

УДК 004.94

Data Modeling in the Solution of Hard-to-Formalise Socio-Economic Problems

Konstantin V. Simonov*

Institute of Computational Modeling SB RAS
Academgorodok 50/44, Krasnoyarsk, 660036, Russia

Mikhail A. Kurako†

School of Space and Information Technologies
Siberian Federal University
Svobodny 79, Krasnoyarsk, 660041, Russia

Alexey A. Kabanov‡

SDTB Science ICT SB RAS
Mira pr. 53, Krasnoyarsk, 660049, Russia

Fedor P. Kapsargin§

Lubov F. Zueva¶

Artem V. Ershov||

Krasnoyarsk State Medical University
Partizana Zheleznyaka st., 1, Krasnoyarsk, 660022, Russia

Svetlana N. Grib**

School of Economics, Management and Environmental Studies
Siberian Federal University
Svobodny, 79, Krasnoyarsk, 660041, Russia

Received 29.11.2018, received in revised form 20.01.2019, accepted 20.02.2019

Algorithms for data modeling in the solution of hard-to-formalize social problems are considered in the paper. They are connected with the healthcare in terms of interconnections and interactions of territories of the Yenisei Siberia where the Krasnoyarsk Krai is the key region. Elements of information system for the analysis of the current state and estimation of the scenarios of future interaction of the territories of the Yenisei economic zone are considered to solve the problems. GIS technologies and modern approaches to model various data are described.

Keywords: data analysis algorithms, databases, GIS, data modelling, neural networks.

DOI: 10.17516/1997-1397-2019-12-2-249-260.

*simonovkv@icm.krasn.ru

†mkurako@sfu-kras.ru

‡weller86@inbox.ru

§kapsargin@mail.ru

¶kapsargina0705@mail.ru

||ershov1989@mail.ru

**sgrib@sfu-kras.ru

Introduction

We consider one of the problems of the socio-economic development of the region, namely, the increasing territorial disparity in terms of the level of medical services for socially significant diseases (Fig. 1). They are urological diseases and, in particular, the kidney stone disease. In the Krasnoyarsk Krai over the past 5 years the incidence of the kidney stone disease was at the level of 29.4% due to various socio-economic and environmental causes. In the Republic of Khakassia and Tyva the incidence was even higher: at the level of 29.8% and 30.1%, respectively.

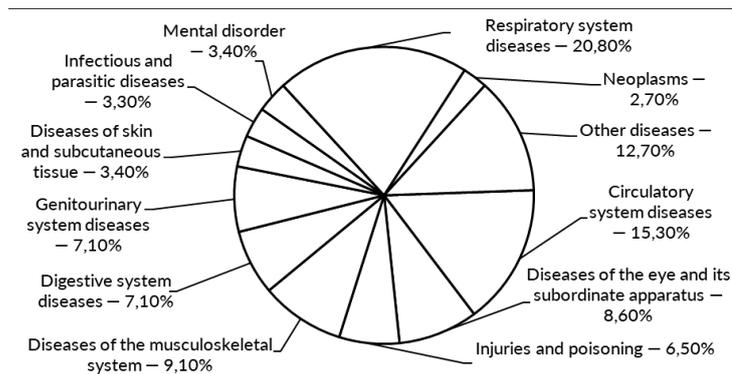


Fig. 1. Structure of diseases in the Krasnoyarsk Krai

The territories of the Yenisei Siberia, where the Krasnoyarsk Krai is the key region, have significant meridional extent. This causes a variety of climatic features and sharp distinctions in socio-economic conditions. There are large industrial facilities on the territory of the region. It causes emissions of harmful substances and leads to soil and groundwater pollutions. At present, the modern multi-level system of rendering assistance to urological patients requires new optimization approaches. The monitoring of urological morbidity is not actively conducted to predict the morbidity and calculate the needs of consumables.

The specific feature of nephrolithiasis in the region includes a relatively large number of calculuses which are difficult to treat by the distant lithotripsy. One should also mention that it is impossible to provide the specialized urological assistance with all the technical capabilities of modern diagnostics and treatment in each settlement. All these result in late hospitalization of patients and a large number of complications, in particular, acute pyelonephritis with nephrolithiasis. Urolithiasis is often accompanied by severe complications with relapses in a large percentage of cases. The importance of the problem is also due to the fact that the disease is more common mainly among people of working age. People with the kidney stone disease accounts for 6% of people with disabilities and 7.6% of patients have one kidney. More than 90% of people with disabilities are people of working age.

Numerous studies devoted to this problem allowed us to highlight the most important reasons contributing to the development of the kidney stone disease in the region. They are environmental factors that include chemical composition of soil and plants, the degree of mineralization of water, climate, gender, age, production and living conditions. Moreover, the geographical distribution of patients, even within the same region, is very heterogeneous. There is a tendency to believe that for each geographic region or locality there is one or several factors that serve as a trigger for the development of the kidney stone disease among the population living in this territory.

1. Modern digital technology and data modeling

The study is to determine intraterritorial factors which contribute to the development of the kidney stone disease and to outline the digital model of patient ecology. The following problems are considered: computer analysis of climatic and environmental features of the region, computer analysis of the survey to determine the regional model of ecology of humans with the kidney stone disease, development of digital diagnostic criteria for the kidney stone disease with the definition of a patient-oriented routing algorithm, expert selection of general metaphylactic measures based on the results of a comprehensive numerical analysis of the chemical composition of stones using IR spectroscopy and IC.

Therefore, the main objective of the research is to determine the factors that contribute to stones formation on the basis of data of quality of soil and water for the Krasnoyarsk Krai; to develop the data base of the patients with the kidney stone disease; to develop the digital questionnaire on the kidney stone disease; to perform computer analysis of preliminary diagnostics with the kidney stone disease determined for the first time; to develop an algorithm for phased diagnosis of the kidney stone disease; to implement the phased diagnostic algorithm for the kidney stone disease; to perform computer analysis of groups of patients with the kidney stone disease who underwent surgical treatment, taking into account the implementation of the phased algorithm; to select general preventive measures based on the results of comprehensive analysis of environmental data for each district of the region; to develop postoperative metaphylaxis on the basis of the chemical composition of stones using IR spectroscopy and IC. As a result, we plan to obtain the accessible and unified approach in the diagnosis and treatment of the kidney stone disease, uniform quality criteria, control of quality criteria. We also plan to change the approach to the diagnosis and treatment of urolithiasis which is associated with the introduction of modern digital methods of early diagnosis into practice implementation of metaphylaxis.

We considered morbidity of the population with the kidney stone disease, the degree of pollution of the air and drinking water, climate conditions, optimization of patients routing, diagnostic measures in accordance with the stages of routing and chemical composition of the extracted stones for 5 districts of the region (Norilsk, Lesosibirsk, Achinsk, Minusinsk, Krasnoyarsk). The obtained results were analysed in order to determine regional features of the progress of the kidney stone disease [1–7].

Currently, X-ray examinations are performed with the Siemens equipment. They include an overview image of the urinary tract, excretory, retrograde and/or antegrade (if indicated) urography. In order to diagnose urological diseases, including urolithiasis, multispiral computed tomography (MSCT) is used with the BRIGSPID multispiral computed tomograph (**General Electrics**). The method is non-invasive and highly informative. It allows one to receive information not only on the structure of the pelvicalyceal system but also on the angioarchitecture of the kidney (presence of large segmental vessels in the puncture zone) and the state of the surrounding tissues.

The research under the "TissueVolume" programme contributed to the determination of the exact location, size and density of the stones to be removed. This information in combination with traditional diagnostic methods allows one to carefully plan the upcoming surgery, especially with percutaneous nephrolithotripsy. Non-invasive nephrolithiasis research methods include the widely used ultrasound scanning, which is performed with the use of **ALOKA Pro Sound SSD 3500** (Japan) ultrasonic device. To determine the component composition in viva DECT was used. The method of ion chromatography was used to control the effectiveness of metaphylaxis.

It should be noted that data from space satellites are also used to determine the level of air pollution in various locations of the region.

2. Computer shell of the information system

Let us briefly consider the software shell of the information system which includes new algorithms developed for data modeling and analysis. The software was developed in Microsoft Visual Studio with the use of C# programming language. The databases are located on the local MS SQL Server at the Ecological Centre for the Rational Development of Natural Resources (Krasnoyarsk). The programme complex has an internal division into modules.

The main form of the information system is shown in Fig. 2. The form contains a list of all municipalities of the region, information on developer and the authority responsible for compiling and maintaining the eco-passport.

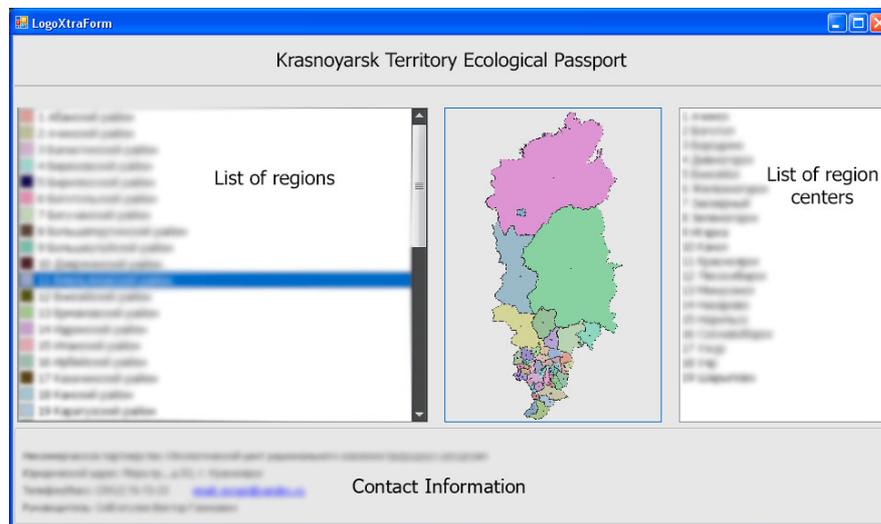


Fig. 2. The main form of the programme complex

The form contains general information on municipal districts, the map of municipal districts with the indication of settlements and three main information blocks: 1 – environmental impact, 2 – the state of environmental components, 3 – environmental and economic indicators. Each block contains several sub-blocks represented as tables in the database. Navigating the database tables, as well as editing the tables (adding, deleting, changing information) is done using the "navigation panel". It is located at the bottom of the table form. The databases of the developed software are designed in such a way that the user can perform an on-line search for the required information using the list filtering. Using table data from the tab "Table", one can build graphs by various digital (not text) columns.

Fig. 3 shows an option of data processing and constructing a graph in the 3D projection. Users can also group together similar content in any of the columns in the table using the context menu. For example, the database "Discharges of Pollutants" in Fig. 4 presents grouped information on the types of wastewater receiver. It should be noted that the components of the environment are automatically assessed on the basis of criteria for assessing the environmental state with the use of the obtained values of the indicators. It allows the comprehensive assessment

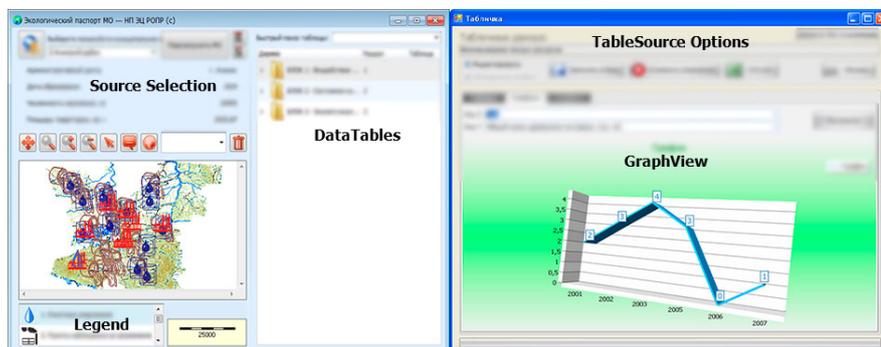


Fig. 3. Data processing (left) and graph construction (right)

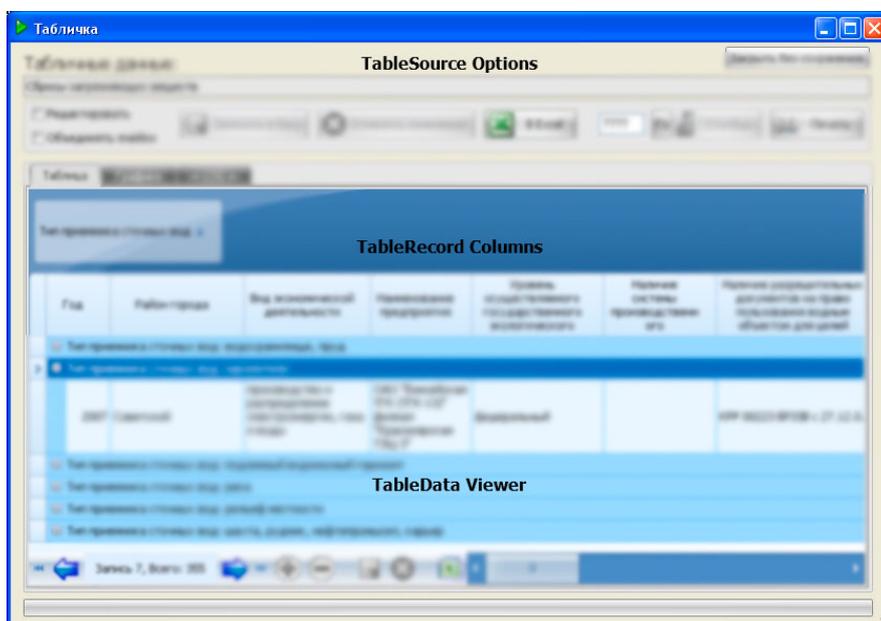


Fig. 4. Grouping of information for solving applied problems

of the environmental situation in any municipal district. It is assumed in the framework of the information system that calculated data are exported to the Excel environment. A library for Excel was developed to construct approximate models, and to export tables. It is also used to perform data analysis within the Matlab computing environment.

3. Models of relationship between morbidity and environmental factors

A method for constructing a predictive model based on approximation of some function of five variables was proposed [8,9]. The following mathematical formalization of the relationship between morbidity and the most influential factors was used

$$Z_i = a_1 T_i + a_2 W_i + a_3 V_i + a_4 \frac{\ln P_i}{\ln V_i} + a_5 \ln C_i,$$

where T_i is the average annual air temperature (in Celcius), W_i is the medical service density (number of doctors per 1,000 people), V_i is the average annual wind velocity (m/s), P_i is the integral indicator of atmosphere pollution (standard units), C_i is the social conditions typical for the city (expert assessment, points).

The unknown coefficients $a_j, j = 1..5$ are calculated by the least square method where the discrepancy between calculated and experimental data to be minimized has the form

$$F(a) = \sum_{i=1}^{12} (Z_i(a) - H_i)^2,$$

where $Z_i(a)$ are estimated data, H_i are actual data on the morbidity of the population.

To minimize the functional $F(a)$ in the volume $D = \{\mathbf{a} \in \mathbb{R}^n \mid \underline{a}_i \leq a_i \leq \overline{a}_i\}$ the PARABOL programme was used [8, 9]. It implements the method of searching for the global extremum of a functional. As a result we obtained the relationship between the morbidity rate of the adult population and considered factors in the form

$$Z = 62.72T + 521.39W + 86.13V + 90.17 \frac{\ln P}{\ln V} - 181.81 \ln C.$$

A comparison between observed morbidity and calculated morbidity is shown in Fig. 5. The obtained results allowed one to forecast the morbidity of various population groups on the basis of expert assessments of social factors.

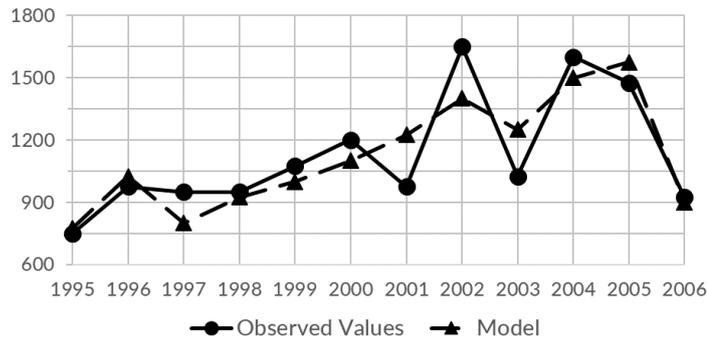


Fig. 5. Comparison between observed morbidity and morbidity obtained on the basis of the proposed model (adults)

An improved methodology for modeling observational data with the use of neural networks was studied. Regression analysis was also performed to assess the relationship between the morbidity and environmental factors on the basis of the observational data analysis [10, 11]. The base function has the form

$$Y_a^t = b_a + c_a \sum_j \sin \left(\varphi_{aj} + \sum_i w_{ji} X_i^t \right),$$

where X are the input dataset (training set), i is the index of input (sample), j is the index of neuron, t is the index of the task in the sample, a is the index of output, $b_k, c_k, w_{pq}, \varphi_{uv}$ are adjustable parameters. Initial data for neural network modeling is shown in Fig. 6. The programme window for setting the neural network parameters is shown in Fig. 7. Fig. 8 presents the results of the comparative analysis of observed data (adolescents and adults) and data obtained with the use of neural network model. Exact match was obtained.

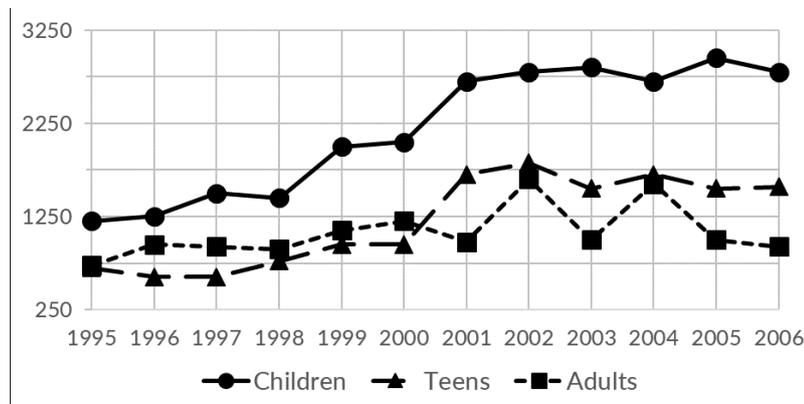


Fig. 6. Initial data for model construction

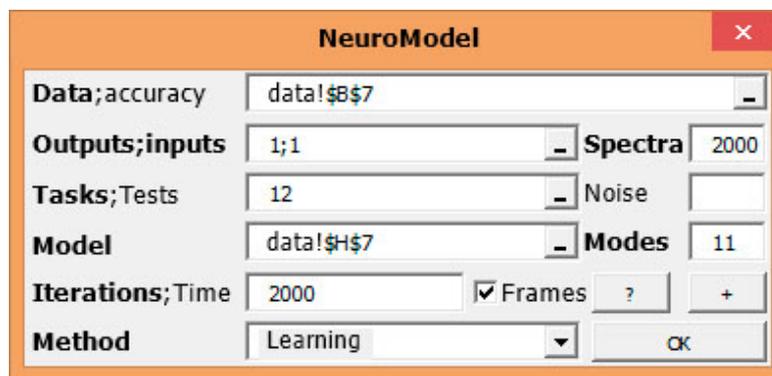


Fig. 7. The programme window for setting the neural network parameters

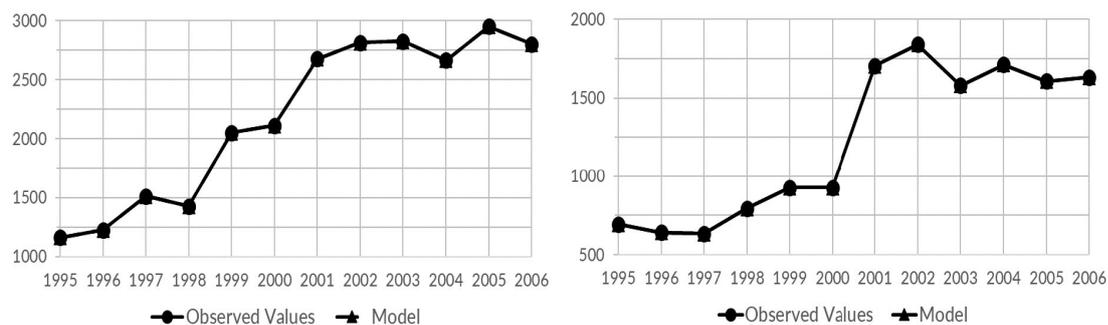


Fig. 8. Comparison between observed data (a – children, b – adolescents) and the results of regression model (exact match)

Numerical experiments were performed to analyze how environmental conditions influence the morbidity of certain population groups. The obtained models reveal the relationship between the morbidity of certain population groups and conditions of the natural and social environment. Additional procedures of the improved methodology allowed us to successfully solve typical problems of environmental monitoring with a given accuracy in the framework of the information system.

4. Choosing the method of treatment of the kidney stone disease with use of neural network

Currently, there is a trend in the medical practice to quantitatively evaluate clinical indicators. Moreover, this applies not only to laboratory data but also to factors that are very subjective in character. Artificial neural networks are used to solve such problems. The main advantage of application of neural networks in medicine is that they are capable to process a large number of parameters which are often impossible for a doctor to identify. This is the main difficulty in diagnosing because much information is the subjective assessment of doctors based on their knowledge and experience [12, 13]. Nevertheless, the impact of these data on the diagnosis or choice of treatment is quite high. Neural networks are able to make decisions based on hidden patterns by filtering large amounts of data [14].

The neural network model is fundamentally different from other mathematical classifiers. Firstly, neurocomputers and neural programmes are self-learning. Secondly, the principles of their work are very similar (in a very simplified form) to the interaction of neurons through synapses [14]. According to the available literature data, neural networks are not used in the treatment of the kidney stone disease. There are not prognostic expert systems that allow choosing the most acceptable method of treatment in each specific case and considering postoperative complications.

The research objective was to develop elements of the expert system for recognition and selection of methods of treatment of the kidney stone disease based on the neural network classifier. The training sample consisted of the data from 510 patients of urological departments of the Krasnoyarsk Interdistrict Clinical Emergency Hospital named after N. S. Karpovich and Road Clinical Hospital at Krasnoyarsk Station of Russian Railways. 352 patients (69%) were men, and 158 patients (31%) were women. 249 patients (48.8%) were hospitalized according to the schedule and 261 patients (51.2%) were hospitalized urgently. The urolithiasis was first determined in 313 (61.4%) patients, recurrent stones were found in 197 (38.6%) patients, in 68 (13.4%) cases the stones were localized in the renal pelvis, in 46 (9.0%) cases the stones were localized in calyces. Stone obstruction of the urethral segment was observed in 16 patients.

The sizes of ureteral stones ranged from 0.4 to 1.2 cm. Stones were localized in the lower third in 174 (34.1%) patients, in the upper third in 132 (25.9%) patients and in the middle third in 90 (17.6%) patients. All patients were examined according to the standard diagnostic procedure. It includes laboratory and X-ray methods. They allow one to choose the optimal treatment that depend on the anatomical and functional state of the upper urinary tract, localization and size of stones and the phase of pyelonephritis. Medical care for patients with the kidney stone disease, algorithms for prediction of nephrolithiasis treatment were realized with the use of the expert neural network system (Fig. 9). Remote shock-wave lithotripsy was performed in 141 (27.6%) cases, conservative therapy was performed in 125 (24.5%) cases. Endourological treatment methods (ureteroscopy, percutaneous nephrolitholapaxy) were used in 110 (21.6%) and 81 (15.9%) cases, respectively. The remaining stones were removed during traditional upfront surgery in 53 (10.5%) cases.

Information about each patient is presented in the form of a vector (Fig. 10). It has 28 components: data from clinical examination, instrumental and laboratory research. Each case has the "class" parameter. It represents the specified method of treatment (0 is the conservative treatment, 1 is the CLT ureteroscopy, 2 is DLT, 3 is the percutaneous nephrolitholapaxy, 4 is the pyelolithotomy, and 5 is the ureterolithotomy). The number of cases of each class consisted of all

patients treated in these urological departments.

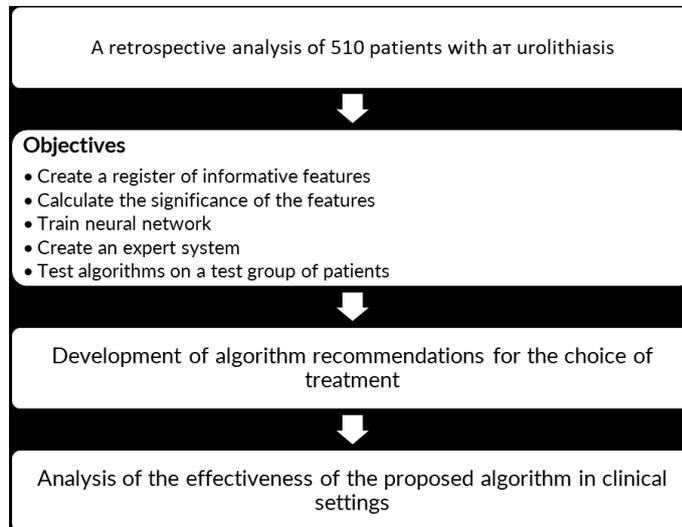


Fig. 9. Flow chart of the expert system

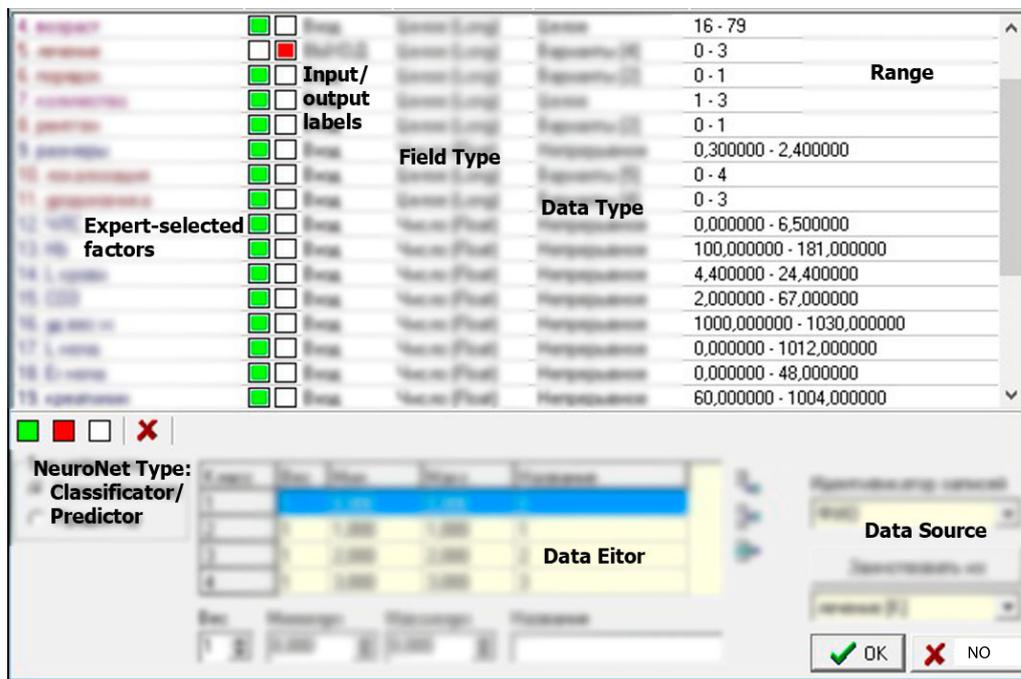


Fig. 10. Process of the neural network training

We used a three-layer neural network in the Panalyzer 5.0 environment. It is the most optimal neural network to solve the problem of producing experience from scientific data in the humanities. Training of the neural network consists of finding patterns and many parameters by comparison with the standards. At the output we get a certain conclusion, which is made on

the basis of the experience gained. Network training for 28 input vectors and 510 records took about 65 minutes.

The next important task was to study the significance of the input parameters for the diagnosis. The trained neural network automatically deleted the least significant input parameters. As a result, we identified 13 parameters that had the greatest influence on the choice of the method of treatment (Fig. 11).

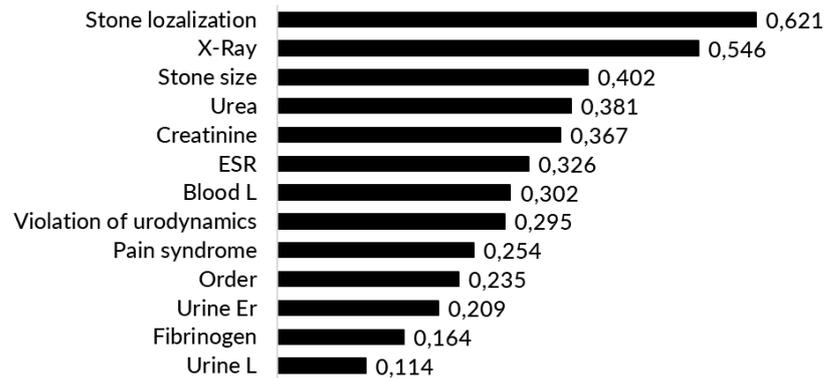


Fig. 11. Average significance of input parameters

The most significant indicators ($p < 0.05$) were the localization and size of calculus, the visibility in the images, the indicators of erythrocyte sedimentation rate, urea and creatinine, and the degree of violation of urodynamics. A distinctive feature of the **Panalyzer 5.0** programme is the ability to create a separate programme (neural network expert system) that allows one to test the data. Testing of the expert system was carried out on 22 examples. In this case, the neural network has identified all the proposed examples correctly. Moreover, the expert system "rejected" two examples that were introduced into the test sample by mistake. In a retrospective analysis of case histories of these patients, it turned out that in both cases the stones of the lower third of the ureter went out independently a day before the ureteroscopy. During surgery the stones in the urinary tract were not detected.

Conclusion

In the course of experimental studies the environmental monitoring data were analyzed. They are presented in the database of the information system that contains specialized information about all municipalities and cities of the Krasnoyarsk Krai as the key region of the Yenisei Siberia. The problem of determining a regression between the morbidity and environmental factors based on the analysis of observational data was solved.

A neural network method for the choice of treatment of urolithiasis (kidney stone disease) is proposed. Training of the artificial three-layer neural network of direct propagation was based on 510 cases and a multidimensional vector characterized by 28 input parameters. The training of the neural network consists in searching regular trends in the set of parameters through comparison with the reference results. Testing of the expert system was carried out on 22 examples. In this case, the neural network has identified all the proposed examples correctly. An expert system was developed that classifies methods of treatment (6 types) of the kidney stone

disease with a degree of confidence of 91 %. The expert system can be used in the regions of the Yenisei Siberia adjacent to the Krasnoyarsk Krai.

The study was supported by Krasnoyarsk Region Science and Technology Support Fund and by Krasnoyarsk Regional Fund of Support Scientific and Technical Activities

References

- [1] A.G.Berezhnoi, A.V.Ershov, S.Yu.Krasnoperov, Analysis of the results of treatment of nephrolithiasis and ureterolithiasis at the modern stage, *Ural Med. J.*, (2017), no. 3, 90–94 (in Russian).
- [2] L.Cadena, F.Cadena, A.S.Kruglyakov, K.V.Simonov, F.P.Kapsargin, Diagnostics of Complex Phenomena on the Basis of Geometrial Analysis Images, Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, 2017, 401–404.
- [3] N.V.Efimova, V.S.Rukavishnikov, P.K.Kaurov, A.N.Perezhogin, Z.A.Zaikova, A.Yu.Gornov, T.S.Zarodnyuk, Environmental factors: experience of complex assessment, Irkutsk Research Centre of Surgery and Traumatology, Irkutsk, 2010 (in Russian).
- [4] A.V.Ershov, F.P.Kapsargin, K.V.Simonov, Neural network assessment of the treatment method for the kidney stone disease, Neural computer science, its application and data analysis: Materials of the XXVI All-Russian seminar (28–30 of September 2018). Krasnoyarsk, ICM SB RAS, 2018, 63–67 (in Russian).
- [5] A.N.Gorban, Neural networks in the personal computer, Novosibirsk, Nauka, 1996 (in Russian).
- [6] F.P.Kapsargin, L.Cadena, S.V.Kirillova, K.V.Simonov, Mathematical models in the problem of human ecology (review), Proceedings of the tenth international conference on financial-actuarial mathematics and event-convergence technologies, Krasnoyarsk, SibFU, 2011, 171–173 (in Russian).
- [7] F.P.Kapsargin, L.Cadena, S.V.Kirillova, K.V.Simonov, Efficient computational technology for constructing mathematical models from observational data, Proceedings of the 10th Regional Scientific Practical Conference of Urologists of Siberia. Current issues of diagnosis and treatment of urological diseases, Barnaul, Altai State Medical University, 2011, 33–34 (in Russian).
- [8] F.P.Kapsargin, K.V.Simonov, L.Cadena, D.V.Zuev, A.V.Ershov, Diagnostic of complicated phenomena on the basis of geometric image analysis, *Vestnik SibSAU*, **46**(2012), no. 6, 77–82 (in Russian).
- [9] V.V.Kruglov, Artificial neural networks. Theory and practice, Moscow, Goriachaia liniia, Telekom, 2001 (in Russian).
- [10] A.A.Solomakha, Modern methods of forecast in medicine, *Materials of the 8th All-Russian seminar "Neural computer science and its applications"*, Krasnoyarsk, 2000, 162–163 (in Russian).

- [11] T.S.Zarodnyuk, A.Yu.Gornov, Global extremum search technology in optimal control problem, *Modern technology. System analysis. Modelling, Irkutsk*, **19**(2008), no. 3, 70–76 (in Russian).
- [12] S.I.Zhestovskaya, F.P.Kapsargin, L.F.Zueva, E.A.Alekseeva, Determination of calculus composition for early postoperative metaphylaxis, Current issues of urology. Collection of scientific papers of the V Congress of Urology of Siberia with international participation, Krasnoyarsk, KASS, 2016, 287–291 (in Russian).
- [13] A.G.Zotin, K.V.Simonov, F.P.Kapsargin, T.Cherepanova, A.S.Kruglyakov, L.Cadena, Techniques for Medical Images Processing Using Shearlet Transform and Color Coding, Favorskaya M., Jain L. (eds). *Computer Vision in Control Systems-4*, **136**(2018), 223–259.
- [14] L.F.Zueva, S.I.Zhestovskaya, F.P.Kapsargin, K.V.Simonov, Early postoperative metaphylaxis based on the calculus component composition, Proceedings of the XVI Congress of the Russian Society of Urologists, Ufa, 2016, 185–186 (in Russian).

Моделирование данных при решении трудноформализуемых социально-экономических задач

Константин В. Симонов

Институт вычислительного моделирования СО РАН
Академгородок, 50/44, Красноярск, 660036, Россия

Михил А. Курако

Институт космических и информационных технологий
Сибирский федеральный университет
Киренского, 26, Красноярск, 660074, Россия

Алексей А. Кабанов

СКТБ «Наука» КНИЦ СО РАН
пр. Мира, 53, Красноярск, 660049, Россия

Федор П. Капсаргин

Любовь Ф. Зуева

Артем В. Ершов

Красноярский государственный медицинский университет
Партизана Железняка, 1, Красноярск, 660022, Россия

Светлана Н. Гриб

Институт экономики, управления и природопользования
Сибирский федеральный университет
Свободный, 79, Красноярск, 660041, Россия

Исследование посвящено разработке алгоритмического обеспечения для моделирования данных при решении трудно формализуемых социально значимых задач, связанных с здравоохранением в разрезе взаимосвязей и взаимодействий территорий Енисейской Сибири, где Красноярский край играет роль ключевого региона. В работе предлагаются элементы информационной системы для анализа текущего состояния и оценка сценариев будущего взаимодействия территорий Енисейской экономической зоны в решении проблем в этой актуальной предметной области. Предлагаются решения в формате ГИС-технологий и современные средства моделирования разнородных данных наблюдений.

Ключевые слова: алгоритмы анализа данных, базы данных, ГИС, моделирование данных, нейросети.