International Journal of Informatics and Data Science Research
ISSN 2997-3961 (Online)
Vol. 2, No. 9, Sep 2025,
Available online at: https://scientificbulletin.com/index.php/IJIDSR

Efficient Method and Algorithm for Preliminary Processing of Medical Data

Juraev Gulomjon Primovich
Acting Associate Professor, Asian Innovation University, Doctor of Philosophy (PhD) in Technical Sciences, Uzbekistan

Article information:
Manuscript received: 4 Jul 2025; **Accepted:** 10 Aug 2025; **Published:** 27 Sep 2025

Abstract: To solve medical problems that are practical, a special type of computer software is utilized. The ideal solution depends on the quality and volume of medical data. Nonetheless, most of the medical records are all based on sick leaves. Often, the collected medical data comprises the type information with a very lowered level of accuracy, non-standard terminology, information that is not incomplete and that is input incorrectly. Frequently, the patient's cardinal activity, behavior and habits are no how reflected in the medical data. Generally, there is no precise mechanism to collect medical data and dischargement of research in this sphere is one of the most necessitious issues ot consider. To solve these problems, algorithms that are based on full and partial selection as well as statistics have been used and these algorithms do not take intrinsic properties of the object into account while it is based on the interdependence of medical data. The algorithm which is proposed in the article eliminates these shortcomings and aims to reveal hidden connections togetherh with relationships between characters based on a random selection of the characters.

This paper addresses topics (or issues) such as the initial processing of medical data generated by medical professionals, reclassifying learning choices/selections and choosing a range of informative characters (signs) that distinguish class objects. The underlined topics (or issues) are solved using algorithms to calculate grades based on Fisher criteria. As a result of primary data processing, five class objects are established (or formed): 1) Progressive angina pectoris; 2) Acute myocardial infarction (MI); 3) Arrhythmic form (AF); 4) Postinfarction cardiosclerosis (PC); 5) Permanent form of atrial fibrillation (PersAfib). Then problems such as the classification using Fisher's criteria and algorithms to calculate grades, the selection of a set (number) of informative characters (signs) that distinguish class objects, etc. are solved and appropriate software are developed. As key result, reference classes are established from the initial data processing. Objects that deviate from their class in the process of formation are excluded from the sample. Using established (or formed) classes, a set (or number) of classifications and indicators (i.e., informative characters) are selected. Initially there were five class objects of 62 characters each, all provided by medical professionals, and the initial processing of the data resulted in the formation of a reference table consisting of a set of 128 objects in first class, 110 in second class, 39 in third class, 76 in fourth class and 26 in fifth class, with 62 characters each.

Keywords: Fisher criterion, informative signs, classification, algorithms for calculating grades, medical data processing, “A” algorithm, adaptive genetic algorithm-based on random search.

I. INTRODUCTION

One of the most common causes of death is cardiovascular disease (CVD), which claims 17.9 million lives worldwide every year. Cardiovascular diseases are presented in groups such as rheumatic heart disease, cerebrovascular disease, coronary heart disease, heart/blood vessel disease, and others. Most of the deaths from these diseases are associated with heart strokes/attacks and in people under 70 years of age.[1]

CVD prevention involves developing a healthy lifestyle, developing targeted programs to prevent and control non-communicable diseases and their risk factors, and improve preventive measures. In view of the above, the state of the current pandemic, the coronavirus COVID-19, the establishment of criteria for determining the status of research objects related to medical and health problems, the assessment of the condition/state of objects, the analysis and the development of algorithms/methods for the formation of a set (or number) of quantitative variables, i.e. the intellectual analysis of medical data, are regarded as one of the important topics of the present [1-17].

In solving the underlined problems, there are size and space constraints in determining the hidden connections and patterns between the characters (or signs) and sets of signs that characterize the research objects. This depends on the solution to the problem of identifying the characters (or signs). One of the main stages in characters (or signs) identification is classification, and the solution to the problem of identifying informative characters (signs) is based on the selection of criteria that determine the content [1-6,12-21].

A frequently used criterion for processing medical data (see [5-8, 11-13] and references therein) is the Fischer criterion, also used in this work. Essentially, this work solves two problems. The first problem concerns the creation of a reference table, and the second problem is choosing the usefulest from the character set that materializes the research objects (i.e., choosing a set of informative characters). The solution to the first problem determines the degree of importance of the signs and objects, as well as the calculation of their contribution to the class [1-5, 9,10]. The solution to the second problem shows the selection of informative characters from the given table, their visualization and identification of the contribution of the symbol set to the formation of classes [1-21].

In this paper, based on medical data analysis, topics such as the initial processing of medical data created (or provided) by medical professionals, the reclassification of learning selection facilities (or learning choices) and the determination of the or meaning of classes, the creation of reference and the selection of a set of informative characters solved using algorithms for computing grades based on Fischer's criteria and an adaptive genetic algorithm-based on random search.

This paper is written like this. Section 2 develops the fundamental concept for both objects classification and identification of informative signs (characters). Full details of the fundamental concept are provided and a very important intermediate point related to the formation of reference tables is clearly addressed. Section 3 deals with a case study. Using the medical data (provided by medical professionals), the fundamental concept developed (in section 2) is applied to solve the practical problems/issues at stake (that are clearly mentioned above). Section 4 contains some final remarks. We present a brief summary of the results obtained and discuss some important and unresolved research questions that are currently viewed as outlooks.

II. MATHEMATICAL EXPRESSION AND THE SOLUTION OF THE ISSUE.

In this section as well as medical professionals which are based on the histories of cases of

[1] https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1

patients with coronary heart disease, "Exacerbation of angina pectoris", "Acute MI", "AF", "Postinfarction cardiosclerosis", " PersAfib." the mathematical manifestation of the curriculum is provided. Afterwards, based on the curriculum that is depicted in these lessons, it was clarified that the implementation of the commencing data realization of processes is based on solving the following problems that are given below.

The initial data processing problems are illustrated in the example using real/concrete medical data. Four problems are examined and resolved: The first is to determine the level of availability for the use of the signs that characterize classified objects. The second is to convert the signs that characterize objects classified by a continuous quantification specified as 0 or 1. The third is to build a reference table (or look-up table) by determining whether the classified objects can belong to their own/other classes; The fourth focus is on identifying a set/number of informative signs that make a clear difference between one another in class objects.

It is assumed that the source-based learning sample is divided and labeled as follows:

$$K_1 = \begin{bmatrix} x_{11}^1 & x_{11}^2 & \dots & x_{11}^N \\ x_{12}^1 & x_{12}^2 & \dots & x_{12}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m_1}^1 & x_{1m_1}^2 & \dots & x_{1m_1}^N \end{bmatrix}, \dots, K_r = \begin{bmatrix} x_{r1}^1 & x_{r1}^2 & \dots & x_{r1}^N \\ x_{r2}^1 & x_{r2}^2 & \dots & x_{r2}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{rm_r}^1 & x_{rm_r}^2 & \dots & x_{rm_r}^N \end{bmatrix} \text{(1a)}$$

The relation (1a) can be expressed into a general form as follows (see (1b)):

$$K_p = \begin{bmatrix} x_{p1}^1 & x_{p1}^2 & \dots & x_{p1}^N \\ x_{p2}^1 & x_{p2}^2 & \dots & x_{p2}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{pm_p}^1 & x_{pm_p}^2 & \dots & x_{pm_p}^N \end{bmatrix} \text{(1b)}$$

Here $p = \overline{1,r}$; the learning selection is described as $K = \bigcup_{p=1}^{r} K_p$ and should contain classes that do not collide with each other. This latter statement is expressed through the following conditions: $K_p \cap K_q = \emptyset, (p \neq q, p = \overline{1,r};\ q = \overline{1,r};)$.

The components of the x_{pi} object are x_{pi}^j- real numbers, Here: j – sign of i – patient belongs to p– class; $p = \overline{1,r}$, $i = \overline{1,m_p}$, and $j = \overline{1,N}$. Further quantities r, m_p, and N are defined as follows: r–overall number of given classes, m_p- p – overall number of the patients in the class and N– overall number of signs.

Issues we are considering, each class is treated as a disease, i.e. K_1-class "Exacerbated angina", K_2-class "Acute MI", K_3-class "AF", K_4-class "Post-infarction cardiosclerosis", K_5-class " PersAfib ".

Here, each class is characterized and will comprise 62 features that are formed by specialists, the names of these features are given in Table 1.

Table 1. Naming/designation of the classes that characterize the class objects.

№	Name of signs	
1.	pain in the region of the heart	x1
2.	chest pain	x2
3.	epigastric pain	x3
4.	pain in the left shoulder blade area	x4
5.	pain in the left shoulder and arm	x5
6.	duration of pain	x6
7.	nitroglycerin helps	x7
8.	pain after exercise	x8

9.	niboli alone	x9
10.	cold sweat	x10
11.	Weakness	x11
12.	shortness of breath on exertion	x12
13.	dyspnea at rest	x13
14.	orthopedic position	x14
15.	swelling in the legs	x15
16.	LV hypertrophy	x16
17.	enlarged liver	x17
18.	feeling of lack of air	x18
19.	Heart rate	x19
20.	systolic blood pressure	x20
21.	diastolic blood pressure	x21
22.	Saturation (SPO2).	x22
23.	Hemoglobin	x23
24.	Erythrocyte	x24
25.	Leukocyte	x25
26.	ESR	x26
27.	blood sugar	x27
28.	Alt	x28
29.	AST	x29
30.	KFK	x30
31.	VSK (min) start	x31
32.	VSK (min) end	x32
33.	PTI	x33
34.	Thrombotest	x34
35.	Fibrinogen	x35
36.	FAK	x36
37.	TPG	x37
38.	Clot retraction	x38
39.	left atrium	x39
40.	MVP	x40
41.	ZSLZH	x41
42.	CRA	x42
43.	DAC	x43
44.	BWW	x44
45.	Ejection fraction (EF%)	x45
46.	Stroke volume (UOml)	x46
47.	total LV contractility	x47
48.	decrease in regional contractility	x48
49.	Hypokinesis	x49
50.	Hypertension	x50
51.	Diabetes	x51
52.	PEAKS	x52
53.	NKII (A)	x53
54.	NKII (B)	x54
55.	COPD	x55
56.	chronic cholecystitis	x56
57.	lung congestion	x57
58.	Increased blood pressure	x58
59.	ECG changes	x59

60.	Pain in the region of the heart	x60
61.	Leukocytosis	x61
62.	Muting heart tones	x62
63.	Violations of local contractility of the left ventricle of the heart	x63

The naming/designation of the signs that characterize the 5 class objects specified above is described in Table 1. Essentially, the following four main tasks are worth mentioning:

Task-1. Determining the extent to which there is availability for the use of the signs that characterizing objects of the learning selection classes that we are considering.

Task-2. Converting features from a continuous quantitative view to view 0 or 1.

Task-3. Solving the problem of classifying objects of class K_p.

Task-4. Selection of sets ($\ell \ll 62$) of features that give differences between classes $K_p, p = \overline{1,5}$. Here, ℓ is a predetermined small number and is read as a number less than 62 to some extent.

III. METHODS AND SOLUTIONS TO PROBLEMS (OR ISSUES) FORMULATED

This section attempts to provide answer to the problems/issues formulated in section II. The main steps to achieve efficient (i.e., accurate and robust) solutions are fourfold and can be presented as follows:

Step 1. "Identification of the reliability of main primary data". This consists of determining the degree of availability for the usage signs of belonging to the class K_p. This stage is carried out separately for each class sequentially, expressed in subparagraphs a), b), etc.

a) Let us enter the following notation: vector $\bar{x}_p = (\overline{x}_p^1, \overline{x}_p^2, \dots, \overline{x}_p^N)$, the average representative objects $p = \overline{1,r}$ of the X_p classes. Its components are calculated according to formula (3).

$$\overline{x}_p^j = \frac{1}{m_p}\sum_{i=1}^{m_p} x_{pi}^j, \left(p = \overline{1,3};\ j = \overline{1,62};\ i = \overline{1,m_p}\right) \quad (3)$$

The results calculated for each class in the sign section are depicted/shown in below (see Fig. 1-5).

Figure 1

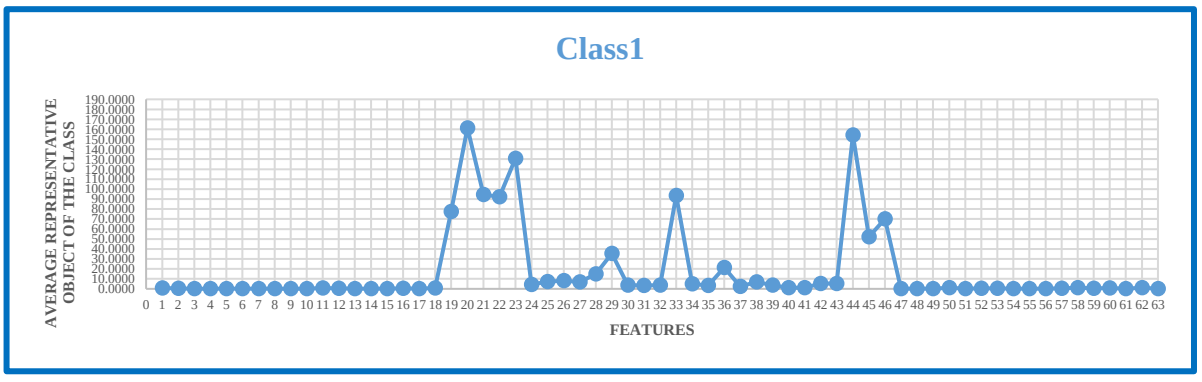

Figure 1. The characteristic values of medium representative object of 1 class are described.

Figure 2

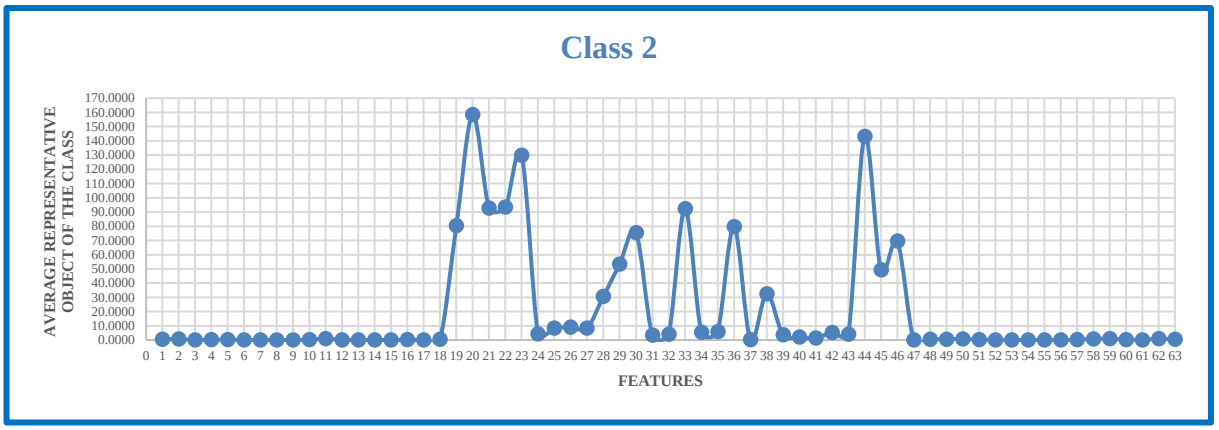

Figure 2. The characteristic values of medium representative object of 2 class are described.

Figure 3

Figure 3. The characteristic values of medium representative object of 3 class are described.

Figure 4

Figure 4. The characteristic values of medium representative object of 4 class are described.

Figure 5

Figure 5. The characteristic values of medium representative object of 5 class are described.

Average representative class objects are accordingly the formula (3) above and their values in the symbols section are revealed in Figure 1-5 given above. The main purpose of determining average representative objects of these classes is used in the following steps to determine the level of use of the beginning data that is generated in cooperation with experts in the field and in the process of calculating formulas 4-5-6.

b) The distance among the x_{pi} and $\bar{x}_p$ objects of the X_p -class is calculated according to (4).

$$\left|x_{pi} - \bar{x}_p\right| = \sqrt{\sum_{j=1}^{N}(\bar{x}_p^j - x_{pi}^j)^2}\,, \left(p = \overline{1,3};\ j = \overline{1,62};\ i = \overline{1,m_p}\right) \quad (4)$$

The reflected position of the objects of each class in the plane is shown in diagram figures 6-10.

Figure 6

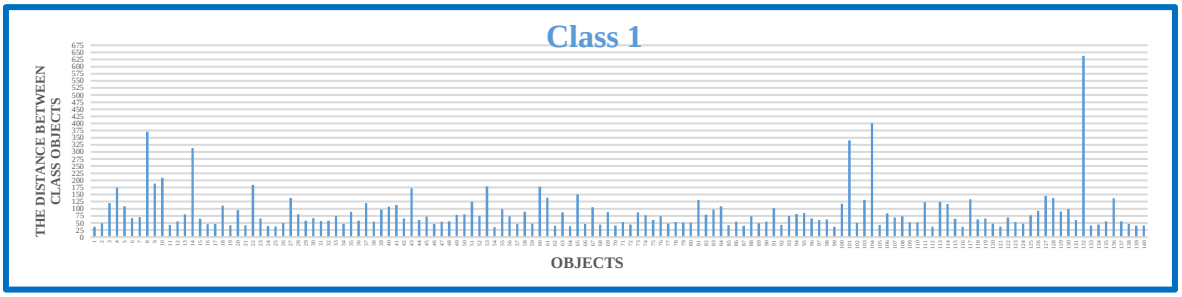

Figure 6. The distance between objects of 1 class and the average representative object of class is depicted.

Figure 7

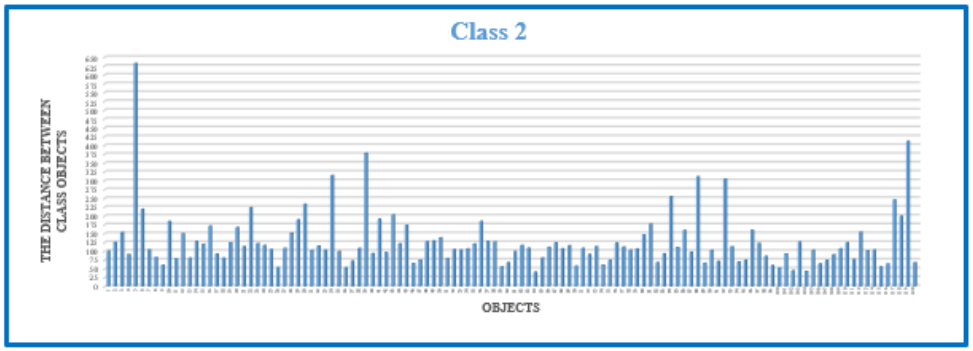

Figure 7. The distance between objects of 2 class and the average representative object of class is depicted.

Figure 8

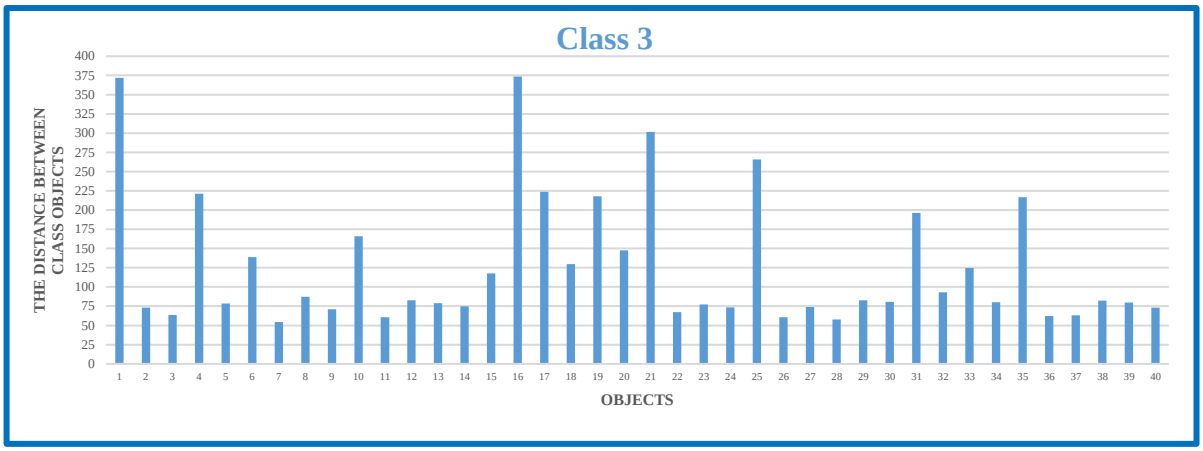

Figure 8. The distance between objects of 3 class and the average representative object of class is depicted.

Figure 9

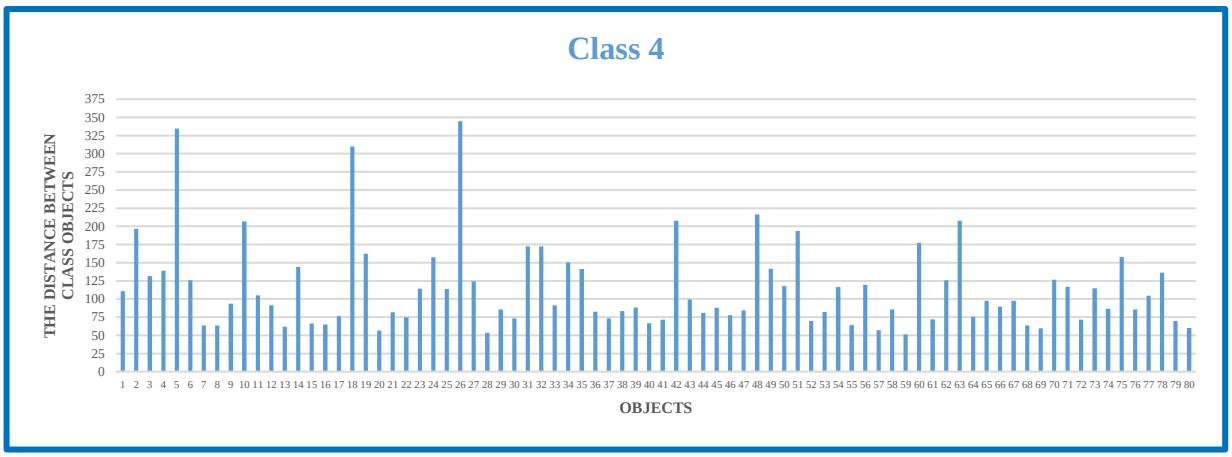

Figure 9. The distance between objects of 4 class and the average representative object of class is depicted.

Figure 10

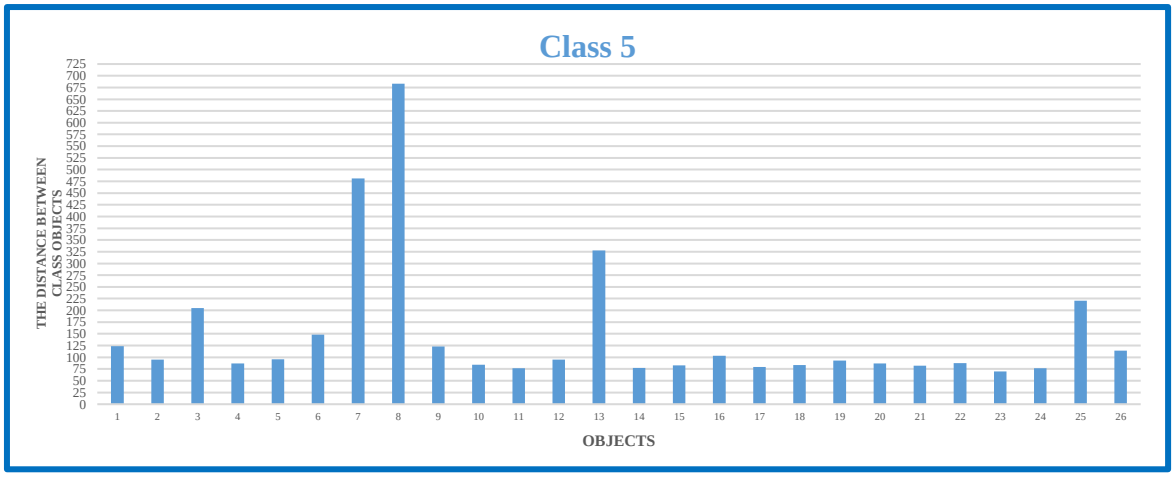

Figure 10. The distance between objects of 5 class and the average representative object of class is depicted.

Surmised from the described above, the smaller the distance between objects of each class and the average representative object of the class that is defined separately for each of the

classes, the more there is consideration of this object to claim to be one of the most efficient objects of its class. Otherly, it is clear that, the greater the distance, the less the contribution of the object to its class.

c) The average quadratic deviation $D(\bar{x}_p)$ of X_p- objects of a class relative to the average object is calculated from:

$$D(\bar{x}_p) = \sqrt{\frac{1}{m_p}\sum_{i=1}^{m_p}|x_{pi} - \bar{x}_p|^2} = \sqrt{\frac{1}{m_p}\sum_{i=1}^{m_p}\sum_{j=1}^{N}(\bar{x}_p^j - x_{pi}^j)^2} \quad (5)$$

$$(p = \overline{1,3};\ j = \overline{1,62};\ i = \overline{1,m_p})$$

The results of the average quadratic deviation $D(\bar{x}_p)$ of each class are shown in Figure 11.

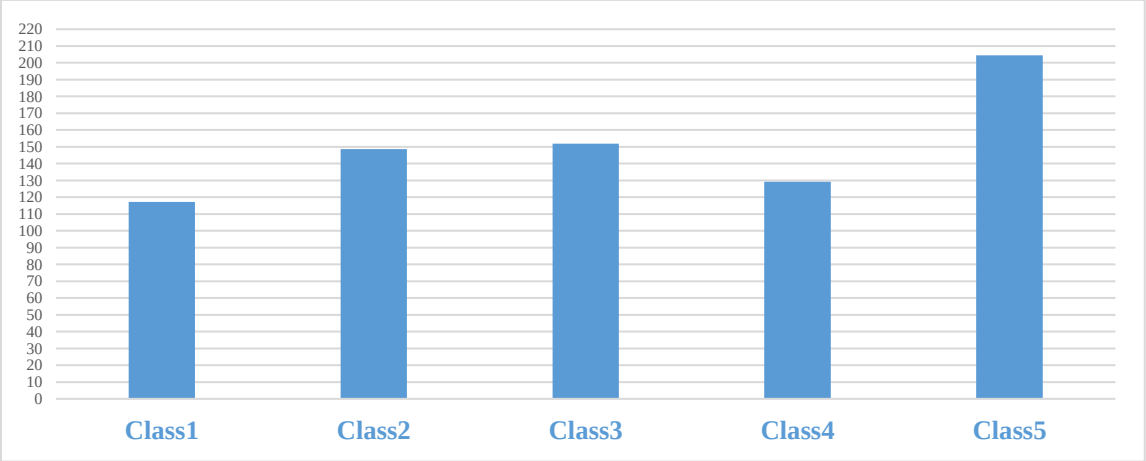

Figure 11. The results of the average quadratic deviation $D(\bar{x}_p)$ of each class.

The average quadratic deviation $D(\bar{x}_p)$ for each of the 5 classes given in Table 2 is calculated and the result values calculated for each class are expressed.

d) Inequality is calculated, indicators are taken into account as a percentage of objects belonging to the class:

$$\|x_{pi} - \bar{x}_p\| \le D(\bar{x}_p), (p = \overline{1,r};\ i = \overline{1,m_p.}) \quad (6)$$

The final phase performance of classes 1-2-3-4-5 with relative to objects is shown in Figure 12.

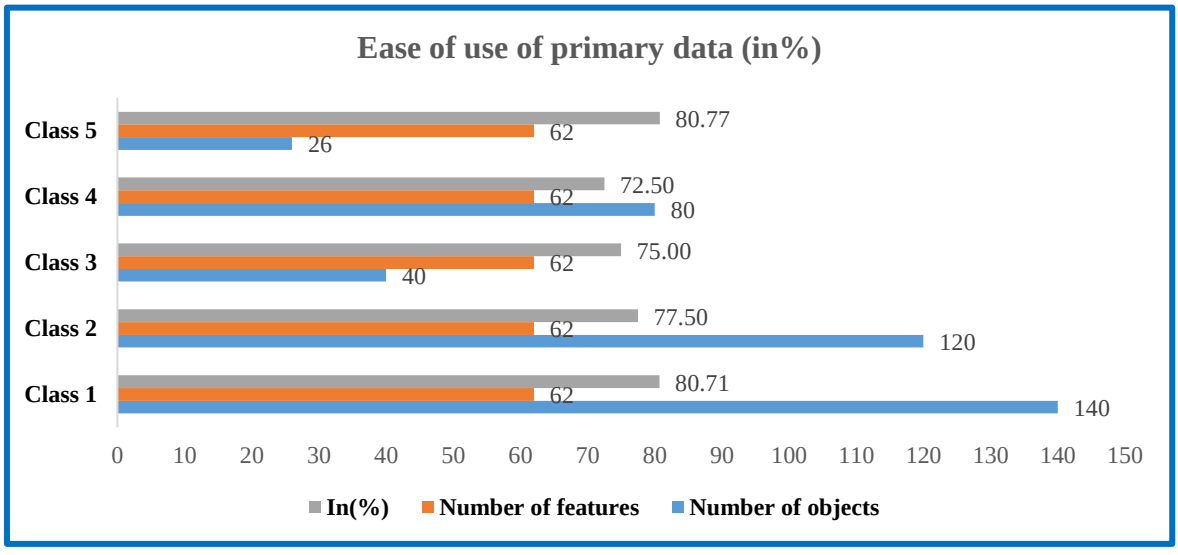

Figure 12. Degree of availability for the usage of the initial data.

According to Figure 3, the degree of availability for the use of the signs that identify each object of class K_p is 80.71% in 1st class, 77.50% in 2nd class, 75.00% in 3rd class, 72.50% in 4th class and in 5th class it's 80.77%.

Chart 12, that is based on these results, clearly pinpoints the usability of the data that is raw. It can be observed that seen the more the percentage of classes tends to be 100%, the more complete and the reliable the data is.

We can see from the results here that the percentage of satisfaction (or fulfillment) of the inequality (4) is considered relative to the 5 classes i.e., the distance $\|x_{pi} - \bar{x}_p\|$ between the objects of each class is calculated as a percentage how many are less than or equal to the mean square deviation $D(\bar{x}_p)$ determined for each class.

Step 2. "Modification to 0 or 1". Outgoing data is given in the form of a continuous form, it performs the transformation of the attributes of class objects into 0 and 1. The process of converting unambiguous attributes or zero values characterizing objects belonging to the K_p class into a value in vector form is carried out by introducing the following designations into the sign class of all objects in each class/in all sign sections. The overall process is organized around the following main bullets:

- Calculation of the components of the vector $\bar{x}_p = (\overline{x}_p^1, \overline{x}_p^2, \dots, \overline{x}_p^N)$ (see formula (7)). In formula (7), $p = \overline{1,r}$ corresponds to the average representative objects of the K_p classes.

$$\overline{x}_p^j = \frac{1}{m_p}\sum_{i=1}^{m_p} x_{pi}^j \,, \left(p = \overline{1,r};\ j = \overline{1,N};\ i = \overline{1,m_{p.}}\right) \quad (7)$$

- Let's consider the vectors $a_p = \left(a_p^1, a_p^2, \dots, a_p^N\right)$ and $b_p = \left(b_p^1, b_p^2, \dots, b_p^N\right)$ and determine their respective components through the formulas (7) and (8).

$$a_p^j = \frac{1}{m_p}\sum_{i=1}^{m_p}(\overline{x}_p^j - x_{pi}^j)^2, \left(p = \overline{1,r}\ ;\ j = \overline{1,N.}\right) \quad (8)$$

$$b_{pi}^j = (\overline{x}_p^j - x_{pi}^j)^2, \left(p = \overline{1,r}\ ;\ j = \overline{1,N.}\right) \quad (9)$$

- The K_p element components of the training sample are converted from real number representation to Bull representation based on the following operations. The study converts the components of the selected K_p elements from the actual number to the bull view based on the following operations:

$$x_{pi}^j = \begin{cases} 1\,, if\ \frac{b_{pi}^j}{a_p^j} \le 1, \\ 0\,, otherwise \end{cases} \quad (10)$$

At the end of the third step, the values of the features that characterize objects of 5 classes are replaced by 0 and 1.

Step 3. "Classification". This consists of solving the classification problem of K_p class objects, i.e., determining whether each object in the class belongs to its own class or to a class other than itself. Here each object belonging to a X_p class has a one-to-one comparison with objects in its own class and in other classes, and the function $\rho_i(x_{p1,}x_{p2})$ of inter-object proximity in the space of informative signs can be described by (11).

$$\rho_i(x_{p1,}x_{p2}) = \begin{cases} 1\ if\ \left(x_{p1}^i - x_{p2}^i\right) = 0, i = \overline{1,N}, \\ 0\ \text{otherwise}\ . \end{cases} \quad (11)$$

The first condition in (11) refers to the degree of analogy betwixt the two objects, while the second condition implies that the objects are different from each other. Further the formula (12) is used to performed the comparative evaluation.

$$\Gamma_j(x_{pj,}x_{pk}) = \sum_{\kappa=1}^{m_p}\sum_{i=1}^{N}\rho_i(x_{pj,}x_{pk}), j = \overline{1,m_p};\ k = \overline{1,m_p};\ j \neq k \quad (12)$$

The assessment is calculated by each class. The large of the middle values of the sum gotten indicates that the object belongs to this class. At this phase, the operation of reclassification objects integrated with classes is fulfilled to define if the class objects pertain to their own class or to either class. This process is done step by step, at each step the objects that do not correspond to the class are removed, and the objects in the classes are executed until they completely find their class, up to 100%. The results of each step are depicted/represented in Figures 13, 14).

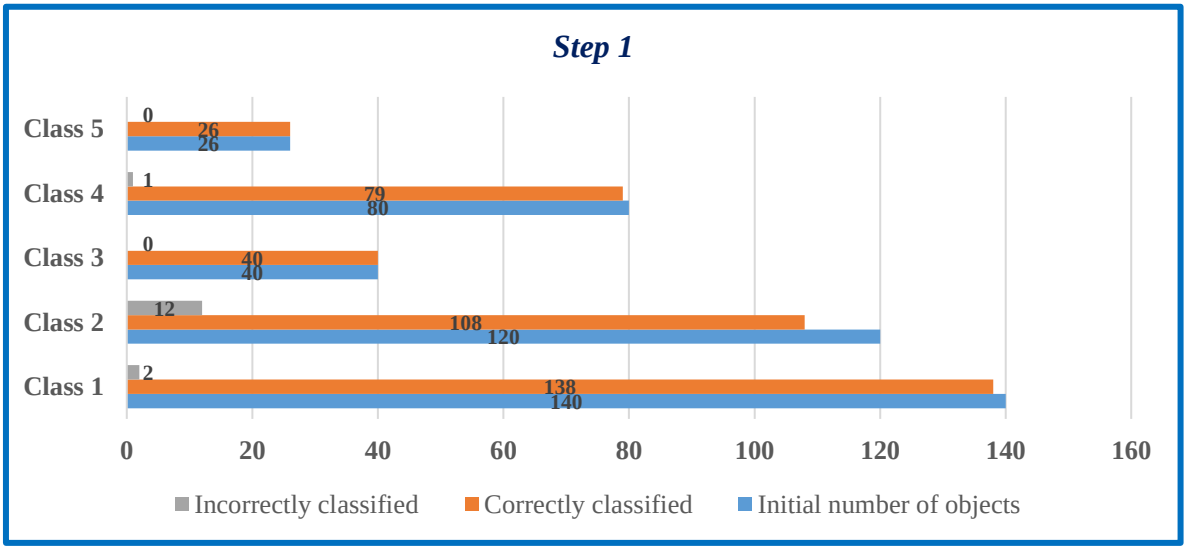

Figure 13. Objects classification: Results obtained in step 1.

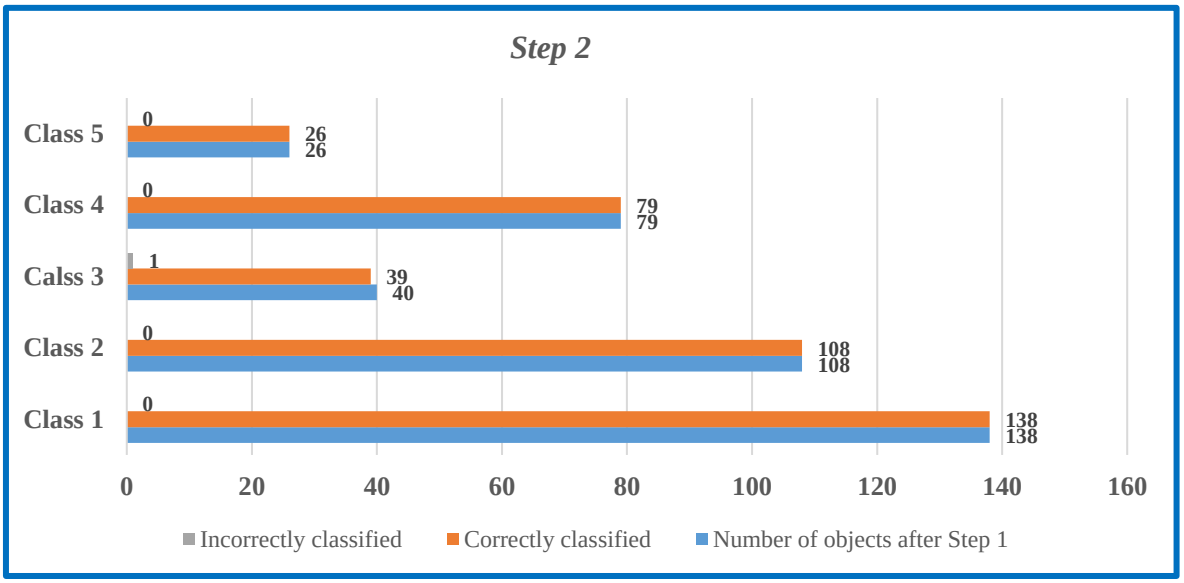

Figure 14. Objects classification: Results obtained in step 2.

Based on the results of the stage, a reference educational sample of subjects is formed, which consists of 138 subjects in grade 1: 108, grade 2: 39, grade 3: 79, grade 4: and grade 5: 26. Using the initial data, 2 objects out of 140 objects in 1st grade, 12 objects out of 120 objects in 2nd grade, 1 object out of 40 objects in 3rd grade, 1 objects out of 80 objects in 4th grade were removed from their class because they were transferred to another class as a result of classification and the 5th grade consisted of 26 objects, and not a single object in the 5th grade was transferred to another class, i.e., the object remained in its own class. It should be worth mentioning that the process leading to the results obtained above is based

on formulas (10) - (11).

The doctor comprehends complaints of the patients, collects anamnestic data and conducts an initial examination. Then, the initial examination fully analyzes the results and on condition it is necessary sends patients more careful study in laboratory and instrumental situations. After passing these examinations, the patient analyzes the results of the evaluation and makes a conclusive diagnosis, taking into account the concomitant diseases that are observed in the patient.

In practical matters, that we are taking into consideration, the class is heeded to be the diagnosis and the object is the patient. It literally means that the curriculum, that is drawn up in collaboration with medical professionals, is divided into elective classes meaning the more subjects there are present in the class, the more likely it is that the diagnosis (classification) made by that doctor will be inaccurate. This may probably be owing to the fact that the doctor acknowledges his issue well. Therefore, a re-classification process is discharged to select a curriculum and objects that consider their class to be incorrect are fully removed from their class. This process is maintained until the time the objects of the class detect own class 100% correctly leading to the formation of a lookup table.

Step 4. "The right selection of characteristics of information". The overall procedure uses the reference table previously created. Since the results of the classification according to the directory are equal to 100%, the essence of the work at the current stage is the choice of useful ℓ from the sets of characters. The method proposed below aims to overcome some of the problems mentioned, where signs can adopt on various kinds/types of values, and the suggested informative criteria for choosing a sign set is based precisely on reduction classification errors. Also, the probability vector used in character selection prevents unreasonable exclusion of objects without selecting important characters.

In analyzing data intellectually, it is important to determine how the objects of learning selections behave in different dimensional sign spaces; i.e., from one dimensional sign space to another dimensional sign space. For example, if the problem of classification as a result of changing the size of the sign space is resolved, some objects may be passed to a different class, but not to their own class. This statement can be expressed mathematically as follows:

$\exists x \in X, \lambda \in \Lambda^{N_1}, \mu \in \Lambda^{N_2}, N_1 \neq N_2$ учун $x|_\lambda \in X_p, x|_\mu \in X_q$, $p \neq q$.

To identify and prevent such cases, we take into account the errors in the classification process when reducing the size of the sign space.

Suppose that the error coefficient $\theta(\ell)|_\lambda$ of classification with respect to the informative $\lambda -$ vector ℓ and the number of misclassified objects that cannot find their class is $\varkappa(\ell)|_\lambda$. We determine the relationship between the error coefficient in the classification and the number of misclassified objects as follows:

$$\theta(\ell)|_\lambda = \frac{\varkappa(\ell)|_\lambda}{m}, (13)$$

In (13) the parameter $m = \sum_{p=1}^{r} m_p$ stands for the total number of objects in the learning selection.

Note 1. If it is significant that objects of a certain classes are not lost in another class, or, inversely, objects of another class fall into these classes, then when defining the error coefficient in the classification, it is necessary to take into account the importance of classes. For example, if the level of importance not to pass an object of class X_p to another class is α_p and vice versa, the level of importance not to pass another class object to this class is lost- β_p. In this case, the number of misclassified objects for class X_p can be determined according to the degree of significance as follows:

$$\varkappa(\ell)|_{\lambda}(X_p) = \alpha_p \cdot t_p^{out}|_{\lambda} + \beta_p \cdot t_p^{in}|_{\lambda}, (14)$$

Where t_p^{out} and t_p^{in} are the number of objects were left out of the X_p class and/or added. Combining (13) and (14) leads to the following expression of the error coefficient of classification:

$$\theta(\ell)|_{\lambda} = \frac{\sum_{p=1}^{r}(\alpha_p \cdot t_p^{out}|_{\lambda} + \beta_p \cdot t_p^{in}|_{\lambda})}{m} \ (15)$$

Definition 1. The system classifier of objects $X|_{\lambda} \in X|_{\Lambda^{\ell}}$ or the vector $\lambda \in \Lambda^{\ell}$ is said to be optimal if $\forall \mu \in \Lambda^{\ell}$ for $\theta(\ell)|_{\lambda} \leq \theta(\ell)|_{\mu}$ is feasible.

Definition.2. if $\theta(\ell) = \min_{\lambda \in \Lambda^{\ell}} \theta(\ell)|_{\lambda}$, then we have the classification error coefficient $\theta(\ell)$ in informative systems denoted $X|_{\Lambda^{\ell}}$.

The function of proximity betwixt objects and the decisive rule are required (or necessary) for the computation (or calculation) of $\theta(\ell)$ and $\varkappa(\ell)$ ($\ell = \overline{1,N}$). We use (10), (11) or (12) as a function (measure) of proximity for the research objects, depending on which values of the signs are to be taken. As a decisive rule, we use the algorithm for calculating the values defined by algorithms for calculating grades**.**

$$\Re(x)|_{\lambda} = \begin{cases} p, \text{arap } \Gamma_p|_{\lambda} - \Gamma_q|_{\lambda} \geq \delta, q \neq p, q = \overline{1,r} \\ 0, \textit{in all other cases} \end{cases} \ (16)$$

In (16), the function of proximity is denoted $\Gamma_q|_{\lambda} = \sum_{y \in X_q} \rho(x,y)|_{\lambda}$, $\rho(x,y)|_{\lambda}$-, and δ stands for a positive number (to be defined).

Note 2. If $\ell \ll N$, the following condition $\lambda \in \Lambda^{\ell}$ optimal vector and $\varkappa(\ell)|_{\lambda}$ should be fulfilled:

$$\theta(\ell) \leq \theta(N) \ (17)$$

Note 3. When trying to meet the previous condition (see Note 1), a problem may occur as follows: there may not be $\ell < N$ that satisfies the condition (16), i.e., every sign in the space of the initial signs is important, and excluding any of the signs can adversely/negatively affect the quality of classification.

A suggestion for a test solution. In such a situation, if the space of signs needs to be reduced, instead of (16) use the following condition:

$$\theta(\ell) \leq \theta(N) + \delta \ (18)$$

The quantity δ corresponds to the allowed mistake, which is fixed/obtained as $\theta = \theta(N)$.

To form information systems that classify objects, we can include the probability vector as follows:

$$p_{\nu} = (p_{\nu}^1, p_{\nu}^2, \dots, p_{\nu}^N) \ (19)$$

Here, p_{ν}^j stands for the probability of storing the j^{th} sign in a system; ν expresses the order of iteration. When $\nu = 1$, we take $p_1 = (p_1^1, p_1^2, \dots, p_1^N)$ for $p_1^j = \frac{1}{N}; j = \overline{1,N}$. If $\lambda \in \Lambda^N$ and $\mu \in \Lambda^k$, then $\lambda - \mu \in \Lambda^{N-k}$.

For all $\mu \in \Lambda^k$ the values of the vectors $\varkappa(N-k)|_{\lambda-\mu}$, $\theta(N-k)|_{\lambda-\mu}$ and $\theta(k)|_{\mu}$ are determined and the following condition is checked:

$\theta(N-k)|_{\lambda-\mu} \leq \theta$. (20)

If the condition (20) is feasible, the probability of the signs corresponding to the vector μ is divided by the probabilities of the signs corresponding to the vector $\lambda - \mu$, and this

corresponds to (21).

$$p_{\nu+1}(\mu) = p_\nu + \frac{(p_\nu,\mu)}{N-k} * p_\nu|_{\lambda-\mu} - p_\nu|_\mu. \quad (21)$$

As a result, the likelihood (or probability) of certain signs participating in the next contest (or competition) increases, while the likelihood of some signs participating decreases.

$$p_{\nu+1} = \frac{1}{T}\sum_{\mu} p_{\nu+1}(\mu) \quad (22)$$

In (22), the quantity T stands for the number of vectors μ satisfying the condition (20).

After each iteration we order them according to the values of the components of the likelihood vector $p_{\nu+1} = (p^1_{\nu+1}, p^2_{\nu+1}, \dots, p^N_{\nu+1})$. For components $p_{\nu+1}$ of the vector there is a sequence of $j_1, j_2, \dots, j_N$ indices, for which one of the following two cases ((23) or (24)) is suitable/appropriate:

$$p^{j_1}_{\nu+1} \geq p^{j_2}_{\nu+1} \geq \dots \geq p^{j_{N-k}}_{\nu+1} > p^{j_{N-k+1}}_{\nu+1} \geq \dots \geq p^{j_N}_{\nu+1} \quad (23)$$

$$p^{j_1}_{\nu+1} \geq \dots > p^{j_{N-k-t}}_{\nu+1} = \dots = p^{j_{N-k+l}}_{\nu+1} > p^{j_{N-k+l+1}}_{\nu+1} \geq \dots \geq p^{j_N}_{\nu+1}, \quad (24)$$

In formulas (23) and (24), t and l are positive numbers.

If the condition (23) can be realized, k signs corresponding to the index $j_{N-k+1}, j_{N-k+2}, \dots, j_N$ are sent out of signs space and assimilated as $N = N - k$.

If the condition (24) can be realized, then there are k signs which correspond to non-zero components of the vector $\mu^* \in \Lambda^k$, which conditions $\theta(k)|_{\mu^*} = \max_{\mu\in\Lambda^{*k}} \theta(k)|_\mu$ are sent out of signs space and assimilated as $N = N - k$. Here, $\Lambda^{*k} = \{\mu \in \Lambda^k | \, \mu^{j_i} = 1, i = \overline{N-k+l+1, N}, \mu^{j_z} \in \{0,1\}, z = \overline{N-k-t, N-k+l}, \mu^{j_g} = 0, g = \overline{1, N-k-t-1}\}$.

This process repeats only $\frac{N-\ell}{k}$ or $\left[\frac{N-\ell}{k}\right] + 1$, here $[*]$ the whole part of the number. If the process repeats only $\left[\frac{N-\ell}{k}\right] + 1$, then in the last cycle $k = N - \ell$ is took, because here, $k \geq N - \ell$ is already observed.

When practical questions are solved, k for $k \ll \ell$ condition should be viable and usually , $k = 3 \div 5$ is recommended. .

Note 3. We now present the overall procedure for selecting $\mu \in \Lambda^k$ vectors. Obviously, in the $\mu \in \Lambda^k$ set the number of informative k vectors is equal to $\frac{N!}{k!(N-k)!}$. The strict order/sequence of selection of these vectors to increase productivity can be defined as follows.

- for μ vector , conditionally, it can be $\begin{cases} \mu_1 = \left\{\underbrace{0, 0, \dots, 0}_{N-k}, \underbrace{1, 1, \dots, 1}_{k}\right\} \\ \text{and} \\ \mu_{\frac{N!}{k!(N-k)!}} = \left\{\underbrace{1, 1, \dots, 1}_{k} \underbrace{0, 0, \dots, 0}_{N-k}\right\} \end{cases}$

It is possible to determine the sequence of these vectors as follow.

For this, in a general way, it is enough to identify the vector coming at $h + 1$place by knowing the vector at h place.

In general, on $\mu_h = (\mu^1_h, \mu^2_h, \dots, \mu^N_h)$ vector components at h - place, the following cases are observed:

Case 1. components with a value of 1 are arranged in series, i.e. there is no component

with a value of 0 between two arbitrary components with a value of 1;

Case 2. Components with a value of 1 or 0 are arranged arbitrarily, i.e. mixed.

Let's first look at what the next vector looks like in case 1..

According to the $\mu_h = (\mu_h^1, \mu_h^2, \dots, \mu_h^N)$ vector components index, beginning from the least index, the component that its value is equal to one is found, and this component is replaced by the previous component, i.e., if the index of component found is j , then it will be replaced by $j-1$. All other componentswith a value of 1 will be pushed to $N-k$.

$$\mu_h^1 = 0, \dots, \mu_h^j = 0, \mu_h^{j+1} = 1, \mu_h^{j+2} = 1, \dots, \mu_h^{j+k} = 1, \mu_h^{j+k+1} = 0, \dots, \mu_h^N = 0 \Longrightarrow$$

$$\mu_{h+1}^1 = 0, \dots, \mu_{h+1}^{j-1} = 0, \mu_{h+1}^j = 1, \mu_{h+1}^{j+1} = 0, \dots, \mu_{h+1}^{N-k+2} = 0, \mu_{h+1}^{N-k+1} = 1, \dots, \mu_{h+1}^N = 1$$

Now let's look at the second case. In this case, unlike in case 1, the process will be started with the largest index (N) . First of all, the largest index satisfying the $\mu_h^j = 0, \mu_h^{j+1} = 1$ condition is determined and they will ne replaced, i.e. beginning from $\mu_h^j = 1, \mu_h^{j+1} = 0$ and $j+2$ all components with a value of 1 are replaced from the N index, i.e.

$$\mu_h^1 = *, \dots, \mu_h^j = 0, \mu_h^{j+1} = 1, \dots, \mu_h^{j+q} = 1, \mu_h^{j+q+1} = 0, \dots, \mu_h^N = 0 \Longrightarrow$$

$$\mu_{h+1}^1 = *, \dots, \mu_{h+1}^{j-1} = *, \mu_{h+1}^j = 1, \mu_{h+1}^{j+1} = 0, \dots, \mu_{h+1}^{N-q} = 0, \mu_h^{N-q+1} = 1, \dots, \mu_h^N = 1.$$

The mechanism of the operation of the algorithm is as following, that is, an arbitrary single character is selected and fixed from the symbols characterizing class objects in the generated look-up table. The classification process is then discharged. Providing that, as a result of classification, objects in classes do not recognize 100% of their class correctly or make up less than 100%, that is, if incorrectly recognized objects are identified in the classes, fixed symbol is returned to its original place. Conversely, if, as a result of classification, the objects of the class possess the correct identification ratio of 100%, the fixed symbol is then removed from the look-up table.

The process continues until the maximum character reduction is obtained during the process of maintenance of the quality of the classes.

Step 1. Initial data : X –learning selection; X_p ($p = \overline{1,r}$) –classes; r –number of classes ; k, ℓ, N.

Step 2. $\varkappa(N)$ and $\theta = \theta(N)|_\lambda$ are calculated for $\lambda \in \Lambda^N$.

Step 3. $i = k$

Step 4. When $\nu = 1, p_\nu^j = \frac{1}{N}; j = \overline{1,N}$ values are assimilated for $p_\nu = (p_\nu^1, p_\nu^2, \dots, p_\nu^N)$

Step 5. $\theta(N-i)|_{\lambda-\mu}$ and $\theta(i)|_\mu$ are calculated for all $\mu \in \Lambda^i$ vectors..

Step 6 for all μ vectors : if $\theta(N-i)|_{\lambda-\mu} \le \theta$ is viable , then $p_{\nu+1}(\mu) = p_\nu + \frac{(p_\nu,\mu)}{N-i} * p_\nu|_{\lambda-\mu} - p_\nu|_\mu$ is assimilated.

Step 7. $p_{\nu+1} = \frac{1}{T}\sum_\mu p_{\nu+1}(\mu)$is calculated, here T - $p_{\nu+1}(\mu)$ is the number of vectors. (3.3.7) or (3.3.8) are checked for $p_\nu = (p_\nu^1, p_\nu^2, \dots, p_\nu^N)$ vector and accordingly i signs are removed from the space of signs.

Step 8 . then $N = N - i$ and $\lambda = (1, 1, \dots, 1) \in \Lambda^N$ are assimilated as new value.

Step 9 . If $i \le N - \ell$, then from step 3 is returned.

10-step. If $i \ge N - \ell$, then it is considered as $i = N - \ell$ and from step 5 is returned.

11-step Output parameter : λ optimal vector .

Therefore, the essence of this algorithm is to select the most useful l from the set of signs that characterize the objects under study, i.e., to solve the problem of evaluating the set of informative signs.

Based on the above steps and algorithms for solving the above four practical problems, a software tool called "Algorithm and software of initial processing of medical data" was developed.

During the operating the module "Adaptive genetic algorithm-based on random search of a set of informative signs" of this software, the objects of the class can find 100% of their class, that is, when the process of minimizing the 62 parameters that characterize the class objects with a margin of error of 0% was reduced from 62 to 10 and the result obtained can be seen in Table 3 below.

Table 2. A set of informative signs.

$l = 15$	
1.	$x_2, x_3, x_4, x_6, x_9, x_{16}, x_{17}, x_{27}, x_{30}, x_{31}, x_{32}, x_{37}, x_{52}, x_{53}, x_{58}.$
2.	$x_1, x_3, x_6, x_9, x_{16}, x_{17}, x_{22}, x_{28}, x_{35}, x_{37}, x_{39}, x_{52}, x_{58}, x_{59}, x_{63}.$
3.	$x_1, x_3, x_7, x_9, x_{10}, x_{17}, x_{22}, x_{35}, x_{36}, x_{39}, x_{47}, x_{53}, x_{58}, x_{59}, x_{63}.$
4.	$x_1, x_3, x_4, x_9, x_{10}, x_{15}, x_{16}, x_{17}, x_{35}, x_{38}, x_{39}, x_{47}, x_{53}, x_{58}, x_{60}.$
5.	$x_3, x_6, x_9, x_{17}, x_{30}, x_{36}, x_{39}, x_{40}, x_{47}, x_{48}, x_{52}, x_{53}, x_{58}, x_{59}, x_{62}.$
6.	$x_3, x_{10}, x_{16}, x_{17}, x_{22}, x_{35}, x_{37}, x_{39}, x_{49}, x_{53}, x_{54}, x_{57}, x_{58}, x_{59}, x_{63}.$
7.	$x_3, x_4, x_9, x_{16}, x_{17}, x_{36}, x_{41}, x_{44}, x_{47}, x_{48}, x_{52}, x_{54}, x_{58}, x_{59}, x_{62}.$
8.	$x_1, x_3, x_7, x_9, x_{10}, x_{17}, x_{44}, x_{50}, x_{52}, x_{53}, x_{54}, x_{57}, x_{58}, x_{60}, x_{61}.$
$l = 14$	
9.	$x_1, x_3, x_4, x_9, x_{10}, x_{16}, x_{17}, x_{34}, x_{37}, x_{47}, x_{50}, x_{52}, x_{53}, x_{58}.$
10.	$x_1, x_3, x_6, x_9, x_{10}, x_{16}, x_{17}, x_{35}, x_{37}, x_{38}, x_{47}, x_{58}, x_{59}, x_{63}.$
11.	$x_1, x_2, x_3, x_6, x_7, x_9, x_{10}, x_{16}, x_{17}, x_{35}, x_{36}, x_{39}, x_{47}, x_{58}.$
12.	$x_3, x_6, x_7, x_9, x_{16}, x_{17}, x_{35}, x_{39}, x_{47}, x_{49}, x_{53}, x_{58}, x_{59}, x_{63}.$
13.	$x_3, x_5, x_9, x_{10}, x_{16}, x_{17}, x_{25}, x_{30}, x_{44}, x_{47}, x_{50}, x_{53}, x_{58}, x_{59}.$
$l = 13$	
14.	$x_3, x_9, x_{10}, x_{16}, x_{17}, x_{35}, x_{47}, x_{48}, x_{52}, x_{53}, x_{57}, x_{58}, x_{63}.$
15.	$x_3, x_9, x_{10}, x_{16}, x_{17}, x_{36}, x_{41}, x_{47}, x_{53}, x_{54}, x_{58}, x_{62}, x_{63}.$
$l = 12$	
16.	$x_3, x_9, x_{16}, x_{17}, x_{35}, x_{39}, x_{47}, x_{52}, x_{53}, x_{58}, x_{62}, x_{63}.$
17.	$x_3, x_5, x_9, x_{16}, x_{17}, x_{30}, x_{47}, x_{52}, x_{53}, x_{57}, x_{58}, x_{63}.$
$l = 11$	
18.	$x_3, x_9, x_{10}, x_{15}, x_{16}, x_{17}, x_{38}, x_{47}, x_{52}, x_{53}, x_{58}.$
19.	$x_3, x_9, x_{10}, x_{16}, x_{17}, x_{22}, x_{47}, x_{53}, x_{58}, x_{59}, x_{63}.$
$l = 10$	
20.	$x_3, x_9, x_{10}, x_{16}, x_{17}, x_{47}, x_{53}, x_{58}, x_{62}, x_{63}.$
21.	$x_3, x_9, x_{10}, x_{16}, x_{17}, x_{36}, x_{47}, x_{53}, x_{58}, x_{62}.$
22.	$x_3, x_9, x_{10}, x_{16}, x_{17}, x_{38}, x_{47}, x_{53}, x_{58}, x_{62}.$

A set of informative signs from l = 15 to l = 10 found in the calculation in **Table 2.**

III.CONCLUSION

The article describes methods and algorithms for solving problems in the following three domains of recognition of pattern, including the formation of academic choice, classification, the choice of a set of symbols that are informative. Based on this theory, the presented methods and algorithms were all applied in solving practical problems that are related to cardiology.

5 cases of coronary heart disease as an object of study are researched in this paper.

By solving the above problems, the following results were achieved:

- According to the first problem, the degree of availability for usage of the signs characterizing the objects of the selection classes we are considering;
- According to the second problem, in the medical data under study, the process of conversion of the signs that characterizing objects in classes from continuous quantitative view to 0 or 1 view was carried out;
- In solving the third problem, it was determined that the objects divided into classes belong to their own class or another class, that is, as a result of the exclusion of objects that did not find their own class, a reference table was formed;
- According to the fourth problem, a significant level of signs was determined from the standard table, that is, based on the results of reclassification, a set of informative signs that clearly distinguish between class objects was extracted.

REFERENCE

1. Yu.I. Zhuravlev Selected scientific works. –M: Magister Publishing House, 1998. – 420 p.
2. Yu, L., H. Liu. Efficient Feature Selection via Analysis of Relevance and Redundancy. – J. Mach. Learn. Res., Vol. 5, 2004, No Oct, pp. 1205-1224.
3. Yu, L.: Toward Integrating Feature Selection Algorithm for Classification and Clustering. IEEE Transaction on Knowledge and Data Engineering 17(4), 491-502
4. Yan, K., D. Zhang. Feature Selection and Analysis on Correlated Gas Sensor Data with Recursive Feature Elimination. – Sensors Actuators, B Chem., Vol. 212, Jun 2015, p. 353-363.
5. Nishanov A.Kh,, Turakulov Kh.A., Turakhanov Kh.V. A decision rule for identification of eye pathologies // Biomedical Engineering (1999) 33(4) 178-179.
6. Nishanov A. Kh,, Turakulov Kh.A., Turakhanov Kh.V. A decisive rule in classifying diseases of the visual system // Meditsinskaia tekhnika (1999) (4) 16-18.
7. Xiang Fang, Lina Wang Feature Selection Based on Fisher Criterion and Sequential Forward Selection for Intrusion Detection // Revista de la Facultad de Ingeniería U.C.V., Vol. 32, N° 1, pp.498-503, 2017.
8. Linhui Sun, Sheng Fu and Fu Wang. Decision tree SVM model with Fisher feature selection for speech emotion recognition // Sun et al. EURASIP Journal on Audio, Speech, and Music Processing (2019) 2019:2 https://doi.org/10.1186/s13636-018-0145-5
9. Yan, X.,Tan, M., Yan,Y., Lü, M. (2012). “Research on Hidden Markov Model-Based And Neural Network-Based Intrusion Detections”. Computer Applications and Software. 29(2), 294–297.
10. JING, X.,WANG, H., NIE, K., LUO, Z. (2012). “Feature Selection Algorithm Based on IMGA and MKSVM to Intrusion Detection”. Computer Science, 39(7), 96–100.
11. Koushal Kumar, Jaspreet Singh Batth. Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms // International Journal of Computer Applications (0975 – 8887) Volume 150 – No.12, September 2016, 1-13.
12. Nishanov, A. Kh., Djurayev, G., & Khasanova, M. A. Improved algorithms for calculating evaluations in processing medical data// Compusoft: An International Journal of

Advanced Computer Technology, (2019), 8(6), 3158-3165.

13. Kamilov M., Nishanov A., Beglerbekov R. Modified stages of algorithms for computing estimates in the space of informative features // International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-8 Issue-6S, April 2019, 714-717.
14. Nishanov A. Kh., Djurayev, G., & Khasanova, M. A. Classification and feature selection in medical data preprocessing// Compusoft: An International Journal of Advanced Computer Technology, (2020), 9(6), 3725-3732.
15. 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.
16. Nishanov A.Kh., Jurayev G.P., Narziyev N.B., Adaptable genetic algorithm for primary processing of medical data based on random search // Шестая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск, Республика Беларусь, 20-21 мая 2020 года, 147-155.
17. Nishanov AH, Akbaraliev BB, Juraev GP Algorithm for the selection of ischemic heart disease symptom complexes based on a randomized search // "Muhammad al-Khwarizmi Avlodlari" scientific-practical and information-analytical magazine, Tashkent, № 2 (12), June 2020, 10-15 p
18. A. Kh. Nishanov, O. B. Ruzibaev, J. C. Chedjou, K. Kyamakya, Kolli Abhiram, Perumadura De Silva, G.P. Djurayev and M.A. Khasanova Algorithm for the selection of informative symptoms in the classification of medical data// World Scientific Proceedings Series on Computer Engineering and Information ScienceDevelopments of Artificial Intelligence Technologies in Computation and Robotics, pp. 647-658 (2020).
19. Nishanov A. Avazov E. Akbaraliyev B. Partial selection method and algorithm for determining graph-based traffic routes in a real-time environment// International Journal of Innovative Technology and Exploring Engineering. ISSN: 2278-3075, Volume-8 Issue-6S, April 2019, 696-698.
20. Nishanov A. Ruzibaev O. Tran N. Modification of decision rules 'ball Apolonia' the problem of classification// 2016 International Conference on Information Science and Communications Technologies, ICISCT 2016 (2016). DOI: 10.1109/ICISCT.2016.7777382.
21. Nishanov A.Kh., Akbaraliev B.B., Samandarov B.S., Akhmedov O.K., Tajibaev S.K. An algorithm for classification, localization and selection of informative features in the space of politypic data// Webology, Volume 17, No 1, 2020, Pages: 341-364.