Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 6, No. 2, September 2023, pp. 197 – 207
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

The Use of Large Databases for Diagnosing Human Diseases at Early Stage

[1]Abdullah M. Al-Ansi1, [2]Vladimir Ryabtsev, [3]Tatyana Utkina
[1]Faculty of Education, Thamar University, Yemen
[2]Department of robotics and specialized computer systems Cherkassy State Technological University, Ukrane
[3]Department of Information Technologies, Cherkassy Branch of Private Higher Education Establishment European University, Ukrane
Email: [1]ebrar.ansi@yahoo.com, [2]t.utkina@chdtu.edu.ua, [3]akimonoke2013@gmail.com

Article Info
Article history:
Received Jul 6th, 2023
Revised Jul 16th, 2023
Accepted Aug 22th, 2023

Keyword:
Features
Human diseases
Intelligent System Eidos
Large databases
Repository

ABSTRACT
The purpose of this article is to demonstrate the ability of the Eidos intellectual system to recognize human diseases at an early stage by processing large databases containing signs of diseases. To study the signs of diseases, it is proposed to use an automated system-cognitive analysis implemented in the Eidos intellectual system. Automated system-cognitive analysis extracts information from large databases and forms knowledge from them that makes it possible to recognize human diseases. In the process of forming models, the amount of information is calculated in the value of the factor by which the modeling object will pass under its influence to a certain state corresponding to the class. This allows for comparable and correct processing of heterogeneous information about observations of the object of modeling, presented in different types of measuring scales and different units of measurement. The results of recognition of the following diseases were obtained with high reliability: chronic kidney disease, lung cancer, breast cancer, liver disease, risks of developing diabetes and stroke. The results of the study can be applied in medical institutions in many countries, since the Eidos system is freely available on the Internet.

Corresponding Author:
Tatyana Utkina,
Department of robotics and specialized computer systems,
Cherkassy State Technological University,
460 Shevchenko Boulevard, Cherkassy 18006, Ukraine.
Email: t.utkina@chdtu.edu.ua
DOI: http://dx.doi.org/10.24014/ijaidm.v6i2.24525

1. INTRODUCTION

Artificial intelligence (AI) systems can process large databases and are rapidly changing the way healthcare is diagnosed today. By using large volumes of medical data, it is possible to obtain reliable results for diagnosing diseases that were previously impossible to establish [1]. The results obtained with the help of AI systems make it possible to provide patients with timely medical care and increase the reliability of diagnosis [2]. Reducing mortality from chronic diseases requires early detection and effective treatment, so medical scientists are using new predictive models to detect diseases at an early stage [3, 4]. There are no best methods for diagnosing diseases in humans, since each method has its own advantages and disadvantages. Currently, Support Vector Algorithms (SVM), logistic regression (LR) and clustering are often used. Convolution artificial neural networks automatically extract key features - knowledge from structured data, which makes it possible to establish a diagnosis of a disease with high certainty [5-7]. Modern intelligent systems perform automated processing of large databases containing signs of diseases, which ensures the diagnosis and classification of human diseases [8-10]. The purpose of this article is to use automated system-cognitive analysis (ASC-analysis) to identify the diseases of patients with the most common chronic diseases at an early stage [11]. To process medical data using ASC analysis, it is necessary to perform a machine learning procedure, i.e., develop representative databases for the types of diseases studied. To achieve this goal, we need free access to large databases and a methodology that will help transform this data into a form which is necessary for working in an artificial intelligence system. Some of these good choices are UCI database repositories, Kaggle and others [12-14]. This depends on how many organs a person has - so many databases can be compiled. But the donors of these databases secured their data and put a "?" instead of real data. Therefore, these databases need to be carefully edited. Missing attribute values should be replaced with empirical data taken from the literature. To achieve this goal, it is necessary to solve the following issues:

1. Conversion of databases from csv-format to intermediate MS Excel files;
2. Input of initial data into the Eidos system;
3. Synthesis and verification of domain models;
4. Application of models for solving problems including identification, forecasting and research of the subject area.

Success at the knowledge extraction stage largely depends on the qualifications of the analyst, who must have knowledge from various fields, including cognitive psychology, systems analysis, mathematical logic, artificial intelligence and so on. The actual gap in prior literature works is represented through lack of smart systems depending on AI that are able to diagnose deseases in early stage. Eidos system introduced in this research in explains in details how to diagnose deseases at early stages based on generative AI system. The main aim of this paper is to introduce easy approach to diagnose deseases at early stages by helping of Edios system including Chronic Kidney Disease Recognition, Lung Cancer Diagnostics, Predicting the Risk of Diabetes, Stroke risk prediction and Diagnosing Breast Cancer. In addition, the main contribution of this work is represented in facilitating and illustrating the methods of diagnoses of every sickness and giving actual evidence how the Eidos system works in every different stage.

2. RESEARCH METHOD

2.1. Formalization of The Subject Area

The database csv-file is converted into xls format using the online converter (https://onlineconvertfree.com/ru/convert-format/csv-to-xls). In the first column of the file, we add its name - number. The other columns are filled with the names taken from the source database. Then we copy the training sample to the folder (d:\Aidos-X\AID_DATA\Inp_data\) and rename it to Inp_data. At this stage of the ASC analysis, classification and descriptive scales and their gradations are developed, which are used to encode the initial data, resulting in a training sample [15-17].

The universal cognitive analytical system Eidos is in full open free access on the website http://lc.kubagro.ru/aidos/_Aidos-X.htm with an open license CC BY-SA 4.0 and this means that everyone can use it, whoever wishes, without any additional permission from the developer of this system, Professor E. V. Lutsenko. To enter the Eidos system, you need login 1 and password 1, as well as set the parameters shown in Figure 1.

A universal interface is used to enter an external database into the Eidos system. The training sample, in fact, is the original data, normalized using classification and descriptive scales and their gradations. The spring 2011 version of the Eidos system provided a training sample size of no more than 100,000 objects, in the current version this restriction has been removed and now the system can work with millions of objects. But there remains a limitation on the dimension of knowledge bases: no more than 4000 classes and 4000 gradations of factors.

The Eidos intellectual system provides the ability to model complex dynamic systems and comparative analysis of implementation options, and also allows you to develop an action plan to achieve the expected result. The Eidos system is a highly effective tool for scientific research in a variety of subject areas that require generalization, systemic multi-parameter identification, forecasting and decision making, comparison and classification.

2.2. Formation and Verification of Models

Synthesis and verification of models in the Eidos system is carried out in mode 3.5. Figure 2 shows that 30% should be left from the original database, removing the least reliable recognition results.

When the training sample is loaded into the Eidos system, then the study should begin with an analysis of the information portrait of features. In the process of compiling the training sample, additional redundant features may get into the descriptive features due to user errors. This can be digital data entered instead of characters, or digital data whose values exceed the specified range of features. Even an extra space in symbolic features is accepted by the Eidos system as a new feature. In multi-word features, use the underscore character (_) to concatenate individual words [18].

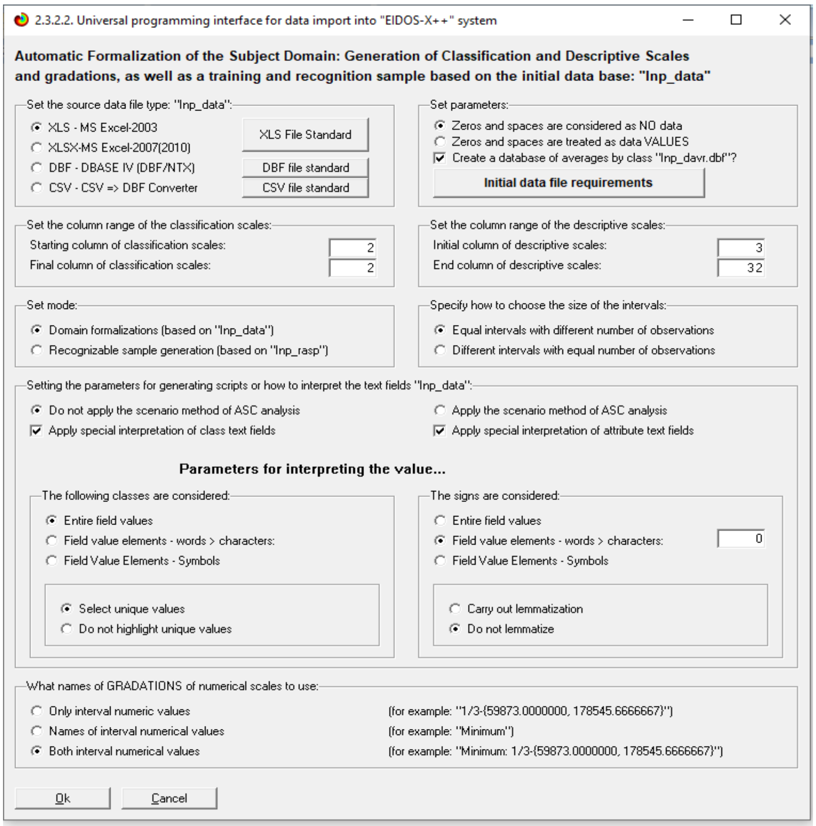


Figure 1. Entering The Database Into The Eidos System

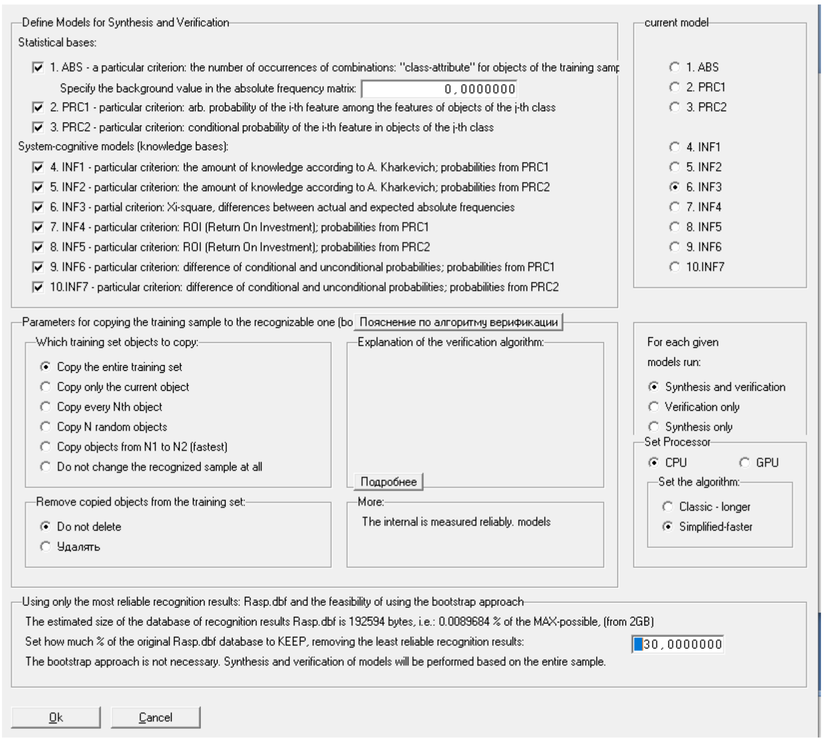


Figure 2. Synthesis and Verification of Models

To verify the models, the number of errors of the 1st and 2nd kind is counted, i.e., non-identification errors and false identification errors. As experience shows, this method gives a quite adequate assessment of the reliability of the model. If all objects are attributed to all classes, then the error of the 1st kind will be equal to zero, however, in this case, the error of false identification will be maximum, since all objects will be assigned not only to those classes to which they actually belong, but also to those to which they do not belong. Similar action, if all objects are not assigned to any of the classes, then the error of the 2nd kind will turn to zero, however, in this case, the non-identification error will be maximum. Thus, it is necessary to choose a certain threshold, such that the average number of errors of the 1st and 2nd kind is minimized, which is implemented in the Eidos system [19]. The Eidos system identifies the objects of the training sample, the classification vector of which is already known, and then calculates the weighted average identification error (integral validity), as well as the identification error with each class (differential validity). If the model has an acceptable level of adequacy, then a decision is made on the possibility of using it in an adaptive mode on objects that are not included in the training sample, but belong to the general population, in relation to which this sample is representative. If the model is not adequate enough, then work continues on the synthesis of an adequate model by increasing the number of classes and factors, as well as updating the descriptions of the training sample objects or increasing their number. To assess the reliability of models in the Eidos system, the F-measure of Van Risbergen and the L1-measure proposed by Professor Lutsenko E.V. are used. The resulting model quality criteria are shown in Figure 3. Model reliability estimation errors depend on the size of the training sample, and for a large sample size, it is logical to assume that the model reliability estimation is performed with sufficient accuracy. Usually, the INF3 model has the highest value of the integral criterion "Amount of knowledge" and it is selected as the current model, and then batch object recognition is performed.

3.4. Generalized form for valid models at different int.crit. Current Model: "INF1"

Model name and private criterion	Integral criterion	S-precision models	S-Completeness models	L1-measure prof. E.V.Lutsenko
1. ABS - a particular criterion: the number of occurrences of combinati...	Correlation of abs.frequencies wit...	0.931	1.000	0.964
1. ABS - a particular criterion: the number of occurrences of combinati...	The sum of the absolute frequen...	0.890	1.000	0.942
2. PRC1 - particular criterion: arb. probability of the i-th feature among t...	Correlation of conditional relative ...	0.931	1.000	0.964
2. PRC1 - particular criterion: arb. probability of the i-th feature among t...	The sum of the conditional relativ...	0.801	1.000	0.889
3. PRC2 - particular criterion: conditional probability of the i-th feature i...	Correlation of conditional relative ...	0.931	1.000	0.964
3. PRC2 - particular criterion: conditional probability of the i-th feature i...	The sum of the conditional relativ...	0.800	1.000	0.889
4. INF1 - particular criterion: the amount of knowledge according to A. ...	Semantic resonance of knowledge	0.999	0.999	0.999
4. INF1 - particular criterion: the amount of knowledge according to A. ...	Sum of knowledge	1.000	0.972	0.986
5. INF2 - particular criterion: the amount of knowledge according to A. ...	Semantic resonance of knowledge	0.999	0.999	0.999
5. INF2 - particular criterion: the amount of knowledge according to A. ...	Sum of knowledge	1.000	0.972	0.986
6. INF3 - partial criterion: Xi-square, differences between actual and ex...	Semantic resonance of knowledge	1.000	1.000	1.000
6. INF3 - partial criterion: Xi-square, differences between actual and ex...	Sum of knowledge	1.000	1.000	1.000
7. INF4 - particular criterion: ROI (Return On Investment); probabilities f...	Semantic resonance of knowledge	1.000	0.981	0.990
7. INF4 - particular criterion: ROI (Return On Investment); probabilities f...	Sum of knowledge	0.991	0.998	0.994
8. INF5 - particular criterion: ROI (Return On Investment); probabilities f...	Semantic resonance of knowledge	1.000	0.981	0.990
8. INF5 - particular criterion: ROI (Return On Investment); probabilities f...	Sum of knowledge	0.990	0.998	0.994
9. INF6 - particular criterion: difference of conditional and unconditiona...	Semantic resonance of knowledge	1.000	1.000	1.000
9. INF6 - particular criterion: difference of conditional and unconditiona...	Sum of knowledge	1.000	1.000	1.000
10.INF7 - particular criterion: difference of conditional and uncondition...	Semantic resonance of knowledge	1.000	1.000	1.000
10.INF7 - particular criterion: difference of conditional and uncondition...	Sum of knowledge	1.000	1.000	1.000

Figure 3. Credibility of System-Cognitive Models

If the training sample includes all objects of the general population, then the reliability of the conclusions will be the highest. If the training sample is very small, then it is unlikely that reliable conclusions about the general population can be made on its basis, since in this case the training sample may not even include examples of objects of all or the vast majority of classes. Under the representativeness of the training sample, we mean its ability to adequately represent the general population, so that the study of the general population itself can be correctly replaced by the study of the training sample. But representativeness depends not only on the volume, but also on the structure of the training sample, that is, on how fully all categories of objects in the general population (classes) are represented and on how fully they are described by features.

3. RESEARCH RESULTS

3.1. Outcomes of Early-Stage Chronic Kidney Disease Recognition

A database developed by Senior Consultant Nephrologist P. Soundarapandian from Apollo Hospital, Managiri, India was used to recognize early-stage chronic kidney disease (CKD) [20]. The missing attribute values were supplemented by data taken from a survey of a 50-year-old woman admitted to the Maisonneure Rosemont hospital with complaints of uncontrolled blood pressure and sudden pain in the left side lasting 24 hours, which occurred 6 weeks before admission. The test results of 400 patients were analyzed. Patient health attributes contain 24 parameters, database size is 400 x 24.

The Eidos system forms 10 models, the principles of which are set out in numerous monographs by Professor Lutsenko, including those in English [21-23]. Books can be downloaded from the Internet for free. The system generates and determines the reliability of ten models. The user selects the most reliable model for the study. In addition, the system checks the reliability of recognition of samples of instances. Instances with low recognition reliability should be changed or excluded from the training set. Professor Lutsenko recommends that the reliability of recognition of sample instances be at least 70%. The variability of the similarity of belonging of sample specimens to the classification scale of the CKD class was determined in the range from 67.88 to 100. The variability of the similarity of the belonging of specimens of the sample to the classification scale of the not-CKD class was determined in the range from 91 to 100. The strengths and directions of the influence of factors on the detection or non-detection of CKD were determined, which allows, during the performance of laboratory tests, to more carefully determine the results of analyzes of those samples that significantly affect the diagnosis [24]. The high manufacturability and the possibility of diagnosing CKD by ASC analysis of the modified database of the UCI repository in the Eidos system were demonstrated. The application of the new research methodology in practice will reduce the time and improve the reliability of the results of diagnosing patients' diseases.

3.2. Outcomes of Lung Cancer Diagnostics using the Eidos System

For the recognition of lung cancer, the UCI repository database was used, containing 15 symptoms of the disease for 309 patients. At first glance, the solution to this problem seems simple, since it requires mainly knowledge and skills to use the standard features of the Internet browser, Word and Excel. However, as experience shows, this is quite a certain complexity due to the large number of information transformation operations and an incomplete database [25]. The training sample contains not only numerical, but also linguistic variables. The boundary positive and negative values of signs for the recognition of lung cancer are shown in Table 1. The similarity variability of belonging of the samples to the classification scale of the Lung Cancel “YES” class was determined in the range from 43.99 to 100. The similarity variability of the belonging of the samples to the classification scale of the Lung Cancel “NO” class was determined in the range from 34.26 to 99.51.

Table 1. Values of Signs For The Recognition of Lung Cancer

No. Fields	Attribute Name	Lung Cancer NO	Lung Cancer YES
B	Gender	F	M
C	Age	21-50	51-75
D	Smoking	NO	YES
E	Yellow fingers	NO	YES
F	Anxiety	NO	YES
G	Peer pressure	NO	YES
H	Chronic Disease	NO	YES
I	Fatigue	NO	YES
J	Allergy	NO	YES
K	Wheezing	NO	YES
L	Alcohol	NO	YES
M	Coughing	NO	YES
N	Shortness of Breath	NO	YES
O	Swallowing Difficulty	NO	YES
P	Chest pain	NO	YES

However, if the Eidos system has detected lung cancer for patient, then we cannot assume that s/he really have it. The Eidos system only claims that such a disease is possible, and we can only be diagnosed correctly in a medical institution.

Regular chest x-rays do not help most people live longer, and x-ray exposure itself is dangerous to health [26]. The proposed methodology and existing toolkit help people know their lung cancer risk at low cost. Participation in the procedure for diagnosing lung cancer of medical specialists is not required. All signs of the disease can be determined by the person herself/himself. On the computer of any enterprise with harmful working conditions, we can install the Eidos intellectual system with the necessary representative training sample. It will be operated by a user trained in simple skills of working in the Eidos system. In the examined people, tendencies to the manifestation of the disease can be revealed. Based on the results of forecasting, appropriate decisions can be made taking into account the high risk of the disease. For example, we can quit smoking or change jobs to eliminate the effects of smoke, dust and gas pollution. A representative training sample, fine-tuned by machine learning, can be transferred to medical institutions or to any firms with harmful working conditions that contribute to the occurrence of lung cancer. The use of innovative technology allows, without additional medical tests, to reduce the time of diagnosis of the disease, significantly increase the efficiency and accuracy of identifying the risk of developing lung cancer at an early stage.

3.3. Outcomes of Predicting the Risk of Diabetes at Early Stage

The aim of the work is to determine the risk of developing diabetes mellitus using ASC analysis and its software tool - the Eidos intellectual system. Signs of diabetes in the early stages are: severe weight loss, thirst, frequent urination, dry mouth, irritability, fatigue [27]. To achieve this goal, a training sample was used containing 17 features of 337 patients [28]. The name of the features of the training sample and their significance obtained by the Eidos system are shown in Table 2.

The levels of similarity of belonging of specimens to the classification scale of gradation Positive obtained in the range from 70,15 to 98,67. The levels of similarity of belonging of specimens to the classification scale of gradation Negative obtained in the range from 78.70 to 100. Timely due attention to the factors that determine the risk of developing diabetes mellitus eliminates negative consequences.

The use of a new innovative technique will reduce the time of diagnosis and increase the reliability of disease detection. The Eidos intelligent system is a further development of e-health that does not require the use of expensive equipment. A user trained in simple skills of working in the Eidos system will be able to perform automated procedures for detecting diseases on a representative training sample [29].

Table 2. Name of Features and Their Significance

No.	Feature name	Feature gradations	Significance of grades features
1	Age	(25-45); (45-65); (65-85)	24.28; 12.24; 12.04
2	Gender	Female; Male	69.83; 69.83
3	Polyuria	(1.9-2.9); (2.9-3.97); (3.97-5)	115.5; 28.25; 87.24
4	Sudden weight loss	No; Yes	113.25; 113.25
5	Weakness	No; Yes	65.89; 65.89
6	Polyphagia	No; Yes	112.42; 112.42
7	Genital thrush	No; Yes	104.7; 104.7
8	Visual blurring	No; Yes	114.08; 114.08
9	Irritability	No; Yes	83.45; 83.45
10	Delayed healing	No; Yes	111.32; 111.34
11	Partial paresis	No; Yes	114.66; 114.66
12	Muscle stiffness	No; Yes	86.54; 86.54
13	Alopecia	No; Yes	65.25; 64.25
14	Obesity	No; Yes	8.77; 8.77
15	Sugar in blood	(85-178); (178-271); (271-357)	115.49; 14.95; 101.37
16	Polydipsia	No; Yes	114.66; 114.66
17	Itching	No; Yes	113.83; 113.83

Early prediction of diabetes plays a vital role in choosing the appropriate treatment procedure for a patient to avoid severe consequences. Thus, by setting a factor for improving the patient's condition, it is possible to improve his condition at minimal cost. If practical changes are made without mathematical research, then large expenditures of monetary and other resources will be required. In addition, the duration of the experiments will be long and the results unpredictable. With computer simulations, there will be only minor costs for the salary of the researcher and the payment of electricity. The positive effect of using automated system-cognitive analysis to predict the condition of patients is obvious. The use of a new innovative methodology will reduce the time of diagnosis and increase the reliability of disease detection.

3.4. Outcomes of Stroke risk prediction

The stroke risk database is from the Kaggle repository, but has been greatly improved [30]. Records reflecting disease manifestation factors in persons under 18 years of age were excluded. Entries that did not contain information on smoking status were also excluded, and information on body mass index was added, because. For those who want to determine the risk of stroke, almost all factors are known, except for blood glucose levels. Although this indicator is not difficult to determine, because there are even portable devices for obtaining blood test results [31]. Thus, the number of patients in the study was reduced to 1000. The following factors were included in the dataset:

1. Gender: male, female;
2. Patient's age;
3. Hypertension: 0 if the patient does not have hypertension, 1 if the patient has hypertension;
4. Heart disease: 0 if the patient does not have heart disease, 1 if the patient has heart disease;
5. Marital status: married/unmarried: yes or no;
6. Type of work: government work, never worked, private, entrepreneur;
7. Place of residence: rural or urban;

8. Average blood glucose level, mg/dl;
9. Body mass index;
10. Smoking status: previously smoked, never smoked, smokes.

The similarity level of belonging of specimens to the classification scale of gradation “Yes” obtained in the range from 70,46 to 100. The similarity level of belonging of specimens to the classification scale of gradation “No” obtained in the range from 70,16 to 86,12. To have a low risk of stroke, patient must be free of heart disease, hypertension, live in a rural area, self-employed, never smoke, have low blood glucose, and have a normal body mass index. In addition, s/he must be single and middle aged. The risk of stroke will be high if the person has heart disease or hypertension, s/he needs to live in the city, has high blood glucose levels, has a body mass index above the norm, be in the public service or s/he smokes, being old and married.

Early prediction of stroke plays a vital role in choosing the appropriate treatment for the patient and avoiding severe consequences. With computer simulation, the costs will be minimal, these are the salary of the researcher and payment for electricity [32, 33]. The positive effect of using automated system-cognitive analysis to predict the condition of patients is obvious. The use of a new innovative technique reduces the time of diagnosis and increases the reliability of the detection of the disease.

Outcomes of Liver Disease Recognition

The UCI repository database was used to recognize liver disease for 583 patients. The database was collected in the northeast of Andhra Pradesh, India [34]. This dataset contains symptoms for 441 men and 142 women. 10 variables are contained as symptoms including age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. While verification of models INF1-INF7, the model INF4 was determined, this has the highest reliability of identification and non-identification of objects.

The similarity variability of belonging to the sample specimens of classification “No” is determined in the range from 34.51 to 98.19. The similarity variability of belonging to the sample specimens “Yes” is determined in the range from 34.72 to 87.62. The results of the study can be used by employees of medical institutions due to the fact that the Eidos intellectual system, which is a tool for ASC analysis, is available on the Internet for free [35]. The introduction of a new methodology will increase the reliability of liver disease recognition and improve people's health.

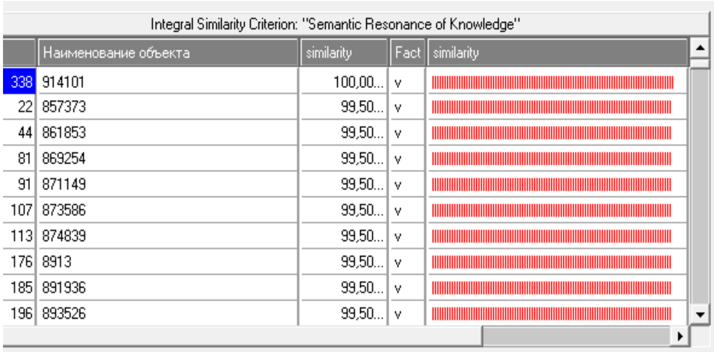
3.5. Outcomes of Diagnosing Breast Cancer

Researchers at New York University Abu Dhabi have developed a new artificial intelligence system based on a convolution neural network that provides radiologist-level accuracy in detecting breast cancer on ultrasound images [36 - 38]. Breast ultrasound has a number of advantages over other imaging modalities, including relatively lower cost, the absence of ionizing radiation, and the ability to evaluate images in real time. However, training a convolution neural network requires a large number of images and special software, so this intelligent system has not yet found wide application in clinics in other countries.

To perform automated system-cognitive analysis, a dataset was created by well-known researchers in the field of general surgery and computer science from the University of Wisconsin-Madison under the direction of Dr. William H. Volberg with the participation of professors W. Nick Street and Olvi L. Mangasarian [39, 40]. A total of 30 features are included in the set: radius_mean, texture_mean, perimeter_mean, area_mean, smoothness_mean, compactness_mean, concavity_mean, concave points_mean, symmetry_mean, fractal_dimension_mean, radius_se, texture_se, perimeter_se, area_se, smoothness_se, compactness_se, concavity_se, concave points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave points_worst, symmetry_worst, fractal_dimension_worst [41, 42].

The variability of sample instances for the gradation of the class DIAGNOSIS-B is shown in Figure 4. The levels of similarity of belonging of specimens to the classification scale of gradation no obtained in the range from 30,63 to 100. The variability of sample instances for the gradation of the class DIAGNOSIS-M is shown in Figure 5.

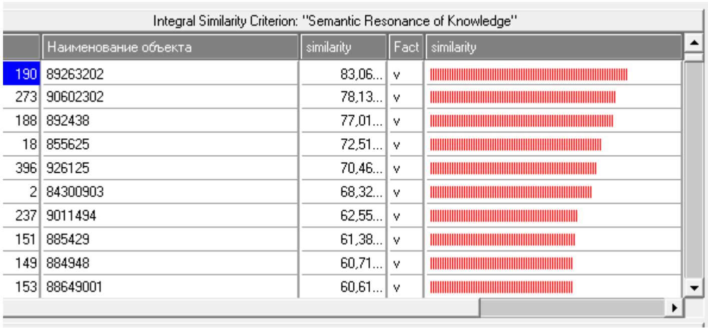
The similarity level of belonging to specimens to the classification scale of gradation “Yes” obtained in the range from 29,95 to 83,06. The distribution of the number of objects of this class DIAGNOSTICS-B according to the degree of similarity of their recognition is shown in the Figure 6. The distribution of the number of objects of class DIAGNOSTICS-M according to the degree of similarity of their recognition is shown in the Figure 7.



Integral Similarity Criterion: "Semantic Resonance of Knowledge"

	Наименование объекта	similarity	Fact	similarity
338	914101	100,00...	v	
22	857373	99,50...	v	
44	861853	99,50...	v	
81	869254	99,50...	v	
91	871149	99,50...	v	
107	873586	99,50...	v	
113	874839	99,50...	v	
176	8913	99,50...	v	
185	891936	99,50...	v	
196	893526	99,50...	v	

Figure 4. The Variability of Sample Instances for The Gradation B



Integral Similarity Criterion: "Semantic Resonance of Knowledge"

	Наименование объекта	similarity	Fact	similarity
190	89263202	83,06...	v	
273	90602302	78,13...	v	
188	892438	77,01...	v	
18	855625	72,51...	v	
396	926125	70,46...	v	
2	84300903	68,32...	v	
237	9011494	62,55...	v	
151	885429	61,38...	v	
149	884948	60,71...	v	
153	88649001	60,61...	v	

Figure 5. The Variability of Sample Instances for The Gradation M

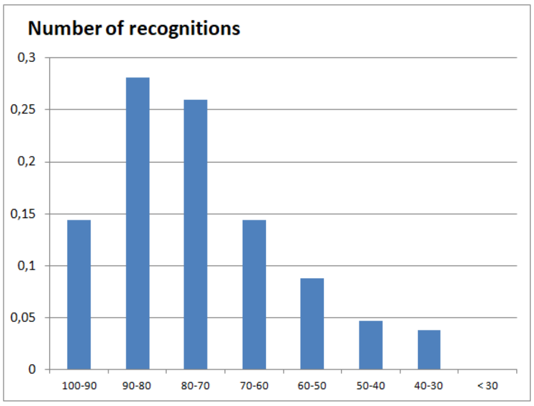


Figure 6. Distribution of Objects of Class DIAGNOSTICS-B According to The Degree of Similarity of Their Recognition

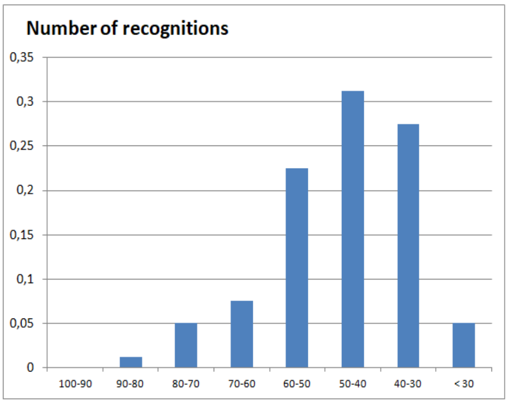


Figure 7. Distribution of Objects of Class DIAGNOSTICS-M According to The Degree of Similarity of Their Recognition

According to the results of the study, presented in the histograms shown in Figures 6 and 7, the average values of recognition similarity were calculated including benign tumors equal to 74,59 and malignant tumors equal to 56,995. In the Eidos system, it is possible to assess the strength of the influence of descriptive scales (factors). Factors that have a low value of the power of influence can be excluded from the study. In the training sample for the classification of breast tumors, the texture_se parameter was first excluded, and then the area-se parameter, but this practically did not affect the similarity of object recognition to the given classes. These changes significantly reduce the complexity of obtaining test results for new patients. The results of the study can be used in breast cancer screening centers because the Eidos system is available free of charge on the Internet and the user interface language is changed to another most common language in the world. The use of new technology for diagnosing breast cancer at an earlier stage will allow women to maintain their health and reduce treatment costs.

4. CONCLUSION

To use the Eidos system for disease recognition, it is necessary to add the patient's characteristics to a new line of the training sample and perform a typical research procedure in the Eidos system. This can be done by a non-medical specialist. It is enough to be a simple computer user and have basic skills in working in the Eidos system. This circumstance, of course, reduces the cost of research. Early diagnosis of diseases is very important, as it helps to prevent the need for complex operations and expensive drugs. It is important to note that intelligent systems do not replace the experience of doctors. However, the powerful complementary role that these systems play as a decision support tool confirms that they must and will be increasingly implemented in clinical practice. The main obstacle that must be overcome when using the Eidos system in medical research is the development of representative training databases that adequately reveal the features of the simulated objects.

The Eidos system has 55 modes of operation, the features of which cannot be described in a short article. More detailed information about working methods in the Eidos system is presented in a series of open lectures, which are posted on the website https://www.patreon.com/user?u=87599532. The Eidos system, like any intellectual system, can be used by any user if he has well-functioning databases, the creation of which is a rather laborious and lengthy process. We have developed representative training databases for detection of stroke, diabetes, early kidney disease, liver disease, lung cancer, breast cancer, detection of human organ tumors, and risk of liver implant failure. Retrospective databases on the detection of human diseases will be distributed free of charge to readers of the journal upon request. Health professionals can save a lot of time when developing training databases.

Patients' diseases were diagnosed by experienced doctors and scientists, including P. Soundarapandian, Dr. William H. Volberg, professors W. Nick Street and Olvi L. Mangasarian, S. Hammerschmidt et al. But they made manual diagnoses of diseases based on the results of tests and interviews with patients based on their knowledge and experience. The reliability of the diagnoses established by doctors and specialists was confirmed using ASC analysis performed using the Eidos system. But with the help of the intelligent system Eidos, the diagnosis of diseases was carried out in an automated way, which reduces the complexity of diagnosing people's diseases.

5. FUTURE DIRECTION

Eidos system is multi-task AI-based system that is able to diagnose different illnesses where authors have given examples in this paper. Interested researchers are incourage to use Eidos system for diagnosing more deseases and give more evidence from different environments. In addition, research results are applicable and easy to replicate in more hospitals and research centers. However, the process of conducting diagnoses is explained in this paper, but it is also possible to gain more benefits explaining the different applications with more reseaerchers and practitioners.

CONFLICT OF INTEREST STATEMENT

The authors declare that the study was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest**.**

ACKNOWLEDGMENTS

The authors are grateful to Professor Lutsenko E.V. for the opportunity to work in the system Eidos.

REFERENCES

[1] Gandomi, A. H.; Chen, F.; Abualigah, L. Big Data Analytics Using Artificial Intelligence. *Electronics*. 2023, 12, 957. https://doi.org/10.3390/electronics12040957

[2] Thakur, S, Rim, T. H., Ting, D. S. J., Hsieh, Y-T. and Kim, T-I. (2023) Editorial: Big data and artificial intelligence in ophthalmology. *Front. Med*. 10:1145522. DOI: 10.3389/fmed.2023.1145522

[3] Breast cancer symptoms: What are the early signs you should look out for? [Electronic resource]. Access mode: https://www.independent.co.uk/life-style/health-and-families/breast-cancer-symptoms-signs-treatment-b2264268.html

[4] James, M. T., Hemmelgarn, B. R., Tonelli, M. Early recognition and prevention kidney disease. *Lancet*, 2010; vol 375, pg. 1296.

[5] Gu, Y., Xu, W., Lin, B., et al. (2022). Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study. *Insights into Imaging*, 13, 124. 2022.

[6] Zheng, D., He, X., & Jing, J. (2023). Overview of Artificial Intelligence in Breast Cancer Medical Imaging. *Journal of Clinical Medicine*, 12(2), 419. DOI: 10.3390/jcm12020419.

[7] Singh, V. K., Rashwan, H. A., Romani, S., et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Syst*. Appl. 2020, 139, 112855.

[8] Loku, L., Fetaji, B., & Krsteski, A. (2018). Automated Medical Data Analyses of Diseases Using Big Data. *Knowledge - International Journal*, 28(5), 1719–1724. https://doi.org/10.35120/kij28051719L.

[9] Zeng, N., Clawson, K., Peng, Y. Editorial: Data-Enabled Intelligence for Medical Technology Innovation, Volume I. *Front Med Technol*. 2022 Jan 14; 3: 841150. DOI:10.3389/fmedt.2021.841150.

[10] Rebolledo-Mendez, J.D. and Vaishnav, R. (2022). Artificial intelligence in precision oncology: The way forward. International Journal of Molecular and Immuno Oncology, 7(3), 77-81. DOI: 10.25259/IJMIO_21_2022.

[11] Lutsenko, E. V. Design and development of intelligent information systems: textbook. - Krasnodar: VTsSKI "Eidos," 2020. – 529 pages.

[12] Chalamet, M. Brain Cancer Data. 2021. [Electronic resource]. Access mode: https://www.kaggle.com/ datasets/michaelchalamet/brain-cancer-data.

[13] Dua, D. and Graff, C. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. 2019.

[14] Al-Ansi, A. M., Jaboob, M., & Awain, A. M. S. B. (2023). Examining the mediating role of job satisfaction between motivation, organizational culture, and employee performance in higher education: A case study in the Arab region. Educ Sci. Manag, 1(1), 30-42.

[15] Lutsenko, E. V. Universal cognitive analytical system “Eidos”. Monograph (scientific edition). Krasnodar: Kuban State Agrarian University after I. T. Trubilina; 2014. 600 p. http://elibrary.ru/item.asp?id=22401787.

[16] Lutsenko, E. V. Synthesis of adaptive intelligent measuring systems using ASC analysis and the Eidos system and system identification in econometrics, biometrics, ecology, pedagogy, psychology and medicine. *Polythematic network electronic scientific journal of the Kuban State Agrarian University (Scientific journal of KubGAU*) [Electronic resource]. - Krasnodar: Kuban State Agrarian University after I. T. Trubilina, 2016. - No. 02 (116). P. 1-60. - IDA [article ID]: 1161602001. http://ej.kubagro.ru/2016/02/pdf/01

[17] Lutsenko E.V. Laboratory workshop on intelligent information systems: Textbook for students of the specialty "Applied Informatics" and other economic specialties. 2nd ed., revised and additional - Krasnodar: Kubanskiy Gosudarstvennyy Agrarnyy Universitet I.m. I. T. Trubilina, 2006. – 318 p. http://elibrary.ru/item.asp?id=21683721

[18] Al-Ansi, A. M., Almadi M., Ichhpujani P., Ryabtsev V. Eidos System Prediction of Myopia in Children in Early Education Stages. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, Vol. 9, No. 2, June 2023, pp. 411-419. ISSN: 2338-3070, DOI: 10.26555/jiteki.v9i2.26292.

[19] Lutsenko, E. V. Design and Development of Intelligent Information Systems. Krasnodar: *VTsSKI “Eidos”,* 2020. – 529 pages.

[20] Al-Ansi, A. M., & Fatmawati, I. (2023). Integration of ICT in higher education during Covid-19 pandemic: a case study. International Journal of Learning and Change, 15(2).

[21] Lutsenko, E. V. Intelligent information systems: Textbook for students of specialty: 351400 "Applied informatics (by industry)." - Krasnodar: KubGAU. 2004. – 633 pages.

[22] Lutsenko, E. V. Information and automated control systems: textbook/ E. V. Lutsenko, V. I. Loiko, V. N. Laptev; commonly. Ed. E. V. Lutsenko. - Krasnodar: KubGAU, 2015. – 450 pages.

[23] Orlov, A. I. Analysis of data, information and knowledge in systemic fuzzy interval mathematics: scientific monograph / A. I. Orlov, E. V. Lutsenko. - Krasnodar: KubGAU, 2022. – 405 p.

[24] Cailhier, J.-F., Nolin L., Hébert M.-J. Thrombotic microangiopathy following renal ischemia and revascularization: apoptosis of endothelial cells in action. *Nephrology Dialysis Transplantation*, Volume 16, Issue 8, August 2001, Pages 1725–1726, https://doi.org/10.1093/ndt/16.8.1725.

[25] Hammerschmidt, S., Wirtz H. Lung cancer: current diagnosis and treatment. *Dtsch Arztebl Int*. 2009 Dec; 106(49): 809-18; quiz 819-20. DOI:10.3238/arztebl.2009.0809.

[26] Stapley, S, Sharp, D., Hamilton, W. Negative chest X-rays in primary care patients with lung cancer. *Br J Gen Pract*. 2006 Aug;56(529):570-3.

[27] Skyler, J. S., Bakris, G. L., Bonifacio, E., Darsow, T., Eckel, R. H, Groop, L, et al. Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes.* 2017; 66: 241–55.

[28] Alam, T. M., Iqbal, M. A., Ali, Y., et al. A model for early prediction of diabetes. *Informatics in Medicine Unlocked*, Volume 16, 2019, 100204, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2019.100204.

[29] Cynthia, Huang, Yizhi, Zhang, Yitian, Hu, Juien, Yang. Visual Analysis and Prediction of Stroke. 36-315 Final Project. [Electronic resource]. Access mode: https://www.stat.cmu.edu/capstoneresearch/spring2021/ 315files/team16.html.

[30] Al-Ansi, A. M. (2022). Reinforcement of student-centered learning through social e-learning and e-assessment. SN Social Sciences, 2(9), 194.

[31] Biswas, M., Kuppili, V., Saba, L. et al. An Artificial Intelligence-Based Bio-Medical Stroke Prediction and Analytical System Using a Machine Learning Approach. *Computational Intelligence and Neuroscience*, Volume 2022, Article ID 5489084, 9 pages. https://doi.org/10.1155/2022/5489084.

[32] Stoot, L. J., Cairns, N. A., Cull, F., et al. Use of portable blood physiology point-of-care devices for basic and applied research on vertebrates: a review. *Conserv Physiol*. 2014 Apr 4; 2(1): cou011. DOI: 10.1093/conphys/cou011.

[33] Al-Ansi, A. M., & Al-Ansi, A. (2023). Enhancing student-centered learning through introducing module for STEM development and assessment. International Journal of STEM Education for Sustainability, 3(1), 22-27.

[34] Lu, H., & Yang, S. (Eds.). (2022). Artificial Intelligence and Robotics: *7th International Symposium, ISAIR* 2022, Shanghai, China, October 21-23, 2022, Proceedings, Part 2 (Vol. 2). Springer Nature Singapore.

[35] Rebolledo-Mendez, J. D. and Vaishnav, R. (2022). Artificial intelligence in precision oncology: The way forward. *International Journal of Molecular and Immuno Oncology*, 7(3), 77-81. DOI: 10.25259/IJMIO_21_2022.

[36] Nagaraj, J. Indian Liver Patient Dataset. [Electronic resource]. Access mode: https://www.kaggle.com/datasets/jeevannagaraj/indian-liver-patient-dataset
[37] Lutsenko, E. V., Ryabtsev V. G. Application of automated system-cognitive analysis for predicting risks in the operation of electrical installations in the agro-industrial complex. *Polythematic network electronic scientific*. 2015. - P. 1-18.
[38] Shen, Y., Shamout, F. E., Oliver, J. R. et al. Artificial intelligence system reduces falsepositive findings in the interpretation of breast ultrasound exams. *Nat Commun,* 12, 5645 (2021). https://doi.org/10.1038/s41467-021-26023-2
[39] Al-Ansi, A. M., & Al-Ansi, A. (2023). An Overview of Artificial Intelligence (AI) in 6G: Types, Advantages, Challenges and Recent Applications. Buletin Ilmiah Sarjana Teknik Elektro, 5(1), 67-75.
[40] Breast Cancer Wisconsin (Diagnostic) Data Set. [Electronic resource]. Access mode: https://archive.ics.uci.edu/ml/datasets/Breast+ Cancer+Wisconsin+%28Diagnostic%29.
[41] Dua, D. and Graff, C. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. 2019.
[42] Jabeen, K., Khan, M. A., Alhaisoni, M., et al. Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors,* 2022, 22, 807. https://doi.org/10.3390/s22030807

BIBLIOGRAPHY OF AUTHORS

Abdullah M. Al-Ansi, Assistant Professor, Ph.D. in Education and Technology. Management, Faculty of Education, Thamar University. Visiting Lecturer at Facuty of Economic and Business, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia. Editor for 6 International Journal and Reviewer for more than 20 International journals with more than 150 reviews. Email: ebrar.ansi@yahoo.com, abdullahalansi@umy.ac.id. ORCiD: https://orcid.org/0000-0002-1682-3154

Vladimir Ryabtsev, Doctor of Engineering Sciences, Professor Information Technologies. Cherkassy Branch of Private Higher Education Establishment European University, Cherkassy, Ukraine. Email: akimonoke2013@gmail.com. ORCiD: https://orcid.org/ 0000-0002-0592-2413

Tatyana Utkina, Ph.D. in Computer Science, Associate Professor. Department of robotics and specialized computer systems, Cherkassy State Technological University, Cherkassy, Ukraine. Email: t.utkina@chdtu.edu.ua ORCID: http://orcid.org/0000-0002-6614-4133