

Development the system for analysis clinical data of patients with diabetes for assistance in creating therapy and estimating efficiency of its application

A A Sirotinin^{1,3}, A A Popov¹ and M G Dorrer^{1,2}

¹ Reshetnev Siberian State University of Science and Technology, Krasnoyarsky Rabochy Av. 31, Krasnoyarsk, 660037, Russian Federation.

² Siberian Federal University, 79 Svobodny pr., 660041 Krasnoyarsk, Russia

³ e-mail: panzer1233000@gmail.com

Abstract. The article analyzes data on the methods of treating patients with diabetes, as well as the possibility of obtaining additional information based on the processing of case histories by neural network methods, which is later used to build a decision-making assistance system for medical specialists. This information system can be used to assess the risks arising in the treatment of patients, allows you to more accurately set the diagnosis, prescribe treatment and monitoring dynamics of the effectiveness measures taken in relation to the patient. Presents solutions related tasks to data collection, processing and normalization both in the form of human analysis and with the help of machine learning systems. Affected by the heterogeneity of medical data. An analysis is made of existing systems with the same desired result in decision making for the medical sector. The approaches to the practical implementation with the observance of the norms and rules for handling personal information are considered. Accurately described the desired target result in achieving the solution of the task that satisfies a number of requirements of specialists using a similar system on their existing clinical data. The most important part of the system is the implementation of an adaptive method that allows specialists to influence the operation of the system by using the data and information at their disposal that is specific to a given region and social standard of living. Further possibilities for the application and development of technology in the future by adding additional methods to the already created system, affecting the efficiency of its work, as well as the possibility of creation new methods.

1. Introduction

Currently, according to etiological classification, four groups of patients are distinguished:

1. Diabetes of the first type or insulin-dependent.
2. Diabetes type two or insulin-dependent.
3. Gestational diabetes characteristic of pregnant women
4. Other forms of diabetes not classified according to three previous groups.

There are also many types of diabetes that are not amenable to the usual classification, they include:

- Genetic defects of beta-cell functions caused by a disorder in a certain group of genes.
- Genetic changes in insulin action in case of insulin receptor gene mutation.
- Diseases of the exocrine pancreas
- Different types of endocrinopathy

If there is such a wide range of possible target groups of diseases, even without taking into account the occurrence of side and concomitant diseases, which the medical specialist also cannot ignore, there is a problem of making a diagnosis and prescribing necessary therapy for patients. In most cases, medical specialists use a rather old model of processing clinical data highlighting the main obvious symptoms of the disease allowing to quickly classify the first or second type of diabetes.

With the above described state of patient diagnosis, there is a need for an expert system capable of taking into account a large amount of data and highlighting important aspects, providing medical personnel with additional data for decision making. In the course of deep iterative training on a large amount of data, it is possible to achieve a result that allows the system to make diagnoses with a small fraction of the error and prescribe therapy for the patient taking into account many factors.

With all the variety of research and observations that currently represent modern technologies, it should be borne in mind that the psychophysical state is a significant part of the biological organism. This group of parameters is responsible for a variety of processes, such as: day regimen, dietary preferences, amount of physical exertion, work efficiency, etc. Consideration of such patient characteristics does not rarely take a large amount of specialist time or goes down to the background as a result of using reliable time-tested therapy methods in combination with modern medications, which in general can worsen the effectiveness and dynamics of treatment [1].

Based on this set of physical and mental parameters of the patient, including behavioral features, it is possible to create a therapy that meets the needs of this particular individual, taking into account the behavioral patterns identified earlier by machine learning systems when analyzing the history of diseases and a general history recorded with the help of specially created testing or recording conversation with the patient and further analysis using the natural language processing module [2].

2. Methods

There are a number of problems that require complex solutions before implementing a complex system of decision-making assistance in the field of medical information technology. Highlight these tasks in sub-paragraphs for a more visual representation of the front of the necessary preparatory work.

2.1. Acquisition and normalization of data.

The main task for the application of data analysis methods is their acquisition, normalization and character, which becomes extremely difficult due to the use of individual record templates and the analysis of clinical data by each attending physician even if they were issued by the same medical institution with the same specialty. The main format of clinical data in most cases is a text record of examination results, analyzes and complaints, coupled with the patient's feelings. Moreover, it would seem that the standardized data issued by modern analyzers of the chemical composition of body fluids can differ dramatically depending on the model, manufacturer and year of equipment. In general, this does not lead to a worsening of the decision by the directly treating staff, not only because of the person's ability to quickly extract important information from the common pool, but also in the work of medical personnel with a known data format characteristic of the equipment used in the current segment.

At the moment, a large amount of information suitable for analysis and normalization is in handwritten form or not at all taken into account due to the fact that the attending staff does not attach importance to some complaints and side effects of the patient's research.

The solution in this situation will be the additional collection of information on the means of specially compiled test with a large number of questions allowing to obtain information about the patient characterizing his social status, psychological state and lifestyle [3].

At the moment, the development of natural language processing technologies gives us the opportunity to record and format the recording of the patient's communication with the attending physician, turning the audio recording into a text document. It is already possible to apply the methods of natural language analysis to the received text document in order to identify important information that can influence the diagnosis, the choice of therapy for the patient and reduce the likelihood of medical error without taking some data into account in a normal conversation [4].

2.2. Regulatory framework.

Developing a decision-making system in the medical field is a rather complicated process due to the large number of regulatory standards in both the industry itself and civil-law relations which create obstacles in collecting data due to the unwillingness of most people to publicize their medical history even if all data are impersonal. In the industry itself, obtaining such volumes of data with a systemic nature will require separate agreements with medical organizations or with government agencies authorizing such activities or requiring all medical institutions to provide this information.

2.3. Obtaining the minimum required amount of clinical data.

This problem is defined as the definition of the minimum necessary data sampling, which makes it possible to start developing a system with further expansion as new data sets become available and methods for their collection. In the early stages of development, with a lack of input data, it is possible to abandon the test sample to increase the total amount of training, which in turn will exclude the possibility of checking the system on a test sample, but this is not a critical drawback of the systems, since its use is implied in symbiosis with a medical specialist. not as an independent decision-making system. Additional options for solving the task are the methods of cross-variation, which consist of removing and adding examples to the training sample and changing the network structure by complicating it or breaking it up into smaller parts of the samples and then applying single-layer networks for their analysis and then combining them into a single complex.

At the moment it is difficult to say that there is a lack of data in this area of research since in most cases we have redundant information coming from the patient in the course of communication with the staff, but not recorded by him in writing or electronic form. Extremely this deficiency can be corrected by the natural language processing system described in paragraph 1.

To implement the method, it is planned to use neural networks and neural network technologies because of the already tested methods used for the analysis of natural language and other similar data models that have proven effective in processing non-standardized data with a large proportion of uncertainties and the likelihood of errors. However, in our case, this is still due to the backward compatibility of the output data of each method, as well as a very wide range of input data unavailable for direct processing by statistical methods without prior human analysis. What ultimately does not negate their use in the system on the ready-made output of neural network algorithms, allowing for better training samples and the submission of additional input data [5].

At present, several leading services providing services such as IBM, Microsoft Azure, Amazon Web Services, TensorFlow can be distinguished in the market of machine learning and neural network technologies. The main task is not to create a “unique” neural network architecture, but to comprehensively apply such technology to solve massive tasks requiring large resources in terms of building and designing processes for receiving, preprocessing and storing data in conjunction with adapting the results of their analysis to health.

Let's look at a simplified scheme for obtaining the desired result in the form of a sequence of activities to achieve the working model of the decision assistance system in Figure 1:

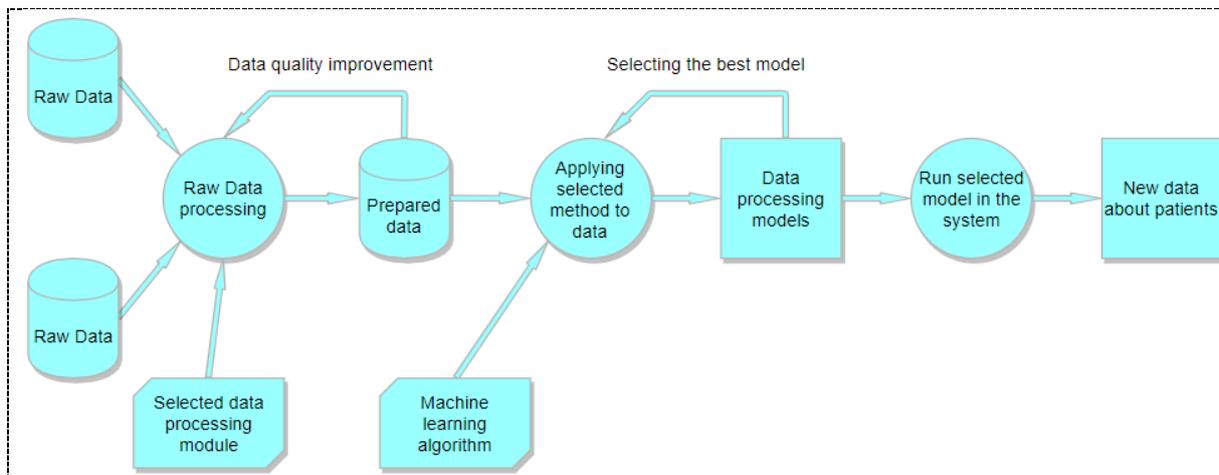


Figure 1. Step-by-step creation of a clinical data processing model

The distribution of information of value for the compilation of training samples in the first stages of designing a decision-making assistance system is carried out in three areas:

1. Anamnesis and clinical information characteristic of the specific patient and the area of the disease.
2. Psychological, consisting in the psychological state of the patient and his predisposition to behavior change, following the instructions coming from the attending physician.
3. The social status of the patient, the availability of social interaction, the opportunity to participate in social health programs, his mobility and the ability to adapt to new living conditions.

In the future, methods of normalization and preprocessing of data are applied to the information collected, depending on their source format and source. This process is an iterative one that allows to achieve the required format and purity of the data most suitable for machine learning algorithms for input [6].

Based on the obtained pre-processed data arrays, a group of models is built, which also using the iterative method of selecting the best solution allows a gradation with the ability to control the significance of the three groups of information collected about the patient during the algorithm.

The subsequent post processing of the obtained results, which consists in comparing the dynamics of radiation during the operation of the system, allows revealing the relationship between the applied therapy methods and the effect of certain input data on their efficiency of use for a particular patient or group with similar clinical history and data [7].

This approach in combining a wide range of diverse information will provide a comprehensive solution to the problems of collecting patient information, as well as an in-depth analysis of the interrelationship of various data that helps in solving practical problems of a therapeutic and prophylactic nature.

From all the above, you can describe the required characteristics of the projected system of decision-making assistance:

1. The ability to take at the entrance of various data of anamnesis and clinical studies.
2. The ability to download data from own specialist data for training the system.
3. The presence of intermediate control points both at the stage of data preparation and the choice of the optimal model for solving the problem, providing the ability to influence the course of training the system by a medical specialist.
4. Availability of an algorithm for improving the results of the decision-making system by correlating the results obtained with the specialist's own decisions, providing the ability to fine-tune.

5. The presence of a clear interface that does not require deep technical knowledge of the work of the system and the methods used by medical personnel.
6. The availability of the possibility of expanding and scaling the system by adding additional modules that allow processing various data different from those already used by the system.

3. Applicability

The use of such methods of analyzing clinical data and the subsequent choice of therapy or other impact on patients leads to an increase in the effectiveness of the treatment of diseases and brings them closer to the calculated way of reducing non-performing actions and instructions for psychological and social reasons. The importance of this technology lies in a wide range of applications focused mainly on long-term treatment and rehabilitation courses, which in turn are the most costly with increasing time and cost.

Taking into account the report of the World Health Organization and the projections to 2030 to increase the impact of chronic long-lasting diseases on the further development of society, as well as addressing the problem of implementation and implementation of long-established practices for reducing the number of non-infectious diseases, especially in areas where there is a shortage of qualified personnel and a streamlined management system for their implementation. On the basis of this information, there is a clear need to develop these methods and means of decision-making assistance, contributing to the improvement of the effectiveness of the activities carried out, as well as providing expert assessment of medical decisions of specialists with a small amount of experience and qualifications [8].

The economic efficiency of the applied system can be achieved by reducing the time spent by a specialist on the analysis and isolation of important information from the patient's history, obtained clinical data in the form of various analyzes and examinations.

4. Future development

The subsequent development of the system will consist not only in improving the core module through an iterative process of selecting training samples and data, but also adding additional solutions capable of collectively forming a consultation of neural networks, where each incoming network is able to work with different data complementing a solution based on anamnesis and psychophysical clinical studies patient's parameters by his behavior model and other social characteristics.

Creating flexible teaching methods with the ability of a specialist to independently create a set of training data specific to his area of work will allow an adaptive system to be able to work in different regions regardless of their characteristics.

It is necessary to understand that this development is based on data on diabetes mellitus, but the methods can also be applied in identifying, treating and managing other diseases depending on setting up and training the system by achieving a standardized model for collecting medical information and bringing it into a single format suitable for use of designed system.

References

- [1] Levitov N D 1964 *About mental states of human*, (Moscow, Prosveshenie) p 20
- [2] Kravcheenko A I 2011 *Sociology 2nd edition processing and addition*] (Moscow, Akademicheskii Proekt) p 508
- [3] Mishina O S 2016 *Organization of Medical Care for Chronic Patients Over the Lifetime / Methodical recommendations* (Moscow, RIO TSNIIOIZ MZ RF) p 40
- [4] Zagoryiko N G 1999 *Applied methods of data and knowledge analysis* (Novosibirsk, Institute of Mathematics Publishing) p 270
- [5] Korobkova S V 2006 *Problems of efficient approximation of multidimensional functions using neural networks* (Proceedings of the Southern Federal University. Technical science) **58** pp. 121–127

- [6] Dydarov S P and Diev A N 2013 Neural network modeling based on perceptron complexes with small volumes of learning assemblies *Mathematical methods in engineering and technology – MMTT-26: collection of works XXVI international scientific conference, sections 6, 7, 8, 9* (Nizhny Novgorod, Nizhny Novgorod State Technical University, 2013) p 114–116
- [7] Huang C and Moraga C A 2004 Diffusion-Neural-Network for Learning from Small Samples *International Journal of Approximate Reasoning* **35** p 137–161
- [8] WHO 2009 *Time to deliver: report of the WHO Independent High-level Commission on Noncommunicable Diseases* [Online] Available:
<https://apps.who.int/iris/bitstream/handle/10665/272710/9789241514163-eng.pdf?sequence=1&isAllowed=y>