Mathematical Model for Clinical Decision Support System Using Genetic Algorithm

Zhenisgul Rakhmetullina
D. Serikbayev East Kazakhstan technical university
Oskemen, Kazakhstan
Zhrakhmetullina@ektu.kz

Raushan Mukhamedova
D. Serikbayev East Kazakhstan technical university
Oskemen, Kazakhstan
Rmukhamedova@ektu.kz

Roza Mukasheva
D. Serikbayev East Kazakhstan technical university
Oskemen, Kazakhstan
Mukashevaru@mail.ru

Elvira Aitmukhanbetova
Astana IT University
Nur-Sultan, Kazakhstan
E.aitmukhanbetova@gmail.com

***Abstract*— Genetic algorithms are used as a basis for many modern decision support systems. With their help it is possible to find the solutions in a shorter time in tasks with limited time. The paper describes the main principles of medical technology. The paper describes the main principles of medical technology, the structure of the medical process for the diagnosis of anemia, which occupies a prominent place in the structure of diseases. The modeling of medical and technological processes for supporting the adoption of medical solutions is described, a mathematical model of decision rules based on algorithms of the genetic type is developed. An algorithm for controlling the technological processes of screening and classification of anemia is proposed.**

Keywords*— medical technology, clinical decision support system, medical-technological process, genetic algorithm, diagnostics, anemia, symptoms of diseases, model of diagnosis, synthesis of decisive rules.

INTRODUCTION

Genetic algorithms belong to the group of optimized and finding algorithms. First works on genetic algorithms were published in 1975 [1], from that moment it is possible to talk about them as about perspective and progressive way to solve many science and technological issues.

Genetic algorithms are widely used in medicine. Most of the medical conclusions can be formulated as a searching for the potential solutions for tasks in specific field [2]. For instance, radiologist, while planning the treatment for patient, tries to find the it among all the possible ways of treating. The search area in medicine is very big and complicated [3]. The solutions depend on clinical tests, that include vast amount of data. Based on this data, it is required to make only one decision. Evolutionary algorithms are used in medicine for making decisions that can be devided into three groups: the study of data: to diagnose and predict; generating and conversion of signals; planning the treatment [4].

Data mining is the process of searching for patterns and trends by analyzing large amounts of data [5]. In medical data analysis, evolutionary algorithms are usually used to find the value of parameters. Many medical data are expressed using images or other signals. Evolutionary algorithms are also used here to correct activity signal processing algorithms by finding their optimal parameters [6]. They can also be used to directly obtain useful information based on the provided data. Evolutionary algorithms are particularly well suited for solving scheduling problems, for example, for a patient undergoing various medical procedures and requiring the advice of various specialists in order to optimize patient waiting times and equipment usage [7-10].

I. MATHEMATICAL MODEL OF DECISION RULES BASED ON ALGORITHMS OF GENETIC TYPE

Decision rule is synthesized, which makes it possible to obtain the general confidence coefficient for a given characteristic subspace from the partial confidence coefficients CC^*w_l. In one of the variants of the classification of anemia, an expression of the type (1) can be used as a carrier (1):

$$Y = \sum_j \alpha^j_{\omega_1} KY^*_{\omega_1}(j), \qquad (1)$$

where a^i_{wl} - is the weight coefficient that determines the contribution of the j-th component to the general solution of the hypothesis ω1. the overall confidence coefficient in the hypothesis ω1 is determined from the relation (2):

$$CC^O_{\omega_l} = \mu_{\omega_l}(Y) \qquad (2)$$

To synthesize the block of fuzzy inference at the second level, the decision rules based on algorithms of the genetic type [11] were used.

As an example, a five-component vector Y is selected, which is present at the output of the first level aggregators. In the fuzzy inference module of the model in Fig. 1, N-1 fuzzy operations r^i_{wi} are performed on the components of

978-1-7281-9090-7/20/$31.00 ©2020 IEEE

the vector Y, where N is the dimension of the vector Y; i is the ordinal number of the fuzzy operation in the vector of fuzzy operations R_{wi}. As components of the vector R_{wj}, the numbers corresponding to fuzzy operations codes introduced over fuzzy sets are used [12].

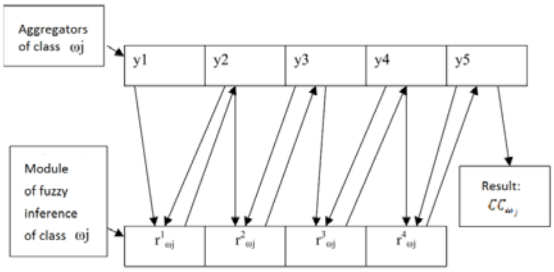

Fig. 1. Model of the second level fuzzy output module

If g fuzzy operations are entered, then, encoding them with decimal numbers from 0 to g-1, we get the model of the fuzzy output module of the second level in the form of a chromosome, the number of genes in which is one less than the dimension of the input vector Y, and the number of alleles is equal to the number of introduced fuzzy operations g on these fuzzy sets. Each gene is an operation code over two fuzzy numbers, and its allele interacts with two components of the vector Y. As a result of this interaction at the i-th position of the vector Y, according to the scheme in Fig. 1, a new fuzzy number is obtained, determined by the recurrence relation (3):

$$Z=r_i(y_i,y_{i+1}); \quad y_{i+1}=Z \qquad (3)$$

where r_i is one of the fuzzy operators corresponding to the allele of the gene standing on the i-th place in the chromosome; y_{i+1} –(i+1)th component of the vector Y; yi is the i-th component of the vector Y; Z is an intermediate variable. It is established that for each class there is some optimal set of alleles from the point of view of minimization of errors of the first or second kind or their ratio, which can be selected in the learning process realized according to the rules adopted in genetic algorithms. The choice of the optimal chromosome structure for a particular class ω_j on training samples consisting of P vectors, in the presence of m possible classes, is proposed by minimizing the criterion represented as the following expression (4)

$$\frac{1}{F_{\omega j}}\beta_{\omega j}\sum_{t=1}^{F_{\omega j}}\left(1-K_{t\omega_j}\right)+\sum_{\substack{j=1\\ j\neq\omega_j}}^{m}\frac{1}{F_j}\beta_j\sum_{t=1}^{F_j}\left(K_{tj}\right)=min \qquad (4)$$

where K_{ij} is the confidence factor in the diagnosis of the j-class for the i-th vector from the training sample of the j-th class, given out by the second-level fuzzy output module; The βj-normalizing coefficient is less than one, taking into account the nonequivalence of the errors of the first and second kind. The index at P points to the training sample of the corresponding class. K_{ij} is calculated by the interaction of the chromosome $R_{\omega j}$ with the i-th vector of the training sample of class j by expression (5):

$$K_{tj}=r_{(N-1)_j}\left(\left(\dots r_{2j}\left(r_{1j}(y_{t1},y_{t2}),y_{t3}\right),\dots\right),y_{tN}\right) \qquad (5)$$

The task in synthesizing the module of fuzzy inference is to determine for each class a chromosome with the maximum fitness function, using the technology of selection and mutations. Because operation (5) is not commutative in the general case, then in the genetic algorithm the mutation process is realized by permutations of the components of the vector Y.

II. ALGORITHM FOR MANAGING TECHNOLOGICAL PROCESSES FOR SCREENING AND CLASSIFYING ANEMIA

For the construction of a decision support system, an algorithm for controlling the processes of screening and classification of anemia was developed, which uses erythrocyte indices: γ - color index, MCV - red blood cell volume, MCH - hemoglobin content in erythrocytes, De - erythrocyte diameter, RDW - erythrocyte volume deviations [13-15].

The scheme of the algorithm for controlling the screening and classification of anemia in five classes is shown in Fig. 2. Calculation of erythrocyte indices is carried out in block 2. Since according to the automated blood analysis data both screening diagnostics and classification of anemia can be carried out, then in the 1st stage in block 3 the mode of operation of the automated system is selected: screening or classification and differentiated diagnostics.

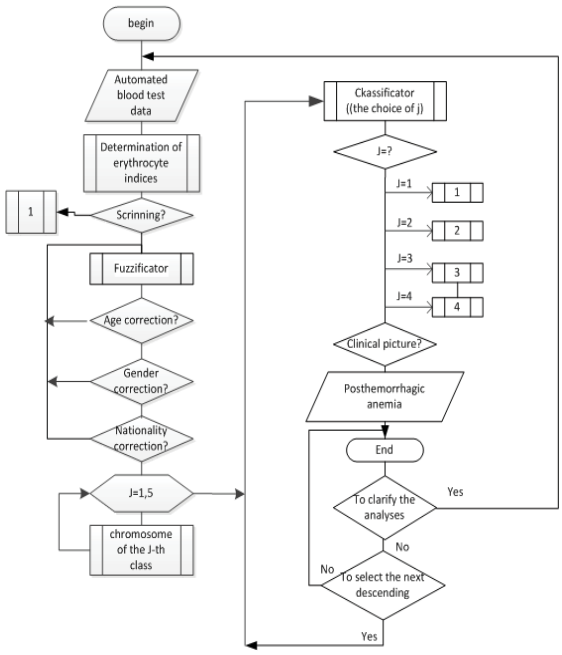

Fig. 2. Algorithm for controlling the process of screening and diagnosis of anemia

When working in the classification mode and differentiated diagnostics in block 4, there is a transition from clear values of erythrocyte indices to their accessory functions.

Since the norm-pathology border in erythrocyte indices is not unambiguous, the control algorithm introduces the possibility of correcting the functions of accessories by age, sex and nationality (blocks 5, 6 and 7). As a result of the operation of block 4 at the input of block 8, we have a fuzzy 5-element vector $\{\mu 1(\gamma), \mu 2(MCV), \mu 3(MCH), \mu 4(De), \mu 5(RDW)\}$.

This vector is subsequently converted into 5 scalar quantities by interacting with 5 operators (chromosomes) in block 9, according to the scheme in Fig. 2. As a result of the interaction at the input of block 10, 5 values of confidence coefficients for 5 types of anemia are formed (6):

$$CCj=R^{j}\{\mu 1(\gamma), \mu 2(MCV), \mu 3(MCH), \mu 4(De), \mu 5(RDW)\}. \quad (6)$$

where j is the anemia class, R_j is the operator corresponding to the chromosome of the j-th class. Block 10 searches for j, which corresponds to the maximum value of KUj. Block 11 transfers control depending on the determined value of j to one of the five algorithms that we developed that implement the two-level classification scheme proposed by us, in which case the data obtained from erythrocyte indices, according to the scheme of the algorithm in Fig. 2, are the results obtained one by one from the characteristic subspaces. The experimental research database model is shown in Figure 3.

In determining the stages of iron deficiency (ID) anemia $(j = 1)$, subspaces synthesized as a result of instrumental studies, analysis of the clinical picture, socio-environmental factors were used. In the diagnosis of megaloblastic anemia $(j = 2)$, a feature subspace is used, the structure of which was synthesized interactively.

In the diagnosis of hemolytic anemia (HA) $(j = 4)$, signs of accelerated destruction of erythrocytes and signs of enhanced erythropoiesis were used as indicative subspaces. In the algorithm for diagnosing aplastic anemia (AA) $(j = 5)$, peripheral blood and bone marrow indices were used as additional indicative subspaces [16].

If $j=3$, that is, there is an assumption about posthemorrhagic anemia (PHA), clinical manifestations of hemorrhage are revealed and, if they are present, a PHA diagnosis is made. In the absence of clinical manifestations either refines the data of laboratory analysis, or stops working with the automated classification system and to clarify the diagnosis, analyze data on other systems: bone marrow, biochemical analysis, and instrumental research.

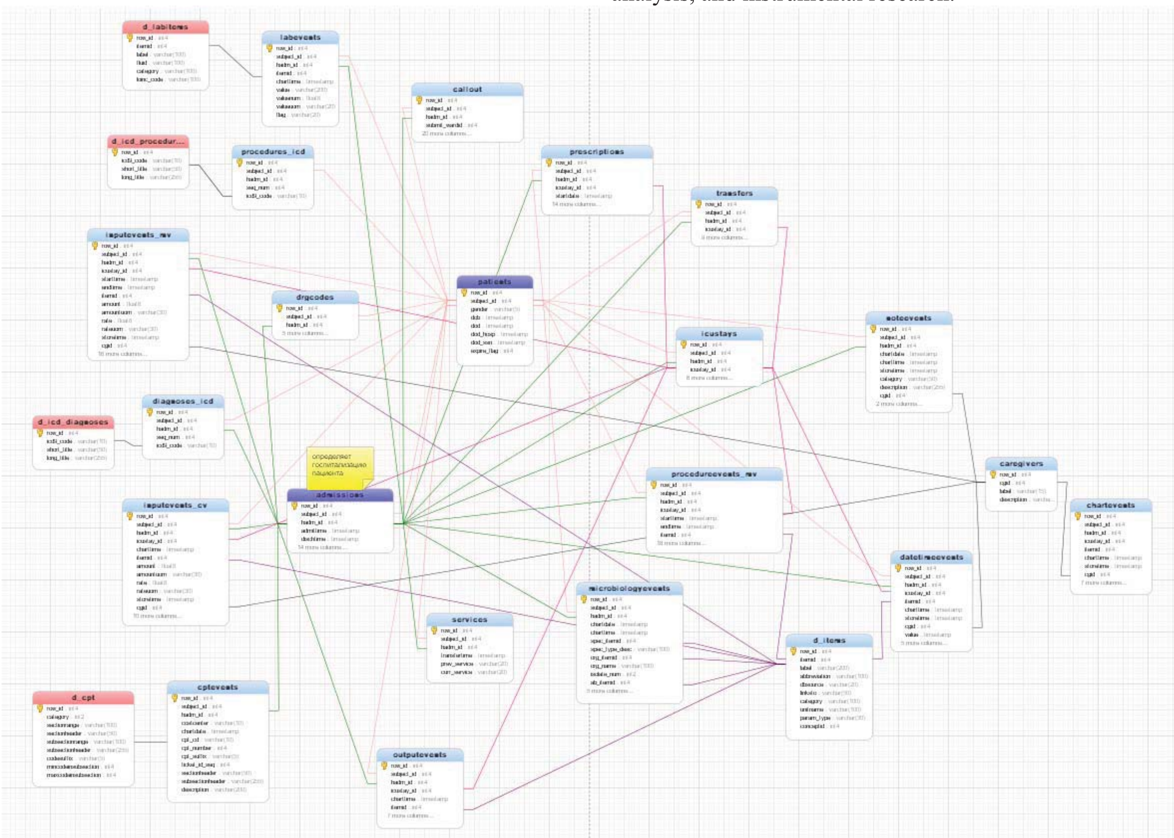

Fig. 3. The structure of the experimental database

III. Result of The Experimental Study

For the training of the system, data from the case histories of the Regional Clinical Hospital were used. At the first stage of training, the synthesis of chromosomes for the classification of anemia was carried out using erythrocyte indices. As a result, for each type of anemia, a chromosome is obtained, which describes a fuzzy output module for an anemia of a particular type. The quality of the classification of anemia was checked by the method of the sliding examination (Table 1).

TABLE I. The quality of anemia classification by the method of the sliding examination

Type of anemia	IDA	B12	PHA	HA	AA	Total	Sensibility
IDA	150	17	11	2	0	180	0,833
B12 deficiency	19	118	4	0	9	150	0,787
PHA	12	4	32	1	1	50	0,64
HA	2	0	3	25	0	30	0,833
AA	0	4	1	1	34	40	0,85
Specificity	0,878	0,917	0,953	0,99	0,976		

Statistical studies showed that the increase in the incidence of anemia is due to iron and B12 deficiency anemia, which allowed for the introduction in the online mode of correction in the synthesis of crucial rules for reducing errors of the 1st and 2nd kind. Correction according to statistical data was carried out by introducing into the minimizing functional (4) the weight coefficients β, which masked the errors in the diagnosis of statistically less frequent types of anemia.

When classifying by types of anemia, the system was tested as a single-level system with the synthesis of decision rules by using a genetic algorithm. Two-level classification system was tested for differentiation of megaloblastic anemia. To this end, at the 1st level, three indicative subspaces were used: data from biological analysis, internal risk (history, constitution) and external risk (environmental factors). Using expressions of the type (1) and (2), the integral carriers SB, SA and SC were defined for the membership functions that determine the confidence in making decisions on these attribute groups.

Using the genetic algorithm, two chromosomes were obtained for the second level fuzzy output module [17,18]. To optimize the structure of the decision blocks at the 1st and 2nd levels, imitational modeling of the output vectors of the 1st level was used, which allowed training the fuzzy control unit of the 2nd level regardless of the 1st, simulating the bimodal distribution at its input with a given degree of expression, using for this beta distribution. The results of differentiation of megaloblastic anemia, see Table. 2.

TABLE II. Analysis of the quality of differentiation of megaloblastic anemia by the method of a sliding examination

Type of anemia	Folic deficiency	B12 deficiency	Total
Folic deficiency	165	9	174
B12 deficiency	4	138	142
Specificity	0,971831	0,948275862	

To clarify the reliability of the work of the decisive rules, they were checked on control samples, the volume of which was determined in accordance with the recommendations adopted in medical practice [16].

For the problem of diagnosing IDA anemia, the volume of the control sample was determined nW = 115 people, for the diagnosis of B12-deficient anemia, the value nB12 = 121 was obtained; for the diagnosis of hemolytic anemia, nA = 128; for the diagnosis of aplastic anemia, nAA = 125 human. An analysis of the quality of the work of the obtained decision rules in the control samples showed that when the IDA was classified on the control sample, the classification error level did not exceed 5% (the probability of a correct classification of RZD = 0.95 with an expert confidence estimate of 0.955).

Conclusion

Medicine was one of the first industries where genetic algorithms were introduced as part of commercial diagnostic systems. In various fields of medicine, the use of genetic algorithms makes a significant contribution, both to the modernization of the medical industry and to the development of the theory as a whole. Thus, within the framework of this study, the process of decision-making in diagnosing anemia was described, a mathematical model of decision rules based on genetic type algorithms was developed, the stages of anemia diagnosis based on the method of synthesis of decision rules, an algorithm for controlling the screening and anemia classification processes, experimental research. When determining the risk of hemolytic anemia, the magnitude of the error did not exceed 7% (RGA = 0.93 with an expert confidence estimate of 0.96). The results of the expert evaluation are quite close to the results obtained on the control sample.

References

[1] Goldberg, D. E., & Holland, J. H. Genetic algorithms and machine learning. Machine Learning 3. Kluwer Academic Publishers - Manufactured in The Netherlands, 1988, pp. 95-99.

[2] Groenier M., Pieters J., Miedema H. Technical Medicine: Designing Medical Technological Solutions for Improved Health Care. Medical Science Educator 2017; Volume 27(4): pp. 621-631.

[3] Bountris, Panagiotis et al. "Development of a clinical decision support system using genetic algorithms and Bayesian classification for improving the personalised management of women attending a colposcopy room." Healthcare technology letters vol. 3,2 143-9. 14 Jun. 2016, doi:10.1049/htl.2015.0051

[4] Pirpinia K. et al. Evolutionary Machine Learning for Multi-Objective Class Solutions in Medical Deformable Image Registration //Algorithms. – 2019. – T. 12. – №. 5. – C. 99.

[5] Song J., Kim K., Lee M. A Big Data Analysis and Mining Approach for IoT Big Data //International Journal. – 2018. – T. 7. – №. 1.

[6] Groenier M., Pieters J., Miedema H. Technical Medicine: Designing Medical Technological Solutions for Improved Health Care. Medical Science Educator 2017; Volume 27(4): pp. 621-631

[7] Poongothai S., Dharuman C., Venkatesan P. A Comparative Study of Hybrid Evolutionary Based Algorithms with Machine Learning Classifiers for the Prediction of Medical Database //Journal of Physics: Conference Series. – IOP Publishing, 2019. – T. 1377. – №. 1. – C. 012026.

[8] Ali M., Siarry P., Pant M. Multi-level image thresholding based on hybrid differential evolution algorithm. application on medical images //Metaheuristics for Medicine and Biology. – Springer, Berlin, Heidelberg, 2017. – C. 23-36.

[9] Dey N., Ashour A. S. Computing in medical image analysis //Soft computing based medical image analysis. – Academic Press, 2018. – C. 3-11.

[10] Liu P. et al. Deep Evolutionary Networks with Expedited Genetic Algorithms for Medical Image Denoising //Medical image analysis. – 2019. – T. 54. – C. 306-315.

[11] I. Uvaliyeva, S. Belginova and A. Ismukhamedova, "Development and implementation of the algorithm of differential diagnostics," 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, 2018, pp. 1-6. doi: 10.1109/ICAICT.2018.8747116

[12] Adlassnig K.-P. Fuzzy Methods in Clinical Research and Patient Care. Procedia Computer Science 102 (12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016. Vienna, 2016. pp. 9-

[13] S. Belginova, I. Uvaliyeva and A. Ismukhamedova, "Decision support system for diagnosing anemia," 2018 4th International Conference on Computer and Technology Applications (ICCTA), Istanbul, 2018, pp. 211-215. doi: 10.1109/CATA.2018.8398684

[14] Zhang Y., Qiu M., Tsai C., Alamri A. Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data. IEEE Systems Journal 2017; Volume 11(1): pp. 88-95.

[15] Davvaz B., Hassani Sadrabadi E., Nieto J., Torres A. Twin Hypercube for Intuitionistic Fuzzy Sets and Their Application in Medicine. International Journal of Analysis and Applications 2017; Volume 15(1): pp. 31-45.

[16] Uvaliyeva I., Belginova S., Ismukhamedova A. Informational and analytical system to diagnose anemia. Proceedings of the Fourth International Conference on Engineering & MIS 2018. Istanbul, 2018. pp. 1-7.

[17] Amirkhani A., Papageorgiou E., Mohseni A., Mosavi M. A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications. Computer Methods and Programs in Biomedicine 2017; Volume 142: pp. 129-145.

[18] Aramideh J., Jelodar H. Application of Fuzzy Logic for Presentation of an Expert Fuzzy System to Diagnose Anemia. Indian Journal of Science and Technology 2014; Volume 7(7): pp. 933-938.