Evaluation of the Informative Features of Cardiac Studies Diagnostic Data using the Kullback Method

Olga Skitsan, Ievgen Meniailov, Kseniia Bazilevych, Halyna Padalko

National Aerospace University "Kharkiv Aviation Institute", Chkalow str., 17, Kharkiv, Ukraine

Abstract

The high rates of development of information technologies today have led to the fact that large amounts of information have been accumulated in various databases. The issue of separating more informative data from less informative data for further analysis and use is important. This determines the relevance of the study. As a result of the study, methods for assessing the informativeness of signs for medical data were analyzed. The Kullback method was chosen as the most appropriate method for medical data. On the basis of the Kullback method, a model for assessing the information content was built and a software package was implemented. For the experimental study, data from 303 patients and 13 features were used. The information content was calculated for various groups of cardiac data. We got that the following signs are the most informative: thal, chest pain type, colored vessels, angina, age. The Kullback method is used to determine the informativeness of a feature that is involved in the recognition of two classes of objects. Also, comparisons of the Kullback method with other methods for assessing the informativeness of features are made.

Keywords 1

Features informativeness, Kullback method, medical diagnostics, data driven medicine, heart disease.

1. Introduction

Many modern medical institutions have information systems for storing various medical data on the health of patients, used by doctors to recognize (diagnose) pathological processes [1-2]. The use of modern information technologies in medicine contributes to the accumulation of huge volumes of medical data stored and processed using medical information systems [3]. This data includes medical knowledge that can be extracted and used to make decisions, for example, in the recognition (diagnosis) of pathological processes [4]. Over the past year, with the COVID-19 pandemic, public health digitalization has become even more important. Information systems are used to analyze and predict epidemic processes [5-7], make decisions regarding diagnostics [8-9], development of recommendation systems [10], study pathogens [11], train medical personnel [12] and their management [13-14]. When analyzing medical data, identifying patterns in these data and extracting them, one has to face the problem of dimensionality. the dimension of the stored data, which is determined by the number of different signs describing the patient's health, is very large and sometimes reaches several tens and hundreds of indicators [15]. Therefore, the problem of reducing the dimension of the feature space and identifying the most informative features is very relevant for medical information systems.

Mathematical processing of the initial medical data allows you to determine the diagnostic value of indicators or their complexes, in the future it helps to build an optimal examination plan, and significantly reduce the number of studies required for diagnostics and ultimately improve the quality

MoMLeT+DS 2021: 3rd International Workshop on Modern Machine Learning Technologies and Data Science, June 5, 2021, Lviv-Shatsk, Ukraine

EMAIL olgaskitsan@gmail.com (O. Skitsan); evgenii.menyailov@gmail.com (I. Meniailov), ksenia.bazilevich@gmail.com (K. Bazilevych); galinapadalko95@gmail.com (H. Padalko).

ORCID: 0000-0003-1913-0000 (O. Skitsan); 0000-0002-9440-8378 (I. Meniailov), 0000-0001-5332-9545 (K. Bazilevych), 0000-0001-6014-1065 (H. Padalko).

of computer diagnosis [16]. There are several reasons for the possibility of transition from a larger number of baseline indicators of the patient's condition to a significantly smaller number of the most informative features. This is primarily a duplication of information due to the presence of links between features [17], low informativeness of individual features [18], a balanced summation of some features [19] and the construction of generalized features [20]. Assessment of the informativeness of diagnostic signs is necessary for their objective ranking in order of importance and determination of the order of their consideration in the process of making a diagnosis.

The complexity of building computer prognostic and diagnostic systems in medicine is due to the fact that a significant part of the information is subjective expert assessments of a doctor based on his knowledge and experience in treating cardiac patients [21]. To model and display such information, the theory of fuzzy logic is used as a way of the most natural description of the nature of human thinking and the course of its reasoning [22]. Various types of data obtained as a result of biochemical analyzes, instrumental studies and other diagnostic methods are used as informative signs.

The aim of the study is to analyze mathematical models for assessing the information content and develop a model for assessing the information content of diagnostic data from cardiac studies.

2. Evaluation of informative features

Any biomedical information processing is dedicated to specific purposes such as research, treatment, breeding, etc. Perhaps the most important goal of medical research is the classification of the object or, in relation to the patient and the disease, diagnosis [22]. And this is obvious, since all further actions depend on the diagnostic results. Historically, the diagnosis was to a certain extent an art multiplied by the experience and intuition of the doctor, and only with the mathematization of medicine, the diagnosis can be formulated as a mathematical problem, and therefore automated [23]. Since to make a diagnosis means to classify an object (to recognize it as belonging to a certain class), the medical problem of diagnostics (classification) becomes a mathematical problem of pattern recognition [24].

To classify an unknown object, that is, to recognize an image, means to determine which class an object belongs to, based on an analysis of the values of its features. With regard to medicine, it is possible to make a diagnosis, that is, to recognize a disease or its absence, only when some of the signs inherent in this object (patient) are obtained and analyzed. Such features are called informative features [25]. Informative features are useful information for this purpose, obtained from the original information.

However, informative features are far from being equivalent to achieve a specific goal, therefore, a very important task is to search and select features that are sufficiently informative to make a reliable diagnosis [26]. To understand what the concept of “sufficiently informative” means, the concept of the informativeness of a feature is introduced. The informativeness of a feature means how much this feature characterizes the psychophysical state of an object, that is, how much the diagnosis depends on it – the result of recognition. There are at least 2 approaches to assessing information content – energy [27] and information [28].

The energy approach is based on the fact that the information content is assessed by the value of the attribute. The features are sorted by size, and the most informative is the one whose value is greater. However, this approach to assessing the information content may turn out to be poorly suited for object recognition. Indeed, if some feature is large in absolute value, but almost the same for objects of different classes, then by the value of this feature it is difficult to assign an object to a certain class. And vice versa - if the feature is relatively small in size, but differs greatly for objects of different classes, then the object can be easily classified by its value. Therefore, the information approach is more suitable for object recognition, according to which information features are considered as a reliable difference between the classes of images in the space of features.

If, when recognizing an object, it must be attributed to one of 2 classes, then the difference in the probability distributions of a feature constructed from samples from 2 compared classes can act as such a significant difference. The method for determining the informativeness is chosen by the researcher himself, depending on the objectives of the study, the number of recognized classes and biomedical data, indicators - the coding method, the sample size, the number of gradations.

So, let ω be a set of objects, $X - \{x_1, x_2, ..., x_n\}$ is a finite set of quantitative features of these objects. For any object $\omega \in \Omega$, its feature description is known $\{x_1(\omega), x_2(\omega), ..., x_n(\omega)\}$ is an n-dimensional vector, and the coordinate of this vector is equal to the value of the *i*-th feature. The set of feature descriptions of objects from a given sample of objects $A \subseteq \Omega$ is given in the form of a matrix of size $|A| \times n$, which is called the "feature-attribute" table.

Let $I(Z)$ be the measure of the informativity of the subset of features $Z \subseteq X$, defined on A. It is necessary to choose all different subsets of the set X some subset $Z^* \subseteq X$ such that

$$I(Z^*) = \max_{Z \subset X}\{I(Z)\}. \quad (1)$$

Requirements for the preparation of data (in accordance with the objectives of the study) for mathematical analysis, to assess the nature of the distribution in the sample under study (preliminary data analysis) dictate to solve the following tasks:

- checking the homogeneity of selected observation groups, including control groups, which can be carried out either by expert advice, or by methods of multivariate statistics (for example, using cluster analysis);
- normalization of variables, that is, elimination of anomalies of indicators in the data matrix (agreement of opinions);
- reducing the dimension of the feature space (by formal methods by assessing the information content);
- standardized description of features;
- construction of classification scales of attributes, i.e. a procedure for identifying and establishing the physical boundaries of the parameters under study and presenting information in quantized form (a certain code number corresponds to each value of a feature).

Let us pose, firstly, the task of determining the assessment of the informativeness of the features - the most important of the above, both at the stage of preliminary) and final analysis in accordance with the objectives of the study; secondly, the problem of estimating the probable distribution of the minimum required number of informative features that provide a given level of reliability of an algorithm (procedure) if the distribution of all feature distributions is known (which is natural for applications where the corresponding statistical data are accumulated)

3. Kullback method application

Cybernetics considers the processes occurring in biological systems, in service systems, in production and, accordingly, their models as information processing processes, studying the quantitative laws of information processes. Therefore, the so-called measures of information in statistics are of particular interest to us. The amount of information can be understood as the amount of eliminated uncertainty.

In particular, the assessment of the informativeness of signs is used in medicine in the diagnosis of numerous diseases. Further treatment of the patient depends on the results of the diagnosis. It is possible to make a diagnosis (that is, to recognize this or that disease or its absence), provided that the characteristics inherent in the object (in medicine – the patient) are analyzed. informative indicators are indicators that make the greatest contribution to the characteristics of the state of the object (patient).

There are numerous methods for determining the informativeness of features. However, unlike other criteria for the statistical significance of differences, the Kullback measure allows one to assess not the reliability of differences between divisions, but the degree of these differences. the method of analyzing signs by assessing the informativeness using the Kullback informative measure has been widely used in medicine when considering individual factors that influence the diagnosis [29]. In this method, the measure of the difference between the two classes, which is called divergence, is assumed as an assessment of the informativeness.

To measure the amount of information N. Wiener and K. Shannon independently from each other proposed logarithmic measures, which were recognized as quantitative measures of information [30-

31]. To the class of similar logarithmic measures belongs and is similarly studied as Kullback's information measure *J* is the discrepancy between statistical distributions 1 and 2. For discrete distributions, this measure is reflected by the formulas:

$$J(x_i \,/\, A_1, x_i \,/\, A_2) = \sum_j lg \frac{P(x_{ij} \,/\, A_1)}{P(x_{ij} \,/\, A_2)} \left[P\left(x_{ij} \,/\, A_1\right) - P(x_{ij} \,/\, A_2)\,\right] \tag{2}$$

$$J(x_i \,/\, A_1, x_i \,/\, A_2) = \sum_j DC(x_{ij}) \cdot \left[P\left(x_{ij} \,/\, A_1\right) - P(x_{ij} \,/\, A_2)\,\right], \tag{3}$$

where DC (x_{ij}) is diagnostic coefficient.

The Kullback method of evaluation of informative measures is based on the calculation of diagnostic coefficients. The diagnostic coefficient is presented in the form of the logarithm of the ratio of the probabilities of manifestation of this feature in the main and control groups ($P(\frac{x_j^i}{A})$ and $P(\frac{x_j^i}{B})$, respectively):

$$DC\left(x_j^i\right) = \log\left(P(\frac{x_j^i}{A})/P(\frac{x_j^i}{B})\right). \tag{4}$$

Diagnostic coefficients are most often ambiguous or single-digit positive or negative numbers. They are positive in the case of a predominance of the probability $P(\frac{x_j^i}{A})$, which is in the numerator, negative – in the case of a predominance of the probability $P(\frac{x_j^i}{B})$. That is, the diagnostic coefficients with the “+” sign speak for a greater likelihood of hypothesis *A* (about belonging to the main group) with the familiar “-” is about a greater likelihood of hypothesis A_2 (about belonging to the control group). Obviously, the coefficients with a positive sign carry positive information, brings the sum of diagnostic coefficients closer to the threshold, which is positive for *A*. Coefficient with a negative sign, on the contrary, “removes” the amount from the threshold. For hypothesis *B*, on the contrary, coefficients with a negative sign bring the sum closer to the threshold, and coefficients with a positive sign – see it from the threshold, since the threshold is a negative value.

It should be noted that the greater the value of the diagnostic coefficient, the more differential diagnostic information, that is, information about the prevalence of the probability of one of the diagnoses, it carries. However, the informativeness of each value of the trait depends on the frequency with which this value occurs in each disease, that is, on the magnitude of $P(\frac{x_j^i}{A})$ and $P(\frac{x_j^i}{B})$. If the diagnostic coefficient for determining the signs x_j^i is too large, but patients with such knowledge are relatively rare, then in the process of diagnosing the role of such a value of the sign x_j^i is too small.

To determine the information that the attribute x_j^i carries, you first need to calculate the amount of information that the values of the attributes x_j^i give. To do this, it is necessary to multiply the *DC* by the difference in the probabilities of this feature when belonging to the main group (hypothesis *A*) and to the control group (hypothesis *B*):

$$DC\left(x_j^i\right)\left[P\left(\frac{x_j^i}{A}\right) - P(\frac{x_j^i}{B})\right]. \tag{5}$$

It should be noted that the difference $\left[P\left(\frac{x_j^i}{A}\right) - P(\frac{x_j^i}{B})\right]$ will be positive if *DC* is positive. The difference will show how, on average, the sum of *DC* will approach the threshold as a result of identifying the symptom x_j^i in the patient.

Similarly, other values of the same signs $x_j^1, x_j^2, \dots, x_j^n$ are calculated. The informative value of the attribute as a whole will be equal to their sum.

The considered method, in comparison with other methods for assessing the information content, is the simplest and most accessible for algorithmization. His adaptation machine is not laborious and does not entail significant computational costs and resources.

The use of the Kullback information method for assessing the differential informativeness of features in medicine was proposed in other works, however, the formula used differed from the Kullback method in that it presented not the difference in probabilities, but their sum, worse reflects the contribution of signs in their approximation diagnostic amounts in the diagnostic threshold. Therefore, in the future, the indicated sum of probabilities was replaced by their difference, more precisely, by a brew. The need for this can also be explained by the fact that in reality there are two thresholds: with a plus for making a decision “A” and with a minus for making a decision “B”. The information content should reflect the average of the two values.

Application of the Kullback method of evaluation of informative measures consists of the following stages:

1) to objectify the division of the general ordered series into a range, we select the following ranges between each other, the right (lower) boundaries of which are round numbers so that the number of ranges is 8-12;
2) to count the number of observations with group *A* and *B* that fall within this range. These are the frequencies of the given symptom;
3) to calculate the relative frequencies (probabilities) in percent, taking as 100% the sum of frequencies *A* in all ranges and the same sum of frequencies *B*;
4) to calculate smoothed (weighted average) frequencies, in fact, all frequency smoothings are calculated according to the formula

$$y_i = (y_{i-2} + 2 \cdot y_{i-1} + 4 \cdot y_i + 2 \cdot y_{i+1} + y_{i+2})/10; \qquad (6)$$

5) to calculate the ratios of smoothing frequencies *A* and *B* in each range;
6) to calculate the smoothing of diagnostic coefficients according to the formula:

$$DC\left(x_j^i\right) = \log\left(P(\frac{x_j^i}{A})/P(\frac{x_j^i}{B})\right); \qquad (6)$$

7) to calculate the informativeness of the features in each range and the final informativeness of the features, obtained by summing the informativeness of all ranges.

4. Results of experiments

The input data is a dataset of information on the diagnostic data of patients based on cardiac studies, their age, gender, type of chest pain, cholesterol level, etc., a complete list of parameters in Table 1.

Table 1
Parameters list

Name of parameter	Description	Data type
Name	ID	Count
Age	Years	Count
Sex	Sex	String
Chest pain type	Pain type	String
Blood pressure	Scores	Count
Cholesterol	Scores	Count
Fasting blood sugar < 120	+/-	0/1
Resting ECG	Normal/Hyper	String
Maximum heart rate	Scores	Count
Angina	+/-	0/1
Peak	Scores	Float
Slope	Flat/Up/Down	1/2/3
Colored Vessels	0/1/2	0/1/2
Thal	Normal/Rev/Fix	String

The most important stage in the creation of any information system is the design of a tool that can implement all the tasks set at the beginning of the project. A diagram of work execution, information exchange, workflow is graphically presented, visualizes a business process model.

The IDEF0 and DFD methodologies were used. Within the framework of the IDEF0 (Integration Definition for Function Modeling) methodology, a business process is represented as a set of function elements that interact with each other, and also show information, human and production resources consumed by each function. Functional model of system is presented in Figure 1.

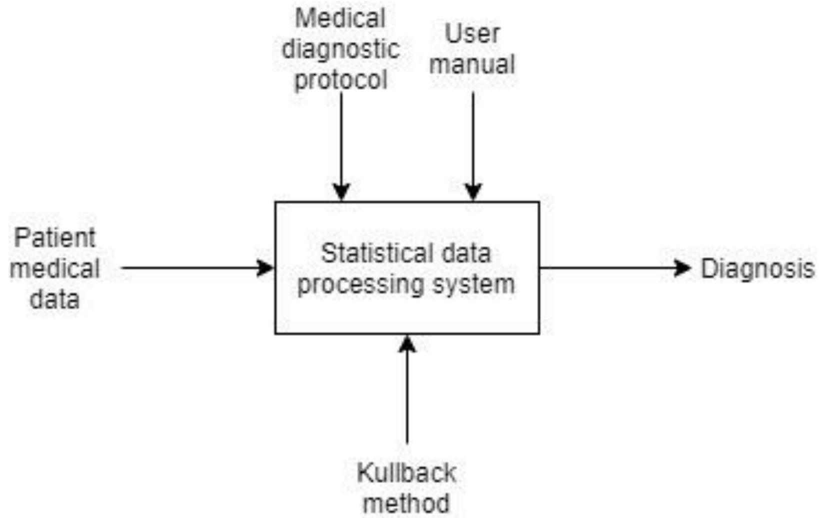

Figure 1: Functional model of system.

Let's consider in more detail the architecture of the project, approaches to solving the assigned tasks and mechanisms for their implementation.

Decomposition of system is presented in figure 2.

For software implementation, the C# programming language was used in the Microsoft Visual Studio environment. To start the software package, you need to download the data presented in the *.csv file (Figure 3).

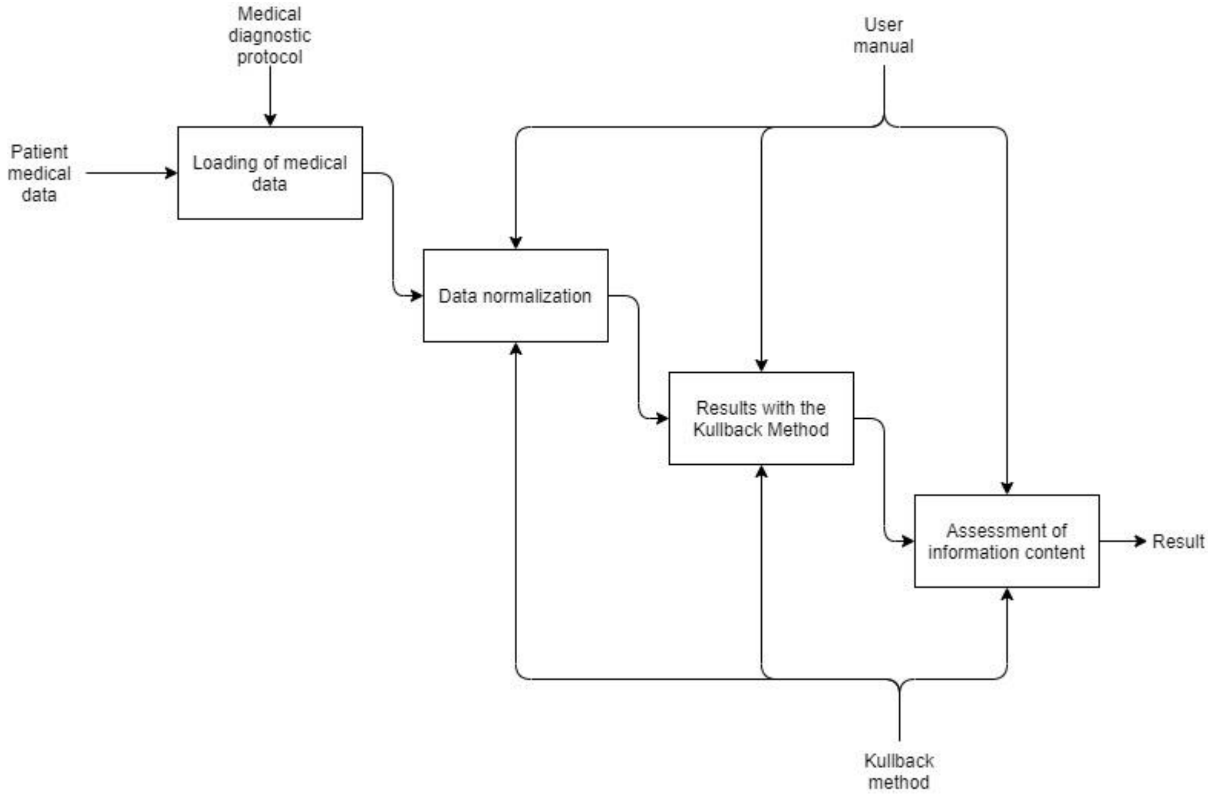

Figure 2: Decomposition of system.

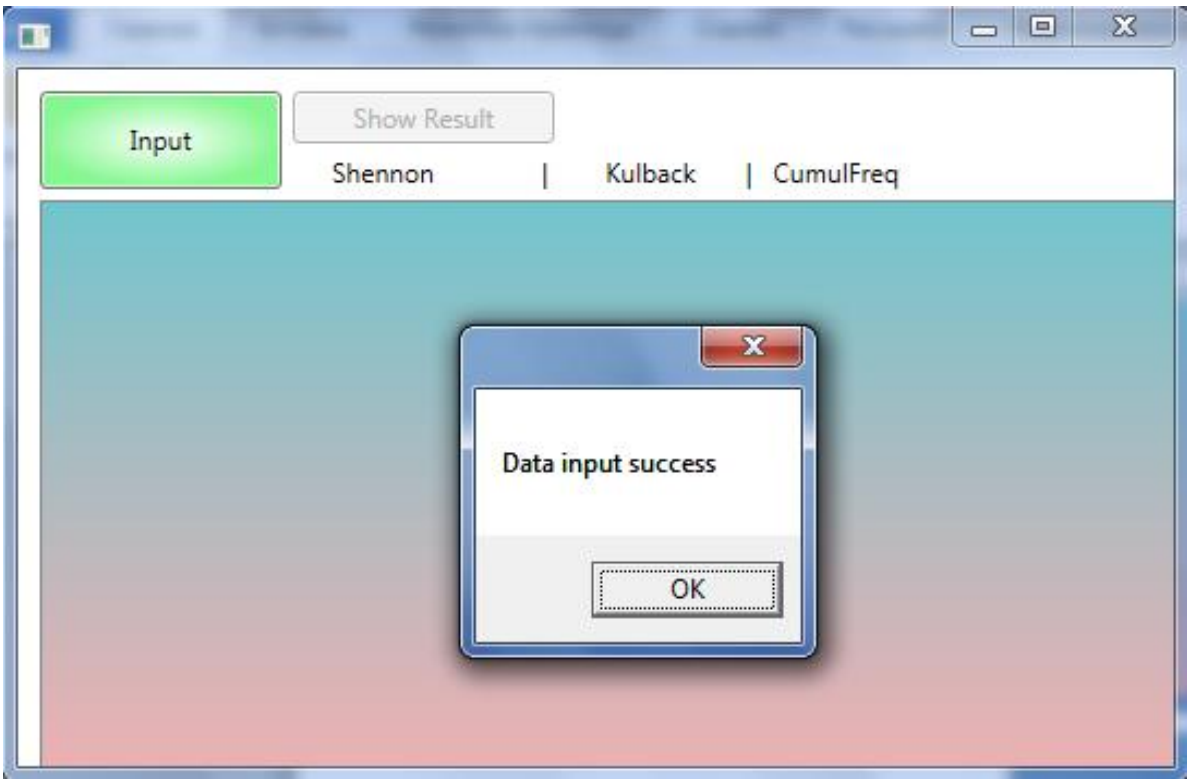

Figure 3: Setting initial parameters for program use.

In total, for example, data from 303 patients and 13 features was taken (their age, gender, type of chest pain, cholesterol level, ECG, blood pressure, maximum pressure, blood sugar level, type and presence of tonsillitis, colored vessels, etc.)

The data is divided into two classes A – “Healthy” and B – “Sick”.

The results of the calculation by the Kullback method for assessing the informativeness of the attribute m = “Patient's age” are shown in Figure 4.

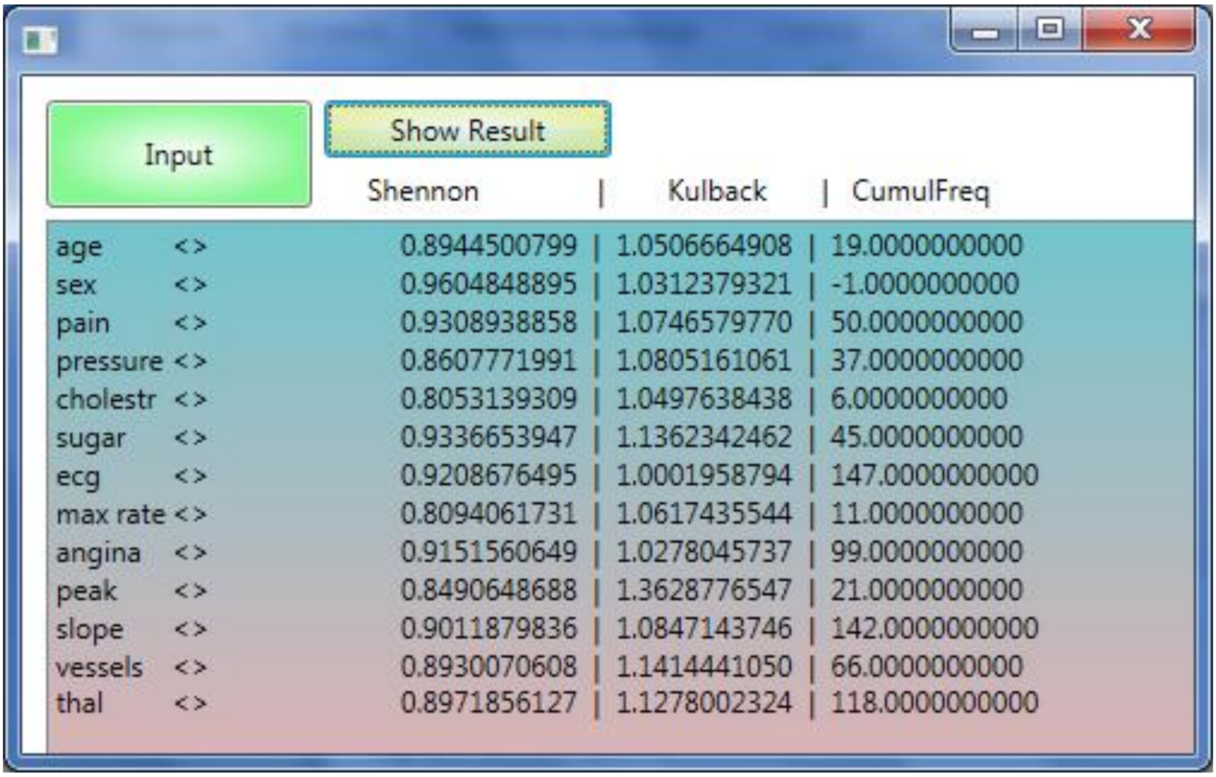

Figure 4: Adjusting the model parameters.

Kullback's method gives an estimate of the informativeness of the studied feature in the form of a value, takes values from 0 to 2. In this case, it is believed that the closer *I(x)* to 2, the higher the informativeness of the feature, on the contrary, the closer *I(x)* to 0, the lower the informative value of *x*. As a result, the information content was calculated for various groups of cardiac data. We got that the following signs are the most informative: thal, chest pain type, colored vessels, angina, age. The Kullback method is used to determine the informativeness of a feature that is involved in the recognition of two classes of objects. Also, comparisons of the Kullback method with other methods for assessing the informativeness of features are made.

5. Conclusions

As a result of the study, methods for assessing the informativeness of signs for medical data were analyzed. The Kullback method was chosen as the most appropriate method for medical data. On the basis of the Kullback method, a model for assessing the information content was built and a software package was implemented. For the experimental study, data from 303 patients and 13 features were used. The information content was calculated for various groups of cardiac data. We got that the following signs are the most informative: thal, chest pain type, colored vessels, angina, age. The Kullback method is used to determine the informativeness of a feature that is involved in the recognition of two classes of objects. Also, comparisons of the Kullback method with other methods for assessing the informativeness of features are made.

6. Acknowledgements

The study was funded by the National Research Foundation of Ukraine in the framework of the research project 2020.02/0404 on the topic “Development of intelligent technologies for assessing the epidemic situation to support decision-making within the population biosafety management” [32].

7. References

[1] V. Yesina, et. al., Method of Data Openness Estimation Based on User-Experience in Infocommunication Systems of Municipal Enterprises, Proceedings of 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology (2019) 171–176. doi: 10.1109/INFOCOMMST.2018.8631897

[2] D. Chumachenko, et. al. On-Line Data Processing, Simulation and Forecasting of the Coronavirus Disease (COVID-19) Propagation in Ukraine Based on Machine Learning Approach, Communications in Computer and Information Science 1158 (2020) 372-382. doi: 10.1007/978-3-030-61656-4_25

[3] M. Mazorchuck, et. al. Web-Application Development for Tasks of Prediction in Medical Domain, 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (2018) 5-8. doi: 10.1109/STC-CSIT.2018.8526684

[4] D. Chumachenko, et. al. On Intelligent Decision Making in Multiagent Systems in Conditions of Uncertainty, Proceedings of 2019 11th International Scientific and Practical Conference on Electronics and Information Technologies (2019) 150-154. doi: 10.1109/ELIT.2019.8892307

[5] Yu. Polyvianna, et. al. Computer Aided System of Time Series Analysis Methods for Forecasting the Epidemics Outbreaks, 2019 15th International Conference on the Experience of Designing and Application of CAD Systems (2019) pp. 7.1-7.4. doi: 10.1109/CADSM.2019.8779344

[6] D. Chumachenko, et. al. On agent-based approach to influenza and acute respiratory virus infection simulation, 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (2018) 192-196. doi: 10.1109/TCSET.2018.8336184

[7] D. Chumachenko, T. Chumachenko, Intelligent Agent-Based Simulation of HIV Epidemic Process, Advances in Intelligent Systems and Computing 1020 (2019) 175-188. doi: 10.1007/978-3-030-26474-1_13

[8] D. Singh, S. Verma, J. Singla, A Comprehensive Review of Intelligent Medical Diagnostic Systems, 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), (2020) 977-981. doi: 10.1109/ICOEI48184.2020.9143043

[9] J. Wang, et. al., The Intelligent Diagnostic System for Common Diseases using the Optimized Medical Knowledge Graph, 2020 International Wireless Communications and Mobile Computing (IWCMC), (2020) 1504-1509. doi: 10.1109/IWCMC48107.2020.9148406.

[10] P. Piletskiy, et. al. Development and Analysis of Intelligent Recommendation System Using Machine Learning Approach, Advances in Intelligent Systems and Computing 1113 (2020) 186-197. doi: 10.1007/978-3-030-37618-5_17

[11] V. Arefiev, et. al. Complete genome sequence of salmonella enterica subsp. enterica serovar kottbus strain Kharkiv, isolated from a commercial pork production facility in Ukraine, Microbiology Resource Announcements 9 (49) (2020) e01171-20. doi: 10.1128/MRA.01171-20

[12] D. Chumachenko, et. al. Intelligent expert system of knowledge examination of medical staff regarding infections associated with the provision of medical care, CEUR Workshop Proceedings, 2386 (2019) 321-330.

[13] N. Dotsenko, et. al. Modeling of the processes of stakeholder involvement in command management in a multi-project environment, Proceedings of 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies 1 (2018) 29-33. doi: 10.1109/STC-CSIT.2018.8526613

[14] N. Dotsenko, et. al. Project-oriented management of adaptive teams' formation resources in multi-project environment, CEUR Workshop Proceedings 2353 (2019) 911-920.

[15] K. Bazilevych, et al. Stochastic modelling of cash flow for personal insurance fund using the cloud data storage, International Journal of Computing 17 (3) (2018) 153-162.

[16] S. V. Georgakopoulos, P. Gallos, V. P. Plagianakos, Using Big Data Analytics to Detect Fraud in Healthcare Provision, 2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME) (2020) 1-3, doi: 10.1109/MECBME47393.2020.9265118

[17] X. Chen, et. al. Named Entity Recognition of Chinese Electronic Medical Records Based on Cascaded Conditional Random Field, 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA) (2019) 364-368. doi: 10.1109/ICBDA.2019.8713244
[18] A. Herasymova, et. al., Development of intelligent information technology of computer processing of pedagogical tests open tasks based on machine learning approach, CEUR Workshop Proceedings 2631 (2020) 121-131.
[19] I. Izonin, et. al. Predictive modeling based on small data in clinical medicine: RBF-based additive input-doubling method, Mathematical Biosciences and Engineering 18(3) (2021) 2599-2613. doi: 10.3934/mbe.2021132
[20] R. Alnashwan, H. Sorensen, A. O'Riordan, Classification of Online Medical Discourse by Modified Co-Training, 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (2019) 131-137. doi: 10.1109/BigDataService.2019.00024
[21] B. Subaeki, et. al. Expert System: Vertigo Disease Diagnosis with Backward Chaining Method, 2020 14th International Conference on Telecommunication Systems, Services, and Applications TSSA (2020) 1-5. doi: 10.1109/TSSA51342.2020.9310820
[22] D. Chumachenko, O. Sokolov, S. Yakovlev, Fuzzy recurrent mappings in multiagent simulation of population dynamics, International Journal of Computing 19 (2) (2020) 290-297.
[23] S. N. Gerasin, et. al., Set coverings and tolerance relations, Cybernetics and Systems Analysis 44 (3) (2008) 333-340.
[24] V. P. Mashtalir, S. V. Yakovlev, Point-set methods of clusterization of standard information, Cybernetics and Systems Analysis 37 (3) (2001) 295-307.
[25] A. A. Joy, M. A. M. Hasan, A Hybrid Approach of Feature Selection and Feature Extraction for Hyperspectral Image classification, 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) (2019) 1-4, doi: 10.1109/IC4ME247184.2019.9036617
[26] S. Madichetty, M. Sridevi, Detecting Informative Tweets during Disaster using Deep Neural Networks, 2019 11th International Conference on Communication Systems & Networks (COMSNETS) (2019) 709-713. doi: 10.1109/COMSNETS.2019.8711095
[27] A. Lykartsis, M. Kotti, A. Papangelis, Y. Stylianou, Prediction of Dialogue Success with Spectral and Rhythm Acoustic Features Using DNNS and SVMS, 2018 IEEE Spoken Language Technology Workshop (SLT) (2018) 838-845. doi: 10.1109/SLT.2018.8639580.
[28] S. Bohdanov, et. al., Forecasting of salmonellosis epidemic proces in Ukraine using autoregressive integrated moving average model, Przeglad epidemiologiczny 74 (2) (2020) 346–354. doi: 10.32394/pe.74.27
[29] J. Qiao, F. Tian, Abnormal Condition Detection Integrated with Kullback Leibler Divergence and Relative Importance Function for Cement Raw Meal Calcination Process, 2020 2nd International Conference on Industrial Artificial Intelligence (IAI) (2020) 1-5. doi: 10.1109/IAI50351.2020.9262170
[30] K. Bazilevych, et.al. Determining the Probability of Heart Disease using Data Mining Methods. CEUR Workshop Proceedings 2488 (2019) 1-12.
[31] S. Wei, D. Wang, Z. Peng, Principal Component Analysis Based Kullback- Leibler Divergence for Die Cracks Detection, 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD) (2020) 224-228. doi: 10.1109/ICSMD50554.2020.9261644.
[32] S. Yakovlev, et. al., The concept of developing a decision support system for the epidemic morbidity control, CEUR Workshop Proceedings 2753 (2020) 265–274.