Available online at: https://ijact.in

Date of Submission	22/04/2020
Date of Acceptance	16/05/2020
Date of Publication	02/07/2020
Page numbers	3725-3732 (8 Pages)

ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

CLASSIFICATION AND FEATURE SELECTION IN MEDICAL DATA PREPROCESSING

[1]Akhram Khasanovich Nishanov, [2]Gulomjon Primovich Djuraev, [3]Malika Akhramovna Khasanova

Professor, Information Technologies Software Department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, DSc
Independent Scientific Researcher, Information Technologies Center at the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Assistant, Department of Faculty and Hospital Therapy No. 2, Tashkent Medical Academy, Republic of Uzbekistan

nishanov_akram@mail.ru, jurayev_g@bk.ru, malikabonuxasanova@gmail.com

Abstract: In this article, the issues such as medical data preprocessing, reclassification of training sets and determining the importance of classes, forming reference tables were solved. As a result of data preprocessing, three objects were formed: 1) Ischemic heart disease. Unstable angina pectoris; 2) Ischemic heart disease. Acute myocardial infarction; 3) Ischemic heart disease. Arrhythmic form. Further, the issues such as classifying, selecting a set of informative features that differentiate between class objects were solved using Fisher criterion and algorithms for an estimated calculation as well as software programs for them were developed. As a result data preprocessing reference classes were formed. Objects that had fallen outside of their class during the formation process were excluded from the training set. A classification and a set of informative features were selected using established classes. Initially, three class objects each containing 62 features were provided by medical professionals and as a result data preprocessing three sets consisting of 131, 115, and 40 objects respectively in three classes were used to form a reference table.

Keywords: Fisher criterion, feature selection, classification, algorithms for an estimate calculation, medical data preprocessing

I. INTRODUCTION

When addressing the issues of medical data mining, we encounter dimensional and space problems in identifying hidden links and regularities between the set of features and the features that characterize the objects of research.

It will depend on finding solutions to the pattern recognition issues. One of the key phases of pattern recognition goes back to the classification and the selection of criteria that determine the content of the problem of informative feature recognition.

The Fisher criterion is widely used for medical data processing [5-8, 11-13], which is also discussed in this article.

The study consists of two important steps. The first one is to create a reference table, and the another one is to choose the most useful set of characteristic features to be investigated, i.e. to solve the issue of selecting a set of informative features.

Finding the solution to the first issue is based on the importance of features and objects as well as their assigning to the classes [1-4, 9, 10].

The solution of the another issue, namely selecting informative features from the given table, their visualization and determining the assignment of a set of features to the formation of classes has been thoroughly investigated by the authors.

Modern analysis of biomedical data requires feature selection methods that could be applied to large-scale feature spaces, function in noisy tasks, discover complex association patterns, flexibly adapt to different problem areas and data types (for example, genetic variants, gene expression and clinical data), and can be computationally computed. To this end, in [13], a set of algorithms for selecting elements in the filter style based on the Relief algorithm, i.e. Relief-based algorithms(RBAs), is considered in the study. The RBA is being introduced and expanded in an open-source environment called ReBATE (a learning environment based on bumping algorithms). A comprehensive study of genetic modeling is offered that compares existing RBAs, proposed by RBA called MultiSURF and other established methods for selecting features, for a number of problems.

In [12], the problems of diagnosis and treatment of cardiovascular diseases, which are often encountered when making diagnostic decisions in the processing of medical data is considered. The classification of heart diseases and the identification of informative features are carried out on the basis of algorithms for an estimate calculation. The main purpose of the study is to solve the issues of inter object remoteness function in a set of informative features that distinguish objects of diagnostic classes, selecting a set of features characterizing the mutual differences of objects as well as identifying the proximity function when diagnosing an unknown object [17-22]. Identifying the level of significance, which are the main stages of the algorithms for an estimate calculation relative to classes [22]. The algorithm for diagnosing unknown object in the space of informative feature is proposed. The proposed theoretical ideas were confirmed in practice. In addition, the decision-making rules in this space and their software were developed [21,22].

Multidimensional data analysis is a challenge for researchers and engineers in the field of machine learning and data mining. Selection of functions provides an effective way to solve this problem by removing unnecessary and redundant data that can help to reduce computation time, improve training accuracy and facilitate understanding of the training model or data. In [14], several commonly used evaluation indicators were studied for selecting features, and then methods for selecting controlled, uncontrolled, and semi-serviced features that are widely used in machine learning problems, such as classifying and clustering were investigated.

When a set of feature objects contains highly correlated features, the SVM-RFE ranking criterion will be biased, making it difficult to apply the SVM-RFE to gas sensor data. The article [15] considers linear and nonlinear SVM-RFE algorithms. After investigating the correlation bias, an improved SVM-RFE + CBR algorithm is proposed that includes a strategy for reducing correlation bias (CBR) in the feature elimination procedure. The ensemble method is additionally studied to increase the stability of the proposed method.

Feature selection is an important stage of data pre-processing, which increases the performance of training algorithms by removing unnecessary and redundant features. In [16], a method for feature selection using the Forest Optimization Algorithm (FSFOA) is proposed. To select more informative features from data sets, the FSFOA method is proposed and implemented on several real data sets and compared with several other methods, including HGAFS, PSO and SVM-FuzCoc. Experimental results show that FSFOA can improve the classification accuracy of classifiers in some selected data sets.

Feature selection is an important task in data mining and machine learning applications that eliminate unnecessary and redundant functions and increase learning productivity. In many real-world applications, it is difficult to collectlabeleddata, while plentiful unlabeled data is easily accessible. This allows researchers to develop methods for selecting objects that use both labeled and unlabeled data to assess the relevance of the object. However, to date, no comprehensive survey has been conducted covering the methods of selecting objects under observation. In [23], the object selection methods under observation are completely studied, and two taxonomies of these methods are presented on the basis of two different points of view, which represent the hierarchical structure of object selection methods under observation. The first point of view is based on the taxonomy of feature selection methods, and the other is based on the taxonomy of observation methods. This solution can be useful for a researcher to get a deep background during observation and choose the right method for selecting objectsbased on their hierarchical structure.

In this article, the issues such as the preprocessing of medical data generated by medical professionals in medical data mining, reclassifying training sets and identifying the importance level of classes, forming reference tables, and selecting informative features that differentiate between class objects are solved based on Fisher criterion using algorithms for an estimate calculation.

II. METHODOLOGY

In this section:

1). The data include preliminary data preprocessing issues and are given in the case of medical issues. The first of the 4 issues in this section is to determine the feasibility of the characteristic features of the objects classified, the second is to convert characters from classed objects into continuous numbers from 0 to 1, the third is to form a reference table by determining whether a class object belongs to its class or to another class, and the fourth issue focuses on determining a set of informative features that clearly differentiate themselves from the objects in the class;

2). Methods for solving these problems are described, which include the **proximity function** that provides the similarity of objects in the space of these informative features and uses algorithms for an estimate calculation based on Fisher criterion;

3). The steps to solve practical problems based on the proposed theoretical data are developed. They describe step-by-step solutions for the class of “ischemic heart disease” on the basis of symptoms and associated objects.

1. Statement of the problem. Let us assume that the curriculum formed on the basis of primary data is divided into the training set classes and presented as follows:

$$K_1 = \begin{bmatrix} x_{11}^1 & x_{11}^2 & \dots & x_{11}^N \\ x_{12}^1 & x_{12}^2 & \dots & x_{12}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m_1}^1 & x_{1m_1}^2 & \dots & x_{1m_1}^N \end{bmatrix} \dots$$

$$\mathrm{K_r} = \begin{bmatrix} \mathrm{x}_{r1}^1 & \mathrm{x}_{r1}^2 & \dots & \mathrm{x}_{r1}^N \\ \mathrm{x}_{r2}^1 & \mathrm{x}_{r2}^2 & \dots & \mathrm{x}_{r2}^N \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{x}_{rm_r}^1 & \mathrm{x}_{rm_r}^2 & \dots & \mathrm{x}_{rm_r}^N \end{bmatrix}.$$

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

This can be summarized as follows:

Here $p = \overline{1,r}$;and the training set is expressed in the form $K = \bigcup_{p=1}^{r} K_p$ and may be represented by classes that do not intersect, that is, $K_p \cap K_q = \emptyset, (p \neq q, p = \overline{1,r};\ q = \overline{1,r};)$ conditions are given.

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

In the issues we are looking at, each class is treated as one type of disease: ClassK_1 "Unstable angina pectoris," class K_2"Acute myocardial infarction," class K_3 "Arithmic form." At the same time, the characteristic feature of each class (type of disease) is formed by experts in the field and consists of 62 features that characterize each class. You can see the names of these features in Table 1 below.

Table: 1

№	Feature Names	
1.	painintheheart	x1
2.	chestpain	x2
3.	epigastricpain	x3
4.	pain in the area of the left shoulder blade	x4
5.	pain in the left shoulder and arm	x5
6.	painduration	x6
7.	nitroglycerinhelps	x7
8.	painafterexercise	x8
9.	pain at rest	x9
10.	coldsweat	x10
11.	weakness	x11
12.	shortness of breath on exertion	x12
13.	dyspneaat rest	x13
14.	positionorthoptic	x14
15.	swellinginthelegs	x15
16.	LV hypertrophy	x16
17.	enlargedliver	x17
18.	feeling of lack of air	x18
19.	Heartrate	x19
20.	systolicbloodpressure	x20
21.	diastolicbloodpressure	x21
22	Saturation (SPO2).	x22
23	hemoglobin	x23
24	erythrocyte	x24
25	leukocyte	x25
26	ESR	x26
27	bloodsugar	x27
28	Alt	x28
29	AST	x29
30	KFK	x30
31	VSK (sec)	x31
32	PTI	x32
33	Thrombotest	x33
34	Fibrinogen	x34
35	Faq	x35
36	TPG	x36
37	Clotretraction	x37
38	leftatrium	x38
39	interventricularseptum	x39
40	WLJ	x40
41	KDR	x41
42	DAC	x42
43	BWW	x43
44	Ejectionfraction (PV%)	x44
45	ImpactVolume (IVml)	x45
46	total contractility of the left ventricle	x46
47	reductioninregionalcontractility	x47
48	Hypokinesis	x48
49	hypertension	x49
50	diabetes	x50
51	PEAKS	x51
52	NKII (A)	x52
53	NKII (B)	x53
54	COPD	x54
55	chroniccholecystitis	x55
56	pulmonarycongestion	x56
57	Highbloodpressure	x57
58	ECG changes	x58
59	Paintheheart	x59
60	Leukocytosis	x60
61	Mutedhearttones	x61
62	Disorders of local contractility of the left ventricle of the heart	x62

Problem 1. Determine the feasibility of the features that characterize the objects of the classes we are considering.

Problem 2. The features that characterize the objects of the classes we are considering should be continuously converted from 0 to 1.

Problem 3. It is necessary to solve the classification problem of the K_p class objects, that is, to define whether the class objects belong to one class or another.

Problem 4. $K_p, p = \overline{1,3}$ requires the selection of $\ell \ll 62$informative features that can clearly distinguish between the objects in the class. Here ℓ is a predetermined small number and is read from 62 to a lesser number.

2. Phases of addressing practical problems:

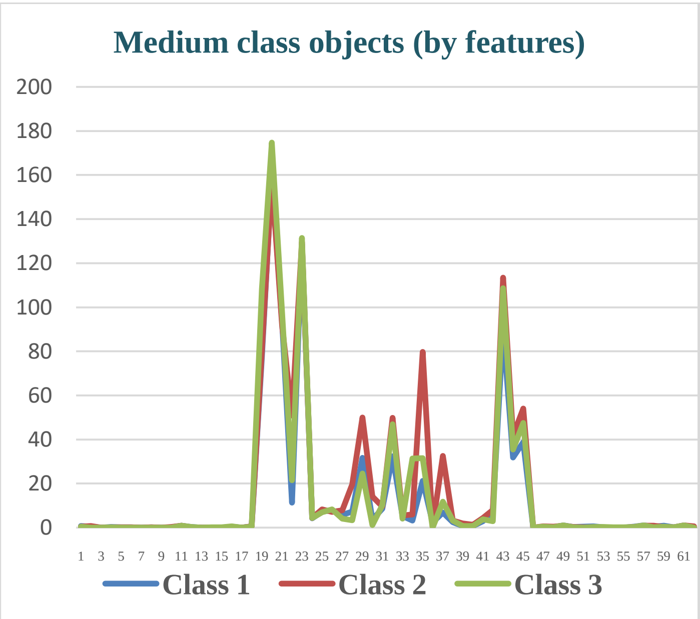
Phase1. Determining the feasibility of characteristic features of each object belonging to the above $K_p$class is done separately for each class in the following order:

a).Let us make the following determinations: : $\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \dots, \bar{x}_p^N)$vector, X_p, the average representative objects of classes , $p = \overline{1,r}$. Compute its components by the following formula:

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

The results calculated in the cross section of each class are shown in Chart 1 below.

Chart 1

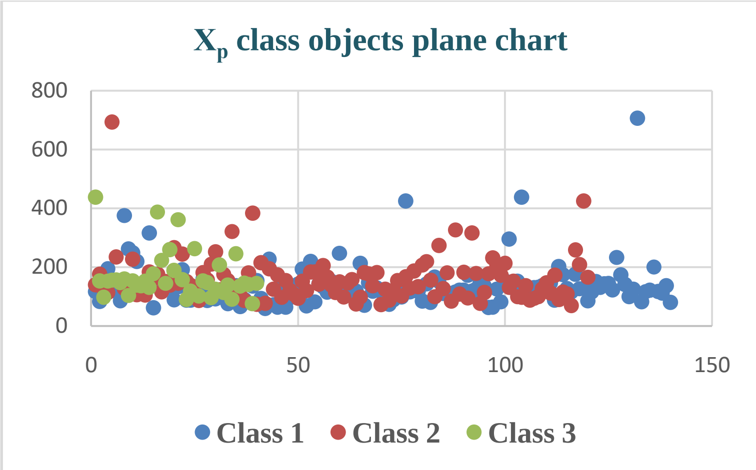


b).Calculate the distance of the X_p class between the objects x_{pi}ва $\bar{x}_p$x by the following formula:

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

The plane chart(Chart 2) for each class object is shown below.

Chart 2



c). The upper limit (exclude) $D(\bar{x}_p)$ of squares taking into account the objects of class X_p is calculated by the following formulas:

$$D(\bar{x}_p) = \sqrt{\frac{1}{m_p}\sum_{j=1}^{m_p}|x_{pi} - \bar{x}_p|^2} = \sqrt{\frac{1}{m_p}\sum_{i=1}^{m_p}\sum_{j=1}^{N}(\bar{x}_p^j - x_{pi}^j)^2}\ . \quad (3)$$

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

The results of the mean squared deviation of each class $D(\bar{x}_p)$ are shown in the table below (Table 2).

Table 2

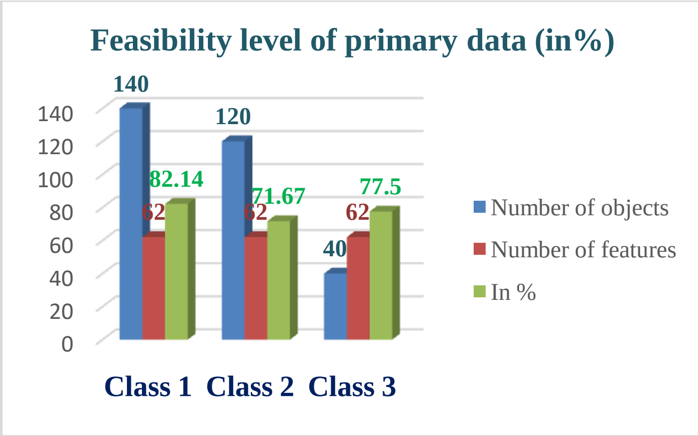
Class 1 ($D(\bar{x}_1)$)	Class 2 $D(\bar{x}_2)$	Class 3 $D(\bar{x}_3)$
159,6286	177,4676	184,4779

d). Satisfy the following inequality and calculate its parameters as a percentage of class objects:

$$|x_{pi} - \bar{x}_p| \le D(\bar{x}_p),\ p = \overline{1,r};\ i = \overline{1,m_p}. \quad (4)$$

At the end of Phase, the index for classes 1-2-3 is shown in the following Chart 3:

Chart 3



According to Chart 3 above, the feasibility level of characteristic features of each object that belongs to the K_p class are 82.14% in Class 1, 71.67% in Class 2, 77.50% in Class 3.

Phase 2: The given initial data is presented in a continuous quantitative form, at which phase the process of converting the feature values of class objects to values of 0 and 1 vector is carried out.

The process of converting the values of zero or one-character features to each of the above-mentioned K_p class objects in vector form is carried out by typing the following symbols in each class and all character sections:

a). $\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \ldots, \bar{x}_p^N)$ vector, K_p mean objects of classes, $p = \overline{1, r}$. Compute its components by the following formula:

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

Let us define the following vectors $a_p = (a_p^1, a_p^2, \ldots, a_p^N)$ and $b_p = (b_p^1, b_p^2, \ldots, b_p^N)$, and calculate its components by the formula:

$$a_p^j = \frac{1}{m_p} \sum_{i=1}^{m_p} (\bar{x}_p^j - x_{pi}^j)^2, \;\; p = \overline{1, r}; \;\; j = \overline{1, N}. \quad (6)$$

$$b_{pi}^j = (\bar{x}_p^j - x_{pi}^j)^2, \;\; p = \overline{1, r}; \;\; j = \overline{1, N}. \quad (7)$$

в). The components of the K_p elements of the training set are converted from the actual numeric form to the view using the following procedure.

$$x_{pi}^j == \begin{cases} equal\ 1, \; if \; \frac{b_{pi}^j}{a_p^j} \le 1, \\ \;\; else\ equal\ to\ 0 \end{cases} \quad (8)$$

At the end of this phase, the characteristic feature values that characterize the 3 class objects are converted to 0 and 1 vector values.

Phase 3. Deciding the classification issue of the objects in the K_p class, that is to determine whether each object in the class belongs to a its own class or a different one.
At the same time, each object belonging to the class X_p is compared with objects in its class and other classes, and the function of inter-object proximity in the space of informative features is r_i (x_ (p1,) x_p2):

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

(9)

The first condition denotes the degree of similarity between the two objects, and the second condition indicates that they are different.
The total of comparative evaluation is based on the following formula:

$$\Gamma_j(x_{pj}, x_{pk}) = \sum_{\kappa=1}^{m_p} \sum_{i=1}^{N} \rho_i(x_{pj}, x_{pk}), j = 1{,}mp;\ k{=}1{,}mp;\ j{\neq}k. \quad (10)$$

Comparative evaluation is calculated for each class, and the largest of the mean values obtained is the attribution of the object to that class.
At this phase, the classification of class objects will be carried out to determine if the class objects belong to their class or to another class. This process is carried out step by step, excluding objects that do not belong to their class at each step, and the objects in the classes are complete until they reach their full class, which is 100%. The results of each step are presented in Charts4, 5, 6, and 7.

Chart 4

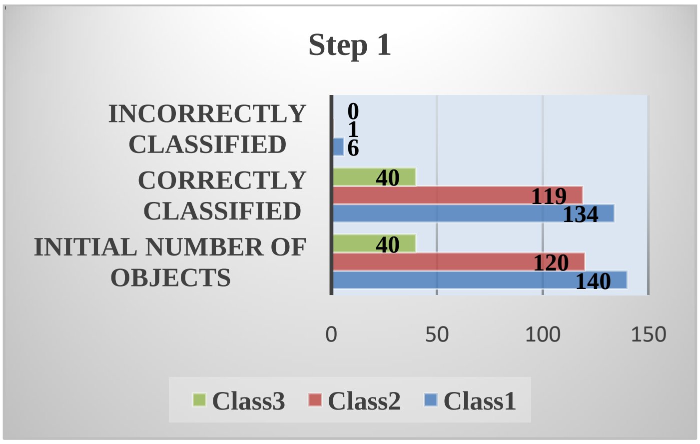


Chart 5

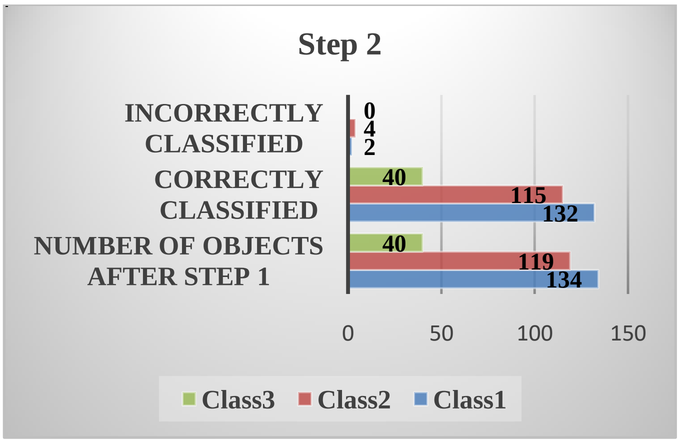


Chart 6

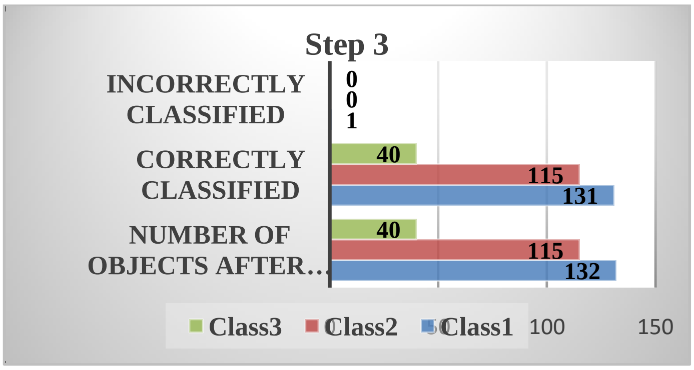
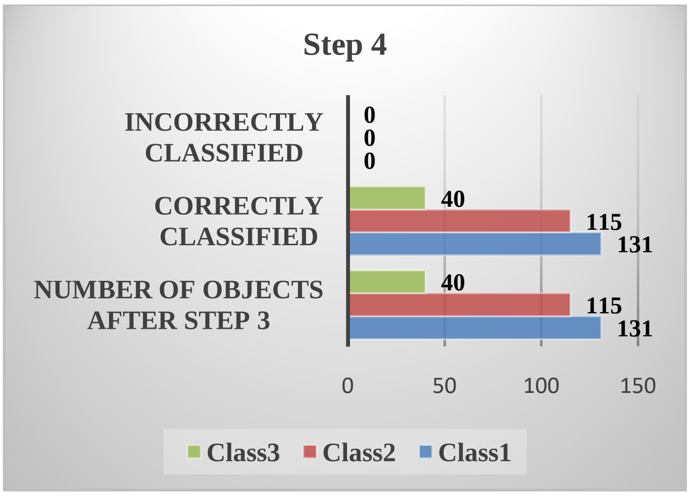


Chart 7



By the end of the phase, there will be a selection of reference training set options of 131 in class 1, 115 in class 2, and 40 objects in class 3.

In Phase 4. A set of informative features is selected using the generated reference table. As you know, the classification results are 100% for the reference table. Now, using the convergence function (9), all of the features will be identified as distinct. According to it, a column with randomly selected features is omitted from the reference table, meaning that in the class we are looking at, there are 62 features in all three classes, and as a result of omitting one of them the classification process is carried out using the remaining 61 characters and the (9) proximity function. If at the end of the process all objects in the calculation find 100% of the class, the column removed from the table will not be redirected, otherwise the column will be returned to its original location by random selection and the process will be returned. The proposed process lasts up to l. If objects have found a different class in their class (switching to another class), the arbitrarily selected symbol will be returned. This process takes place between 62 features in the issue we are looking at, and at the end of the process, the remaining features are distinguished as informative features.

Therefore, the essence of the work performed at this stage is to select the most useful ℓ element from the set of features that characterize the objects under investigation, i.e. the selection of informative symbols. We found it necessary to provide the software codes of the process to be summarized below.

```
public static void
genInformativ(HashMap<String,
ClassObjects> classObjectsHashMap) {
if (noInfor == null) noInfor = new
ArrayList<>();
tempInformativClassObjectsHashMap =
new HashMap<>();
int countParam =
classObjectsHashMap.values().iterator(
).next().getCountParam();
Random random = new Random();
while (true) {
int max = 0;
int sum = 0;
noInfor.clear();
while (true) {
int i = random.nextInt(countParam) +
1;
if (noInfor.contains(i)) continue;
tempInformativClassObjectsHashMap.clea
r();
for (ClassObjects classObjects :
classObjectsHashMap.values()) {
                    ClassObjects
tempClass =
genTempClassObject(classObjects, i,
noInfor);
tempInformativClassObjectsHashMap.put(
tempClass.getClassName(), tempClass);
                }
yaqinlik(tempInformativClassObjectsHas
hMap);
if
(foiz(tempInformativClassObjectsHashMa
p) == 0) noInfor.add(i);


if (max == noInfor.size()) sum++;
if (max <noInfor.size()) {
                    max =
noInfor.size();
                    sum = 0;
                }
if (sum >20) break;
```

```
            }
if (noInfor.size() >50)
printInfor(countParam, noInfor);
if (noInfor.size() >52) break;
        }
}
```

During the operation of this part of the software, seven sets of 10 informative features were identified, yielding the same result for $\ell = 10$. The result is given in Table 3.

Table3

No.	Set of Informative Features
1.	$x_1, x_7, x_{16}, x_{22}, x_{27}, x_{36}, x_{41}, x_{51}, x_{57}, x_{62}$
2.	$x_1, x_7, x_{22}, x_{27}, x_{36}, x_{41}, x_{44}, x_{51}, x_{57}, x_{62}$
3.	$x_1, x_7, x_{22}, x_{27}, x_{36}, x_{41}, x_{47}, x_{51}, x_{57}, x_{62}$
4.	$x_1, x_7, x_{10}, x_{22}, x_{27}, x_{36}, x_{41}, x_{51}, x_{57}, x_{62}$
5.	$x_1, x_7, x_{22}, x_{27}, x_{34}, x_{36}, x_{41}, x_{51}, x_{57}, x_{62}$
6.	$x_1, x_2, x_7, x_{22}, x_{27}, x_{36}, x_{41}, x_{51}, x_{57}, x_{62}$
7.	$x_7, x_{22}, x_{27}, x_{34}, x_{36}, x_{41}, x_{47}, x_{51}, x_{57}, x_{62}$

By analyzing the results, seven 10 feature sets provided by the program have beenseparated from a set of common features and is considered as the final result. As can be seen from the table, the following informative feature set, which is common to the whole set of features, is 9, and its results are summarized in Table 4 below.

Table 4:

Set of Informative Features
$\boldsymbol{x_1, x_7, x_{22}, x_{27}, x_{36}, x_{41}, x_{51}, x_{57}, x_{62}}$

As a result of the classification of objects using a selected informative feature set, it was found that the reference table objects had 100% their class.

III. CONCLUSION

The article investigated 3 cases of ischemic heart disease as a research object.

The following results were achieved by addressing the above:
first, the feasibility of the characteristic features of the objects of training set classes has been determined;

secondly, in the medical data we are looking at, a converting process of the characteristic features of the objects in the classes has been carried out from continuous quantitative to 0 or 1;

thirdly, it was determined that the classification of objects belonged to their own class or to another class, that is, as a result of exclusion of objects that did not find their own class, the reference table was formed;

fourthly, a significant level of features has been identified from the reference table, i.e., based on the results of the reclassification, a set of informative featuresthat clearly distinguish each other within the class objects has been separated.

IV. REFERENCES

[1]. Yu.I.Zhuravlev,“Selected scientific works”, –M: “Magistr” Publishing house, 420 p., 1998.

[2]. L.Yu, H.Liu. “Efficient Feature Selection via Analysis of Relevance and Redundancy”. – J. Mach. Learn. Res., Vol. 5, pp. 1205-1224, No Oct.2004.

[3]. L.Yu, “Toward Integrating Feature Selection Algorithm for Classification and Clustering”, IEEE Transaction on Knowledge and Data Engineering 17(4), 491-502.

[4]. K.Yan, D. Zhang, “Feature Selection and Analysis on Correlated Gas Sensor Data with Recursive Feature Elimination”, Sensors Actuators, B Chem., Vol. 212, pp. 353-363, Jun 2015.

[5]. A.Kh. Nishanov, Kh.A. Turakulov, Kh.V.Turakhanov, “A decision rule for identification of eye pathologies”, Biomedical Engineering, 33(4) 178-179, 1999.

[6]. A. Kh. Nishanov, Kh. A. Turakulov, Kh.V.Turakhanov, “A decisive rule in classifying diseases of the visual system”, Meditsinskaia tekhnika, (4) 16-18, 1999.

[7]. F.Xiang, W.Lina,“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]. S.Linhui, F.Sheng and F.Wang, “Decision tree SVM model with Fisher feature selection for speech emotion recognition”, Sun et al. EURASIP Journal on Audio, Speech, and Music Processing https://doi.org/10.1186/s13636-018-0145-5, 2019.

[9]. X.Yan, M.Tan, Y.Yan, M.Lü, “Research on Hidden Markov Model-Based And Neural Network-Based Intrusion Detections”, Computer Applications and Software, 29(2),294–297, 2012.

[10]. X.Jing, H. Wang, K. Nie, Z.Luo, “Feature Selection Algorithm Based on IMGA and MKSVM to Intrusion Detection”, Computer Science, 39(7), 96–100, 2012.

[11]. K.Koushal, S.B. Jaspreet, “Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms”, International Journal of Computer Applications (0975 – 8887) Volume 150 – No.12, 1-13, September-2016.

[12]. A.Kh. Nishanov, G.P. Djurayev, M.A. Khasanova,“Improved algorithms for calculating evaluations in processing medical data”, COMPUSOFT: An International Journal of Advanced Computer Technology,India, Vol 8, No 6 - 2019.

[13]. M. Kamilov, A.Nishanov, R.Beglerbekov, “Modified Stages of Algorithms for Computing

Estimates in the Space of Informative Features", International Journal of Innovative Technology and Exploring Engineering, 8(6), 2019.

[14]. M.Sulaiman, J.Labadin, "Feature Selection Based on Mutual Information", Publisher: Institute of Electrical and Electronics Engineers (IEEE), pp: 1-6, 2015.

[15]. D.Xuelian, L.Yuqing, W.Jian, Z.Jilian, "Feature Selection for Text Classification: A review" Multimedia Tools and Applications, 1007/s11042-018-6083-5, 2019.

[16]. S.Ma and J.Huang, "Penalized Feature Selection and Classification in Bioinformatics", Briefings in Bioinformatics. doi: 10.1093/bib/bbn027, 2008.

[17]. L.Sun, , S.Fu and F.Wang, 2019) 'Decision Tree SVM Model with Fisher Feature Selection for Speech Emotion Recognition', Eurasip Journal on Audio, Speech, and Music Processing, doi: 10.1186/s13636-018-0145-5, (1)2019.

[18]. R. S. B.Krishna andM.Aramudhan, "Feature Selection Based on Information Theory for Pattern Classification", in 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies, doi: 10.1109/ICCICCT.2014.6993149, ICCICCT 2014.

[19]. V. V.Mokshin, *et al.* "Parallel Genetic Algorithm of Feature Selection for Complex System Analysis", in *Journal of Physics: Conference Series*. doi: 10.1088/1742-6596/1096/1/012089, 2018.

[20]. S.Solorio-Fernández, J.A.Carrasco-Ochoa and J.F.Martínez-Trinidad, "A Review of Unsupervised Feature Selection Methods", *Artificial Intelligence Review*. doi: 10.1007/s10462-019-09682-y, 2019.

[21]. R.J.Urbanowicz*et al.*"Benchmarking Relief-Based Feature Selection Methods for Bioinformatics Data Mining", *Journal of Biomedical Informatics*, 85.doi: 10.1016/j.jbi.2018.07.015, 2018.

[22]. J.Cai*et al.*"Feature Selection in Machine Learning: A New Perspective", *Neurocomputing*, 300. doi: 10.1016/j.neucom.2017.11.077, 2018.

[23]. M.Ghaemiand M. R. Feizi-Derakhshi, "Feature Selection Using Forest Optimization Algorithm", *Pattern Recognition*. Elsevier Ltd, 60, pp. 121–129. doi: 10.1016/j.patcog.2016.05.012, 2016.