APPLICATION AND ANALYSIS OF PRELIMINARY PROCESSING OF MEDICAL IMAGES FOR SOLUTION OF PRACTICAL PROBLEMS

Akhram Khasanovich Nishanov[1], Gulomjon Primovich Juraev[2], Malika Khamdamjanovna Maksudova[3], Malika Akhramovna Khasanova[4]

[1]Information Technologies Software Department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan. E-mail: nishanov_akram@mail.ru

[2]Information Technologies Center at the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan. E-mail: jurayev_g@bk.ru

[3]Department of Faculty and Hospital Therapy No.2, Tashkent Medical Academy, Tashkent, Uzbekistan. E-mail: malika_aziz@mail.ru

[4]Department of Faculty and Hospital Therapy No.2, Tashkent Medical Academy, Tashkent, Uzbekistan. E-mail: malikabonuxasanova@gmail.com

The article describes the stages of the software package “Preliminary processing of medical images” (PPMI) based on logical mathematical and algorithmic software, in particular, theoretically applied in solving practical problems using mathematical models of primary processing of medical emblems. The results obtained by the software suite were then analyzed.

The software package assesses the state of medical diagnosis in the following two cases related to the field of cardiology (formation of an informational character space from a set of characteristics characterizing medical diagnostic objects in the initial processing of medical emblems and classification of medical diagnostic objects from the identified informational character space; constructing a decisive rule for determining which object belongs to a given class) and preliminary processing of medical data (preliminary processing of medical data, reclassification of selected elective objects and determination of the importance of classes, formation of reference tables, a set of informational signs distinguishing class objects) selection, determining the effectiveness of the selected set of informative features) and methods and algorithms for solving practical problems.

Keywords: proximity function, distance function, normalization of initial data values, informative features, classification, estimation calculation algorithms, initial processing of medical data, flexible genetic algorithm based on random search, sliding window method.

I. Introduction

One of the important tasks in modern medicine is the rapid introduction of modern information and communication technologies in medicine, ie the initial processing of large amounts of medical data, algorithms and software for diagnostic processes, automation of medical processes, differentiation and computer diagnostic systems.

The implementation of these important tasks primarily involves the development of methods and algorithms for the initial processing of large volumes of medical data.

Firstly, the training, formed on the basis of information provided by experts in the field, depends on the quality of the selection. This is because the effectiveness of the research work to be carried out in the next stage is based on the properties and characteristics of the objects of study selected.

Secondly, on the basis of training selection formed by experts in the field, scientific research will be conducted on the development of methods and algorithms for the initial processing of data, the improvement of existing ones. The results obtained are analyzed in conjunction with industry experts or experts and we can see that the result is an effective method or algorithm depending on whether it is positively evaluated.

Specifically, a number of studies are conducted in this area, for example, the analysis and integration of research option data [1-5], normalization of educational sample objects in terms of parameters [6-11], the formation of a reference table based on educational selection [12-18], classification, selection of an informative character set that classifies objects, determining the importance levels of class objects, characters, and classes in the informative character space [19-40].

In order to solve the above-mentioned problems, two practical issues related to the field of cardiology, such as the assessment of medical diagnoses and the initial processing of medical data, were solved, and a software package entitled “Preliminary processing of medical images” (PPMI) was created.

II. The main part

The main part consists of 2 sections: a) problem statement; b) methods and algorithms for solving problems.

a). This first section consists of two items, which are intended to address the following two practical issues in the field of cardiology, i.e the assessment of the state of medical diagnosis and the initial processing of medical

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

data: 3 issues in the first paragraph 1) Identify a set of informative signs that give a clear distinction between class X_p diseases; 2) assessment of the contribution of Class X_p objects to the formation of their class; 3) to establish a decisive rule in determining the identity of an unknown object in the diagnosis of the disease, ie which object belongs to the class of the infected person; 5 issues in the second paragraph 1) normalization of the values of the signs characterizing the class objects in the initial data; 2) transfer of the characteristics of class objects from the continuous quantitative view to 0 or 1 view in the initial data; 3) Solve the problem of classification of objects of class K_p that is, determine whether the objects divided into classes belong to their own class or another class; 4)clearly each other distinguishing $K_p, p = \overline{1,5}$ class objects selecting a set of $\ell \ll 63$ informative images; 5) to determine the quality and error of recognition of the classification of objects of K_p class;

b). The second section describes methods and algorithms for solving problems. To solve practical problems in the first case, it is necessary to classify estimates using algorithms for calculating values, including a distance function that provides the difference between two objects in space and a similarity function between an unknown object and objects of a class. Selection of a set of informative features, a decision rule for determining an unknown object when diagnosing a disease, that is, which object belongs to the class of a sick person; normalization of symbolic values of primary data in solving practical problems of the second case, determination of the level of reliability of primary data, conversion of symbolic values of objects according to Fisher's criterion from continuous quantitative representation to 0 or 1, selection of an informational set of symbols based on a flexible genetic random search algorithm. The effectiveness of the sliding window method determines the quality of recognition and the degree of classification error.

Based on the proposed theoretical data, two practical problems in the field of cardiology were solved, and at the same time a software package on "Preliminary processing of medical images" was created.

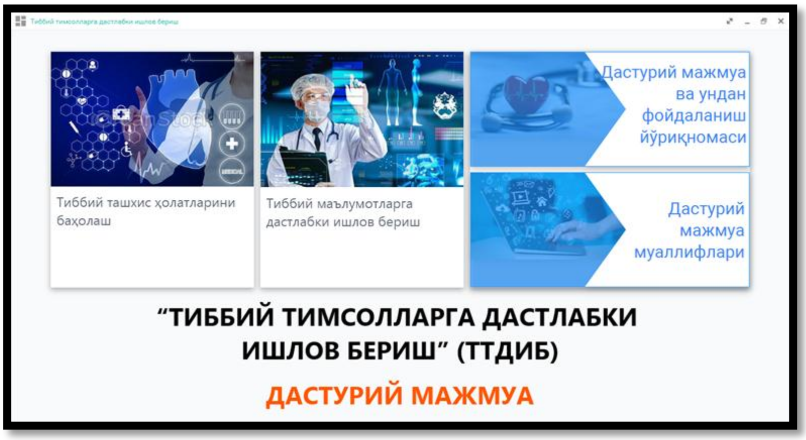

Figure 1. View of the main window of the software package.

The software package includes modules "Assessment of medical diagnostics", "Preliminary processing of medical data", "Software package and instructions for its use", "Authors of the software package".

I. The module "Assessment of medical diagnoses" works on the following principle.

a) "Preliminary data reading". The working mechanism of this module reads the initial data prepared in the Excel program, i.e. the training selection appears in the program window and indicates that the next step should be taken (Figure 1.1).

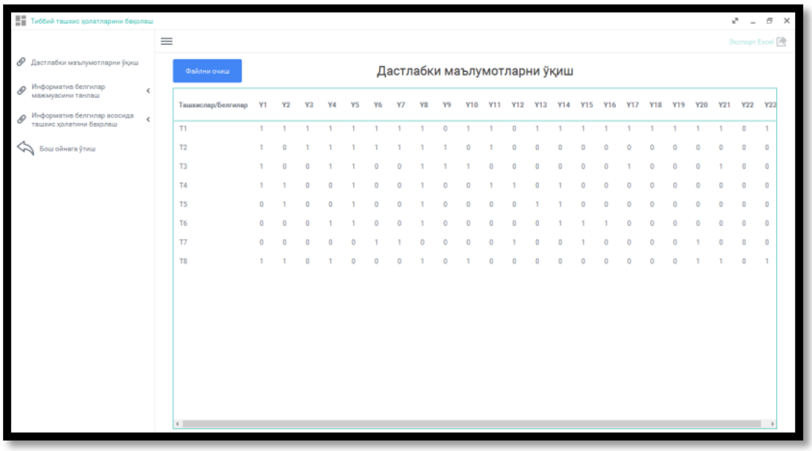

Figure 1.1. The appearance of the "Initial data reading" module window is described.

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

The activity of this module is one of the most important first steps in the formation of educational selection based on primary data. Therefore, it is necessary to determine the issue.

1.1. The statement of the problem. Suppose that the training selection set is expressed in the form $x_{p1}, x_{p2}, \dots, x_{pm_p} \in X_p, p = \overline{1,r}$. Where x_{pi}- N - is the space vector of the dimensional characters, each object is considered in the space of the dimensional characters $x_{pi} = (x_{pi}^1,\ x_{pi}^2, \dots x_{pi}^N), i = \overline{1, m_p},\ N$ - denoting $X_p, p = \overline{1,r}$ set of classes, which consists of m_p from $x_{p1}, \dots, x_{p_{m_p}}$ objects.

Problem 1.1. Within X_p class diseases, it is necessary to identify a set of informative signs that make a clear distinction between them.

Problem 1.2. An assessment of the contribution of X_p class objects to the formation of their own class is required.

Problem 1.3. In diagnosing a disease, construct a decisive rule for determining which of the unknown object, i.e., the object to which the infected person belongs, belongs to the class.

The above issues were studied in the class of diseases of the cardiovascular system "Myocardial infarction". Myocardial infarction class consists of 8 diagnoses and 23 symptoms that characterize each diagnosis.

1.2. Solving set practical problems.

The solution of the set practical problems is carried out by means of the following modules of a software package.

b) "Selection of information sets". In this module, the process of identifying a set of informative signs that give a clear distinction between X_p class diseases is carried out on the basis of the following algorithm:

Step 1. Based on the initial data generated by medical professionals, a training selection in the form of a matrix is formed. The path elements of the matrix are diagnostic objects, and the column elements are the symbols of the objects.

Step 2. The distance function between objects in this boules informational character space is introduced as $\theta_i(x_{p1,} x_{p2})$ in this boules informational character space as follows:

$$\sigma_i(x_{p1,} x_{p2}) = \begin{cases} 1 \ \ if \qquad (x_{p1}^i - x_{p2}^i) \neq 0, i = \overline{1,N}. \\ \\ 0\ if\ not\ (x_{p1}^i - x_{p2}^i) = 0, i = \overline{1,N}. \end{cases} \qquad (1.1)$$

The first condition states that there is no similarity between the two objects according to boules signs, while the second condition means that they have similarities to each other.

In this process, the problem of comparative evaluation is considered, ie the process of comparison of each diagnostic object with the remaining 7 diagnostic objects is carried out at the intersection of the column elements of the matrix (Figure 1.2).

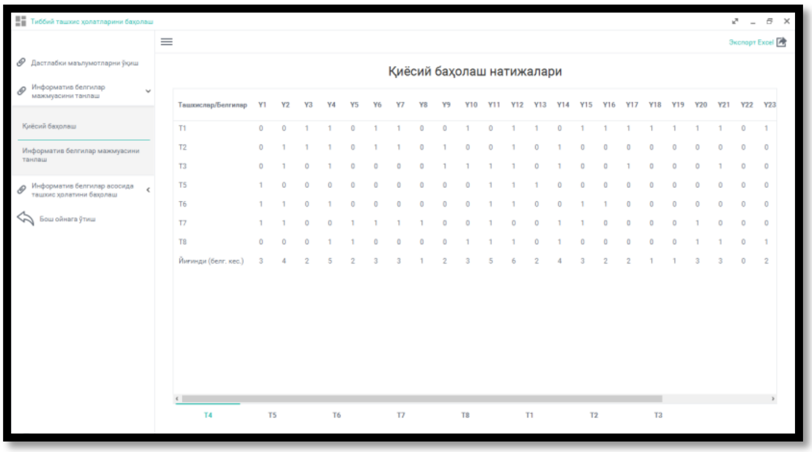

Figure 1.2. The results of the comparative evaluation are described.

Step 3. In boules informative character space, the estimation of the magnitude of the difference between an arbitrary j-diagnostic object and all other diagnostic objects is calculated on the basis of the following formula.

$$\Gamma_j(x_{pj,} x_{pk}) = \sum_{\kappa=1}^{m_p} \sum_{i=1}^{N} \theta_i(x_{pj,} x_{pk}) , j = \overline{1, m_p};\ k = \overline{1, m_p};\ j \neq k . \ (1.2)$$

Based on the given formula, the results of the comparative evaluation are calculated in the form of the sum of the column elements. This indicates the level of comparative importance of the diagnostic object mark for the individual characters relative to the $X_p, p = \overline{2,8}$; diagnostic objects, i.e., a set of informative features is defined for each diagnostic object. The most informative character is said to be the largest of the values in the cross section of the sum of the column elements. At the final stage of the process, a set of informative signs is identified for each diagnostic object (Figure 1.3).

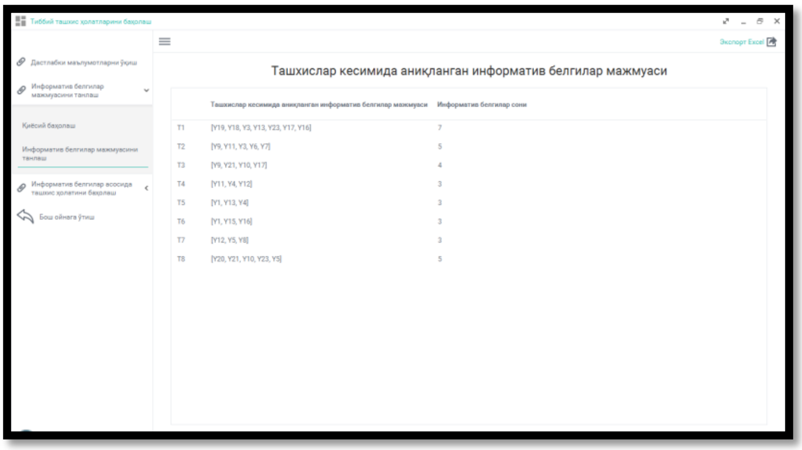

Figure 1.3. The diagnostic section describes a set of informative signs identified.

The final results of this module are presented in Table 1 below.

Table 1.

Diagnosis	A set of informative signs identified at the intersection of diagnoses	Number of informative features
T1	x19, x18, x3, x13, x23, x17, x16	7
T2	x9, x11, x3, x6, x7	5
T3	x9, x21, x10, x17	4
T4	x11, x4, x12	3
T5	x1, x13, x4	3
T6	x1, x15, x16	3
T7	x12, x5, x8	3
T8	x20, x21, x10, x23, x5	5

Table 1. The diagnostic section describes a set of informative signs identified.

c) "Assessment of diagnostic status on the basis of informative signs". This module requires you to solve problem 3 above, and the process is as follows:

1*) enter the value of the characteristics of the unknown diagnostic object (Figure 1.4);*

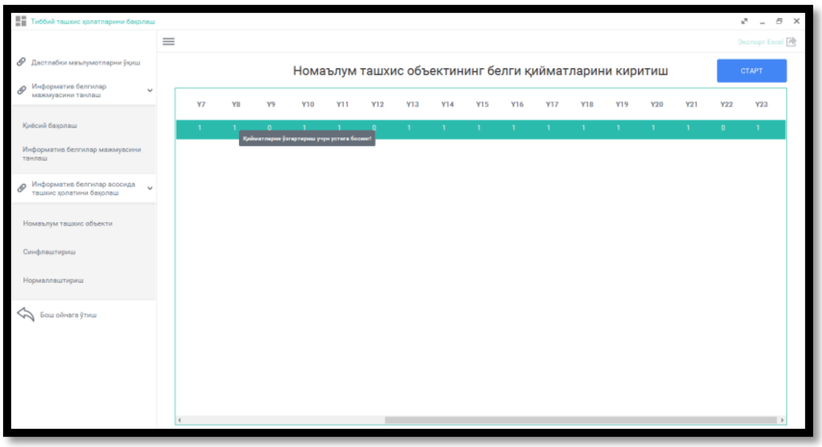

Figure 1.4. The window for entering character values of an unknown diagnostic object is described.

A study selection in the form of a matrix is formed on the basis of the identified informative signs for the class of diseases studied and the sign values of the unknown diagnostic object W entered.

2) classification. This process is implemented based on the following algorithm:

Step 1. The training, formed in the form of a matrix, is carried out separately for each type of disease in the research option, that is, the columns reflecting the identified informative signs are separated, and the rest are discarded.

The process of comparison from the resulting matrix is carried out by entering the function of proximity between objects $\rho_i(x_{p1,}x_{p2})$ in boules informative character space as follows:

$$\rho_i(x_{p1,}x_{p2}) = \begin{cases} 1 \quad if \qquad (x_{p1}^i - x_{p2}^i) = 0, i = \overline{1,N}. \\ 0 \qquad\qquad\qquad\qquad if\ not, \end{cases} \qquad (1.3)$$

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

The first condition indicates the degree of similarity between the two objects, while the second condition indicates that they differ greatly from each other, that is, these components are not similar to each other.

In the resulting matrix, when the evaluation is performed on the cross section of the road elements, the sum of the differences in the cross section of the parameters of the diagnostic object X_p relative to the unknown object W is calculated based on the following formula:

$$\Gamma_j(x_{pj,}x_{pk}) = \sum_{\text{K}=1}^{m_p} \sum_{i=1}^{N} \rho_i(x_{pj,}x_{pk}) \,, j = \overline{1, m_p};\ k = \overline{1, m_p};\ j \neq k\,. \quad (1.4)$$

We can see the classification results in the following window (Figure 1.5).

Figure 1.5. The classification results are described.

Step 2: Normalization of the obtained results with respect to the number of informative signs identified for class X_p diseases is carried outdiseases is carried out.

Step 3. Based on the final results, the unknown $w=(w^1,w^2,...,w^N)$ is determined by calculating which of the given diagnostic objects is more similar to which of the given diagnostic objects by the following formula.

$$\Gamma_w(w, x_{pk}) = \sum_{\text{K}=1}^{m_p} \sum_{i=1}^{N} \rho_i(w, x_{pk}) \,, k = \overline{1, m_p}; \quad (1.5)$$

If inequality $\Gamma_w(w, x_{pi,}) > \Gamma_w(w, x_{pj,})$ (1.6) is satisfied, then $w=(w^1,w^2,...,w^N)$ object is considered to have a higher degree of affiliation to the i-diagnostic object than others (Figure 1.6).

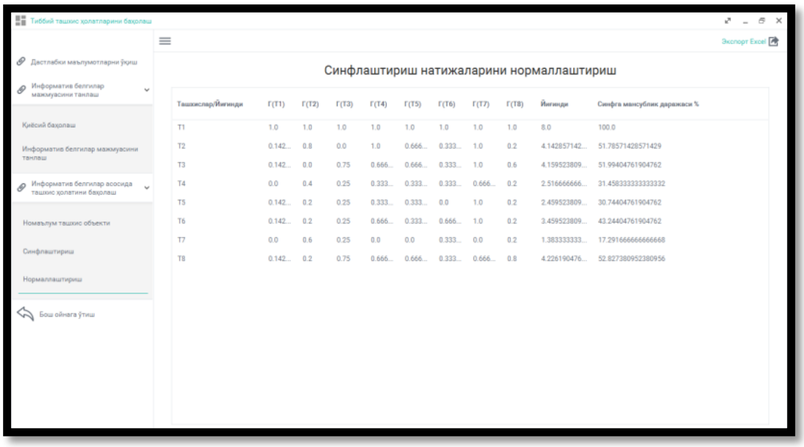

Figure 1.6. The normalization values of the classification results are described.

When using the module "Assessment of the diagnostic status on the basis of informative signs" of the software package, each diagnostic object was taken as an unknown object and tested, each diagnostic object was found 100%.

II. The module "Preliminary processing of medical data" works on the following principle.

2.1. Preliminary data reading. The working mechanism of this module reads the initial data prepared in the Excel program, i.e. the training selection appears in the program window and indicates that the next step should be taken (Figure 1.1).

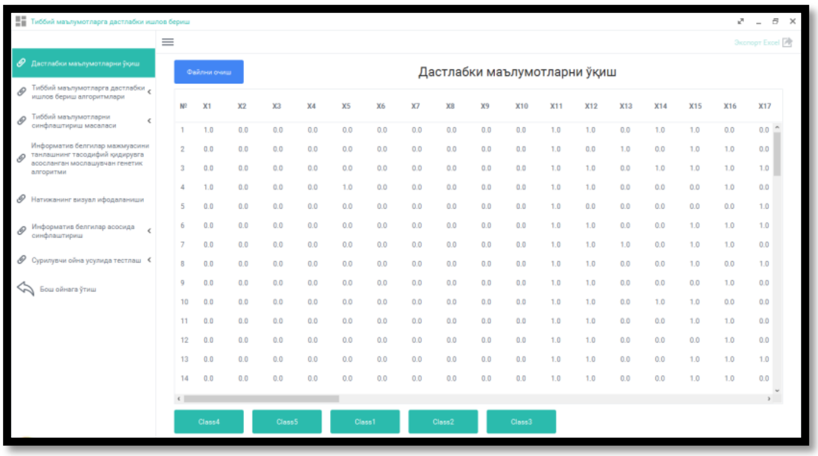

Figure 2.1. The initial data reading window is described.

2.1. The statement of the problem. Suppose that the curriculum formed on the basis of the initial data is divided into elective classes and they are given as follows:

$$K_1 = \begin{bmatrix} x_{11}^1 & x_{11}^2 & \cdots & x_{11}^N \\ x_{12}^1 & x_{12}^2 & \cdots & x_{12}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m_1}^1 & x_{1m_1}^2 & \cdots & x_{1m_1}^N \end{bmatrix} \ldots K_r = \begin{bmatrix} x_{r1}^1 & x_{r1}^2 & \cdots & x_{r1}^N \\ x_{r2}^1 & x_{r2}^2 & \cdots & x_{r2}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{rm_r}^1 & x_{rm_r}^2 & \cdots & x_{rm_r}^N \end{bmatrix}.$$

This can be expressed in general view as follows:

$$K_p = \begin{bmatrix} x_{p1}^1 & x_{p1}^2 & \cdots & x_{p1}^N \\ x_{p2}^1 & x_{p2}^2 & \cdots & x_{p2}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{pm_p}^1 & x_{pm_p}^2 & \cdots & x_{pm_p}^N \end{bmatrix}$$

Here $p = \overline{1,r}$; and research option $K = \bigcup_{p=1}^{r} K_p$ are given in the form that they consist of non-intersecting classes, ie $K_p \cap K_q = \emptyset, (p \neq q, p = \overline{1,r};\ q = \overline{1,r};)$ conditions are given.

Similarly, the components of the object x_{pi} are x_{pi}^{j}- real numbers, which are read as follows: j sign of i patient related to p class. Where $p = \overline{1,r};, i = \overline{1,m_p};, j = \overline{1,N}$; and r-is the total number of given classes, m_p- p is the total number of patients in the p-class and N is the total number of characters.

Problem 2.1. Let us normalize the character values that characterize the objects of the elective elective classes we are looking at.

Problem 2.2. Let the characters that characterize the objects of the elective elective classes we are looking at be converted from a continuous quantitative view to a 0 or 1 view.

Problem 2.3. It is necessary to solve the problem of classification of objects of K_p class, that is, to determine whether the objects divided into classes belong to their own class or to another class.

Problem 2.4. $K_p, p = \overline{1,5}$ class objects are required to select a set of $\ell \ll 63$ informative features that clearly distinguish them from each other. Here ℓ is a pre-given small number, read as a number that is somewhat smaller than 63.

Problem 2.5. Let determine the recognition quality and error of classification of K_p class objects.

In the above issues, each class is considered as a type of disease, ie K_1 class - Exacerbated stenocardia, K_2 class - Acute myocardial infarction, K_3 class - Arrhythmic form, K_4 class - Post-infarction cardiosclerosis, K_5 class - Permanent form of atrial fibrillation. In this case, the character space of each class (type of disease) is formed by experts in the field, and consists of 63 characters that characterize each class.

2.2. Stages of solving set practical problems:

Step 1. The process of normalization of initial data values was considered an important part of the initial processing of data, the more qualitatively this process is carried out, the greater the efficiency of subsequent results.

With this in mind, at this stage, the most commonly used method of normalization of the character values characterizing the class objects in the curriculum we are considering is used and is performed as follows:

a) find the maximum $x_{pi}(max)$ of the character values that characterize the objects of each class;

b) The maximum value found of the character values characterizing the objects of each class is divided by $x_{pi}(max)$, ie.:

$$x_{pi}(normalisation) = \frac{x_{pi}}{x_{pi}(max)} \qquad (2.1)$$

At the end of the phase, the character values characterizing the class objects are normalized in the research option formed on the basis of the initial data we are looking at (Figure 2.2).

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

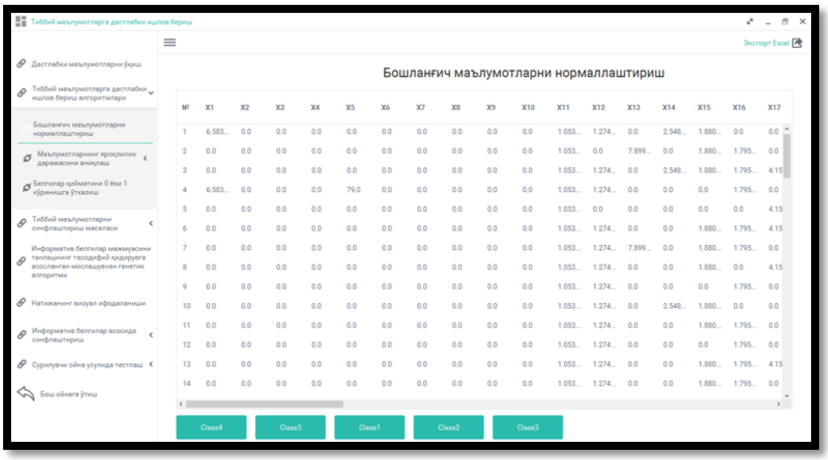

Figure 2.2. The results of the normalization of the initial data are described.

Step 2. Determining the level of usability of the characters that characterize each object belonging to the above K_p class is done separately for each class in the following sequence:

a). Lets enter definitions to the followings: Vector $\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \ldots, \bar{x}_p^N)$, average representative object of K_p classes, $p = \overline{1,r}$. Let its components be calculated by the following formula:

$$\bar{x}_p^j = \frac{1}{m_p} \sum_{i=1}^{m_p} x_{pi}^j, p = \overline{1,5};\ j = \overline{1,63};\ i = \overline{1, m_p}. \quad (2.2)$$

As a result of the calculation, the average representative objects of the K_p classes were identified(Figure 2.3).

Figure 2.3. The results are determined by the average representative object of the class.

b) Let be calculated the distance between $x_{pi} and\ \bar{x}_p$ objects of K_p class:

$$|x_{pi} - \bar{x}_p| = \sqrt{\sum_{j=1}^{N} (\bar{x}_p^j - x_{pi}^j)^2}, p = \overline{1,5};\ j = \overline{1,63};\ \ i = \overline{1, m_p}. \quad (2.3)$$

In this module, the distance between objects of 5 classes is calculated $|x_{pi} - \bar{x}_p|$ (Figure 2.4).

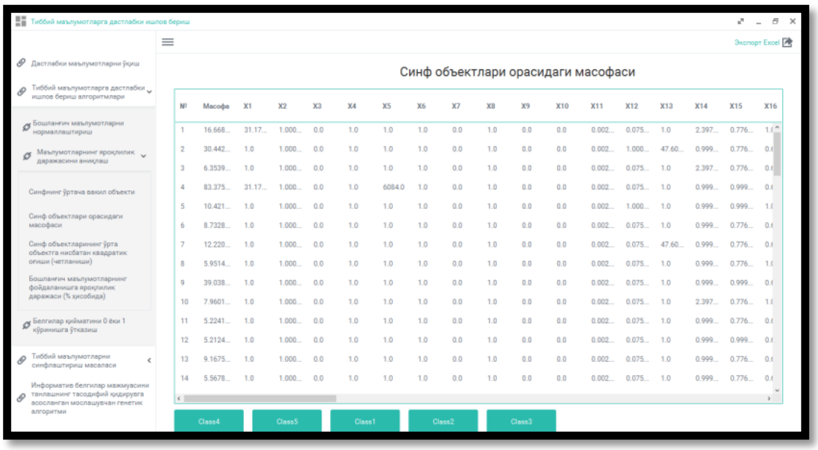

Figure 2.4. The distance between the objects of the class is described in the $|x_{pi} - \bar{x}_p|$ calculation results.

c). K_p-class quadratic deviation (avoidance) objects elative to the middle object $D(\bar{x}_p)$ is calculated by the following formula:

$$D(\bar{x}_p) = \sqrt{\frac{1}{\mathrm{m_p}}\sum_{i=1}^{m_p}|x_{pi} - \bar{x}_p|^2} = \sqrt{\frac{1}{m_p}\sum_{i=1}^{m_p}\sum_{j=1}^{N}(\bar{x}_p^j - x_{pi}^j)^2}\ . \ (2.4)$$

$$p = \overline{1{,}5};\ j = \overline{1{,}63};\ i = \overline{1, m_p}.$$

The mean square deviation (avoidance) of each class $D(\bar{x}_p)$ was calculated(Figure 2.5).

Figure 2.5. The mean squared deviation $D(\bar{x}_p)$ values of each class are described.

d) the following inequality is satisfied and its indicators are calculated as a percentage of class objects:

$$|x_{pi} - \bar{x}_p| \leq D(\bar{x}_p),\ p = \overline{1{,}5};\ i = \overline{1, m_p}.\ (2.5)$$

At the end of the calculation, the results of the percentage of the usability level of the initial data are reflected (Figure 2.6).

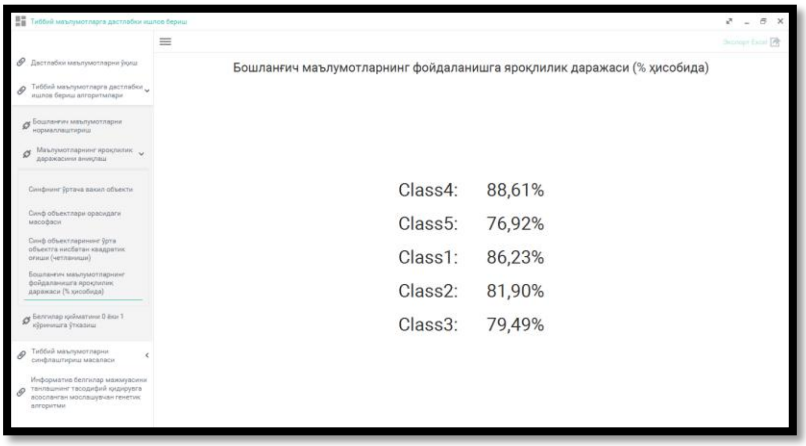

Figure 2.6. The results of the usability level of the initial data are described.

Step 3. The initial data given are given in continuous quantitative form, at which stage the process of replacing the character values of the class objects with values of 0 and 1 is carried out.

The process of converting the values of zero or one-character characters that characterize each object belonging to the K_p class to a value in vector form is performed by entering the following notation in the character class of all objects in each class and all character sections.

a). $\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \dots, \bar{x}_p^N)$ vector, average representative objects of K_p classes, $p = \overline{1{,}5}$. Its components $\bar{x}_p^j$ are calculated Formula 2.2 as above.

b). Let us define vectors $a_p = (a_p^1, a_p^2, \dots, a_p^N)$ and $b_p = (b_p^1, b_p^2, \dots, b_p^N)$, in the following form and its components are calculated by the following formula:

$$a_p^j = \frac{1}{m_p}\sum_{i=1}^{m_p}(\bar{x}_p^j - x_{pi}^j)^2,\ p = \overline{1{,}5};\ j = \overline{1{,}63}. \qquad (2.6)$$

$$b_{pi}^j = (\bar{x}_p^j - x_{pi}^j)^2,\ p = \overline{1{,}5};\ j = \overline{1{,}63}. \qquad (2.7)$$

c). Components of the K_p elements of a research option are converted from a real number view to boules view based on the following operations.

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

$$x_{pi}^{j} = \begin{cases} 1\ equal, if\ \frac{b_{pi}^{j}}{a_{p}^{j}} \leq 1, \\ 0\ equal, if\ not \end{cases} \qquad (2.8)$$

At the end of this step, the values of all the characters that characterize the 5 class objects given are converted to values of 0 and 1 (Figure 2.7).

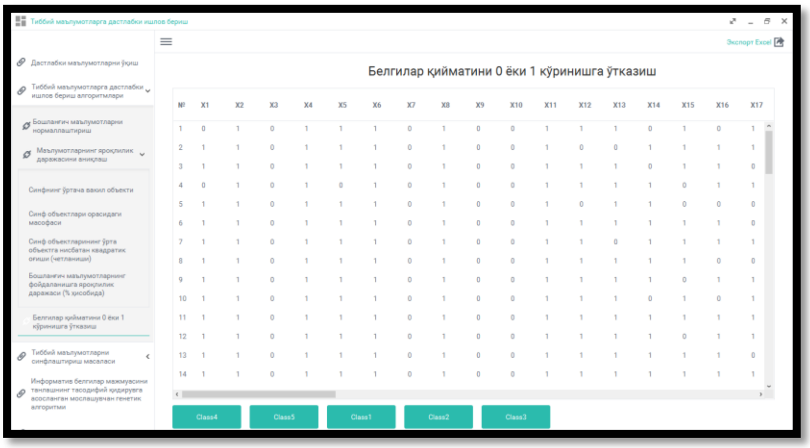

Figure 2.7. The results obtained by converting the character values to 0 and 1 are described.

Step 4. The problem of classifying K_p class objects is solved. In this case, each object belonging to the class K_p is compared one by one with objects of its own class and others, the process of comparative evaluation of the function of interdependence $\rho_i(x_{p1,}x_{p2})$ in the space of informative elements is carried out using formula 1.3.

The sum of the comparative evaluation is calculated for each class using formula 1.4, which means that the object with the largest of the average values of the sums obtained belongs to that class.

At this stage, the process of reclassifying the objects divided into classes is carried out, and it is determined whether the class objects belong to their own class or to another class. This process is done step by step, at each step objects that do not belong to their class are removed, and the process is carried out until the objects find their exact class for 100%. By the end of the stage, a standard research option of 138 objects in the 1st class, 105 objects in the 2nd class, 39 objects in the 3rd class, 79 objects in the 4th class and 26 objects in the 5th class will be formed (Figure 2.8-2.9).

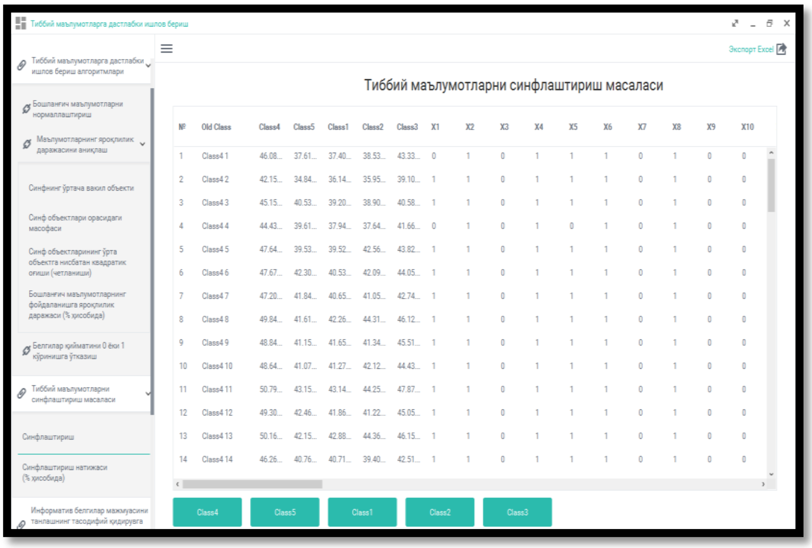

Figure 2.8. The results of the problem of classifying medical data are described

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

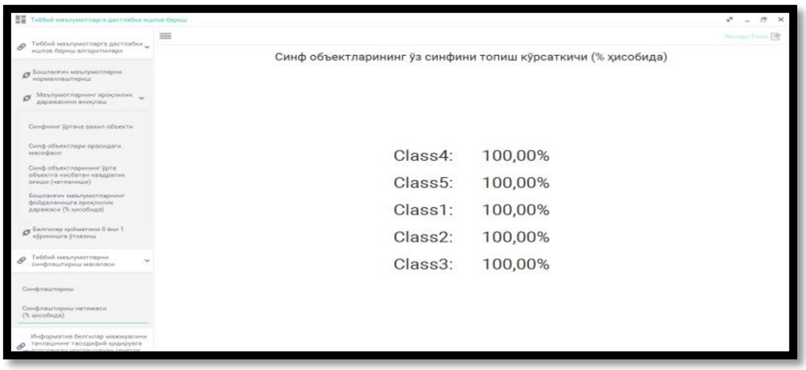

Figure 2.9. The results of the percentage calculation of the class objects' own class finding index are described.

Step 5. An informative element set is selected using the generated standard reference table. It is known that the classification results are 100% for the reference table.

Assume that the result error is θ when classifying the objects of study option on the basis of a reference table. That is, let the ratio of the number of objects lost in their η class to the total number of study choices be defined as θ.

Let the mechanism of operation of the algorithm "A" be defined as follows. Using the proximity function proposed by the authors [2,32,35-36,38,40] (formula 1.3), the levels of importance of all the elements are determined separately. According to it, a column with a randomly selected element is omitted from the reference table on the basis of a random selection. If at the end of the process all the objects involved in the calculation find their class at least with θ accuracy, then the column removed from the reference table is not re-placed in the table, otherwise the column is returned to its place, another column is removed from the reference table and the process is repeated. The proposed process takes up to ℓ elements. During the classification process, if objects find another class from their own class (move to another class), the randomly selected and discarded element is returned to its place. This process is performed between N elements and is separated from the remaining elements at the end of the process as informative elements (Figure 2.10).

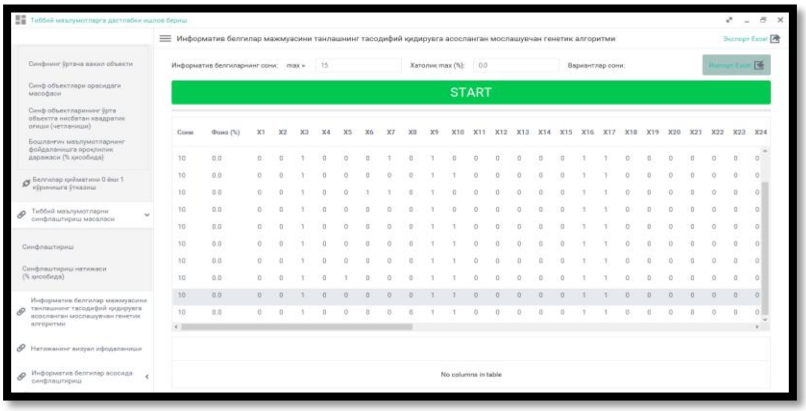

Figure 2.10. The window of the module "Flexible genetic algorithm based on random search for the selection of a set of informative traits" is described.

Therefore, the essence of the work done at this stage, the choice of the most useful ℓ from the set of elements that characterize the objects under study, that is, the choice of the means of informative elements is solved.

Let $\ell \ll N$ be the number of elements in the tool to be selected; The vector components of $p = (p^1, p^2, \dots, p^N)$ are the probabilities that a single element must be removed from the element set. Initially, $p^j = \frac{1}{N}, j = \overline{1, N}$, that is, the probability of removing an arbitrary sign from the set is mutually equal.

Step 1. Algorithm "A" is used in all sections of the given study sample. As a result, the error coefficient $\theta(N)$ is determined. Typically, the number of error-detected objects is divided by the total number of objects to determine the error coefficient $\theta(N)$.

Step 2. Randomly, $p^j = \frac{1}{N}, j = \overline{1, N}$ with probability, one element is selected from N elements and it is excluded from the set. Then, the algorithm "A" is run on the N-1 element section. As a result, the error coefficient $\theta(N-1)$ is determined for N-1 elements.

Step 3. If $\theta(N) > \theta(N-1)$, then proceed to step 2. This process is repeated $h\ (h \le \ell)$ times. Even if $h = \ell$ is done when the process is repeated, still the next step is taken. Similarly, if the equation is not fulfilled, then the process continues until all the elements are considered one by one. Once all the elements have been reviewed, the next step is taken.

Step 4. Suppose $h = \ell$, then $p^j = \frac{1}{2N-h}$, $j = \overline{1, h}$; $p^j = \frac{2}{2N-h}$, $j = \overline{h+1, N}$; probability that one element is selected from N elements and it is excluded from the set. Just like in the second step, the process is repeated, except that the probabilities between the previously selected elements that remain in the system and those that do not remain in the system, which must be randomly selected, differ. Then, the algorithm "A" is run on the N-1 element section. As a result, the error coefficient $\theta(N-1)$ is determined for N-1 elements.

Step 5. If $\theta(N) > \theta(N-1)$, then proceed to step 4. This process is repeated h $(h \leq \ell)$ times and so on.

The proposed algorithm consists of two important parts: In the first part genetics is a flexible algorithm based on a random search for the selection of a new set, while in the second part the h-element medium affecting the generation of mutated objects is identified.

During the operation of the module "Flexible genetic algorithm based on random search for the selection of information sets of elements" of this software package, the objects of the class find their class for 100%, i.e. the error is 0% and when the process of maximum reduction of 63 parameters characterizing the class objects was carried out, it was achieved to reduce them from 63 up to 10. (Table 2).

Table 2

Group of sets of informative elements	Set of informative elements
№1	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{47}, x_{49}, x_{53}, x_{58}, x_{62}.
№2	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{47}, x_{53}, x_{58}, x_{62}, x_{63}.
№3	x_3, x_7, x_9, x_{16}, x_{17}, x_{52}, x_{53}, x_{58}, x_{62}, x_{63}.
№4	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{35}, x_{47}, x_{50}, x_{58}, x_{63}.
№5	x_3, x_6, x_7, x_9, x_{16}, x_{17}, x_{52}, x_{53}, x_{58}, x_{63}.
№6	x_3, x_9, x_{16}, x_{17}, x_{47}, x_{52}, x_{53}, x_{58}, x_{62}, x_{63}.
№7	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{36}, x_{47}, x_{53}, x_{58}, x_{62}.
№8	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{38}, x_{47}, x_{53}, x_{58}, x_{62}.
№9	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{32}, x_{47}, x_{53}, x_{58}, x_{62}.
№10	x_3, x_5, x_9, x_{10}, x_{16}, x_{17}, x_{38}, x_{47}, x_{53}, x_{58}.
№11	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{47}, x_{48}, x_{53}, x_{58}, x_{62}.
№12	x_3, x_9, x_{10}, x_{16}, x_{17}, x_{44}, x_{47}, x_{53}, x_{58}, x_{62}.

Table 2. The selected $l = 10$ groups of informative element sets are described

Step 6. At this stage, in order to distinguish the most efficient group of information elements out of the group of 12 information sets of 10 elements, i.e. the highest quality of recognition and, conversely, the lowest level of classification error, the method of "sliding window" has been held 100 times.

In carrying out this process, first of all, a group of information sets selected from the module of the software package "Studying of the selected set of information signs" is read. (Figure 2.11).

Figure 2.11. A window for training the selected informative element set group.

Taking into account the fact that the testing process was conducted 100 times, 85% of the reference curriculum for each test was selected as the teaching curriculum and 15% as the test curriculum (Figure 2.12).

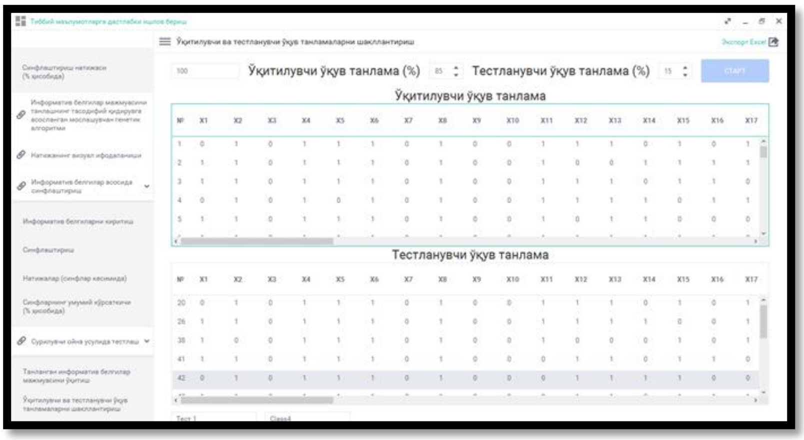

Figure 2.12. The process of distinguishing between teaching and test learning choices is described.

Step 7. Determine the quality and error rate of the classification. At this stage, the classification process was carried out based on the algorithm for calculating grades on the basis of teacher and test curricula formed, and as a result, the quality and error of recognition of classroom objects were determined.

The fact that the testing process is performed 1 time is not sufficient to determine the recognition quality and error of the classification. For this reason, it was tested 100 times (Figure 2.13).

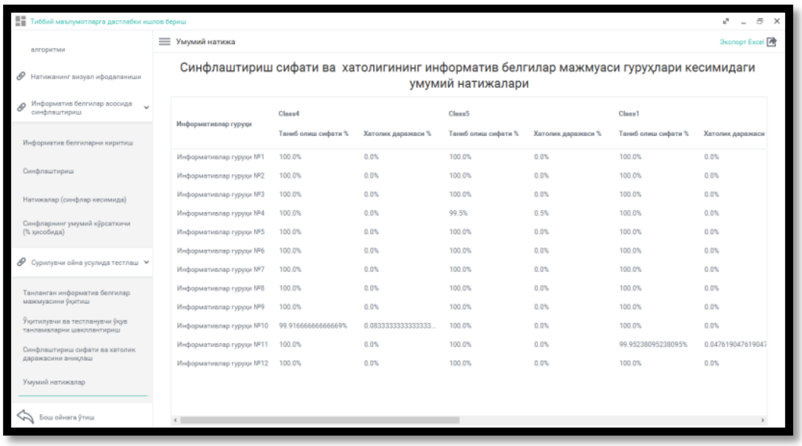

Figure 2.13. The general results of the recognition quality and error of classification are described.

At the end of the phase, the overall results of the classification quality recognition error and error in the testing process performed by the software suite are presented in Table 3.

Table-3

Group of informative elements set	**Class 1**		**Class 2**		**Class 3**		**Class 4**		**Class 5**		**Medium valve of recognition percentage %**	**Average value of error coefficient %**
	Recognition percentage %	Error coefficient %	Recognition percentage %	Error coefficient %	Recognition percentage %	Error coefficient %	Recognition percentage %	Error coefficient %	Recognition percentage %	Error coefficient %		
№1	100	0	99,5	0,5	100	0	100	0	100	0	99,9	0,1
№2	100	0	98,313	1,688	100	0	100	0	100	0	99,6625	0,3375
№3	100	0	100	0	100	0	100	0	100	0	100	0

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

№4	100	0	98,438	1,563	100	0	99,8333333	0,16666667	99,75	0,25	99,6042	0,395833
№5	100	0	100	0	100	0	100	0	100	0	100	0
№6	100	0	100	0	100	0	100	0	100	0	100	0
№7	100	0	99,813	0,188	100	0	100	0	100	0	99,9625	0,0375
№8	100	0	100	0	100	0	100	0	100	0	100	0
№9	100	0	100	0	100	0	100	0	100	0	100	0
№10	100	0	99	1	100	0	99,8333333	0,16666667	100	0	99,7667	0,233333
№11	100	0	100	0	100	0	100	0	100	0	100	0
№12	100	0	100	0	100	0	100	0	100	0	100	0

III. Conclusion

The article develops the following methods and algorithms for solving the following two practical problems in the field of cardiology, namely, the assessment of the state of medical diagnosis and the initial processing of medical data:

a) improved algorithms based on the calculation of estimations for the formation of the information space from the set of elements characterizing the objects of medical diagnosis and the classification of medical diagnostic objects from the identified information space in the initial processing of medical data;

b) normalization of element values of primary data in the initial processing of medical data, determination of the degree of validity of primary data, conversion of element values characterizing class objects from continuous quantitative view to 0 or 1, reclassification of study sample objects and determination of class significance, formation of reference tables, informative elements methods and algorithms for solving problems such as the selection of the set, determining the effectiveness of the selected informative element set.

Also, along with the description of the theoretical basis of these methods and algorithms, the principle of operation of the software package "Preliminary processing of medical emblems" (PPME), created on the basis of the developed methods and algorithms, the results were analyzed. According to the analysis, the results were positively assessed by experts in the field.

REFERENCES:

[1]. Lbov G.S. Methods of processing different types of experimental data / Novosibirsk: Nauka, Siberian Branch, 1981. – 160 p.

[2].Yu.I. Zhuravlev Selected scientific works. –M: Magister Publishing House, 1998. – 420 p.

[3].Zhang, L., Luo, M., Liu, J., Li, Z., Zheng, Q. Diverse fuzzy c-means for image clustering //Pattern Recognition LettersVolume 130, February 2020, Pages 275-283.

[4].Santra, D., Basu, S.K., Mandal, J.K., Goswami, S. Rough set based lattice structure for knowledge representation in medical expert systems: Low back pain management case study//Expert Systems with ApplicationsVolume 145, 1 May 2020, Номер статьи 113084

[5].Xiong, Y., Zuo, R. Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine//Computers and GeosciencesVolume 140, July 2020, 104484

[6].Li, P., Cheng, H. Track recognition algorithm based on neural network for rail transit//International Journal of Computers and Applications 2020.

[7].Gai, J., Shen, J., Wang, H., Hu, Y. A Parameter-Optimized DBN Using GOA and Its Application in Fault Diagnosis of Gearbox//Shock and Vibration, Volume 2020, 2020, Номер статьи 4294095

[8].Raja, P.S., Thangavel, K. Missing value imputation using unsupervised machine learning techniques/Soft Computing 24(6), с. 4361-4392,2020.

[9].Wang,D., Tian,F., Yang,S.X.,(...), Jiang,D., Cai,B. Improved deep CNN with parameter initialization for data analysis of near-infrared spectroscopy sensors/Sensors (Switzerland) Volume 20, Issue 3, 1 February 2020, Номер статьи 874, 20(3),874, 2020.

[10].Basiri,M.E., Kabiri,A. HOMPer: A new hybrid system for opinion mining in the Persian language/Journal of Information ScienceVolume 46, Issue 1, 1 February 2020, Pages 101-117.

[11].Lou, P., Jimeno Yepes, A., Zhang, Z., (...), Li, C., Wren, J. BioNorm: Deep learning-based event normalization for the curation of reaction databases/BioinformaticsVolume 36, Issue 2, 15 January 2020, Pages 611-620.

[12].Si, L., Xiong, X., Wang, Z., Tan, C. A Deep Convolutional Neural Network Model for Intelligent Discrimination between Coal and Rocks in Coal Mining Face/Mathematical Problems in Engineering Volume 2020, 2020, Номер статьи 2616510

[13].Fu, S., Liu, X. A new method to solve the problem of facing less learning samples in signal modulation recognition/Eurasip Journal on Wireless Communications and Networking Volume 2020, Issue 1, 1 December 2020, 8.

[14].Wei, D., Chen, T., Li, S., (...), Zhao, Y., Li, T. Adaptive dictionary learning based on local configuration pattern for face recognition/Eurasip Journal on Advances in Signal Processing Volume 2020, Issue 1, 1 December 2020, 20.

[15].Ala'raj, M., Majdalawieh, M., Abbod, M.F. Improving binary classification using filtering based on k-NN proximity graphs/Journal of Big Data, Volume 7, Issue 1, 1 December 2020, 15.

[16].Mishra, G., Vishwakarma, V.P., Aggarwal, A. Constrained L1-optimal sparse representation technique for face recognition/Optics and Laser TechnologyVolume 129, September 2020, 106232.

[17].Kibbey, T.C.G., Jabrzemski, R., O'Carroll, D.M.Supervised machine learning for source allocation of per- and polyfluoroalkyl substances (PFAS) in environmental samples/ChemosphereVolume 252, August 2020, Номер статьи 126593

[18].Shen, Z., Man, Z., Cao, Z., Zheng, J. A new intelligent pattern classifier based on structured sparse representation/Computers and Electrical EngineeringVolume 84, June 2020, Номер статьи 106641

[19].Kamilov M.M., Nishanov A.Kh., Juraev G.P. Algorithm for the classification of medical data in the space of informative signs using the functions of range and proximity // Scientific and technical journal. "Chemical Technology. Control and management ". - Tashkent, 2018, No. 2, Pages 143-150.

[20]. Nishanov A.Kh., Juraev G.P., Narziev N.B. Application of evaluation algorithms for myocardial infarction // Scientific-practical and information-analytical journal "Muhammad al-Khwarizmi avlodlari", Tashkent-2 (4) 2018, Pages 40-44.

[21].Nishanov A.Kh., Juraev G.P., Egamberdiev N.A. About the application of software based on the calculation algorithms of medical data classification// Proceedings of the international conference "Current issues of production of import-substituting products based on the use of local raw materials in the Fergana Valley", Namangan, October 27-28, 2018, Pages 470-473.

[22].Nishanov A.Kh., Juraev G.P., Kasanova M.Kh. Improved algorithms for calculating evaluations in processing medical data // Compusoft: An International Journal of Advanced Computer Technology, 8(6), June-2019 , 3158-3165. Available from:https:/ https://ijact.in/index.php/ijact/article/view/899

[23].Nishanov A.Kh., Juraev G.P., Egamberdiev N.A. Methods and algorithms for solving the problem of classification based on the calculation of values in the identification of medical emblems // "Muhammad al-Khwarizmi avlodlari", scientific-practical and information-analytical journal, Tashkent-1 (7) 2019, Pages 20-27.

[24].Nishanov A.Kh., Juraev G.P. About the algorithm for solving cardiological diagnostic problems in the identification of emblems// materials of the V. International scientific and practical conference "Global science and innovations 2019: Central Asia", Astana-2019, Pages 3-6.

[25].Nishanov A.Kh., Juraev G.P. Problems based on medical diagnoses in the identification of emblems and algorithms for their solution // Materials of the Republican scientific-technical conference on "The role of information and communication technologies in the innovative development of sectors of the economy", Tashkent, March 14-15, 2019, Pages 101-103.

[26].Nishanov A.Kh., Juraev G.P., Mamasaidov U.E. Stages of solving some problems of identification of logos // Materials of the Republican scientific-technical conference on "Innovations in the development of information and communication technologies", Karshi, April 15-17, 2019, Pages 216-218.

[27].Nishanov A.Kh., Juraev G.P., Azatov B.A. Stages of solving some problems of preliminary data processing // V. International scientific and practical conferences "Science and education in the modern world: CHALLENGES of the XXI century", Nur-Sultan, Kazakhstan December 10-12, 2019, I Volume, Pages 23-27.

[28].Nishanov A.Kh., Juraev G.P., Methods and algorithms of classification in the initial processing of data // Problems of computer science and energy. Journal of Uzbekistan, 2019, №6, Pages 19-26.

[29].Nishanov A.Kh., Juraev G.P. On one classification algorithm using the ε-threshold // Questions of technical and physical and mathematical sciences in the light of modern research: collection of articles. Art. by mater. XXIII-XXIV Int. scientific-practical conf. No. 1-2 (19). - Novosibirsk: SibAK, 2020. Pages 15-20.

[30].Nishanov A.Kh., Juraev G.P. Solving stages of classification in the preliminary treating of medical examples // Technical sciences: problems and solutions, collection of articles. Art. by mater. XXXIII International Scientific and Practical Conference, No. 2 (30). - Moscow, 2020 Pages 69-76.

[31]. Nishanov A.Kh., Juraev G.P., Narziev N.B. Algorithm for solving the problem of classification based on e-threshold value // "Muhammad al-Khwarizmi avlodlari", scientific-practical and information-analytical magazine, Tashkent-1 (11) / 2020, Pages 25-29.

[32].Nishanov A.Kh., Akbaraliev B.B., Juraev G.P. Algorithm for the selection of symptomatic complexes of ischemic heart disease based on a flexible random search // "Muhammad al-Khwarizmi avlodlari", scientific-practical and information-analytical journal, Tashkent-2 (12) / 2020, Pages 10-15.

[33].Nishanov A.Kh., Juraev G.P., Azatov B.A. Stages of assessment of symptoms in the initial processing of medical data // Materials of the Republican scientific-technical conference on "Modern problems of information technology", Urgench, June 5, 2020, Pages 496-499.

[34].Nishanov A.Kh. and Ruzibaev O.B., Chedjou J.C., Kyamakya K. and Kolli Abhiram and Perumadura De Silva, Juraev G.P., Khasanova M.A. Algorithm for the selection of informative symptoms in the classification of medical data // World Scientific Proceedings Series on Computer Engineering and Information Science Developments of

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 03, 2021

Artificial Intelligence Technologies in Computation and Robotics, pp. 647-658 (2020), https://www.worldscientific.com/doi/abs/10.1142/9789811223334_0078.

[35].Nishanov A.Kh., Juraev G.P, Narziyev N.B., Adaptable genetic algorithm for primary processing of medical data based on random search // Sixth International Scientific and Practical Conference "BIG DATA and Advanced Analytics. BIG DATA and high-level analysis ", Republic of Belarus, Minsk, May 20-21, 2020, Part 1., Pages 147-155.

[36].Juraev G.P. An adaptive genetic algorithm based on an arbitrary assessment of a complex of informative features // Problems of technical and physical and mathematical sciences in the light of modern research: collection of articles. Art. by mater. XXVIII int. scientific-practical conf. No. 6 (22). - Novosibirsk: SibAK, 2020. Pages 16-19.

[37].Nishanov A.Kh., Juraev G.P., G.P., Khasanova, M.A. Classification and feature selection in medical data preprocessing// Compusoft: An International Journal of Advanced Computer Technology, 9(6), June-2020, 3725-3732. Available from: https://ijact.in/index.php/ijact/article/view/1166

[38].Nishanov A.Kh., Akbaraliev B.B., Juraev G.P., Khasanova M.A., Maksudova M.Kh., Umarova Z.F. The algorithm for selection of symptom complex of ischemic heart diseases based on flexible search // Journal of Cardiovascular Disease Research, 2020; 11(2): 218-223. Available from: http://www.jcdronline.org//fulltext/10.31838jcdr.2020.11.02.37.pdf?1617617887

[39].Juraev G.P. About the program of assessment of medical diagnostic cases based on algorithms of calculation of estimates // XII International scientific-practical conference «GLOBAL SCIENCE AND INNOVATIONS 2020: CENTRAL ASIA», December 17, 2020, Pages 33-39

[40].Nishanov A.Kh., Akbaraliev B.B., Juraev G.P. A symptom selection algorithm based on classification errors // International conference on information science and communications technologies applications, trends and opportunities, http://www.icisct2020.org/, 4-6 November, 2020.