

Formalization of the diagnostic problem of cancer in automated systems

The article presents that the most promising mathematical apparatus for determining the effective method of treatment or necessary diagnosis is a method of logical conclusion, which also allows the usage of elements of fuzzy logic when improving the work of the software system. Application of fuzzy logic will allow to build evaluation systems based on natural-language expert statements about causal relationships.

The introduction of information technology in medical diagnostics is one of the most urgent branches of industry todayowadays due to lack of specialists, and the list of requirements for decision-making contains more than a hundred parametres. This led to the creation of a wide range of medical information systems (MIS), which conventionally distinguish seven levels of generations / types (from automated medical record accounting systems to intelligent decision support systems in medicine).

Analysis of existing information program with function of oncological diseases

More than 10 different software systems were considered, among which the following were highlighted (more detailed analysis is presented in the work [1]):

- 1) Oncology. Magazine PAS. Herzen;
- 2) Medicine Live. Doctor's Assistant(communiation, analyzes, International Classification of Diseases (ICD), drugs);
- 3) Doctor's Guide: ICE, Registry of Medicines (RLZ);
- 4) Doctor's Assistant;
- 5) First Aid.

The comparative analysis of software applications was decided to be conducted according to the following parameters which, according to experts, are most relevant in the diagnosis of malignant neoplasms: medical calculators, a reference book, PubMed, the International Classifier of Diseases, the TNM Handbook, medical news, TLMs, the possibility Attaching files with analysis results.

The analysis of these applications is summarized in Table. 1, in which the numbers in the columns correspond to the order of examination of the software applications, which is presented above, +/- is the presence / absence of the corresponding parameter.

As you can see, almost all systems have drawbacks and do not provide sufficient list of necessary tools.

The new system, which is developed in cooperation with experts from the National Cancer Institute, takes into account these disadvantages, and for the greater distribution of the system, the possibility of multi-level usage of program system is provided not only by professionals, but also by ordinary citizens with the possibility of remote monitoring of treatment stages for patients (this function is necessary not only for the doctor, but also for close relatives).

Table 1.

Comprehensive analysis of available software tools to support decision-making in
diagnosing oncological diseases

Parameter	Application				
	1	2	3	4	5
Medical Calculators	-	+	+	+	-
Handbook of analyzes	-	+	+	+	+
PubMed	+	+	+	-	-
International Classification of Diseases	-	+	+	-	-
TNM Directory	+	-	+	-	-
Medical News	-	+	+	+	-
Medicines Register	-	+	+	-	-
Attaching files with analysis results	-	-	-	-	+

Presentation of the treatment process in the decision support software package for the treatment of lungs' cancer

Cancer treatment consists of two parts: diagnosis and direct medical intervention. In each country, the Ministry of Health chooses one of two treatment options: use international protocols or develop their own. Ukraine uses international protocols that describe all stages of diagnosis and treatment of cancer.

In [1] was presented a detailed diagnosis algorithm and determination of the disease stage of lungs' cancer based on clinical and morphological classifications.

Clinical classification (cTNM, clinical - "c") is based on data obtained from clinical, radiological, ultrasound, endoscopic, morphological and other appropriate examinations performed before treatment. A pathological classification (pTNM) is based on clinical data from cTNMs, which complement and refine as a result of surgical intervention and histological studies.

Grouping by stages is performed after evaluating T-, N-, M- and / or pT-, pN-, pM-categories. This information - TNM / pTNM is recorded in the medical documentation fixed in the cancer-registry. The clinical stage is the basis for choosing a treatment and evaluating its effectiveness. The pathologist study of the tumor provides information to determine the prognosis of the disease.

Although this algorithm has been tested for several decades, some studies may still be out of date, but additional data is needed to improve the reliability of automatic diagnosis and treatment appointment.

That is why the program application provides both scripted and arbitrary movements based on the treatment algorithm that can be represented as a tree of interconnected actions (each action is represented as a separate window or a group of windows with fields for entering obligatory and optional parameters).

Thus, the formation of the passage of procedures in the program system can be functionally alike to computer games using nonlinear plots, various solutions for same tasks, unscrupulous order of solutions, multivariate methods of treatment [2-4].

Formalization of oncological disease's diagnostic problem

Histogenetic classification of tumors. Malignant tumors are classified based on their origin according to the tissue from which they form:

1) cancer (carcinoma, malignant epithelioma) - is derived from epithelial tissue (has ecto- or endodermal origin);

2) sarcoma - comes from the connective tissue that occurs from the intermediate layer (mesoderm);

3) carcinosarcomas - originate from epithelial and connective tissue at the same time;

4) hemoblastosis (leukemia, malignant lymphoma) - a tumor of hematopoietic tissue;

5) aphids (carcinoids, paragangliomas, small cell lung cancer, medullary thyroid cancer, thymos, pheochromocytomas, chemodecomas) - originate from neuroendocrine cells (APUD-systems);

6) tumors of the endothelium and mesothelium - tumors from blood and lymphatic vessels, synovial and serous membranes;

7) tumors of the nervous tissue;

8) tumors from embryonic remains;

9) trophoblastic tumors;

10) mixed tumors.

In 1952, The International Anti Cancer Union (UICC) proposed the classification of tumors by the TNM system, which is based on three components:

- T (tumor) - distribution of the primary tumor;

- N (nodulus) - the state of regional lymph nodes;

- M (metastasis) - presence or absence of distant metastases. Both clinical (cTNM or TNM) and pathomorphological (pTNM) classification options are used. Today, the VI edition of the TNM (2002) is used.

Clinical Classification:

T - primary tumor (defined before treatment):

Tis - pre-invasive carcinoma (carcinoma in situ);

T0 - the primary tumor is not determined;

T1, T2, T3, T4 - size (distribution) of the primary tumor;

Thx - Not enough data to evaluate the primary tumor.

N - regional lymph nodes:

N0 - there are no signs of affect of regional lymph nodes;

N1, N2, N3 - degrees of damage to regional lymph nodes;

Nx - insufficient data to assess the status of regional lymph nodes.

M - distant metastases:

M0 - no signs of distant metastases;

M1 - there are distant metastases (the M1 index can be supplemented depending on the location of the metastases with the symbols: lungs - pulm, bones - oss, liver - hep, brain - bra, lymph nodes - lym, bone marrow - mar, pleura - ple, skin - ski, other organs - oth);

Mk - not enough data on distant metastases.

Pathomorphological classification:

Categories pT, pN, pM correspond to categories T, N, M. *

G - histopathological gradation
 Gx - degree of differentiation of the tumor can not be determined;
 G1 - high degree of differentiation;
 G2 - an intermediate degree of differentiation;
 G3 - low degree of differentiation;
 G4 - undifferentiated tumor.
 L - lymphatic vessels invasion:
 L0 - no signs of invasion;
 L1 is an invasion of superficial lymphatic vessels;
 L2 - deep lymphatic vessels invasion;
 Lx - insufficient data to estimate lymphatic vessels infestation.
 V - vein invasion:
 V0 - the veins do not have tumor cells;
 V1 - tumor cells found in the lumen of the veins;
 Vh - vein invasion can not be evaluated.
 C - level of reliability of diagnostic methods:
 C1 - data obtained when performing standard diagnostic methods;
 C2 - data obtained using special diagnostic methods;
 C3 - data obtained with the use of experimental surgical intervention (in biopsy and cytological study);
 C4 - data obtained after radical surgery and morphological study of the operating material;
 C5 - data obtained after autopsy.
 R - residual tumor:
 Rh - insufficient data to determine the residual tumor;
 R0 - residual tumor absent;
 R1 - residual tumor is determined microscopically;
 R2 - residual tumor is determined macroscopically.
 Categories G, L, V, C, R are optional. Their usage is not obligatory.
 The symbol p before categories T, N, and M indicates that each of these categories is proved by histological confirmation, and for categories T and N is compulsory complete surgical removal, for the category M there is sufficient histological verification in any way, complete surgical removal of remote metastasis is not mandatory.

Deductive logical conclusion and structure of application's windows

To implement a deductive logical conclusion, the program requires a search by model. In other words, for the conclusion with help of the rule "if x is A, then at is B" it is necessary first of all to check that there is a fact "x is A" in knowledge base. In the knowledge base, you must store information about all valid values of x. This leads to a catastrophically increased time of designing the knowledge base. For example, if the state of an object is described by a ten-coordinate vector, and each coordinate can take one of four possible values, then the complete knowledge base for this object should consist of $4^{10} = 1\ 048\ 576$ rules.

This method involves the usage of hierarchical principles of knowledge in the tasks of assessing the effectiveness level of the method of treatment or the need

for diagnosis. The usage of this principle can overcome the so-called "curse of dimension".

With a large number of state parameters, the construction of a system of utterances about the unknown dependence of the "input-output" becomes rather complicated. This is due to the fact that human memory can hold no more than 7 ± 2 concepts-signs [6] at same time. Regarding this, it is advisable to hold a classification of input variables and construct a derivative tree on it, which will determine the system of embedded one in each other knowledge of smaller dimension. An example of such a tree for thirteen input variables is shown in Fig. 5

From the example we can see that knowledge of the form $D = f(x_1, x_2, \dots, x_{13})$, which connect the inputs $x_1 \div x_{14}$ with the output D , are replaced by the sequence of relations:

$$D = f_1(y_1, w, z),$$

$$y_1 = f_2(x_1, x_2, x_3), y_2 = f_2(x_7, x_8, x_9), y_3 = f_3(x_{11}, x_{12}),$$

$$z = f_4(y_2, y_3, x_{10}),$$

$$w = f_5(z, x_4, x_5, x_6),$$

where y_1, y_2, y_3, z, w are intermediate linguistic variables.

Due to the hierarchucal principle one can take into account practically an unlimited number of state parameters that have influence on the estimation. When constructing a conclusion tree, it is necessary to try to ensure that the number of arguments (input arrows) in each node of the tree satisfies the number of 7 ± 2 .

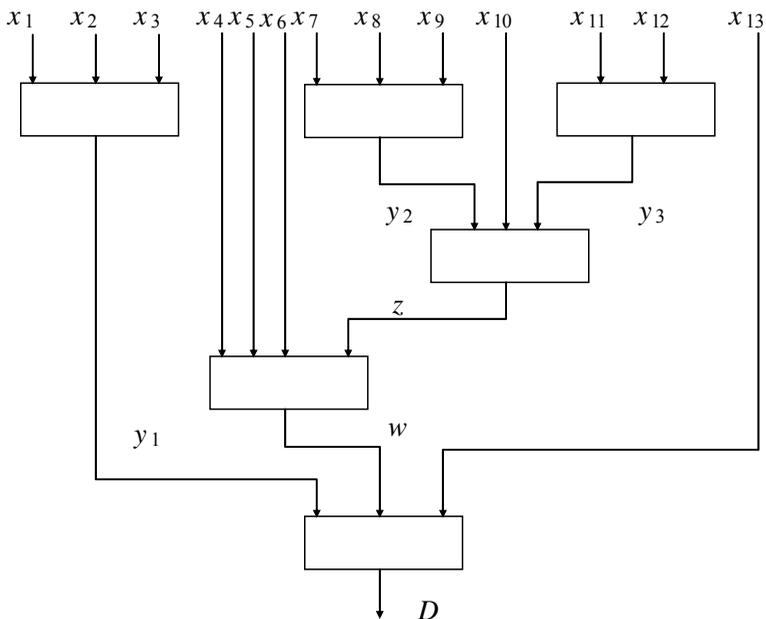


Fig. 1. Example output tree

The expediency of comparative representation of expert knowledge is due not only to the natural hierarchy of objects, but also the necessity to take into account the parameters of the state as knowledge accumulation on the object. In addition, usage of the hierarchy principle allows you to simplify the rules and significantly reduce their number [5, 6].

The program realization of output tree presented on fig. 2.

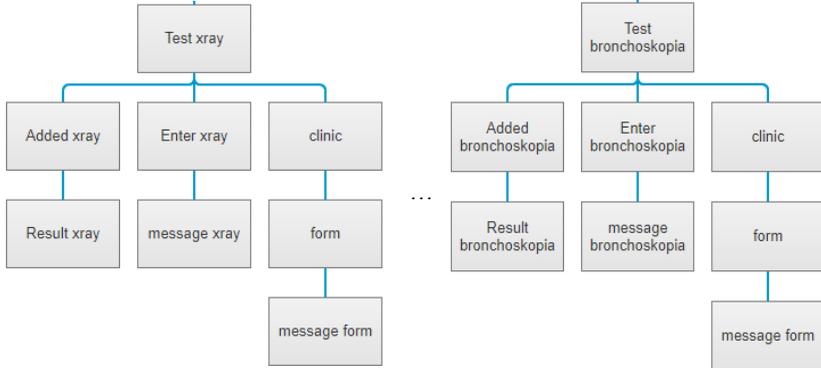


Fig. 2. Some forms of application that display the necessary diagnostic procedures and regulate data entry to the system

Conclusions

As a result of solving the tasks set in the study, the appropriateness of the chosen path in the organization of the work of the program support system for decision-making in the treatment of lungs' cancer was shown. On the basis of the presented data it is envisaged to obtain a set of new scientific results that will allow to consistently substantiate the methods of determining the necessary diagnosis in patients' examination and effective treatment.

Based on the review of existing software solutions for decision support in the diagnosis of lungs' cancer, the feasibility of introducing a new software system that will use modern approaches to work with all groups of users of the system has been proven. Construction of this system is not possible in the development of a mathematical model that describes the relationship of treatment methods and necessary diagnostic procedures with patient parameters. The complexity of obtaining a mathematical model is due to its dimensionality, multicriteria and uncertainty in a number of states, and today it is solved through the usage of expert opinions or a priori protocols for diagnosis and treatment.

References

1. Головач Ю.Ю. Підходи до організації роботи програмного комплексу підтримки прийняття рішень при лікуванні раку легенів/ С.Б. Артамонов, Ю.Ю. Головач // Вісник інженерної академії України. – 2018. – № 1. – С. 128-134.
2. Bulletin of National Cancer Registry of Ukraine 2018; Available from: http://www.ncru.inf.ua/publications/BULL_19/index_e.htm

3. Cancer Research UK survival data. 2010; Available from: <http://info.cancerresearchuk.org/cancerstats/types/lung/survival/index.htm#source14>
4. Cancer Research UK. Cancer Stats. 2010; Available from: <http://info.cancerresearchuk.org/cancerstats/types/lung/index.htm?script=true#mortality>. Crawford, S. M., Sauerzapf, V., Haynes, R., Zhao, H., Forman, D., Jones, A. P. Social and geographical factors affecting access to treatment of lung cancer. *Br J Cancer*. 2009 Sep 15;101(6):897-901.
5. Artamonov E.B. Concept of creating a software environment for automated text manipulation. // Artamonov E.B., Zholdakov O.O. – Scientific journal “Proceedings of the National Aviation University”. – K.: NAU. – 2010. – Вип. 3 (44). – P. 111-115.
6. Артамонов Є.Б. Формування адаптивних динамічних сценаріїв у комп’ютерних системах навчання / Артамонов Є.Б., Панфьоров О.В. // Технологічний аудит та резерви виробництва. – Х.: НТУ «ХП». – 2016. – № 6/1(32). – С. 66-71. (DOI: 10.15587/2312-8372.2016.86342). – <http://journals.uran.ua/tarp/issue/viewIssue/5179/1884>