing diseases, monitoring the condition of patients, determining the treatment program, etc. Therefore, the development of new and improvement of existing methods of the patient stratification in the medical monitoring systems is timely and necessary.

Objective. The goal of intelligent diagnostics of patient’s state in the medical monitoring systems – reducing the likelihood of adverse states based on the choice of an individual treatment program:

− reducing the probability of incorrectly determining the state of the patients when monitoring patients;
− obtaining stable effective estimates of unknown values of treatment actions for patients (corresponding to the found state);
− the choice of a rational individual treatment program for the patients, identified on the basis of the forecasted state.

Method. Proposed methodology, which includes the following computational intelligence methods to patient’s stratification in the medical monitoring systems:

1) method of cluster analysis based on the agent-based approach – the determination of the possible number of patient’s states using controlled variables of state;
2) method of robust metamodels development by means artificial neuron networks under a priori data uncertainty (only accuracy of measurements is known) in the monitoring data: a) a multidimensional logistic regression model in the form of analytical dependences of the posterior probabilities of different states of the patients on the control and controlled variables of state; b) a multidimensional diagnostic model in the form of analytical dependences of the objective functions (quality criteria of the patient’s state) on the control and controlled variables of state;
3) method of estimating informativeness controlled variables of state at a priori data uncertainty;
4) method of robust multidimensional models development for the patient’s state control under a priori data uncertainty in the monitoring data in the form of analytical dependencies predicted from the measured values of the control and controlled variables of state in the monitoring process;

5) method of reducing the controlled state variables space dimension based on the analysis of the variables informativeness of the robust multidimensional models for the patient’s state control;

6) method of patient’s states determination based on the classification problem solution with the values of the control and forecasted controlled variables of state with using the probabilistic neural networks;

7) method of synthesis the rational individual patient’s treatment program in the medical monitoring system, for the state identified on the basis of the forecast.

Proposed the structure of the model for choosing the rational individual patient’s treatment program based on IT Data Stream Mining, which implements the «Big Data for Better Outcomes» concept.

**Results.** The developed advanced computational intelligence methods for forecast states were used in choosing the tactics of treating patients, to forecast treatment complications and assess the patient’s curability before and during special treatment.

**Conclusions.** Experience in the implementation of “Big Data for Better Outcomes” concept for the solution of the problem of computational models for new patient stratification strategies is presented. Advanced methodology, computational methods for a patient stratification in the medical monitoring systems and applied information technology realizing them have been developed. The developed methods for forecast states can be used in choosing the tactics of treating patients, to forecast treatment complications and assess the patient’s curability before and during special treatment.

**KEYWORDS:** Information Technology, Data Stream Mining, Medical Monitoring Systems, Machine Learning Methods, Mathematical Models and Methods for Patient Stratification.
is mean square deviation of variables \( \bar{x}_k \);
\( \sigma^*_{X} \) is required standard deviation of the input data;
\( X_{\text{max}}^{(0)} \) is maximum value of the input data variables;
\( X_{\text{min}}^{(0)} \) is minimum value of the input data variables;
\( D_k \) is training dataset;
\( L \) is dimension of the training dataset (state’s space);
\( M \) is search variables set;
\( M \) is variables subset from \( D_k \) set;
\( D_{F_i} \) is dispersion of the target function, in decibels;
\( D_{F_i}^{(0)} \) is dispersion of the base (linear) model;
\( D_{F_i}^{(\text{mod})} \) is dispersion of the other model;
\( D_{\lambda_x} \) is subset of informativeness variables;
\( \Sigma_x \) is the covariance matrix of variables \( X_l \) and \( X_n \);
\( \sigma_{X_l} \) is the standard deviation of \( X_l \);
\( r_{ab} \) are the correlation coefficients between \( X_l \) and \( X_n \) \( (l=1...L, n=1...L) \);
\( E_t \) is a signal energy;
\( \lambda_0 \) is the informativeness coefficients;
\( D_0 \) is monitoring results;
\( Q \) is dimension of monitoring data;
\( \Pi^* \) is design and regime parameters;
\( U^* \) is control variables;
\( \Phi^* \) is phase variables;
\( t \) is forecast moment;
\( T_l \) is lower forecast moment;
\( T_u \) is upper forecast moment;
\( \varepsilon_{i}^0 \) is relative error;
\( \Psi \) is control state model;
\( \Omega \) is diagnostic model;
\( W^* \) is quality criteria of the patient’s condition;
\( t \) is dimension of time series \( q^{*} \);
\( \bar{X}^* \) is vector of observed symptoms of unidentified precedent;
\( X_{l}^* \) are principal components projections of \( \bar{X}^* \);
\( A \) is number of the principal components;
\( c_{l} \) are the classes centers \( (l, k=1...A) \);
\( p(R_{x_0}) \) is the probability distribution density agreeing to the Student’s \( t \)-law;
\( \{ \bar{X}_m \} \) is average projections value principal components vector of the observed symptoms of layer samples element;
\( \Gamma (\cdot) \) is gamma function;
\( K_l \) is number of precedents in \( l^{\text{th}} \) class, \( l=1...A \);
\( K_k \) is number of precedents in \( k^{\text{th}} \) class, \( k=1...A \);
\( \Sigma_{\text{pooled}} \) is the combined correlation matrix for the considered scenarios (for classes);
\( P(R_{x}) \) is a prior probability of the classes realization;
\( X \) is quasisolution of MV-problem;
\( t_i \) is Student’s coefficient for considered function \( f_i \);
\( R_{OP} \) is Romanovsky’s coefficient for considered function \( f_i \);
\( M_{[\cdot]} \) is mathematical expectation with significance level \( \alpha \);
\( x^*_{p} \) is required variables values \( x_{p} \) for prototype;
\( \sigma^*_{x} \) is required mean deviation values of variables \( x_{p} \) for prototype;
\( \sigma_{x^*_{p}} \) is mean square deviation of variables \( x_{p} \in X^0 \);
\( \sigma^*_{f_i} \) is the mean deviation values of decision making criterion \( f_i \) for prototype;
\( \sigma_{f_i} \) is mean square deviation of decision making criterion \( f_i \in F \);
\( \gamma \) is a regularization parameter;
time is the value of the survival time for the patient after undergoing treatment.

**INTRODUCTION**

The actual problems of modern medicine – the problems of assessing the states, forecasting the outcome of diseases, the effectiveness of treatment methods, assessing the likelihood of complications in patients – can be solved based on the use of advanced machine learning methods and information technologies that implement them. That’s why developing an applied information technology for patient stratification in medical monitoring systems based on advanced machine learning methods is an actual scientific and practical task.

An MMS consists of monitoring hardware for the patient’s condition, decision makers (physicians) and CDSS. The aim of such systems – to ensure continuous observation, information collecting, data processing and analyzing patient’s condition, forming recommendations for treating [1, 2].

National Aerospace University named by N. Zhukovsky “Kharkiv Aviation Institute” for more than 15 years has been developing applied information technology for patient stratification in medical monitoring systems based on machine learning methods. The university has developed its own software to accomplish the assigned problems and has qualified personnel who participate in international projects in the field of applied mathematics, statistics, and machine learning. In cooperation with the Kharkov National Medical University, the developed software is being tested for stratification of patients with prostate cancer, squamous cell carcinoma of the head and neck [3–14]. The developed methods for forecasted states are also planned to be verified on the data obtained in various oncological pathologies and to use them in choosing the tactics of treating patients, to forecast treatment complications and assess the patient’s curability before and during special treatment.

The material and technical support and personnel base of the Kharkov National Medical University allows col-
The aim of the investigation is to develop advanced methodology for solving the synthesis problem of patient’s individual treatment program in the medical monitoring systems by means of computational intelligence methods for the medical information analysis and applied information technology realizing them by means of new robust estimation methodology (M-estimation) based on the concept of invariance of the theory of optimal control and apply it to solving nonlinear multidisciplinary problem of under uncertainty.

Let us input data \( \left( \bar{X}^{(0)}, \bar{F} \right)_p, \; p = 1, \ldots, P \) are the results of the anamnesis (personal, visual examination, laboratory data), these are statistical data (dataset) from the observed patients, accumulated by medical institutions, as well as the individual results of patient monitoring, considered as training pairs. An input vector and output vector is of dimension \( H_0 \) and \( H_{k-1} \), respectively.

The data model, such as information from the ambulatory card, formed by experts (physicians).

Statistical data (dataset) from the observed patients had transferred to the database storage Big Data from medical organizations.

Each patient must be able to contribute their own monitoring results to the database.

Data processing and diagnostic results transmission carried out by using Internet resources, which is available to every user, including the use of CDSS on remote servers.

Structuring new mathematical statements and developing computational methods is necessary for solving the synthesis problem of a patient’s individual treatment program in the medical monitoring systems by means of computational intelligence methods, in stochastic formulation (MV-problem).

On the result of processing the input data, you need to find the following:

- the required number of controlled state variables for the classification of states (single or multiple diseases, and their corresponding disease stages);
- results of the patient’s condition classification (single or multiple cases and their corresponding disease stages);
- estimates of survival rates and clinical effectiveness of the treatment;
- quality of life assessments for patients during treatment and in the follow-up period;
- results of informative monitored condition variables synthesis for patients individually at the current point in the monitoring process;
- results of predicting the time series of monitored state variables (including medical influences);
- justification of the time intervals choice for measuring values of controlled variables during monitoring depending on the disease stage;
- the results of predicting patient conditions that cannot be assessed using current clinical, laboratory and in-
instrumental methods (e.g., hormone resistant prostate cancer);
– selection of an individual treatment regimen (medical interventions) depending on the patient’s condition in order to minimize the risk of serious adverse events.

Such developments compared to existing methods will ensure a reduction in the probability of incorrectly determining the state of the systems (errors of the third kind in classifying the state of the patients), as well as obtaining stable effective estimations of the unknown values variables (corresponding to the found state).

2 REVIEW OF THE LITERATURE

Over the past decade, as a result of cooperation, the authors have developed the methods for solving such problems:
– formation of a subset of controlled variables of state, the values of the quantities of which are check in by measuring instruments [3, 9];
– patient’s classification of state results [4, 6, 7, 9];
– informativeness evaluations of the variables of for different stages of patients’ diseases [4, 5, 9];
– robust metamodels: multidimensional diagnostic model, multidimensional logistic regression [6, 8, 9, 11];
– robust multidimensional models of control the states of the patients, evaluation of forecasted values of the controlled variables of state based on medical monitoring data for patients [11, 5];
– cluster analysis results – the number of recognizable states [5, 9, 12];
– patient’s state classification results, using the values of the control and forecasted controlled variables of state [9, 11];
– the results of the synthesis of a rational individual treatment program for the patients for the state, determined on the basis of a forecast (determination of control variables (medical actions), that ensure the implementation of the treatment program) [9, 11, 14].

Numerical research was carried out with the help of the computer program “Non-linear evaluation methods in multicriteria problem of robust optimal designing and diagnostics of systems under parametric a priori uncertainty (methodology, methods, techniques and computer systems of support and decision-making implementing them)” (ROD&IDS) [13] developed by the authors.

The volumes of medical information about various diseases and their course are extremely large, and machine learning methods make it possible to process the accumulated information, take into account millions of different factors, social, territorial, demographic, genomic, etc., and make it possible to identify the unique features of each patient [15, 16]. It has been proven that the automation of the collection processes and further analysis of medical data allows to increase the accuracy of early diagnosis, prediction of the disease’s development and the treatment effectiveness assessment [17]. For example, the Frost & Sullivan agency notes that artificial intelligence technologies increase the accuracy of diagnoses by 30–40%, and pathologist Andy Beck from Harvard Medical School believes that the further use of artificial intelligence technologies will reduce the errors rate in diagnosis by 85% [18].

Modern technologies are also used to choose the most effective treatment strategy. Recently, scales (systems) for objectifying the assessment of clinical-physiological and laboratory parameters have been used to choose treatment tactics, the scope of anesthetic support and surgical intervention, predict the frequency of probable postoperative complications, lethality, and the treatment effectiveness [19]. The priority of such a strategy is that after receiving the sum of objective indicators, the doctor converts them into a score, which is ranked into numerical and staging corridors. This makes it possible to comprehensively assess the patient’s condition at the moment, monitor his condition and carry out appropriate treatment. The correctness and effectiveness of treatment with this method depends on the qualifications and experience of the doctor treating the patient [20].

The choice of mathematical methods for the description and research of biological and medical objects depends both on the specialist’s individual knowledge and on the specifics of the tasks to be solved [21].

The following methods are most often used to solve the classification problem of the medical facilities state:
– binary classification (decision trees and random forests) [22];
– artificial neural networks (ANN) [23];
– multidimensional logistic regression [24];
– naive Bayesian classifier [25];
– support vector machines [23, 26].

Obviously, the use of each of these methods separately does not allow solving the general problem of synthesizing an individual patient treatment program in a medical monitoring system.

So, there is a need to structure the system model of medical decision-making as a sequence of interrelated tasks and the corresponding system model of decision synthesis. In other words, it is necessary to create a methodology for the synthesis of solution to the problem of making medical decisions in general.

The work proposes methodology for solving the synthesis problem of a patient’s individual treatment program in the medical monitoring systems by means of computational intelligence methods for the medical information analysis.

3 MATERIALS AND METHODS

Medical Monitoring System is a set of monitoring states hardware, information technology Big Data tools (which contains CDSS), patients and decision-makers (doctors) who are in a communicate relation with each other and united with the purpose of managing and organizing the process of systematic or continuous observation, collection, processing and information investigation about the object state (of patient), its functioning (of various organs) and development for a certain period of time.
Its system is created and regulated by the monitoring entities (physicians) to ensure full, timely and accurate information and appropriate organization of effective functioning and control of the object of diagnosis (patient).

The Context diagram of “Big Data for Better Outcomes” concept implementation in the medical monitoring systems is shown in Fig. 1.

![Figure 1 – Context diagram of “Big Data for Better Outcomes” concept implementation in the medical monitoring systems](image)

Generalized methodology for solving the problem of synthesis problem of patient’s individual treatment program in the medical monitoring systems by means of computational intelligence methods as a result of its decomposition may be given as a sequence of processing techniques of structured and unstructured data of large volumes and considerable diversity using developed software. The solving result is sequence of steps to obtain mathematical model $\hat{F}^{(K+1)}(\hat{X}^{(0)})$ in order to solve diagnosing problems.

Patient stratification is carried out in two steps: Data preprocessing and Monitoring the state of the patients. Each stage is described below.

1. Data Preprocessing (Preliminary data preparation is performed by a doctor):

   1.1 Formation of a subset of controlled variables of state $\{\hat{X}^{(0)}, \hat{F}\}_{p=1}^{P}$, the values of the quantities of which are checked in by measuring instruments.

   1.2 A total dataset of alternatives is generated. Each alternative includes subsets of control variables and controlled variables of state; objective functions. The total dataset includes subsets of alternatives corresponding to different states of the patients, including the healthy state.

   1.3 Data cleaning from anomalous values of the quantities (outliers). Normalization of data.

   1.4 Cluster analysis – the determination of the possible number of states (states of the patients) using control and controlled variables of state. We are looking for an answer to the question whether there is enough data from medical research to be able to recognize different states and the corresponding stages of diseases with the maximum a posteriori probability? If the number of identified clusters coincides with the specified one and the distances between the clusters are statistically significant, then the subset of monitored variables of state can be considered complete. Otherwise, the diagnostic system, which is used for monitoring of patient’s states, should be equipped with new measuring instruments.

Based on the agent-based approach and in accordance with the chosen measures of intra-cluster distance, permissive elite selection rules are proposed for the formation of clusters, the selection of the best of them, and also for the selection of elements into clusters in the process of solution synthesis.

The result of solving such a problem is the number of clusters, as well as the number of elements in them.

The c-means clustering method was chosen as the basis. Let the sample of data be considered as $X = \{X_j \}_{j=1}^{k*}$ and $P = \sum_{j=1}^{k*} P_j$ is the total number of elements.

It is necessary to find $\{k_j, P_j\}_{j=1}^{k*}$.

Four measures of intra-cluster distance were used as a metric for the clustering data:

$$d(\hat{x}_{jp}, \hat{c}_j) = \frac{d^1(\hat{x}_{jp}, \hat{c}_j)}{\left(MD^2(\hat{x}_{jp}, \hat{c}_j) + \sum_{j=1}^{k} w_{jp}MD^2(\hat{x}_{jp}, \hat{c}_j) - D_{KL}(\hat{x}_{jp}, \hat{c}_j)\right)}.$$  

Let us define the average measure of the intra-cluster distance:

$$M(\hat{c}_j) = \frac{1}{P_j} \sum_{p=1}^{P_j} d(\hat{x}_{jp}, \hat{c}_j).$$

Also the loss function is defined as

$$LF(X) = \frac{1}{k} \sum_{j=1}^{k} M(\hat{c}_j).$$

Then the research problem statement will take in the form:

$$\begin{cases}
Z = \{k_j, P_j\} \\
\hat{Z} = \arg \min_{X \in D_0} LF(X)
\end{cases}$$

It is necessary to determine the number of clusters and distribute the data among clusters so that the value of the loss function is minimal.

To correct the centers of clusters, we use the expression

$$\hat{c}_j = \frac{\sum_{p=1}^{P_j} w_{jp} \hat{x}_{jp}}{\sum_{p=1}^{P_j} w_{jp}},$$


DOI 10.15588/1607-3274-2023-1-3
where \( w_{jp} = \frac{\rho(\tilde{x}_{jp}, \tilde{c}_j)}{\sum_{p=1}^{P} \rho(\tilde{x}_{jp}, \tilde{c}_j)} \) is the membership matrix with the Cauchy distribution
\[
\rho(\tilde{x}_{jp}, \tilde{c}_j) = \frac{1}{\pi \eta^2 \left[ 1 + MD^2(\tilde{x}_{jp}, \tilde{c}_j) / \eta^2 \right]}
\]

Data clustering algorithm based on agent-based approach:

1. set \( k_j^{(n)} > k^*, P_j^{(n)} = \text{int} \left( \frac{N}{k_j^{(n)}} \right) \) and randomly generate cluster centers \( \{ \tilde{c}_j \} \);
2. using the selected measure with cluster distance \( d(\tilde{x}_{jp}, \tilde{c}_j) \), choose \( P_j^{(n)} \) the nearest neighbors for each \( j \)th cluster;
3. for each \( j \)th cluster, using the \( \{ w_{jp} \} \) and \( \{ \rho(\tilde{x}_{jp}, P_j) \} \), the cluster centers \( \{ \tilde{c}_j \} \) are corrected;
4. for each \( j \)th cluster, using the selected measure within the cluster distance \( d(\tilde{x}_{jp}, \tilde{c}_j) \) and \( P_j^{(n)} \) the nearest neighbors are chosen. Delete the points that are duplicated: \( P_j^{(n)} \rightarrow P_j \);
5. for each \( j \)th cluster the average measure of the intra-cluster distance \( M(\tilde{c}_j) \) is calculated, and also \( LF(X) \) is calculated;
6. elite selection. Find the cluster with the largest \( M(\tilde{c}_j) \) and delete it;
7. \( k_j^{(n+1)} = k_j^{(n)} - 1; \quad P_j^{(n+1)} = \text{int} \left( \frac{N}{k_j^{(n+1)}} \right) \);
8. back to step 2, if \( k_j^{(n+1)} > 1 \).

1.5 Development of robust metamodels at a priori data uncertainty in the monitoring data (given that the results of measurements of variables of states are random variables – only accuracy of measurements is known):

a) a multidimensional logistic regression in the form of analytical dependences of the posterior probabilities of different states of the patients on the control and controlled variables of state;
b) a multidimensional diagnostic model in the form of analytical dependences of the objective functions (quality criteria of the patient’s state) on the control and controlled variables of state.

To development of robust metamodels as initial information used the vector function is given by a training sample \( \left\{ \tilde{X}(0), \tilde{F}_p \right\}, \quad p = 1 \ldots P \). We must approximate the given set. The problem can be solved with a resultant mathematical mechanism, which may give any value of the vector function \( \tilde{F}_p^{(K+1)}(\tilde{X}(0)) \), represented by this training set at a fixed input vector within the range, limited by the input data.

A multilayer feedforward artificial neural network (MFFN) and radial basis function network (RBF), used for data approximation, is a parallel distributed processor, which is capable of saving acquired knowledge and processing information between local processor elements (neuro-elements or neurons), bound by special links (synaptic links).

To provide parameter stability (robustness) and informative capability of statistical systems and processes models on the basis of learning ANN at the a priori input data uncertainty and also practically sufficient data approximation, it is reasonable to use advanced deep learning methods – stable (robust) statistical assessment of their parameters with adaptive learning rate as the ANN learning method.

Student and V. I. Romanovsky are used as a smoothing functional when choosing a rational solution, which provides a stable (robust) estimation of the searching values with parametric uncertainty of the input data, as well as sufficient, from a practical point of view, accuracy of data approximation in problems of improving systems.

The function (MV-problem) was used as a scalar convolution of selection functions, considering \( f_i = F_i^{(2)}, x_h = X_h^{(0)} \):

\[
MV = \frac{1}{2P \delta \zeta} \sum_{p=1}^{P} \sum_{i=1}^{P} \left[ f_{fi} \left( \frac{\mu_j(f_{i,p}) \Delta_{f_i,p}}{f_{j} + \sigma_{f_i,p}^0} \right)^2 + \beta_{f_{j}} f_{fi} \left( \frac{\xi_j f_{i,p} - n_{\alpha}}{\sqrt{2n_{\alpha}}} \right)^2 \right]
\]

here \( l = H + 1, \xi = [0.95, 0.99], \)

\[
f_{fi}(d_i) = 1 - \exp \left[ -\frac{L_{fi,d_i}}{4 \cdot L_{fi,d_i}} \right], \quad L_{fi,d_i} > 4, (d_i > 0);
\]

\[
\Delta_{f_i} = F_i^{(K+1)}(\tilde{X}(0)) - f_i(\tilde{X}(0)) \quad \sigma_{f_i}^0 = \frac{\sigma_{f_i}}{\sigma_{f_i}},
\]

\[
\frac{\xi_j f_{i,p} - n_{\alpha} + 3}{\sqrt{2(n_{\alpha} - 3)}} \quad \frac{\xi_j f_{i,\alpha}^2 - 1 + 3/n_{\alpha}}{n_{\alpha} - 4}
\]

For fitness functions \( f_i(\tilde{X}(0)) \) in the expression for scalar convolution of selection functions MV the meanings of relative values are calculated by formulas:
2.1 The values of the control and controlled variables of state corresponding to the patient’s state at a given time are measured.

2.2 Determination of the state for which the maximum a posteriori probability of its realization corresponds with the observed values of the quantities based on the solution of the classification problem, which allows to determine the disease stages that are not recognized by modern biomarkers (e.g., hormone-resistant stage in prostate cancer).

2.3 Estimation informativeness controlled variables of state at a priori data uncertainty, the synthesis of the set of informative controlled variables of state according to the patient’s state (disease stage) for the reduction of the variables of state space dimension, using multidimensional diagnostic model, i.e. searching informative subset \( D_{M,i} \) of minimal dimension where \( D_{M,i} \subset D_L \).

The set of input data \( X_i(x_i) \), where \( X = \{ x_i \} \), \( i = 1 \ldots L \) is presented as a Taylor series, while retaining only the terms of the first infinitesimal order in the expansion. For the dispersion of an arbitrary gotten linear function of several random variables estimate holds:

\[
D_{F_i} = \left( \frac{\partial F_i}{\partial x_1} \right)^T \Sigma x F_i = \ldots = L \sum_{i=1}^{L} \left( \frac{\partial F_i}{\partial x_1} \right)^2 \sigma^2 + \ldots + L \sum_{i=1}^{L} \sum_{l=1}^{L} r_{in} \frac{\partial F_i}{\partial x_l} \frac{\partial F_i}{\partial x_n} \sigma_{x_l} \sigma_{x_n}.
\]

Let us define the signal energy by the expression

\[
E_i = \sum_{h=1}^{H_i} \left( \frac{\partial F_i}{\partial x_h} \right)^2 \chi_{h0}^2 + \ldots + \sum_{h=1}^{H_i} \sum_{n=1}^{H_i} r_{in} \left( \frac{\partial F_i}{\partial x_l} \right) \left( \frac{\partial F_i}{\partial x_n} \right) \sigma_{x_l} \sigma_{x_n}.
\]

The informativeness coefficients (contribution weight \( \chi_{h0} \)) into \( F_i(2) \) are defined by

\[
\lambda_{ih} = \frac{D_{F_i(2)} \chi_{h0}}{E_i}, L \sum_{i=1}^{L} \lambda_{ih} = 1.
\]

2.4 Development of robust multidimensional models of control the state for the patients at a priori data uncertainty in the monitoring data in the form of analytical dependencies predicted from the measured values of the control and controlled variables of state in the monitoring process.

2.5 Forecasting multidimensional time series of controlled variables of state based on multidimensional models of control of the state for the patients:
where \( t = T_1 \ldots T_2 \) is a limited set of \( \Gamma^0, U^0, \Phi^0 \) and creating time series.

It is required to obtain a functional dependence, that will be reflect relationship between the next and previous values of the time series which satisfies the system preferences of DM, for a given forecast horizon:

\[
q^0(t+1) = F(q^0(t+T_2-1), \ldots, q^0(t-T_1)) + e^0.
\]

Controlled process mathematical model:

\[
\Delta \Phi^0 = \Psi(\Delta \Phi^0, \Delta U^0)
\]

\[
\Delta U^0 = \Omega(\Delta \Phi^0, \Delta U^0)
\]

where the first expression is recurrent mathematical model for monitoring and the second expression is a diagnostic model:

\[
X^0(t+1) = \ln \left( \frac{q^0(t+1)}{q^0(t-T_1)} \right), i = 1 \ldots I, t = (-T_1+1) \ldots T_2.
\]

2.6 Reducing the dimension of the space of controlled variables of state based on the analysis of the informativeness of the variables of the robust multidimensional models of control the state for the patients (Sensitivity Analysis). Estimating the rank of time series cointegration.

2.7 Determination of the states of the patients based on the solution of the classification problem and the values of the control and forecasted controlled variables of state is conducted using probabilistic neural networks. An input layer \( X^*_1, \ldots, X^*_A \) elements are the principal component vector projections of the monitored state variables \( X^* \). The layer of samples \( c_1^*, \ldots, c_A^* \) are the classes centers of the training samples. The number of exemplars equals the number of classes in the training set. The layer of input and the samples are entirely meshed structure. An element activity of the samples layer was determined by the dependence that related to the probability distribution density agreeing to the Student’s t-law (that is proper for the limited samples):

\[
c_{ik} = p(R_k) = \left( \frac{n+1}{2} \right)^{\frac{n+1}{2}} \sqrt{\frac{\pi n \Gamma(n/2)}{\Gamma(n/2)}},
\]

here \( \Gamma \) is gamma function with \( n=K_i+K_{k-2} \);

\[
t_{ik} = \frac{\sqrt{\sum_{l=1}^{K_i} MD^2_{ik}}}{K_i + K_k},
\]

where \( MD^2_{ik} \) is Mahalanobis distance from an unidentified precedent (it is supposed that it appropriate to \( l \) class) to the \( k \) sample:

\[
MD^2_{ik} = \frac{1}{\lambda} \left( X^* - \bar{X}_k \right)^T \Sigma^{-1}_{pooled} \left( X^* - \bar{X}_k \right),
\]

here \( \Sigma_{pooled} \) is determined because of belonging of the precedent to one, or another class:

\[
\Sigma_{pooled} = \frac{1}{K_i + K_k - 2} \left( (K_i - 1) \Sigma_i + (K_k - 1) \Sigma_k \right).
\]

The output layer \( k^* \), \( p(R_{k^*}) \) is a discriminator of the threshold value indicating the class to which belongs an unknown precedent.

The values of t-Student statistics depend on the selection of the degree of proximity of the precedent to the samples (when solving the classification problem), as well as between the samples (when analyzing the significance of the distance between classes). Therefore, it is necessary to further forming of supplementary statistical decision rule selecting a single guide basis. According to the maximum likelihood rule as a criterion for the transition from one state to the other state of patients, Bayes formula is used:

\[
\forall k = 0 \ldots A - 1: \frac{p(X^*_k + 1)}{p(X^*_k)} \geq 1.
\]

Which is true provided that

\[
\frac{P(R_k)}{P(R_{k+1})} \frac{p(X^*_k + 1)}{p(X^*_k)} \approx 1.
\]

2.8 Synthesis of a rational individual treatment program for the patients in the medical monitoring system, for the state identified on the basis of the forecast.

The control variables \( \tilde{X} = \{M[X_0] \sigma^0 \} \) estimation problem can be presented like a MV-problem.

The MV-problem provides an efficient stable (robust) estimation of the required quantities under parametric input data uncertainty.

The quaisolution of this problem is

\[
\tilde{X} = \arg \inf_{\tilde{X} \in D_X} MV(\tilde{X}, f, R_f),
\]

where the scalar convolution of selection functions \( MV \) is
Here \( \gamma = 10^{-7} \ldots 1.0 \), \( I = H_{k+1} \), \( \Delta f_i = M_\alpha [ f_i - f^*_i ] \),

\[
\Delta x_p = M_\alpha [ x_p - x^*_p ], \quad \chi^2_{f_i} = n_\alpha \frac{[ f_i - M_\alpha ( f_i ) ]^2}{\sigma_{f_i}^2},
\]

\[
\chi^2_{x_p} = n_\alpha \frac{[ x_p - M_\alpha ( x_p ) ]^2}{\sigma_{x_p}^2}, \quad \sigma^0 = \frac{\sigma_{f_i}}{\sigma_{x_p}}, \quad \sigma^* = \frac{\sigma_{x_p}}{\sigma_{x_p}}.
\]

For the MV-problem solution the computing method based on the memetic algorithm is applied. In this method the evolutorial neighborhood method and the randomize path laying method together. The evolutorial algorithm has next parameters, which change from epoch to epoch: real coding operator, fitness function and relaxation function.

The proposed method application gives the stable (robust) estimation of desired values under parametric input data uncertainty and lowers the difficulty of the quasisolution technique. The structure of the model for the choice of a rational individual treatment program for the patients based on IT Data Stream Mining, which implements the “Big Data for Better Outcomes” concept is shown in Fig. 2.

4 EXPERIMENTS

In that work, as an implementation of the concept “Big Data for Better Outcomes” for stratification of patients, the following tasks were solved:

- formation of a subset of controlled variables of state, the values of the quantities of which are check in by measuring instruments;
- robust metamodels: multidimensional diagnostic model;
- informativeness evaluations of the variables of states for different stages of patients’ diseases.

The authors considered the task of assessing the increased survival rate of cancer patients who received chemoradiotherapy.

The multidimensional diagnostic model – model for estimating the survival of patients was built using a data set obtained in the real medical practice of the authors.

The patients with squamous cell carcinomas of the head and cancer (stages III, IVa, IVb) was included in study and did a comprehensive examination, including the collection of anamnestic data, general clinical physical examination, computed tomography of the head, neck and chest organs before the start of radiation therapy and 1 month after its completion, laboratory general clinical and biochemical blood tests. All patient received course of radiation therapy 5 Gy per fraction till 6–7 weeks (32–35
fractions) with chemomodification Cisplatin 100 mg/m² at 1, 21, 42 days in SI “Grigoriev Institute for Medical Radiology and Oncology of NAMS of Ukraine”. This set contains records of 57 patients on 9 indicators (parameters) that characterize the current state of each patient: age, localization, stage, tumor (T) and lymph (N) nodes status (TNM classification, 8 edition), response (per RECIST 1.1 criteria), tumor marker squamous cell carcinoma’s antigen “before” and “after” the course treatment, survival time. All indicators have a numerical type, but some of them are enumerable, for example, the “localization” indicator takes four values from 1 to 4. The objective function for each record is “survival time”, the other parameters are attributed to the input data.

First, it is necessary to normalize the data. Since the inputs are different in value and content, the normalization for them will be different. The parameters “localization”, “stage”, “T” and “response” are normalized according to the formula

\[ f_i^0 = \frac{f_i}{f_{\text{max}}}, \]

here \( f_{\text{max}} = 4 \).

The age indicator is normalized as follows:

\[ \text{Age}_\text{norm} = \frac{\text{Age}_{\text{current}} - \text{Age}_{\text{max}}}{\text{Age}_{\text{max}}}, \]

here we take age as the maximum value \( \text{Age}_{\text{max}} = 85 \).

The indicators of tumor markers “before” and “after” are reduced to one parameter of tumor markers as follows:

\[ \text{Oncomarker} = \ln \frac{\text{marker}_{\text{after}}}{\text{marker}_{\text{before}}}. \]

Let’s denote the objective function by \( \text{Suv\_time} \), its values are also pre-normalized by the formula

\[ \text{Suv\_time} = \frac{\text{time}}{\text{targ\_et\_time}}, \]

here \( \text{targ\_et\_time} = 25 \).

The results of the patient survival assessment are a functional relationship between the input data and the survival time. Using a robust meta-model based on a radial-basis neural network, survival time estimation values are calculated that correspond to the input data.

The structure of the simplest radial-basis neural network includes three types of neuron layers. RBF with one hidden layer \( (K=1) \). At the input layer RBF has \( H_0 = 7 \) neurons, at the hidden layer - \( H_1 = 44 \) neurons and \( H_2 = 1 \) network outputs.

In addition, the task of determining the most influential (informative) parameters for assessing patient survival was solved.

Numerical research was carried out with the help of the computer program “Non-linear evaluation methods in multicriteria problem of robust optimal designing and intelligence diagnostics of systems under parametric a priori uncertainty (methodology, methods, techniques and computer systems of support and decision-making implementing them)” (ROD&IDS) developed by the authors [13].

## 5 RESULTS

To train the neural network and test its operation, the data set of the patient’s survival time was divided into two parts: training (47 records, 82%) and test (10 records, 18%).

As a result of training, the relative error was 1.5%.

The relative error on the test sample was 15.4%. The trained neural network can then be used to provide survival estimates for new patients (Table 1).

It was found by means of sensitive analysis of variables of the model for each patient from the training sample that on the average the indicators: T, age of the patient, response and stage of the disease are more likely to be informative (Table 2).

Therefore, to assess the survival of patients, it is necessary to pay attention to these indicators.
6 DISCUSSION

Thus, the goal of the study was achieved – developed advanced methodology for solving the synthesis problem of the patient’s individual treatment program in the medical monitoring systems by means of computational intelligence methods for the medical information analysis and applied information technology realizing them have been developed.

Its application allows:

– reducing the likelihood of adverse states based on the choice of an individual treatment program;
– reducing the probability of incorrectly determining the state of the patients (errors of the third kind in classifying the state of the systems) when monitoring patients;
– obtaining stable effective estimates of unknown values of treatment actions for patients (corresponding to the found state);
– the choice of a rational individual treatment program for the patients, identified on the basis of the forecasted state.

The developed software, which is described below in information technology, is being tested for stratification of patients with prostate cancer, squamous cell carcinoma of the head and neck [3–11]. The developed methods for forecasted states are also planned to be verified on the data obtained in various oncological pathologies and to use them in choosing the tactics of treating patients, to forecast treatment complications and assess the patient’s curability before and during special treatment.

The material and technical support and personnel base of the Kharkov National Medical University allows collecting data on the course of treatment of cancer patients receiving chemoradiation treatment, using the whole range of non-invasive diagnostic methods, a wide range of tumor markers. Working with patients is regulated by the Bioethics and Deontology Committee and complies with international GCP standards.

As part of the implementation of projects, an innovative strategy is proposed for choosing a rational individual tactics for treating patients based on forecasted states identified using robust stratification methods, which will improve clinical results and prevent complications.

It is also planned to collect data on patients with cancer who have undergone COVID-19 before starting treatment and to assess the effect of the latter on the timing of the start of radiation and chemotherapy, depending on the severity of post Covid’s syndrome.

CONCLUSIONS

The scientific novelty obtained results in the fact that an advanced methodology for solving the synthesis problem of a patient’s individual treatment program in the medical monitoring systems by means of computational intelligence methods for the medical information analysis and applied information technology realizing them have been developed.

The practical significance of obtained results is that forecast states can be used in choosing the tactics of treating patients, to forecast treatment complications and assess the patient’s curability before and during special treatment.

Prospects for further research are to study application possibilities of the proposed methodology for the choice of a rational individual treatment program for the patients with various chronic diseases, identified on the basis of the forecasted state by means of IT Data Stream Mining.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Karazin Kharkiv National University “Information process modeling in the complex and distributed systems” (state registration number 0121U109183) and research projects of Grigoriev Institute for Medical Radiology and Oncology of the National Academy of Medical Sciences of Ukraine “Development of programs for personalized control of the absorbed dose during radiation therapy of genital, head and neck tumors using in vivo dosimetry” (state registration number 0117U001046, 2017 – 2019) and “Optimizing topometrical preparation for radiation therapy of head and neck cancer patients” (state registration number 0119U103013, 2020 – 2022).

REFERENCES


Received 31.10.2022. Accepted 22.12.2022.
Актуальність. У сучасній медичній практиці все більше впроваджується автоматизація і інформаційні технології для діагностування захворювань, моніторингу стану пацієнта, визначення програм лікування тощо. Тому розробка нових і удосконалення існуючих методів стратифікації пацієнтів з використанням контрольованих змінних станів.

Метод. Розроблено методологію, яка включає такі методи обчислювального інтелекту для стратифікації пацієнтів у системах медичного моніторингу, як:
1) метод кластерного аналізу на основі агентного підходу – визначення можливої кількості станів пацієнтів з використанням контролювання змінних станів;
2) метод побудови робастних моделей за допомогою штучних нейронних мереж при априорній невизначеності данних (відома лише точність вимірювань) для даних моніторингу стану пацієнтів: а) багатовимірна логістична регресійна модель, яка виглядає аналітичних залежностей апостеріорних імовірностей різних станів пацієнтів від контролюванних змінних станів; б) багатовимірна діагностична модель у вигляді аналітичних залежностей імовірностей (критерій якості стану хворого) від контролюванних змінних станів;
3) метод оцінки інформативності контролюванних змінних станів при невизначеності априорних даних;
4) метод побудови багатовимірних моделей контролю стану пацієнтів при априорній невизначеності даних у даних моніторингу на основі аналізу залежностей, що прогнозуються за виміряними значеннями контролюванних змінних станів у процесі моніторингу;
5) метод зменшення розмірності простору контролюванних змінних станів на основі аналізу інформативності змінних робастних багатовимірних моделей управління станом пацієнта (аналіз чутливості);
6) метод визначення станів пацієнтів на основі виришення задач класифікації за значеннями контролюних та прогнозованих контролюваних змінних стану з використанням імовірних нейронних мереж;
7) метод синтезу рациональної індивідуальної програми лікування хворих у системі медичного моніторингу для стану, виявленого на основі прогнозу.

У роботі запропонована структура моделі вибору рациональної індивідуальної програми лікування пацієнтів на основі IT Data Stream Mining, яка реалізує концепцію «Big Data for Better Outcomes».

Результати. Розроблені передові методи обчислювального інтелекту для прогнозування станів використовувалися при виборі тактики лікування пацієнтів, прогнозуванні ускладнень лікування та оцінці відповідності пацієнта до та під час спеціального лікування.

Висновки. Представлено досвід впровадження концепції «Big Data for Better Outcomes» для вирішення проблеми розробки передових методологій нових стратегій стратифікації пацієнтів. Розроблено передову методологію, методи обчислювального інтелекту для стратифікації пацієнтів у системах медичного моніторингу та прикладну інформаційну технологію, що їх реалізує. Розроблені передові методи прогнозування станів можуть бути використані при виборі тактики лікування хворих, прогнозуванні ускладнень лікування та оцінці відповідності використаного програмного забезпечення до та під час спеціального лікування.

КЛЮЧОВІ СЛОВА: інформаційна технологія, методи прогнозування, програмна систима медичного моніторингу, методи статистичної обробки даних.

УДК 519.876.5:681.518.2

ВИЧИСЛИТЕЛЬНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ МЕТОДЫ СТРАТИФИКАЦИИ ПАЦИЕНТОВ

Бакumenko N. C. – канд. техн. наук, доцент, доцент кафедры теоретической и прикладной системотехники Харьковского национального университета имени В. Н. Каразина, Харьков, Украина.

Стрелец В. Е. – канд. техн. наук, доцент кафедры теоретической и прикладной системотехники Харьковского национального университета имени В. Н. Каразина, Харьков, Украина.

Угрюмов М. Л. – д-р техн. наук, профессор, профессор кафедры теоретической и прикладной системотехники Харьковского национального университета имени В. Н. Каразина, Харьков, Украина.

Зеленский Р. А. – ассистент кафедры урологии и детской урологии Харьковской медицинской академии последипломного образования, Харьков, Украина.

DOI 10.15588/1607-3274-2023-1-3
Аннотация

Актуальность. В современной медицинской практике все больше внедряются автоматизированные и информационные технологии для диагностики заболеваний, мониторинга состояния пациентов, выбора программы лечения и др. В связи с этим разработка новых и усовершенствование существующих методов вычислительного интеллекта для статификации пациентов в системах медицинского мониторинга является своевременной и необходимой задачей.

Метод. Разработана методология, включающая следующие методы вычислительного интеллекта для статификации пациентов:
1) метод кластерного анализа на основе агентного подхода – определение возможного количества состояний пациентов с использованием контролируемых переменных состояния;
2) метод построения робастных метамоделей с помощью искусственных нейронных сетей при априорной неопределенности данных мониторинга (известна только точность измерений) в данных мониторинга: a) многомерная логистическая регрессионная модель в виде аналитических зависимостей апостериорных вероятностей различных состояний пациентов от контрольных и контролируемых переменных состояния; b) многомерная диагностическая модель в виде аналитических зависимостей целевых функций (критерий качества состояния больного) от контрольных и контролируемых переменных состояния;
3) метод оценки информативности контролируемых переменных состояния при неопределенности априорных данных;
4) метод построения робастных многомерных моделей контроля состояния пациентов при априорной неопределенности данных мониторинга в виде аналитических зависимостей, прогнозируемых по измеренным значениям контрольных и контролируемых переменных состояния в процессе мониторинга;
5) метод уменьшения размерности пространства контролируемых переменных состояния на основе анализа информативности переменных робастных многомерных моделей управления состоянием пациентов;
6) метод определения состояний пациентов на основе решения задачи классификации по значениям контрольных и прогнозируемых переменных состояния с использованием вероятностных нейронных сетей;
7) метод синтеза рациональной индивидуальной программы лечения больных в системе медицинского мониторинга, для состояния, выявленного на основе прогноза.

В работе предложена структура модели выбора рациональной индивидуальной программы лечения пациентов на основе IT Data Stream Mining, которая реализует концепцию «Big Data for Better Outcomes».

Результаты. Разработанные усовершенствованные методы вычислительного интеллекта для прогноза состояний использовались при выборе тактики лечения пациентов, для прогнозирования осложнений лечения и оценки излечимости пациента до и во время проведения специального лечения.

Выходы. Представлен опыт реализации концепции «большие данные для лучших результатов» для решения задачи разработки передовых методологий для новых стратегий статификации пациентов. Разработана передовая методология, методы вычислительного интеллекта для статификации пациентов в системах медицинского мониторинга и реализующие их прикладные информационные технологии. Разработанные усовершенствованные методы прогноза состояний могут быть использованы при выборе тактики лечения больных, для прогнозирования осложнений лечения и оценки излечимости больного до и во время проведения специального лечения.

Ключевые слова: информационная технология потокового анализа данных, системы медицинского мониторинга, методы машинного обучения, математические модели и методы статификации пациентов.

Литература


