DEVELOPMENT OF THE STRUCTURE OF AN AUTOMATED SYSTEM FOR DIAGNOSING DISEASES

Abstract: Today, the importance of information support for various medical technologies has increased significantly. The use of modern information technologies is becoming a critical factor in the development of most branches of knowledge and areas of practice, so the development and implementation of information systems is an urgent task. The clinical decision support system provides clinicians and stakeholders with individualized patient assessments and recommendations to assist in the clinical decision-making process. Knowledge-based information systems are widely used in medicine around the world. Modern technical capabilities make it possible to reach a qualitatively new level of presentation of the course of the disease, namely, based on appropriate mathematical models, to model the typical development of the pathological process in a particular disease, to speed up the process of diagnosing and receiving recommendations on treatment protocols.

In Kazakhstan, there is no variety of decision support systems in medicine, especially in the process of diagnosing diseases. The purpose of this study is to develop the structure of an automated system for diagnosing diseases. The Unified Modeling Language (UML) is used as a design tool. The structure of the automated system is presented; the main components and key terms are considered. A mathematical model of diagnostics is shown (in the example of diseases of intestinal and pancreatic insufficiency) based on decision-making methods with fuzzy initial data. A diagram of the data flows of the system is presented. Thus, the paper proposes the structure of an automated system that will contribute to high-quality diagnostics due to an effective method of system organization and the use of fuzzy set theory methods.

Keywords: IT in medicine, decision support system, automated system for diagnosing diseases, automated system structure and diagrams.
Introduction

The increasing complexity of business processes in medicine, the use of expensive resources, and the problems of quality and standardization of medical services require the use of effective management techniques that cannot be implemented without information technology and modern software. Medical information systems are of interest to heads of medical institutions, heads of departments, healthcare professionals, and medical specialists. All of them are constantly in need of all kinds of statistical data, operational and reliable information about the incidence, diagnosis, treatment protocols, and many other data. This is what ensures the relevance of studying the issues of automation in the medical field.

An automated knowledge-based information system is a set of programs that performs the functions of an expert (doctor) in solving problems from a certain subject area. Such systems give advice, analyze, give consultations, diagnosis, and can give some recommendations. The practical application of expert systems in enterprises contributes to the work efficiency and professional development of specialists. The main advantage of the system for diagnosing diseases is the ability to accumulate knowledge, store it for a long time, and update it. The accumulation of knowledge allows improving the skills of specialists.

Moreover, the practical application of artificial intelligence in medicine is based on the developed system, which allows improving the quality and saving decision-making time, as well as finding out the most probable diseases during the diagnosing.

Literature review

One of the key areas of modern medicine is diagnostics, without which no medical diagnosis is made. Today, serious changes are taking place in this area, leading to fundamental changes in the approaches and principles of treatment. Modern diagnostic methods make it possible to quickly diagnose patients, monitor the development of the disease, assess the risk of factors affecting its severity, and monitor the prescription and intake of drugs, treatment protocols and prescribed tests.

Medical expert systems enable a specialist not only to check the correctness of their own diagnostic assumptions, but also to turn to a computer for advice in difficult diagnostic cases. In most cases, expert systems are required to make a diagnosis, select the best treatment strategy or optimal drugs. Dialogue with expert systems, as a rule, is resorted to by young inexperienced medical specialists, doctors who do not work in their specialty or in emergency cases with a lack of time and a large number of factors influencing the decision [31]. To date, there are many expert systems that operate on the basis of various methods and are used in many areas of medicine. Examples of the use of expert systems in medicine are not isolated; they are used in numerous areas of healthcare.

The MYCIN expert system, developed at Stanford in the mid-1970s, was one of the first to address the problem of making decisions based on unreliable or insufficient information. All reasoning of the MYCIN expert system was based on the principles of control logic corresponding to the specifics of the subject area. Many of the expert systems development techniques used today were first developed as part of the MYCIN project [1].

The E-MYCIN system («Empty MYCIN», i.e. empty MYCIN) is a domain-independent version of the MYCIN system, i.e. this is the MYCIN system, but without a specific medical knowledge base. According to the developers, E-MYCIN may well serve as a «skeleton» for creating consulting programs in many subject areas, since it has many software tools that facilitate the task of the designer of a particular expert consulting system. It is especially useful for solving deductive problems, such as diagnosing diseases, which are characterized by a large number of unreliable input measurements (symptoms, laboratory test results, etc.), and the decision space containing possible diagnoses can be quite clearly delineated [2].
Experts from the Institute of Experimental and Clinical Medicine in Novosibirsk created the Cardiologist expert system for diagnosing cardiovascular diseases. The expert system «Cardiologist» determines the diagnosis of the patient according to the entered symptoms, prescribes a course of treatment and prevention, allows you to create and modify knowledge. The Pascal language and the Delphi 7 development environment were chosen to develop the Cardiologist expert system [3].

Currently, there are also online versions of medical expert systems, where any user can establish a diagnosis with a certain degree of probability. One such system is EasyDiagnosis by MatheMEDics [4]. This system is a free English-language service, where it is possible to establish a diagnosis of a disease or the cause of an ailment using a test.

The DXplain system is an example of an intelligent clinical decision support system used to assist in the diagnostic process and contains symptoms, laboratory data and procedures in its knowledge base that link them to a list of diagnoses. It provides support and rationale for differential diagnoses and follow-up studies [5].

The PEIRS system [6] interprets and comments on chemical pathology reports. The system has a built-in automatic machine learning module that allows the pathologist to create new rules without the participation of a knowledge engineer. Currently, 2300 such rules have been created. It takes about a minute to build each new rule. Every day, the system comments on 100 reports in the field of arterial blood gas composition, glucose tolerance test, etc.

An analysis of scientific publications in the literature on the development of automated medical diagnostic systems shows that developments and research in this area have been carried out all over the world in various directions for more than 50 years [7-13]. The dynamics of publications on this topic show post significant growth, especially in recent years.

Kazakhstan, in this matter, is trying to keep up with the leading countries of the world: new, more informative tests are constantly appearing in domestic diagnostics, more useful and significant research for doctors is being carried out, information systems are being introduced and modernized [14-18]. The MEDintel system is a domestic development based on artificial intelligence [19]. Medintel is the first healthcare decision support system created by the international medical community. The program is free and open to improvement. MEDintel is currently developed for the web platform. In the future, it will be available for Android and iOS platforms.

**The purpose of the article**
The aim of the research is to develop the structure of an automated system for diagnosing diseases.

To achieve this goal, the following tasks are defined:
- Create the structure of the automated system;
- Give details of the main components and key terms of the system;
- Design the system using UML diagrams, such as sequence and use case diagrams;
- Provide a literature review of similar systems used;
- Expound the mathematical method used in the system inference engine.

**Components and key terms**
The developed system being is as a program that performs actions similar to those performed by a doctor during a diagnosis, making certain conclusions in the course of issuing advice and recommendations. The main differences of the system from other software products are the use of not only data, but also knowledge, as well as a special mechanism for deriving decisions and new knowledge based on existing ones. Knowledge in our system is presented in a form that can be easily processed on a computer. In our system instead of an algorithm for solving a problem, there is a knowledge processing algorithm.
Therefore, the application of the knowledge processing algorithm can lead to obtaining such a result when solving a specific problem, which was not provided. Moreover, the knowledge processing algorithm is not known in advance and is built in the course of solving the problem based on heuristic rules. In knowledge-based systems, the rules (or heuristics) by which problems in a particular subject area are solved are stored in a knowledge base. Problems are posed to the system in the form of a set of facts describing a certain situation, and the system, using the knowledge base, tries to draw a conclusion from these facts.

The proposed system is characterized by the presence of the following features:

- **User communication (interface);**
- **Ability to self-learning;**
- **Possesses adaptability.**

*User communication:* characterizes the way of interaction (interface) of the end user with the system. Moreover, there is a possibility of choosing symptoms from the list and the degree of symptom’s severity.

*Ability to self-learning:* manifests itself in the possibility of automatic extraction of knowledge for solving problems from the accumulated experience of specific situations.

*Possesses adaptability:* it is the ability to develop a system flexible to changes.

Knowledge based system depends on various applications, the presence of the following symptoms:

It creates the mechanism of thinking of a person (doctor) who is able to determine the correct diagnosis of the patient. The main attention is paid to the reproduction by computer means of the methodology for solving the problem, which is used by the user.

- **The system, in addition to performing computational operations, generates certain considerations and conclusions based on the database of treatment protocols by the Ministry of Health of the Republic of Kazakhstan**
- **According to the feature explained above, the architecture of the proposed system follows.**

The automated knowledge-based information system includes several components: a solver (inference procedure), a dynamically changing knowledge base, a language processor and knowledge engineering. There are the following requirements:

- **The system should be able to conduct a dialogue (communicate) in a language that is convenient for the user (doctor), and, in particular, acquire new knowledge during the communication.**
- **The system should be able to solve a problem in an understandable to the user way.**

The architecture of the automated knowledge-based information system presented in Fig. 1.

![Figure 1. Overall system architecture](image-url)
An automated knowledge-based information system's user should to explain what problems need to learn in order to produce the right solutions. Figure 1 is the structure and architecture of an automated knowledge-based information system that shows relations between knowledge base, inference, and language processor components.

The knowledge based is the knowledge that is a reflection of the field of medicine formalized using a certain method of knowledge representation, as well as the actions performed when diagnosing diseases of internal organs. The knowledge component "fills" the system with the knowledge that allows it to solve problems from the problem area itself, with knowledge described in treatment protocols approved by the Ministry of Health of the Republic of Kazakhstan. It contains rules, facts, and descriptions of diagnosis [20]. In case of the current study, the database includes a control part, which consists of questioning the patient. As a result, the system generates the most probable disease (diagnosis of the patient) for a given condition. Additionally, the system generates the list of further tests to pass to confirm the diagnosis.

The language processor is designed to provide a comfortable interaction between the user (physician) and the system. It implements procedures for morphological, syntactic and semantic control of requests entering the system and bringing them to a form "understandable" by a computer. When issuing response information, the reverse operation is carried out – the conclusion is "translated" into a limited natural language that is understandable to the user. Language processor is used for problem oriented-communications between the user and the system. This communication is supplemented by graphics [20]. A significant feature of the system is that it gives recommendations with further analysis tests.

The interference engine is a part of the system that analyzes the information that enters the system and the information it receives and the formation (inference) on its basis of new agreements (diagnosis and list of basic tests) in response to a request to the system. In other words: solver, deductive machine, and inference engine. Further conclusions were created by performing this task. The interference engine is the Brain of the automated knowledge-based information system. This component is basically a computer program that processes the knowledge base to achieve the goal by communicating through the user interface [21].

The structure shown in Fig. 1 is called structural static automated knowledge-based information system. A system of this type is used in applications where changes in the world (temperature, weather, wind, and climate etc.) can be ignored [22].

Generally, automated knowledge-based information system requirements can be illustrated and modeled in some UML diagrams like use case, class, and sequence. The interaction between the user, developer and the system illustrates in use case diagram (see Figure 2).

![Use case diagram of the proposed system](image)
Correct system design helps developers easily generate the code and understand user needs. Therefore, there are UML diagrams created and explained. The use case diagram (see fig. 2) illustrates a visual representation of the possible scenarios of using the proposed system. It illustrates how a user will perform actions and interact with the system, like noting symptoms and getting result. Also, there are possible actions of the system: do inference or access to the knowledge base. The developer, presented in the diagram above, usually controls user and system data. Using case diagrams can help establish the complexity of the system. It does so by specifying which functions become requirements that will make it to the development stage [23]. Use case diagrams written in natural language, which helps easily understand the diagram.

The following sequence diagram (see fig. 3) displays the relationships between the objects and describes what those objects do. Diagrams are useful in many stages of system design. In the analysis stage, the diagram can help to understand the requirements of the system domain and to identify its components. Illustrate the scenarios or a series of steps performed in the response. During the implementation phase of system development, diagrams help to convert the models into code [24, 25].

![Figure 3. Sequence diagram of proposed system](image-url)

The system processes the incoming information, as well as responds to user actions. As shown in Figure 3, to incoming information systems include all of the symptoms, the answers to the survey questions are recorded in the system. The system processes the information entered. All the entries are recorded in the respective tables. So, for example, to determine the best diagnosis for the patient should receive questionnaires, and then fill in the data following a pattern. Results will be displayed on the page result.jsp, as well as the table user, the cell result in a system database. After processing the input information system allows the user
to get relevant results and other relevant information about specific diagnoses. The project involves the development of a computerized system of information processing. An automated system processes the input data. Input data – is the primary documents (in this case the patient’s symptoms), which enters into the system. Furthermore, data flow diagram shows the way information flows through a process or system [26].

To develop a medical diagnostic system, we considered the general procedure for making decisions in medicine based on operating with certain knowledge. This knowledge is taken from various fields and can formally be represented as a database, knowledge base, and decision-making rules. The functional scheme of a medical diagnostic system is shown in fig. 4.

Based on the functional diagram shown in Fig. 4, we will develop the structure and functional blocks of the proposed automated system for diagnosing diseases (see Fig. 5)

The central components of the automated system are the Blocks «Management of databases and knowledge bases» and «Decision rules (mathematical model) ». The block Blocks «Management of databases and knowledge bases» acts in relation to other components as a content subsystem. The block contains a set of units of data and knowledge, which are formalized using a certain method of representing knowledge about the objects of the subject area, their relationship, actions on objects, and the uncertainty with which these actions are carried out.

Figure 4. Scheme of the functioning of the medical diagnostic system
The database and knowledge management block allows efficient use of data and knowledge through independent access to them from the inference engine, and also allows the use of medical data independently of the data obtained during patient interviews.

The traditional procedure for questioning a patient by a doctor is as follows: the patient tells the doctor about his subjective symptoms, the doctor identifies the most significant among them, taking into account the proximity of the symptoms of one of the diseases. Then, using fundamental medical knowledge and his own experience, he asks in more detail about the symptoms associated with a probable illness, and generates a conclusion.

What is needed as knowledge, in this case, is general relationships between diseases and symptoms, and some measure of such relationship is obviously needed both for the disease in terms of its symptoms and for the symptoms in terms of the disease. Block «Decision rules (mathematical model) » allows simplifying this procedure.

**The mathematical diagnostic model**

When organizing and conducting an expert assessment with the participation of doctors and patients, a number of problems arise related to the peculiarity of the human procedure for assessing and diagnosing. To solve these problems, to apply the fuzzy set theory tools in the process of establishing a diagnosis is proposed. A mathematical diagnostic model has been developed (using the example of intestinal and pancreatic insufficiency) based on decision-
making methods with fuzzy initial data. The model is used in the inference engine. To develop a mathematical model for the differential diagnosis of intestinal and pancreatic insufficiency, we use a diagnostic table [27]. It is convenient to present the utility matrix of intestinal and pancreatic insufficiency based on a diagnostic table in the specific form (table 1).

Table 1. Table of alternatives of enteric and pancreatic insufficiency

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2(A)</th>
<th>X2(B)</th>
<th>X3</th>
<th>X4</th>
<th>X5(A)</th>
<th>X5(B)</th>
<th>X6(A)</th>
<th>X6(B)</th>
<th>X7</th>
<th>X8(A)</th>
<th>X8(B)</th>
<th>X9</th>
<th>X10</th>
<th>X11</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>HX</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>X</td>
<td>X</td>
<td>HX</td>
<td>HX</td>
</tr>
<tr>
<td>A2</td>
<td>OCH</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>X</td>
<td>HX</td>
<td>X</td>
<td>X</td>
<td>HX</td>
<td>HX</td>
</tr>
</tbody>
</table>

In the table of alternatives, which are enteric and pancreatic insufficiency, A1 and A2 are indicated, and the criteria (symptoms) are X1, X2,..X12.

The degree of belonging of the criteria, which are linguistic variables, is set qualitatively: characteristically, not characteristically, very often, often, etc. Qualitative signs presented in the table by letter designations: X – characteristically; HX is not characteristically; OЧ – very often; Ч-often; НЧ – not often. For the convenience of solving the problem, we introduce the following simplifications: the qualitative features presented in the form of HX (not characteristically) and OЧ (not often) replaced by one sign of P (rarely), which, in our opinion, is advisable. Since the sign of OЧ (very often) is close to the sign of X (characteristically), then we replace it.

Then the utility matrix of enteric and pancreatic insufficiency will take the form presented in Table 2.

Table 2. Table of matrix of enteric and pancreatic insufficiency

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2(A)</th>
<th>X2(B)</th>
<th>X3</th>
<th>X4</th>
<th>X5(A)</th>
<th>X5(B)</th>
<th>X6(A)</th>
<th>X6(B)</th>
<th>X7</th>
<th>X8(A)</th>
<th>X8(B)</th>
<th>X9</th>
<th>X10</th>
<th>X11</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>P</td>
<td>X</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>X</td>
<td>P</td>
<td>X</td>
<td>X</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>A2</td>
<td>X</td>
<td>P</td>
<td>X</td>
<td>Ч</td>
<td>Ч</td>
<td>Р</td>
<td>Р</td>
<td>Р</td>
<td>Р</td>
<td>Ч</td>
<td>Р</td>
<td>Р</td>
<td>Р</td>
<td>Р</td>
<td>Р</td>
</tr>
</tbody>
</table>

To create a mathematical model for the differential diagnosis of enteric and pancreatic insufficiency, we use the utilities indicated in the table and the mathematical method of the theory of fuzzy sets [28]. The method allows you to work with data that is set qualitatively. Linguistic variables (characteristic, often, rarely) can be represented as subsets:

\[ X = \{0.5/9; 1.0/10\}; \quad Ч = \{0.5/7; 1.0/8;\}; \quad Р = \{0.5/2; 1.0/1;\} \]

The expert sets the index, which accounts for the maximum degree of ownership.

To develop a diagnostic model, we take the symptoms (conditional patient). We believe that the patient has severe abdominal pain, having the nature of herpes zoster, vomiting, anemia, and there are diffuse changes in the pancreas.

Based on the data of the diagnostic table, a matrix is formed:

\[
U = \begin{bmatrix}
  u_{11} & \cdots & u_{1m} \\
  u_{21} & \cdots & u_{2n} \\
  \vdots & \ddots & \vdots \\
  u_{m1} & \cdots & u_{mn}
\end{bmatrix}
\]

\[
\tilde{U}_y = \bigcup_k \mu_{u_k} (x_k) / u_k
\]

Where \( u_k = \mu \sim (x_k) \), representing a fuzzy set, consist of fuzzy subsets defined:
Set $\vec{u}_j$ simplify as follows:

$$\mu_{\tilde{u}_j}/\mu_\tilde{u} = [\mu(u_j)/\mu_\tilde{u}]/[\mu_{\tilde{u}_j}/\mu_\tilde{u}].$$ (2)

Minimization reduces the risk of misdiagnosis.

Optimizing set $\tilde{U}_i$, which is like the intersection of sets $U_{im}$ and $U_i$ and $\mu_{\tilde{u}_i} = \min \{ \mu_{\tilde{x}_i}, \mu_{\tilde{u}_i} \}$ allows you to calculate the best alternative, which means a probable disease with this complex of symptoms.

$$A^* = \nu_{A_0}(a_o) = \max_{A_o} \nu_{A_0}(a)$$ (4)

Where $\nu_{A_0}(a) = \max_{u_{jo}} \nu_{A_0}(a_k)$

The patient's condition is recorded in a fuzzy set:

\[ X = \{0.5/X1; 0.6/X3; 0.7/X4; 0.3/X11; 0.8/X8(б)\} \]

Define fuzzy utility alternatives:

$$U_1 = \{0.5/P, 0.6/P, 0.7/P, 0.3/X, 0.8/P\} \quad U_2 = \{0.5/X, 0.6/Ч, 0.7/Ч, 0.3/P, 0.8/X,\}$$

Given that some element appears several times, we obtain:

$$U_1 = \{0.98/P, 0.3/X\} \quad U_2 = \{0.9/X, 0.3/P, 0.88 /Ч\}$$

Then:

$$U_1 = \{0.98/P, 0.3/X\} = \{0.98/\{0.5/2; 1.0/1\}; 0.3/\{0.5/9; 1.0/10\}\}$$

$$U_2 = \{0.9/X, 0.3/P, 0.88 /Ч\} = \{0.9/\{0.5/9; 1.0/10\}; 0.3/\{0.5/2; 1.0/1\}; 0.88/\{0.5/7; 1.0/8\}\}$$

Minimize:

$$U_1 = \{0.5/2; 0.98/1; 0.3/9; 0.3/10\} \quad U_2 = \{0.5/9; 0.9/10; 0.3/2; 0.3/1; 0.5/7; 0.88/8\}$$

Using $u_{max} = \sup Y$, where $Y=S(U_1) \bigcup S(U_2) = \{1,2,7,8,9,10\}$; $u_{max} = 10$;

Using $u_{max}$, we are going to find Max from:

$$U_1 = \{2.10/2; 1.1/10; 9.10/9; 10.10/10\}$$

$$U_2 = \{9.10/9; 10.10/10; 2.10/2; 1.1/10; 7.10/7; 8.10/8\}$$

Then the optimizing sets:

$$U_1 = \{\min(0.5,0.2)/2; \min(0.98,0.1)/1; \min(0.3,0.9)/9; \min(0.3,1)/10\} = \{0.2/2; 0.1/1; 0.3/9; 0.3/10\}$$

$$U_2 = \{0.5/9; 0.9/10; 0.2/2; 0.1/1; 0.5/7; 0.8/8\}$$

Therefore, with the above symptoms, the most likely disease: pancreatic insufficiency, which coincides with the findings of experts.

The main advantage of making a diagnosis using a mathematical model is its objectivity, which can eliminate the over diagnosis that occurs most often during the initial examination. The correspondence of the obtained results with the results of diagnosing by other methods, in particular by the Bayes method [29, 30], indicates the applicability of this method for the diagnosis of the disease.
So, we conclude that for a given symptom complex and the patient’s condition, a disease is observed – pancreatic insufficiency, which coincides with the diagnosis made by the decision-making method above with qualitative uncertainty and expert conclusions.

**Conclusion**

In the era of the mass use of information technologies and their penetration into all spheres of human life, it is natural to strive to use information systems to support complex and ambiguous types of human activity. One of them is the activity of a doctor, whose main point of work is the adoption of diagnostic and therapeutic decisions.

Making diagnostic and therapeutic decisions is often difficult, especially for novice physicians or when decisions need to be made in situations related to related medical specialties. Therefore, automated medical systems allow the doctor not only to check their own diagnostic assumptions, but also to turn to the information system for advice in difficult diagnostic cases.

The article presents the structure and describes the main functional blocks of an automated system for diagnosing diseases of internal organs using fuzzy set theory tools. It is shown that the use of mathematical models can improve the reliability and adequacy of automated systems in medical diagnostics.

The originality and novelty of the results are determined by the fact that the proposed system allows diagnosing the disease using a block of fuzzy conclusions. The article presents the structure of an automated system for diagnosing diseases, its main components and key terms. The Unified Modeling Language is used as a design tool. The data flow diagram illustrates how the user should navigate the system. In this paper, the sequence diagram models the process of system design. Use case diagram shows the possible scenarios.

The use of the proposed automated system will change the efficiency of the work of medical specialists and will allow you to quickly receive an expert opinion on a possible patient’s disease, a list of confirmatory tests, and treatment protocols.

**Reference**