Design of medical database for medical decision support system in laboratory diagnosis of acute leukaemia

To cite this article: V G Nikitaev et al 2019 J. Phys.: Conf. Ser. 1189 012029

View the article online for updates and enhancements.
Design of medical database for medical decision support system in laboratory diagnosis of acute leukaemia

V G Nikitaev, A N Pronichev, E V Polyakov, I O Kudryavtseva
National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe shosse 31, 115409, Moscow, Russia.

E-mail: VGNikitayev@mephi.ru

Abstract. This article discusses aspects of the design and physical implementation of the medical database for the subsequent development of medical information system (hereinafter, MIS), capable of intellectually supporting the doctor in the laboratory diagnosis of acute leukemia. The database will allow you to store huge amounts of information on patient histories, laboratory studies and structure the data at the request of the doctor. Also, the main stages of database design are described in detail, taking into account the specifics of the selected subject area. The question of choosing the optimal database management system (hereinafter, DBMS) is highlighted.

1. Introduction
Acute leukemia - neoplasm of the hematopoietic system, which is characterized by a significant variety of cellular composition.

According to various sources in Russia, oncological diseases of the lymphatic and hematopoietic system occupy 5-7 place in the structure of cancer incidence. According to data for 2015, 15 406 cases of malignant lymphoma and 10 026 cases of leukemia (4.7% of all cancer cases) were diagnosed in the country. The increase in morbidity over the past 10 years is 17.28% (about 2% per year) [1-2].

The incidence of leukemia varies in different countries from 2.3 to 11.1 per 100 thousand population, the highest rates are in developed industrial countries with a white population. In the United States, about 81,000 lymphoma neoplasms and 54,700 leukaemia cases were detected in 2015 (5th place in the structure of cancer incidence and an average annual increase of about 1%) [3-6].

According to FAB-classification, there are two types of leukemia: acute lymphoblastic leukemia and acute myeloblastic leukemia. Acute leukemia is a rapidly progressive disease that affects mostly unformed cells (not yet fully developed). Acute lymphoblastic leukemia is most common in children, while acute myeloblastic leukemia mainly affects adults [7-8].

Recent additions to the classification system proposed by the World Health Organization (WHO) require additional assessment of blasts using flow cytometry [9].

Timely determination of the type of leukemia significantly helps in providing appropriate treatment for a particular type. Morphological examination of blood and bone marrow is conducted to confirm the presence of leukemic cells. To classify pathological cells into specific types and subtypes of leukemia, a clinical laboratory physician studies cells under a light microscope in search of anomalies [10].
Along with the visual analysis performed by the doctor for the classification of cells, there are methods and means of digital image processing that allow to recognize objects in the image and automate the process of diagnosis of acute leukemia [11-14].

One of the problems in the development of such systems is a limited database of input data – images of blood cells. A number of developers use in their research images of blood cells from Internet resources. For example, AcuteLymphoblasticLeukemiaImageDatabaseforimageprocessing - ALL-IDB contains about 400 microscopic images of blood in acute lymphoblastic leukemia (image size 2592 x 1944, JPEG format color depth 24-bit) [15]. Leukemia-images.com, hematologyatlas.com, ONKODIN With, EuropeanLeukemiaNet exists along with above-stated resource.

The main disadvantage of these resources is the small volume of samples.

The aim of the work is to develop tools for the formation of knowledge base for laboratory diagnosis of acute leukemia using computer microscopy.

The creation of such a knowledge base in close cooperation with morphologists is a promising direction for several areas: automated systems of diagnosis, training and testing.

2. Conceptual model of interaction of the doctor with the medical information system

Information on a variety of laboratory studies of various patients in the detection of acute leukemia is accumulated in modern medical institutions over the years. In this regard, the question arises about the effectiveness of the use of historical data, as the process of their search and analysis is very time-consuming.

In the world of information technology, attempts are being made to automate the actions of a doctor and to provide work with information obtained over a long period. Therefore, the creation and operation of medical information systems (MIS) is one of the current challenges.

It should be noted that MIS is a system with a number of functional capabilities – search, storage and processing of information, which is a set of software and hardware, database and knowledge base [11-20].

The conceptual scheme of interaction of the doctor with the medical information system and the basic elements of the system are given on Figure 1.

![Figure 1. Conceptual scheme of doctor's interaction with medical information system and its main elements.](image-url)
Almost all such systems are associated with the functionality of machine storage and organization of large data sets, on the basis of which all the problems of the subject area are solved. Databases allow you to save data on laboratory studies of patients up to specific images of blood cells. Also, in the course of laboratory studies, the database should be replenished and updated if necessary. To do this, the data in it should be clearly structured and organized in accordance with the selected subject area. This is done by properly dividing the design process into stages (discussed below).

To manage databases, namely the information stored in them, use common DBMS, which provide the ability to manipulate data. DBMS is an integrating link between the graphical shell, which interacts with the end user (doctor), and the database, which stores the results of laboratory studies, in this case - images of blood cells, especially the system settings for visualization of cells, etc.

The upper-level part of the MIS is the interface added by the developers to the existing scheme of interaction between the DBMS and the database. With the help of the interface and a personal computer, the doctor works with an electronic data archive. MIS allows not only to provide the user with an archive when analyzing the results of the study, but also to increase the speed of the diagnostic process.

3. Design and development of medical database

Database design is the creation of a database model, which describes the entity, their characteristics (attributes), the relationship between them. To achieve the maximum accuracy of displaying the requirements of the subject area, the design is divided into several stages: the construction of a conceptual database model, the construction of a logical model and the construction of a physical model.

The first stage: building a conceptual model. For this stage, the subject area (Medicine > Oncology > Acute leukaemia) was studied, as well as the main objects for which the information is supposed to be stored in the database, and the conceptual links between them were determined.

The structure of the projected knowledge base is based on the algorithm of laboratory research. The proposed Toolkit provides carrying out marking of objects in microscopic images, take into account shooting conditions, system configuration, data myelograms, leukocyte, immunophenotyping.

Figure 2 shows the conceptual model of the database.

Figure 2. A conceptual model of the database.
The second stage: building a logical model. Logical database design has the following tasks:

1. select the main (key) attributes of each object, which can uniquely identify a particular instance of the object;
2. defining the types of links between object instances;
3. the introduction of additional facilities to ensure proper design and integrity of the database (optional).

The logical model of the considered database is shown in Figure 3.

The main objects were identified for the developed database for which it is necessary to store information, and the relationship between them, based on the goal: to save data on laboratory studies of patients, as well as to preserve the technical parameters of the equipment used and shooting modes.

The third stage: building a physical model.

This stage is the final link in the design of an effective database. At this level of design it is necessary to choose the optimal DBMS in which the database will be implemented, and to determine the data types for each attribute (attribute) of the corresponding database object.

Stage four: physical implementation of the database.

![Logical database model](image)

Figure 3. Logical database model.

4. Conclusion
The level of conceptual and logical models of the database is presented. These are the main components of the design of medical information system for the diagnosis of acute leukemia.
The next stage of design is the construction of a physical database model. The final stage of database development is the physical implementation of the medical information system with the filling of the system with real data and the implementation of user requests.

Acknowledgments
The reported study was funded by RFBR according to the research project № 18-29-09115.

References
[1] Davydov M I and Aksel E A 2014 Statistics of malignant neoplasms in Russia and CIS countries in 2012 (original Russian title: Statistika zlokachestvennyh novoobrazovanij v Rossii i stranah SNG v 2012g) (Moscow publishing group RONC) 226
[2] Ivanilov A K 2014 (in Russian) Young scientist 2 337-339
[3] Matsuda T and Niino M 2016 Japanese journal of clinical oncology 46(3) 290-290
[4] Facts 2016-2017 The incidence, prevalence and mortality data in Facts 2016-2017 reflect the statistics from the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program, Cancer Statistics Review (CSR) 1975-2013
[5] Ferlay J 2013 European Journal of Cancer 49 1374-1403
[6] Ferlay J et al. 2012 Cancer Incidence and Mortality Worldwide: IARC CancerBase 10
[7] Amin M M et al. 2015 Journal of medical signals and sensors 5(1) 49
[8] Reta C et al. 2015 PloS one 10(6) e0130805
[9] Swerdlow S H et al. 2016 Blood 127(20) 2375-2390.
[10] Frenkel M A 2014 (in Russia) Haematopoiesis immunology 12(1-2) 18
[11] Shivhare S and Shrivastava R 2012 International Journal of Scientific & Technology Research 1(4) 125
[12] Madhukar M, Agaian S and Chronopoulos A T 2012 Proceedings of SPIE - The International Society for Optical Engineering vol. 8295
[13] Rawat J et al. 2015 Procedia Computer Science 70 748-756
[14] Neoh S C et al. 2015 Scientific Reports vol 5
[15] Singh G, Bathla G and Kaur S. 2016 Int J Appl Eng Res 11(10) 7087-7094
[16] Nikitaev V G 2015 Measurement Techniques 58(4) 68
[17] Nikitaev V G 2015 Measurement Techniques 58(2) 215
[18] Nikitaev V G 2015 Measurement Techniques 58(6) 215
[19] Zaytsev S M et al. 2017 Journal of Physics: Conference Series. 798(1) 012133
[20] Chin Neoh S et al. 2015 Scientific Reports 5