# Intelligent Processing of Medical Information for Application in the Expert System

E. V. Chebanenko[1], L. A. Denisova[2]
Omsk State Technical University
Omsk, Russia
[1]evchebanenko@gmail.com, [2]denisova@asoiu.com

***Abstract*— The article presents an approach to processing and analyzing medical information to identify groups of patients by their compliance rate to therapeutic intervention. The data was analyzed using such intellectual information technology as a method of fuzzy clustering; it is applied to weakly formalized initial data. There were used results of empirical study on quality of life of patients with hypertensive disease and factors of their compliance rate. Compliance rate to therapeutic interference was estimated at quantitative equivalent by three criteria. There were considered criteria of the expected effectiveness of medical support and lifestyle modifications. Also there were discussed criteria of the expected effectiveness of drug therapy. The result of this research is to determine clusters for categories of patients who have similar type compliance of the therapy in the aggregate by three criteria. Moreover, based on the clusters analysis results, functions of patients belonging to groups with the most similar type of compliance rate were obtained. Finally, there were examined the development prospects of advisory expert systems for a cardiologist when prescribing drug therapy.**

***Keywords*— *criterion of the medical support effectiveness; compliance to treatment; method of fuzzy clustering; expert system***

## I. INTRODUCTION

At present medical expert systems based on modern information technologies find broader application as in medical studies as in real clinical practice [1-3]. Medical organizations develop new approaches to therapeutic procedures purpose and support. Collected and generated data require fast and effective processing.

Medical community is noted to pay much attention to issues of patients' compliance rate to medical support as it influences the treatment quality and result. Moreover, it is relevant as developing predictive models of diseases as working out ways of influencing the patients' attitudes to the applied therapy. Also it is important to develop approaches to estimate effectiveness and rationality of medical interference [1–3].

The article examines an approach to processing and analyzing data about patients' compliance rate to therapeutic interference through applications of fuzzy logic mathematical tool [4–14]. Fuzzy intellectual technologies of knowledge processing and using are usually applied when there are uncertainties and incompleteness of knowledge about object, unclear descriptions.

A method of fuzzy clustering [6–10, 13, 14] is used, its possibilities are revealed being applied to weakly formalized initial data. The problem of clustering is known to be in dividing extensive set of objects onto separate groups with common properties. The peculiarity of the algorithm of fuzzy clustering (if compare with clear methods) is that obtained due to objects dividing clusters correspond to fuzzy sets. The same time an element of the sampling can relate to several clusters with a certain degree (in contrast with a clear clustering where an element belongs only to one cluster).

Advantage of using fuzzy clustering method is that there is possibility of quantitative estimating level of patients' belonging to groups when prescribing medical support and obtaining information for additional analysis.

## II. THE PROBLEM STATEMENT

Numerous clinical researches in the therapy of patients with chronic diseases prove that one of important issues is the patient's attitude to the ongoing therapy, value of positive and negative effects, and, finally, to the treatment result [1, 3].

For ascertaining the issue since 2003 professor N.A. Nikolaev has carried out empirical study on learning life quality of the patients with hypertensive disease and factors of their compliance rate to the treatment [1]. The patient's compliance rate to constant drug therapy and expected efficiency of the treatment was predicted on the questionnaires results.

Individual features of compliance rate of the patients with arterial hypertension may be identified, quantitatively estimated and used for prediction and monitoring of the expected effectiveness of the treatment.

Indices of the expected effectiveness offered by N.A. Nikolaev are used [1]. Following the methods one can estimate compliance to therapeutic interference in a quantitative equivalent by three criteria:

- expected efficiency of life image modifications ($I_{EUWL}$);
- expected efficiency of medical therapy ($I_{EMT}$);
- expected efficiency of medical support ($I_{EMS}$).

The study subject is to identify categories of patients who have the most similar type of therapy compliance in the aggregate by three criteria of the expected effectiveness.

**978-1-5386-5824-6/18/$31.00 ©2018 IEEE**

The data analysis of patients' compliance rate to therapeutic interference using method of fuzzy clustering has been carried out. Moreover, based on the cluster analysis results, functions of patients' compliance to groups with the most similar type have been determined.

Finally, there were considered development prospects of advisory expert systems for a cardiologist when prescribing drug therapy.

## III. THEORY. AN ALGORITHM IDENTIFYING PATIENTS' GROUPS BY COMPLIANCE RATE TYPE

To solve the set task of identifying groups of patients having the most similar properties of compliance rate and considering all three criteria, an algorithm of fuzzy c-average is used.

The peculiarity of the fuzzy *c*-average algorithm is that data are represented as fuzzy sets, and each element can belong to several clusters with a certain degree of belonging [10, 13 and 14].

Regarding the task, the cluster structure is a matrix of fuzzy partitioning of patients:

$$M = \left[\mu_{ij}\right] = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2c} \\ \dots & \dots & \dots & \dots \\ \mu_{l1} & \mu_{l2} & \dots & \mu_{lc} \end{bmatrix} \quad (1)$$

Where every *i*-th line ($i = \overline{1,l}$) expresses the degree of belonging of the *i* -th patient to the cluster $A_j$ ($j = \overline{1,c}$), formed according to the degree of compliance for each of three criteria. So, *l* is a number of patients, and *c* is a number of clusters. Compliance rate is $\mu_{ij} \in [0,1]$. The details of the implemented algorithm are considered.

*Step 1.* The number of clusters is determined (one considers *c*=3, interprets initially clusters as groups according their compliance) and a weighting coefficient of the elements (one considers *m*=2, then the parameter is used for calculating clusters center). Also, an algorithm termination condition is considered, starting from the specified accuracy of the solution (one considers $\varepsilon$=1e-5).

*Step 2.* The fuzzy partition matrix *M* based on the following considerations is generated. All objects are to be distributed for each cluster according to the equation

$$\sum_{j=1,c} \mu_{ij} = 1,\ i = \overline{1,l},$$

moreover, no cluster is an empty set or contains all the elements:

$$0 < \sum_{i=1,l} \mu_{ij} < l,\ j = \overline{1,c}.$$

*Step 3.* Cluster centers are calculated with an equation:

$$C_j = \sum_{i=1,l} ((\mu_{ij})^2 M_i) / \sum_{i=1,l} (\mu_{ij})^2,\ j = \overline{1,c},$$

where $M_i \in A_j$ is i-th object of the j-th cluster.

*Step 4.* Distance between objects from the matrix *M* and clusters centers is calculated. With respect to the data under study, each row of the matrix contains the compliance indices of the individual patient; each column of the matrix contains value of the indices $I_{EUWL}$, $I_{EMT}$, $I_{EMS}$, respectively.

When determining the distance between objects, the Euclidean metric is used, and it is geometric distance:

$$D_{ij} = \sqrt{\left\|M_i - C_j\right\|^2};\ j = \overline{1,c};\ i = \overline{1,l}.$$

*Step 5.* Fuzzy partitioning matrix elements of patients are calculated

$$\mu_{ij} = \left( D_{ji}^2 \times \sum_{k=1,c} \frac{1}{D_{ki}^2} \right)^{-1}.$$

*Step 6.* Matrices of the fuzzy partitioning of patients of this and previous steps are compared, provided $\left\|M - M'\right\| < \varepsilon$, the algorithm is completed, otherwise there is transition to *Step 3*. So $M'$ is a matrix of fuzzy partitioning on the previous iteration of the algorithm.

To carry out the cluster analysis the software procedure *fcm* of computational tools MATLAB / FuzzyLogicToolbox [10-12] was used.

## IV. RESULTS OF EXPERIMENTAL STUDIES

### A. Initial data

The initial data are formed on the basis of Professor N.A. Nikolayev's questionnaire [1], it helps to estimate such a qualitative parameter as compliance in a quantitative equivalent. The study examined the criteria for the expected effectiveness of medical support, lifestyle modification and drug therapy.

The expected effectiveness of the treatment was measured quantitatively. At the same time its values in the range 1-1.99 points indicate good efficiency, 2-3.99 points indicate satisfactory efficiency, and 4 points or more indicate an unsatisfactory expected effectiveness of interventions.

The study included 160 patients with cardiovascular diseases from 40 to 85 years of age. Data for the study group of patients are partially presented in Table 1.

TABLE I. TABLE OF COMPLIANCE INDICES TO THERAPY

| **Patients** | $I_{EMS}$ | $I_{EUWL}$ | $I_{EMT}$ |
|---|---|---|---|
| 1 | 1,57 | 1,24 | 1,50 |
| 2 | 1,57 | 1,69 | 1,19 |
| 3 | 4,59 | 3,46 | 2,00 |
| 4 | 1,50 | 1,38 | 1,64 |
| … | … | … | … |
| 160 | 1,36 | 3,00 | 1,07 |

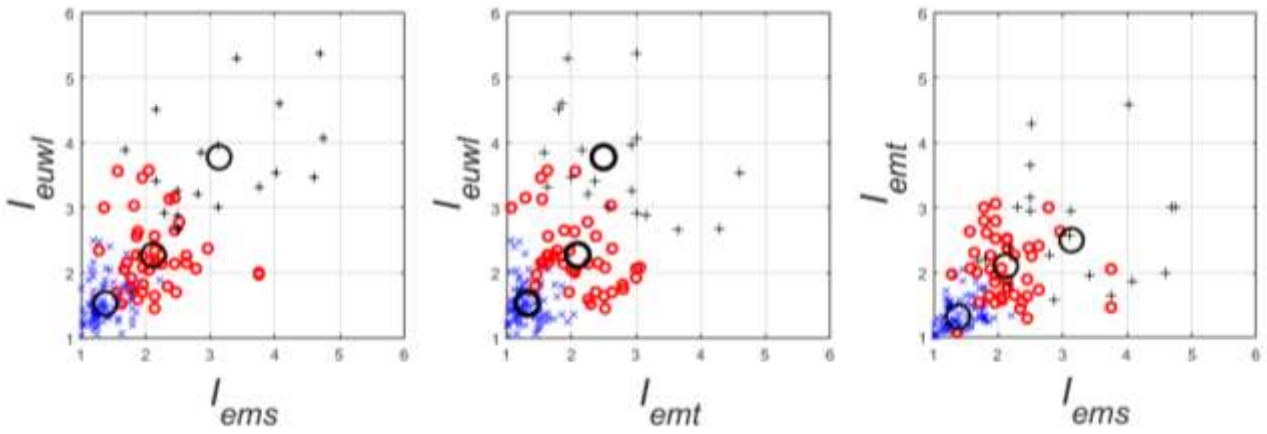


Fig. 1. Results of clustering by expected criteria. Degrees of the expected compliance to therapy are indicated: x is high, o is satisfactory, + is unsatisfactory; symbol o means clusters center

### *B. The results of cluster analysis*

For cluster analysis, individual characteristics of patients' compliance rate (to be identified and quantified) were used for grouping by the expected effectiveness of treatment. The results of the cluster analysis are shown in Fig. 1.

This approach demonstrates possibility of identifying a group of patients with similar criteria, to classify it without preliminary information; and later to develop functions of belonging and due to them draw a conclusion about new patients' compliance.

### *C. Compliance functions extraction*

Based on cluster analysis and identified clusters centers there are developed functions of terms belonging to patients with varying degrees of compliance to treatment (Fig. 2).

There are identified three groups of patients with clearly identified belonging to a certain kind of treatment. According to methodology [1] out of expected efficiency criteria (indices) index $I_{EMT}$ has the most important value; index $I_{EUWL}$ has an average one, and index $I_{EMS}$ has the least value.

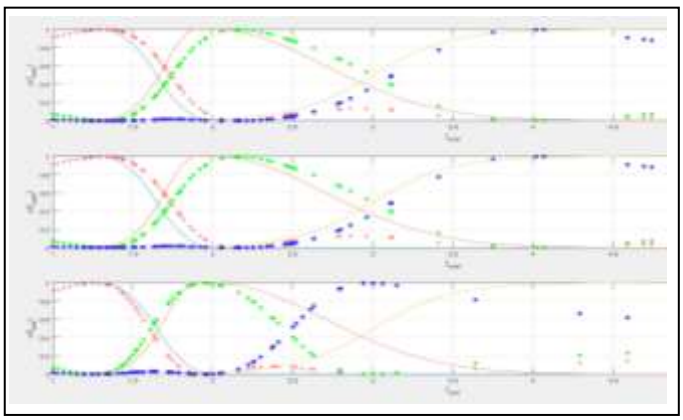

Fig. 2. Functions of belonging by expected efficiency criteria (based on the clustering results)

Every fuzzy cluster can be mapped to a fuzzy rule [9, 12 and 13], and it is to be performed in future for developing an expert system. Accordingly, the degree of compliance of every input variable (i.e. value of the compliance index) will be considered when forming every fuzzy rule.

Functions of terms belonging are obtained by projecting the corresponding cluster belonging degrees, i.e. projecting of matrix rows of fuzzy partitions $M$ (it is obtained according to equation (1)) on the axis of expected efficiency indices.

## V. THE DISCUSSION OF THE RESULTS

The proposed fuzzy system can be used to estimate the level of patients' compliance in primary treatment, as a consultative support system.

Cluster analysis for processing quantitative estimation of patients' compliance is an initial stage for developing predictive models of the proposed therapy and predicting outcomes and developing an expert consultative system for prescribing patient-centered therapy by a cardiologist.

Analyzing the results of clustering, the following groups of data to apply additional analysis are to be considered, and, as a result, to determine recommended settings for prescribing medical support:

- "disputable" objects that have an equal degree of belonging to several clusters;
- "emissions", that is, objects with extremely low compliance to all available clusters;
- objects directly belonging to the cluster, but with a pronounced predominance of one or two indices. For example, a patient has high compliance to the drug therapy, but the same time he has extremely low compliances to modify lifestyle and medical support.

Based on the obtained results, it seems expedient to formulate and to solve the task of modernizing the realized method of fuzzy clustering, that can reflect all the specifics of the data under study, as well as to consider this specificity with subsequent extraction of fuzzy rules.

## VI. CONCLUSION

The method for identifying groups of cardiac patients according to the degree of compliance to therapeutic intervention was presented and investigated. During the study, the individual characteristics of compliance rate were considered to be identified, quantified and used to predict and monitor the expected effectiveness of treatment.

An analysis of data on patients' compliance to treatment by using the fuzzy clustering method was performed. As a result of the study, three groups of patients were formed according to the compliance rate. Information on the border sets of each criterion is obtained. The membership functions for each indicator are developed for further use in the expert system.

The use of the proposed method of processing medical information in the aggregate with techniques designed to be used in clinical practice develops intelligent systems for cardiac patients.

Such a consulting expert system of support will significantly improve the efficiency of the cardiologist when prescribing drug therapy.

## REFERENCES


[1] Nikolaev N.A. Guide to clinical trials of internal diseases. Moscow. Publ. Russian Academy of Natural History. 2015. (In Russian).

[2] Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC) J . Hypertens 25. 2007. Pp.105-1187.

[3] Gallup G. The Gallup poll: Public opinion 1978. Wilmington, Delaware. Scholarly Resources. 1979. p. XLIV.

[4] Zadeh L.A. Fuzzy sets Information and Control. 8. 1965. Pp 338-53.

[5] Sugeno M., Yasukawa T. Fuzzy-logic-based approach to qualitative modeling. IEEE Transactions on Fuzzy Systems. 1993. No. 1. Pp 7–31.

[6] Ren M., Liu P., Wang Z., Yi J. A Self-Adaptive Fuzzy c-Means Algorithm for Determining the Optimal Number of Clusters. Journal of Computational Intelligence and Neuroscience. 2016. Pp. 5-8.

[7] Kotov A., Krasilnikov N. Clustering of data. URL: http://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf. Reference date: 05.02.2017. Pp 13-5.

[8] Jiangsheng Y. Clustering Methods Applications of Multivariate Statistical Analysis. Beijing: Institute of Computational Linguistics. URL.http://www.archive.org/web/20030422001200/icl.pku.edu.cn/yujs/papers/pdf/clster.pdf. Reference date.: 07.03.2017. Pp. 15–20.

[9] Roubos J., Setnes M. 2006 Learning Fuzzy Classification Rules from Labeled Data. Delft. Delft University of Technology. URL:http://www.researchgate.net/publication/222657076_Learning_fuzzy_classificaion_rules_from_labeled_data/links/09e4150879582c877f000000. Reference date: 05.05.2017. Pp 7–13.

[10] Shtovba S. Designing fuzzy systems using MATLAB. Moscow. Hot line-Telecom. 2007. Pp. 187-223. (In Russian).

[11] Denisova L.A., Nadtochii P.N., Raskin E.M Realization of the Automatic Control System with Fuzzy-Logic Static Error Compensation in the Computer-Aided Design Environment Teprol. Automation and Remote Control. 76. No. 1. 2015. Pp 157-165.

[12] Meshcheryakov V., Denisova L. Computer-aided design of the fuzzy control system using the genetic algorithm. Dynamics of Systems, Mechanisms and Machines. Dynamics 2016. 2016.P. 7819000.

[13] Zhang M., Zhang W., Sicotte H., Yang P. A new validity measure for a correlation-based fuzzy c-means clustering algorithm. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009. P. 4 3865-8.

[14] Knyazev N.L., Denisova L.A. The detection methods of dynamic objects. J. Phys. Conf. Ser. 2018. P. 944 012051.