UDK - 004.414.28

K. S. Cherneha, B. I. Tymchenko,
N. O. Komleva, PhD.

DECISION SUPPORT SYSTEM FOR AUTOMATED MEDICAL DIAGNOSTICS

Abstract. *This article is dedicated to the decision support, which has become one of the most promising and rapidly developing areas of application of modern intellectual and information technologies in various areas of professional activity, including problems in the automated medical diagnostics.*

Modern level of artificial intelligence development allows us to develop programs that help to analyze all the data collected on the patient for their reasonable interpretation. The basic idea of medical decision support system is to build logic diagnosis setting process, correct from the medical point of view, and transparent for doctor.

For diagnosing automation, the state of bronchopulmonary system DSS DiaSpectrEx was developed. The analysis of changes in qualitative and quantitative composition of air exhaled by the patient is the source of information about the damage of the respiratory tract, inflammatory processes and the effectiveness of the treatment.

DiaSpectrEx system allows improving efficiency of pulmonary diseases diagnosing methods through the use of modern computerized equipment, machine learning algorithms and data processing. Development of high-precision medical decision support system enables non-invasive computerized pulmonary diagnostics.

Keywords: *support system, medical decision, neural networks, machine learning, classification, software, respiratory system*

К. С. Чернега, Б. И.Тимченко,
Н. О. Комлевая, канд. техн. наук

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ АВТОМАТИЗИРОВАННОЙ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

Аннотация. *Представлен подход к диагностике заболеваний легких с помощью конденсата влаги выдыхаемого воздуха. Предложен новый подход, который решает проблему автоматизированных интеллектуальных диагностик с использованием методов машинного обучения. Наш метод работает в режиме реального времени, и достигает точности около 95%.*

Ключевые слова: *система поддержки; медицинское решение; нейронные сети; машинное обучение; классификация; программное обеспечение; дыхательная система*

К. С. Чернега, Б. І. Тимченко,
Н. О. Комлева, канд. техн. наук

СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ АВТОМАТИЗОВАНОЇ МЕДИЧНОЇ ДІАГНОСТИКИ

Анотація. *Представлений підхід до діагностики захворювань легенів за допомогою конденсату вологи повітря, що видихається. Запропоновано новий підхід, який вирішує проблему автоматизованих інтелектуальних діагностик з використанням методів машинного навчання. Наш метод працює в режимі реального часу, і досягає точності близько 95%.*

Ключові слова: *система підтримки; медичне рішення; нейронні мережі; машинне навчання; класифікація; програмне забезпечення; дихальна система*

Introduction

Problems with managing the flow of information lead to difficulties in monitoring the functioning of the medical institution and without the ability to effectively analyze key performance indicators is difficult to efficiently and quickly make the right managerial and medical decisions.

To solve the problems of operational and strategic management in difficult conditions used decision support system (DSS) to ensure complete and objective analysis of objective activity [1]. Support for decision-making in different areas of professional activity has become one of the most promising and rapidly developing areas of application of modern Information Technologies and artificial intelligence.

Different methods are used in the DSS to analyze the proposals and developments. These may include information search, data mining, knowledge search in databases, reasoning based on precedents, simulation, evolutionary computation and genetic algorithms, neural networks, situational analysis, cognitive modeling, etc. Limitation of time resources, the lack of ability to attract a large number of competent experts, incomplete information about the patient's condition, the high dynamics of the disease, the variability of diseases and the emergence of new ones very often characterize taking the medical solutions.

The aim of this work is to improve the methods of diagnosing pulmonary diseases using laser-correlation analysis of the air exhaled by the patient. For this purpose the following tasks are solved: the analysis of the design features of the automated systems of medicine; formation requirements for decision support system (DSS) for highly accurate pulmonary diagnosis based on the spectral composition of exhaled air; design and implementation of DSS software using ensemble classifier build of neural networks. Therefore, the development and improvement of information technology problem of computer decision support in medicine becomes relevant [2].

Design features of medical automated systems

Medical, technical and technological aspects can be separated from the basic aspects of medical activity information. The medical aspect is the proper preparation of medical data and knowledge (formalization, unity of terminology, standardization), creating an interface common information base structure, the construction of mathematical models of medical and biological processes (physiological and pathological), etc. [4]. Development of theoretical models of data representation and knowledge to solve relevant medical problems and the specific hardware and software implementation of an information base based on developed models compose the technical aspect of the problem. The technological aspect is the coordination of the technical system built with the technological scheme of the diagnostic and treatment process (figuratively speaking, a "recipe for the implementation" of the system in the diagnostic and treatment process).

Problem of functionality and fitness of information systems and decision support system is relevant, it is expressed in the fact that the system is more functional, and then it is more complex and therefore less suitable for practical use. To alleviate the problem, go to the development and implementation of information systems and decision support system to system analysis positions, one of which phases is to create a model that takes into account the organization structure that planned introduction (medical staff, patients, technicians, office management system, laboratory and technical tools, etc.) and a full set of necessary functions (record keeping, diagnostics, providing background information for physicians, providing medical support decision-making, effectiveness analysis of treatment) [8].

As a result, a set of interconnected modules should be developed, allow you to manage all the activities of medical institutions of any type of medical, administrative, laboratory and diagnostic, financial, warehouse, flow control patients. The modular delivery system software will enable a comprehensive automation and control of individual business processes. Thanks to its modular structure, the system in each case will be configured to the specific needs of the institution and therefore will not require business process reengineering.

Decision support systems in medicine

The present level of artificial intelligence development allows us to develop programs that help to analyze all the data collected on the patient for their reasonable interpretation. This solves the problem of redundancy of information, difficulties of its coherent analysis and mutual influences symptoms and diagnoses at each other. The basic idea of medical decision support system is to build logic diagnosis setting process, correct from the point of view of medicine, and transparent to the physician [10].

The basic requirements for medical DSS are:

• definition of basic and associated diseases;

• contradictions elimination in the patient's symptoms and diagnoses;

• situations detection where all possible diagnoses have approximately equal coefficients of confidence;

• identification of the summation effect - when a few “harmless” diagnoses, existing at the same time the patient can result in life threatening or cause complications;

• providing high speed – the time to enter information about the patient, the wording of the request to the DSS and the formation of the diagnosis is strictly limited in the clinical setting.

An analysis of the causes has been identified classifications of the causes of diagnostic errors, which can successfully resist by using the DSS.

Objective causes of diagnostic errors are:

• insufficient level of science development;

• low level of medical institutions equipment;

• a large amount of practical load doctor;

• atypical course or rare diseases;

• low level of medical staff training;

• laboratory error and instrumental investigations.

Among the subjective reasons, diagnostic errors are the following:

• low qualification of doctor, general methodological unpreparedness;

• excessive faith in the laboratory and instrumental data (their underestimation or overestimation);

• poor organization of the doctor work;

• consultative defects;

• logical error diagnostics.

Decision support system for precision pulmonary diagnostics using artificial neural networks.

For the purpose of automation of the bronchopulmonary system state diagnosing developed by the authors DSS DiaSpectrEx. This information source is about the damage of the respiratory tract, inflammatory processes and the treatment effectiveness is the analysis of changes in qualitative and quantitative composition of exhaled air by the patient.

DiaSpectrEx system allows improving the diagnosing methods efficiency of pulmonary diseases through the use of modern computerized equipment, machine learning algorithms and data processing. Development of high-precision medical decision support system enables non-invasive computerized pulmonary diagnostics.

This medical system ensures the registration and authorization of users. The system provides the ability to make major personal, medical, and other patient data. After examining the spectral composition of the condensate of moisture in exhaled air (CMiEA) patient records system CMiEA data surveyed.

There are no systems that use CMiEA data to diagnose pulmonary diseases in Ukraine.

The program can generate the patient's medical card, with following modes: the creation of a new card with a unique number, view the existing map, complement the already established data map, edit, print, save to the database, search the map database according to different criteria – simple and composite. The system does not distribute personal information about the patient, all data is encrypted [3].

The program performs the processing and analysis of patient – CMiEA data integrity, provides further spectral analysis CMiEA data, checks spectra for compliance with the normal distribution law, calculates the values of diagnostic features in accordance with the laid down procedures and algorithms. State of the patient's respiratory system is identified by the calculated values of diagnostic features. Doctor starts to work in case of impossibility of automatic identification; a doctor relates the patient with a diagnostic of existing group or creates a new group for the patient. Each group should have its own description, is stored in the corresponding database tables.

Since the system can work DiaSpectrEx has 4 categories of users: doctor, consultant, administrator and operator. Each user has its limitations in the use of the system.

The main function of the doctor's diagnosis is followed by treatment regimens. If necessary, the patient's diagnostic data can be used to refine the description of a suitable medical diagnostic group.

Consultants may be involved according to the decision of the doctor in the process of diagnosis and development of treatment regimens. It involves the removal of the work of one or more consultants with the remote service. With the participation of consultants in the course of a few of their work should be carried out independently of each other, each consultant does not see the results of the other, thus achieving independence of the assessment. The main functions of the remote control of the consultants are the correct definition of diagnostic classes and verification of data on the reliability and consistency. In the case of inaccuracy or inconsistency of data consultant requires a re-examination.

The last word in the diagnosis and forming a treatment regimen given to the doctor who looks at all provided by consultants and DSS results, diagnoses and prescribe treatment (Fig. 1).

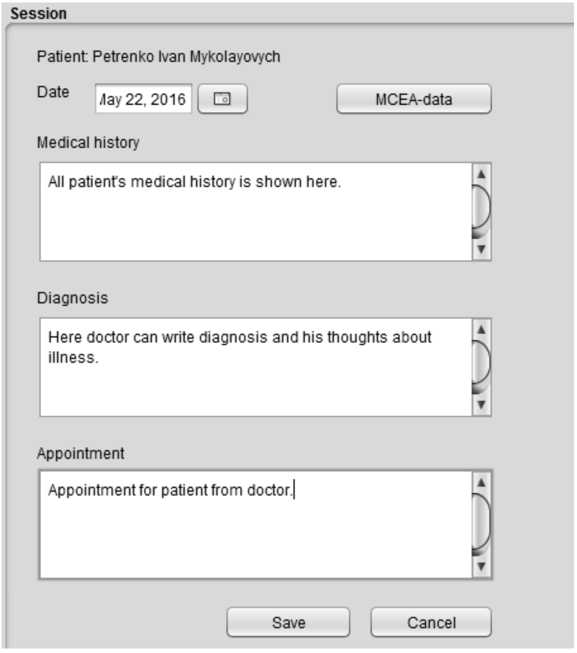

Fig. 1. Fields for doctor's writing

The functions of the operator applies to record the patient data, the medical card formation, perform calculation of diagnostic features, as well as the calculation and verification of spectra normality CMiEA patient data. All health workers can change the patient data, but update the medical record may only doctor.

The main function of the administrator is to register new health workers and update data on the diagnostic groups.

In order to improve patient care provided for the convenience of the possibility of remote work with them. This includes the sampling procedure at home and transporting them to the medical center for further processing. Also, the patient can be sent its diagnostic conclusion; this patient uses contact details (mobile number, e-mail).

In the Fig. 2 there is functional scheme of diagnostics.

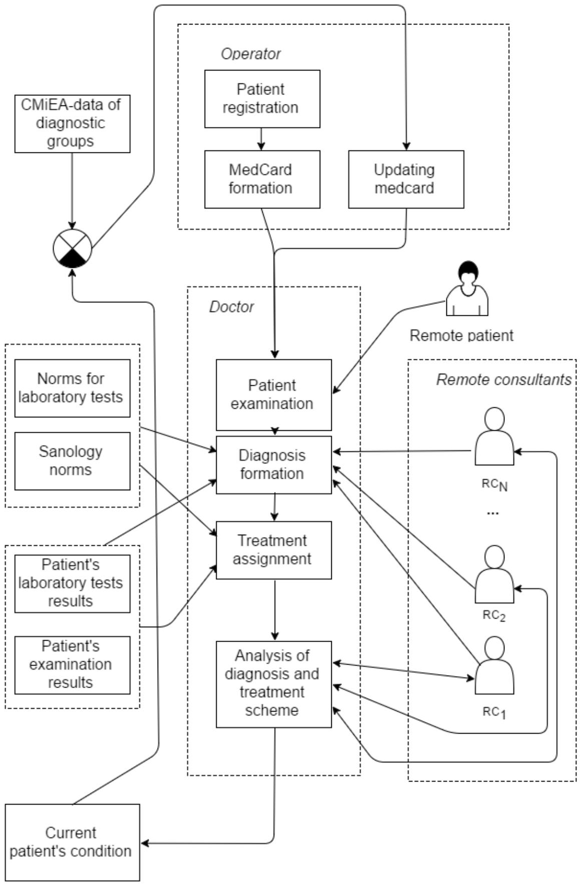

Fig. 2. Functional scheme of diagnostics

The diagnostic report should meet the following basic criteria:

• completeness – the diagnosis should be fully comprehensive, transparent, admitting of no misinterpretation; on the basis of the diagnosis must be clearly delivered in medical terms, what treatment to offer patients;

• integrated – diagnosis should fit into the overall picture of the patient's health, not conflict with other values of its medical and physiological parameters;

• warranty – the decision must be carefully checked, with the presence of good obosnovatelnoy base.

Since the developed software product is intended for use paramedics who have computer equipment at the user level, it is lightweight and easy to use, that is, has a convenient user-friendly interface.

Neural network approaches

DiaSpectrEx system diagnosis results are based on the classification of the input CMiEA data vectors. CMiEA results are data vectors of 32 parameters that describe spectral densities of parts in exhaled air [6].

The physical meaning of the method of obtaining data respiratory metabolism is to measure spectral characteristics of a monochromatic coherent radiation due to light scattering as it passes through the system of dispersed nanoparticles suspended in a biological fluid. This method is based on the analysis of the spectrogram of CMiEA and uses the original spectroscopic equipment. Identification and analysis of differentially significant spectral shifts CMiEA to determine the condition of the respiratory system in diseases of bronchitis and pneumonia (Fig. 3).

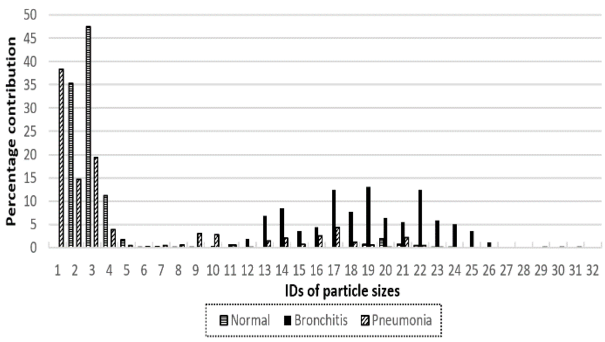

Fig. 3. Comparison of particles size CMiEA of different diagnostic groups

Contemporary DSS use different classification methods; each of them has its own advantages and features of the application. The main ones are the classification using decision trees, Bayesian classification, classification using artificial neural networks (ANN), classification by support vector machines; statistical methods such as linear regression, nearest neighbor search, CBR-classification method, and genetic algorithms approach [5]. These approaches belong to naïve, so classifiers can be built relatively easy.

At early stages of development, we used statistical classification – the linear discriminant analysis (LDA). It showed poor performance and several drawbacks, which are critical for medical systems:

Relatively low accuracy (< 90 %) that is lower than other diagnosis methods

Inability to add new classes into model, that is critical in medical applications.

However, using LDA confirmed the assumption about the differences between samples in different groups of disease. From this was concluded that diseases could be considered as disjoint classes [9].

Due to the named constraints of method, the other one was developed. As the problem is linearly solvable, neural networks can be used with good chance to overcome statistic approaches. Multilayer perceptron (MLP) is a good classifier for linearly divided data, so we decided to use this type of neural network [7].

According to [7], optimal layer sizes were selected as following:

$$Ni = (Nin + Nout) * 0.66$$

Where Ni – number of neurons in hidden layer, Nin – number of neurons in input layer, $Nout$ - number of neurons in output layer.

Activation functions are selected with exhaustive search over the given set, that includes [8]:

- ehreshold;
- symmetric threshold;
- sigmoid;
- stepwise sigmoid;
- symmetric sigmoid;
- stepwise symmetric sigmoid;
- gaussian;
- symmetric Gaussian;
- stepwise Gaussian;
- elliot function;
- symmetric Elliot function;
- linear function;
- symmetric linear function;
- symmetric sine;
- symmetric cosine;
- sine;
- cosine.

Testing of suitability of functions is performed on small subset of training data with sparse network, as we only need to estimate the performance of network with these functions.

The initial version of the classifier consisted of a single neural network trained to separate all the available disease classes. For three known classes we built the multilayer perceptron with number of inputs that corresponds to the number of CMiEA factors for features and three outputs for classes. This approach has shown the complexity of neural network learning because of the large amount of factors and overfitting problems under conditions of small samples number in groups of diseases. This method showed the accuracy on the level of ~73 %, which is lower than accuracy of LDA, thus the method was rejected.

To deal with overfitting, deep decision tree was built. Combining neural networks and decision trees structure, another algorithm appeared. Classification tree is an ordered graph that consists of classifiers in its nodes and class labels in its leaves. Thus, each neural network is trained as a "one-vs.-all" binary classifier for each of the data classes. This method allows us to achieve ease of training and to get rid of the network overfitting. Consequently, each ANN in a tree vertex distinguishes only one disease class out of all data. Unlike the approach of one ANN, approach with separate neural networks added the ability to use different activation functions and training algorithms for each classifier. This has significantly improved the accuracy and speed of learning. In addition, it became possible to determine data that do not fit into any of the known classes. All the classifiers executed one by one, when the classifier of level that is closer to top answered "false".

The main disadvantage of this method is that we need to use exhaustive search over all the nodes, in order to find the optimal configuration for the best performance. In addition, if the wrong structure is found, amount of false negatives enlarges dramatically, up to 70 %, that is also unacceptable in medical diagnosis, i.e. patient will be diagnosed as "normal", while having, for example, pneumonia.

As the usage of deep decision tree is risky, more robust ensemble classifier was found. Flat or wide decision tree excesses all of these drawbacks. Wide decision trees consist of the same classifiers as a deep decision tree but they are organized in a different way. Instead of serial execution of classifiers, this algorithm executes all the classifiers in parallel. Majority function then executes on results of classification, so the classifier with the strongest output is selected (Fig. 4)

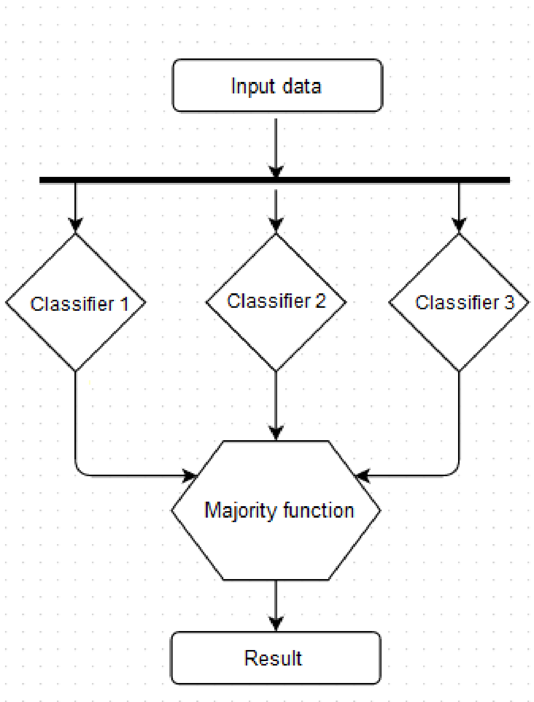

Fig. 4. Schematic view of wide decision tree

In case when all classifiers return doubtful or similar results, simple distance check is executed calculating Mahalanobis distance from sample to the class centroid. Inverted distance is, thus, the measure of similarity. Result is marked as doubtful and the probabilities for every class then returned. The comparison plot between centroid of class and the classification result is shown on Fig. 5, and Fig. 6. The correlation between factors is clearly seen.

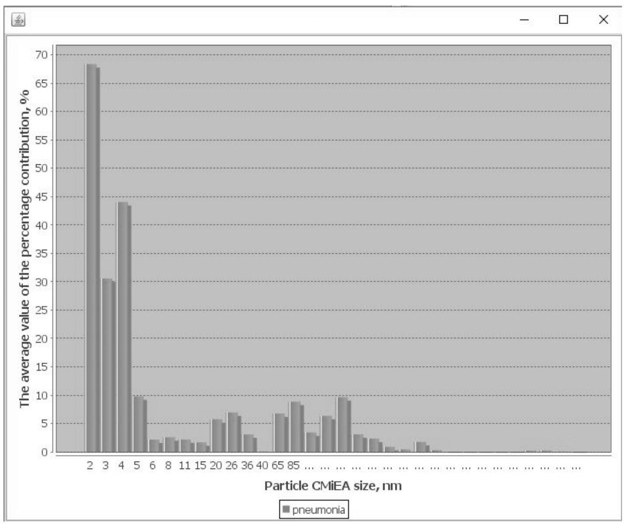

Fig. 5. CMiEA of pneumonia

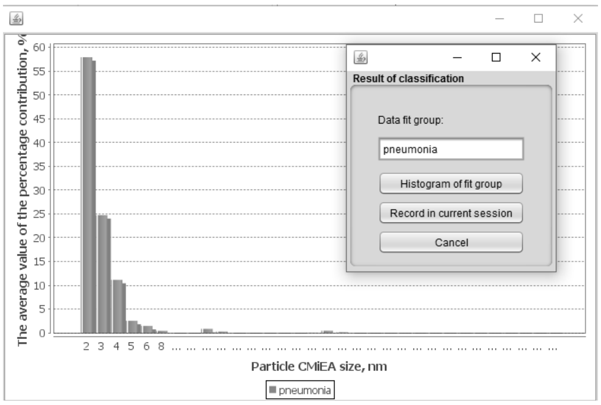

Fig. 6. Result of classification (pneumonia)

On test data, this approach showed the best results for known data with accuracy above 95 %. Test data is represented by the CMiEA data that were diagnosed a priori by doctor. In this case, result of diagnosis by our system for 72 samples out of 76 coincided with manual diagnosis by doctor. These samples consisted of 15 healthy patients, 37 ill with bronchitis patients and 24 ill with pneumonia patients.

Out of these diagnoses, four were false: for one system returned “Unknown” result, also there was one false negative (patient was treated healthy, while he was ill with pneumonia), and for two samples, there were wrong cross-diagnoses (bronchitis treated as pneumonia and vice versa). Sample amount is low to estimate the precise percentage of false positives and true negatives.

In addition, it could recognize test samples for uncertain data, where all other tried algorithms showed false negative result.

In addition, decision tree classifiers allow simple addition of new data samples in runtime, without the necessity of retraining the whole ensemble from scratch. This ability opens new perspectives in medical diagnostics, while the program learns with every new cured patient to make classification better.

Conclusions

The intelligent system requires constant accumulation and analysis of the experience of the diagnosis – a constant additional training the neural network. This will improve the accuracy of expertise evaluation. Thus, the DSS was developed in order to diagnose pulmonary diseases. This system can be used for primary diagnosis and, later for monitoring of the patients’ treatment progress.

New approach showed the best results for known data with accuracy above 95%. In addition, authors investigate questions interpretation of the diagnostic results, of technical analysis and filtering of input sample data, develop software tools to resolve situations of using poor quality of the input information, and others.

References

1. Guilan Kong, Dong-Ling Xu, and Jian-Bo Yang, (2008), Clinical Decision Support Systems: a Review on Knowledge Representation and Inference Under Uncertainties, *International Journal of Computational Intelligence Systems* (in English).

2. Josceli Maria Tenórioa, Anderson Diniz Hummela, et al., (2011), Artificial Intelligence Techniques Applied to the Development of a Decision Support System for Diagnosing Celiac Disease, *International Journal of Medical Informatics* (in English).

3. Komlevaya N., and Cherneha K. Sredstva dlya kompleksnoy zaschityi informatsii v spetsializirovannoy diagnosticheskoy sisteme, [Tools for Comprehensive Protection of Information in a Specialized Diagnostic system], (2015), *Tr. Shestnadtsatoy Mezhdunarodnoy Nauchno-prakticheskoy Konferentsii “Sovremennyie Informatsionnyie i Elektronnyie Tehnologii”* - Odessa, Ukraine (In Russian).

4. Komlevaya N., and Komlevoy A. Avtomatizatsiya diagnostirovaniya sostoyaniya dyihatelnoy sistemy, [Automatization of Pulmonary System Diagnostics], (2012), *Tr. Trinadtsatoy Mezhdunarodnoy Nauchno-prakticheskoy Konferentsii “Sovremennyie Informatsionnyie i Elektronnyie Tehnologii”,* Odessa, Ukraine (In Russian).

5. Komlevaya N., Komlevoy A., and Tymchenko B. Sravnitelnyiy analiz dvuh podhodov pri reshenii zadachi klassifikatsii, [Comparative Analysis of Two Approaches to Solving the Problem of Classification], (2014), *Nauko-tehnichniy Zhurnal “Radioelektronni i Komp’yuterni Sistemi”,* Kharkov, Ukraine (In Russian).

6.Komlevoy A., Bazhora Yu., and Cherniavskyi V., (2014), The Differential Analysis of Seasonal Changes of the Moisture Condensate Macromolecular Structure of the Exhaled Air According to Laser Correlation Spectroscopy Data, *British Journal of Science,* *Education and Culture*. London University Press (In English).

7. Martin Riedmiller, (1994), Advanced Supervised Learning in Multilayer Perceptrons - From Backpropagation to Adaptive Learning Algorithms, *Computer Standards and Interfaces* (In English).

8. Sibi P., and Allwyn-Jones S., (2013), Analysis of Different Activation Functions Using Back Propagation Neural Networks, *Journal of Theoretical and Applied Information Technology* (In English).

9. Kazemzadeh R.S., Sartipi K., and Jayaratna P., (2010), A Framework for data and Mined Knowledge Interoperability in Clinical Decision Support Systems, *International Journal of Healthcare Information Systems and Informatics* (In English).

10. Tymchenko B., (2015), Machine Learning Approach in Problems of Pulmonary Diagnostics, *Materials IV Student Conference of Computer Science*, Erfurt, Germany (In English)

11. Bazhora Yu.I., Komlevoy A.N., Chesnokova M.M., Nalazek A., and Zhukov W., (2013), Respiratory System Estimation at the Healthy Children and Children with Bronchitis with the Use of Laser Correlative Spectroscopy, *Journal of Health Sciences,* Vol. 3, No. 7(In English).

Received 06.06.2016

Cherneha
Kyrylo Sergeyovych, bachelor, Odessa National Polytechnic University,1, Shevchenko avenue. Odessa, Ukraine, +380934487179.
E-mail: kirill-chernega@mail.ru

Tymchenko
Borys Igorevych, bachelor, Odessa National Polytechnic University, 1, Shevchenko avenue Odessa, Ukraine, +380664471907.
E-mail: tim4bor@gmail.com

Komleva
Natalya Olegovna, PhD., docent at System Software Department, Odessa National Polytechnic University, 1, Shevchenko avenue Odessa, Ukraine, +380664428574.
E-mail: nokoml@yandex.ua