Biology and Medicine
ISSN: 0974-8369

Zakharov et al., Biol Med (Aligarh) 2016, 8:6
DOI: 10.4172/0974-8369.1000329

Research Article Open Access

Scientific Data Analysis Using Neural Networks as Exemplified in Defining the Factors Impacting the C-reactive Protein Level

Alexander A. Zakharov*, Evgeni A. Olennikov, Tatyana I. Payusova, Tatyana I. Petelina, Natalia A. Musikhina and Ludmila I. Gapon

10 Republic Str., 625003 Tyumen, Russian Federation

***Corresponding author:** Zakharov AA, 10 Republic Str., 625003 Tyumen, Russian Federation; E-mail: aazaharov@yandex.ru

Received: May 5, 2016; **Accepted:** July 20, 2016; **Published:** Aug 27, 2016

Abstract

The article describes the approaches, involving the usage of medical information systems as a tool for clinical and biological studies. The example describes the technology and technique to automate the medical and clinical studies, conducted to define the factors, impacting the level of inflammatory response marker—C-reactive protein level for the patients with ischemic heart disease. The toolset includes the artificial neural network to analyze the overall situation based on the analysis of various parameters and their interdependencies. Technology and technique are implemented as software module of medical information system of the branch of Institute of Cardiology "Tyumen Cardiology Center".

Keywords: Artificial neural networks; Scientific data analysis; C-reactive protein level

Introduction

During the recent years, there are significant qualitative changes in the medical automation in the Russian Federation. These changes are related to the shift from separate local software products for specific purposes (patient accounting, medical equipment data processing, etc.) to medical information systems (MIS), which cover all these functions as a whole [1].

Though there are a lot of MIS in the market, they basically support three major objectives of information support for medicine:

- accounting and planning of services and resources (doctors' schedule, medicines, medical services etc.)
- electronic maintenance of a medical case, including the functionality of the advising systems [2-4];
- reports generation for insurance companies, Federal Compulsory Medical Insurance Fund, Medical Information and Analytical Center (MIAC),etc.

Thus MIS act as accounting systems, relieving the medical specialist from routine tasks.

Materials and Methods

It is important to note that one of the results of MIS implementation is the accumulation of digital medical data about a patient, ready for automated processing and for scientific studies [5,6]. It provides a technological possibility for qualitative MIS shift from accounting system to scientific medical information system (SMIS). SMIS preserves all MIS functions and is used by doctor-researcher, providing the functionality to gather and analyze the required data. Usually the required data are exported from MIS database via SQL queries and then are processed in statistical modules. This approach doesn't require MIS modification but has its disadvantages: researcher should know the database structure and have SQL skills. "Manual" work also implies the data formatting for specific application.

Alternative approach is to analyze data directly in MIS. Following tasks should be done to make it possible.

1. Data for study should be described in meta-tables in advance. Thus a doctor will be able to define the required data in own terms without knowing the database structure;
2. Develop "scientific" software modules for data analysis with standard statistical and/or neural network methods;
3. Develop the toolset to define the query criteria and module (or set of modules) for subsequent query processing and data export.

This approach doesn't require the modification of existing software modules in MIS, but requires having the new modules to be developed and included in MIS. It should be noted that the doctors, experienced in mathematical statistics modules, can continue using them. But in the meantime they receive new possibility to analyze data directly in MIS using well-known statistical methods and neural networks (NN), which are currently very promising for medical data analysis. An artificial neural network (ANN) refers to a mathematical model inspired by biological neural networks [7].

We analyzed the data set based on the Tatyana Petelina's, Natalia Musikhina's, Ludmila Gapon's scientific research. This research has been approved by the branch of Institute of Cardiology "Tyumen Cardiology Center." We obtained written informed consent from all participants in the study.

Some description of data set is represented in Tables 1 and 2.

In this study we use NN models. We can present following reasons to use NN for the analysis of medical and biological data [8,9]:

1. NNs are trainable. During its education NN can discover complicated dependencies between input and output parameters even when statistical methods don't show that;
2. You can use NN during process formalization to solve the issue that is hard or impossible—when you can't consider all the conditions impacting the solution;
3. In case of missing a priori knowledge about the data or in case of large number of parameters, NNs usually provide the faster and more efficient solution which is usually not worse than the same, provided by statistical methods after a thorough analysis of the data structure;

Mean age (full years)	Mean smoking experience (full years)	Mean ischemic heart disease experience (full years)	Mean diabetes experience (full years)	Mean systolic pressure (mm or mercury)	Mean diastolic pressure (mm or mercury)	Mean very little density lipoprotein (mol/l)	Mean glycated hemoglobin (%)	Mean calculation ratio (Units)	Mean atherogenic index (Units)	Mean glomerular filtration rate (kg/ml/min 31.73 m^2 as per MDRD)	Mean ligand CD 40 l (ng/ml)
61	4.49	8.35	2.03	144.03	89.19	0.78	5.92	28.25	0.57	90.10	3,45

Table 1: The description of data set (patients with stable angina)

Mean age (full years)	Mean smoking experience (full years)	Mean ischemic heart disease experience (full years)	Mean diabetes experience (full years)	Mean systolic pressure (mm or mercury)	Mean diastolic pressure (mm or mercury)	Mean very little density lipoprotein (mol/l)	Mean glycated hemoglobin (%)	Mean calculation ratio (Units)	Mean atherogenic index (Units)	Mean glomerular filtration rate (kg/ml/min 31.73 m^2 as per MDRD)	Mean ligand CD 40 l (ng/ml)
60	5.29	7.32	1.74	140.24	85.85	0.77	5.62	26.39	0.58	84.06	3.51

Table 2: The description of data set (patients with unstable angina)

4. Significant accumulated experience in development and implementation of the artificial neural networks makes it possible to get the skills of NN operation and start using it in practice after a relatively short period of time.

Artificial neural networks (ANNs) are processors that are trained to perform a particular task, such as a classification, prediction, or control task. ANNs are designed to perform work analogous to functions typically implemented within the cerebral cortex [10].

ANNs consist of an interconnected group of artificial neurons, and they process information using a connectionist approach to computation [7]. ANNs employ nonlinear mathematical models to mimic the human brain's own problem-solving process, by using previously solved examples to build a system of "neurons" that makes new decisions, classifications, and forecasts [11]. ANN is a complex and flexible nonlinear system with exclusive properties consisting of robust performance in dealing with noisy or incomplete input patterns, high fault tolerance, and the ability to make generalizations on the basis of the input data computation [7]. ANN is often applied to model complex relationships between inputs and outputs or to find patterns in data. In clinical medicine, ANN models have been applied in the diagnosis of diseases such as myocardial infarction [12]. ANN models have also been successfully used to predict trauma mortality and in clinical decision-making in the management of traumatic brain injury patients [13,14]. A previous study compared the LR and ANN models used in the prediction of living setting after hip fracture [15]. Study using artificial neural network was undertaken to develop mortality prediction models in children admitted to pediatric intensive care unit [16].

Results

In this study NN is used to analyze the factors with nonlinear impact on the level of inflammatory response marker—C-reactive protein (CRP) [17] level for the patients with ischemic heart disease (IHD) and with various manifestation of the disease—stable and unstable angina. List of factors are represented in Table 3.

MIS of the Tyumen Cardiology Center and some included software modules with implemented neural networks of various architecture was used as a toolset for the study.

Algorithm for the development of the neural network to define (assess) the factors impacting the inflammatory response

Factor	Measure
Patient's gender;	Male, female;
Stent model;	Cypher (Cordis), Taxus Liberte Monorail, Endeavor, Palmaz medium, Xience V, Promus;
Total cholesterol;	mmol/l;
Triglycerides level;	mol/l;
Very low density lipoprotein;	mol/l;
Lipoprotein (a);	mg/dl;
Apolipoprotein A-I;	mg/dl;
Apolipoprotein B;	mg/dl;
Atherogenic index;	Units;
Interleukin-6;	pg/ml;
Interleukin-8;	pg/ml;
CD40 receptor;	ng/ml;
Ligand CD40-L;	ng/ml;
Calculation ratio;	Units;
Tissue inhibitor of matrix metalloproteinases;	ng/ml;
Glycated hemoglobin;	%;
Patient age;	full years;
Ischemic heart disease experience;	full years;
Diabetes experience;	full years;
Smoking experience;	full years;
Patient's body mass index;	kg/m^2;
Creatinine;	mcmol/l;
Glomerular filtration rate;	kg/ml/min 31.73 m^2 as per MDRD;
Microalbuminuria;	mg/l;
Systolic pressure;	mm or mercury;
Diastolic pressure;	mm or mercury.;
Hereditary tainted;	not tainted, tainted by arterial hypertension (AH) or IHD, tainted by AH and IHD;
Type 2 diabetes;	Yes/no
Smoker/nonsmoker;	Yes/no
Diagnosis;	IHD without AH, IHD with AH;
Obesity;	no obesity, stage 1, stage 2, stage 3;
Disaggregant intake on prehospital stage;	Yes/no
Statins intake on prehospital stage.	Yes/no

Table 3: List of factors

marker—C-reactive protein level for the patients with ischemic heart disease using neural network is presented below:

STEP 1. Data preprocessing, including the definition of ordinal and categorical variables, coding and normalizing.

STEP 2. Development of several neural networks with various architecture. The number of hidden layer nodes may affect the performance of the ANN classifier [18].

STEP 3. Teaching the resultant neural networks using various education techniques.

STEP 4. Definition of the most efficient neural network.

STEP 5. Sensitivity analysis of the variables to define predictor impact on final value.

Software implementation of the described algorithm helps to define the CRP level for the patients with IHD of various manifestations. Neural network with MLP 81-56-2 architecture (multilayer perceptron) has shown the maximum efficiency for the first group of patients (with stable angina): teaching efficiency 98.25; testing efficiency 60. Cross entropy was used as error function and logistic function was chosen as transfer function. Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was used to teach the neural network.

Neural network with MLP 74-51-2 architecture (multilayer perceptron) has shown the maximum efficiency for the second group of patients (with unstable angina): teaching efficiency 100; testing efficiency 50. Cross entropy was used as error function and logistic function was chosen as transfer function. BFGS algorithm was used to teach the neural network. The architecture of artificial neural network is considered on Figure 1.

Sensitivity analysis has shown that CRP level for both patient groups (with stable or unstable angina) is impacted by the same factor

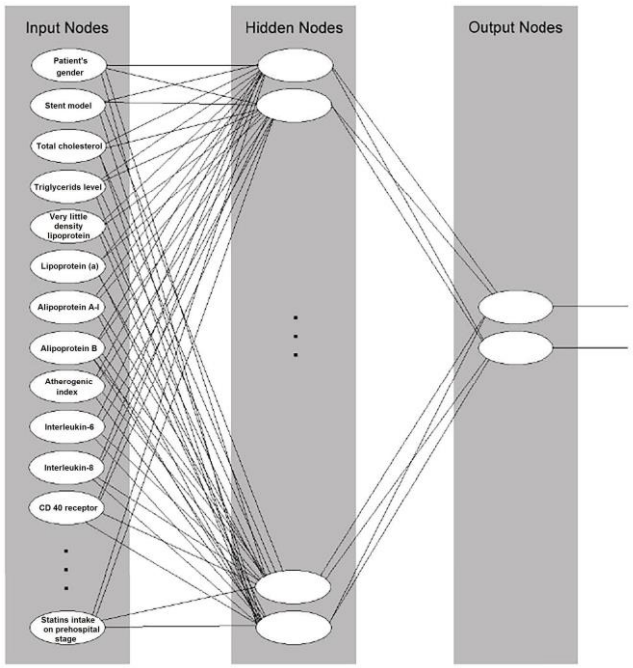

Figure 1: The architecture of artificial neural network

Factor	Patients with stable angina (Units)	Patients with unstable angina (Units)
Patient's gender	7.33	32.15
Stent model	7.73	37.18
Smoking experience	11.95	5.13
Disaggregant intake on prehospital stage	9.49	5.81
IHD experience	5.99	6.02
Diabetes experience	3.47	26.55
Very low density lipoprotein	3.29	3.26
Hereditary background	2.83	17.48

Table 4: Results of the sensitivity analysis, conducted via artificial neural networks

to the different extent. Analysis of the implemented neural networks has shown the impact of these factors on CRP level for the patients with stable or unstable angina. Analysis results are summarized in Table 4.

Discussion

Based on derived results, it was concluded that the comparative analysis of inflammatory response marker—C-reactive protein level for the patients with ischemic heart disease with originally stable and unstable angina can define the dependencies and cause-effect relations between IHD factors.

Comprehensive meta-analysis of the results and CRP studies [18-21] has shown that this marker has the same or even higher impact on vessel wall than blood pressure or cholesterol level. Constantly increased CRP is a serious risk factor for recurrent vessel damage.

CRP level can vary significantly among the patients with stable IHD [22].

CRP increase for the patients with high total cholesterol and low-density lipoprotein cholesterol (LDL-cholesterol) rapidly increases the risk of myocardial infarction. Such patients need treatment with surgery more often; acute myocardial infarction is also more frequent for such patients than for the same patients group with unstable angina and lower CRP level.

Thus the algorithm to analyze the factors impacting the CRP level for the IHD patients via the neural networks helps to define the hidden nonlinear dependencies and discover the predictor impact on final value. Algorithm is implemented in the module of the branch of Institute of Cardiology "Tyumen Cardiology Center" medical information system.

Conclusion

The results show that dynamic control over the provided laboratory markers for the patients from both groups will have high forecast value to develop and implement the measures for the timely correction of medicine treatment and thus for decreasing the cardiovascular complications risk.

References

1. Jel'janov, M.M. Medicinskie informacionnye tehnologii. Katalog. Vyp. 13 [Medical information technology. Catalogue. 13]. Moscow, 2013. 300 p. (in Russian). Medical Information Technology. Capital Press.
2. Zakharov A, Olennikov E, Petuhov A (2007) Information model of electronic patient records. TSU Herald 5: 97-101.

3. Petuhov A, Olennikov E, Zakharov A (2008) Models and methods for multi-level output component advising subsystem consisting of electronic health record. Bull Orel State Tech Univ 2008: 153-158.

4. Zakharov A, Olennikov E, Petuhov A (2008) Mathematical methods of evaluation of research results in the diagnostic subsystem consisting of electronic health record. TSU Herald 6: 145-152.

5. Zakharov A, Nesterova O, Olennikov E (2009) Problems of information retrieval for research in medical information systems. TSU Herald 6: 215-219.

6. Zakharov A, Nesterova O, Olennikov E (2010) Information retrieval algorithm in medical archives based on context-time ontology. TSU Herald 6: 177-182.

7. Tang Z-H, Liu J, Zeng F, Li Z, Yu X, *et al*. (2013) Comparison of prediction model for cardiovascular autonomic dysfunction using artificial neural network and logistic regression analysis. PLoS ONE 8: e70571.

8. Evdokimenkov, V.N. Komp'juternye tehnologii sbora, obrabotki i analiza dannyh mediko-biologicheskih issledovanij: Uchebnoe posobie [Computer technologies for collecting, processing and analyzing data for biomedical research: schoolbook]. Moscow, 2005. 436 p. (in Russian). Computer Technologies for Collecting, Processing and Analyzing Data for Biomedical Research: Schoolbook. Moscow.

9. Eghov A, Chechyotkin V (1997) Neural networks in medicine. Open Systems. 4: 34-37.

10. Schuman CD, Birdwell JD (2013) Dynamic artificial neural networks with affective systems. PLoS ONE 8: e80455.

11. Terrin N, Schmid CH, Griffith JL, D'Agostino RB, Selker HP (2003) External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. J Clin Epidemiol 56: 721-729.

12. Baxt WG (1991) Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 115: 843-848.

13. Harrison RF, Kennedy RL (2005) Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. Ann Emerg Med 46: 431-439.

14. Ding W, Zhou L, Bao Y, Yang Y, Lu B, *et al*. (2011) Autonomic nervous function and baroreflex sensitivity in hypertensive diabetic patients. Acta Cardiol 66: 465-470.

15. Ottenbacher KJ, Linn RT, Smith PM, Illig SB, Mancuso M, *et al*. (2004) Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture. Ann Epidemiol 14: 551-559.

16. Chan CH, Chan EY, Ng DK, Chow PY, Kwok KL (2006) Application of artificial neural networks to establish a predictive mortality risk model in children admitted to a paediatric intensive care unit. Singapore Med J 47: 928-934.

17. Gusev D, Ponomar E (2006) The role of C-reactive protein and other markers of inflammation in the acute phase of atherosclerosis. Clin Med 5: 25-30.

18. Bagher-Ebadian H, Jafari-Khouzani K, Mitsias PD, Lu M, Soltanian-Zadeh H, *et al*. (2011) Predicting final extent of ischemic infarction using artificial neural network analysis of multi-parametric MRI in patients with stroke. PLoS One 6: e22626.

19. Shishkin V, Polyakov A (2006) C-reactive protein as a prognostic factor in patients with coronary heart disease. Ukrainian J Cardiol 1: 14-17.

20. Shalnev V (2006) Markers of inflammation in the pathogenesis of coronary heart disease. The role of C-reactive protein. Ambulance 1: 54-61.

21. Kim KH (2016) Prognostic factors for extended cecectomy in complicated appendicitis. Biol Med (Aligarh) 8: 265.

22. Madu AJ, Abuknesha N, Ghebremeskel K (2015) The role of lipids in inflammation: review of the evolving pathogenesis of sickle cell disease. Biol Med (Aligarh) 7: 244.