METHODOLOGY FOR BUILDING A MEDICAL EXPERT SYSTEM FOR DISEASE DIAGNOSIS

Magrupov T.M.
Tashkent state technical university named after Islam Karimov, Uzbekistan
Kuchkarova N.A.
Tashkent state technical university named after Islam Karimov, Uzbekistan

https://doi.org/10.5281/zenodo.7226027

Abstract. *This paper proposes an algorithm for constructing expert systems (ES) in comparison with a physician's decision to determine epilepsy in children. The main characteristics of electroencephalography are defined. A decision tree method (DTM), which provides a therapeutic and diagnostic process, is proposed. A method of medical data analysis is proposed, which makes it possible to determine the type of epilepsy in children. The main provisions of a medical expert system for diagnosing epilepsy, based on the decision tree method, are developed. The specialist performs a visual analysis of the EEG and concludes that the EEG is normal. A brief note is given on the analysis of medical data, which allows to determine the type of epilepsy in children. The aim of this study is to develop expert systems for the diagnosis of epilepsy, as well as to create new methods based on the solution tree. The proposed algorithmic and software for determining epilepsy in children with reliability and efficiency of 92 % determines the type of the disease. In a study of 370 patients, we received 186 patients (the rest are healthy), in which the diagnoses coincided with the doctor's diagnoses. The software was created in the Python language, and a Dataset of patients with their disease parameters was created.*

Keywords: *expert system, electroencephalography, epilepsy, decision tree, method, algorithm, program.*

МЕТОДОЛОГИЯ ПОСТРОЕНИЯ МЕДИЦИНСКОЙ ЭКСПЕРТНОЙ СИСТЕМЫ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ

Аннотация. *В статье предложен алгоритм построения экспертных систем (ЭС) в сравнении с решением врача об определении эпилепсии у детей. Определены основные характеристики электроэнцефалографии. Предложен метод дерева решений (DTM), обеспечивающий лечебно-диагностический процесс. Предложен метод анализа медицинских данных, позволяющий определить тип эпилепсии у детей. Разработаны основные положения медицинской экспертной системы диагностики эпилепсии, основанной на методе дерева решений. Специалист выполняет визуальный анализ ЭЭГ и приходит к выводу, что ЭЭГ в норме. Дана краткая заметка об анализе медицинских данных, которая позволяет определить тип эпилепсии у детей. Целью данного исследования является разработка экспертных систем диагностики эпилепсии, а также создание новых методов на основе дерева решений. Предложенный алгоритм и программное обеспечение для определения эпилепсии у детей с достоверностью и эффективностью 92 % определяет тип заболевания. В исследовании 370 пациентов мы получили 186 пациентов (остальные здоровы), в которых диагнозы совпали с диагнозами врача. Программное обеспечение было создано на языке Python, и был создан Dataset пациентов с их параметрами заболевания.*

Ключевые слова: *экспертная система, электроэнцефалография, эпилепсия, дерево решений, метод, алгоритм, программа.*

INTRODUCTION

Currently, there is a need to apply modern information and communication technologies in solving problems of creating medical diagnostic systems. One of the approaches to the creation of medical diagnostic systems is the theory of building expert systems (ES). The use of ES makes it possible to improve the quality of medical diagnostics in clinical hospitals and medical institutions.

MATERIALS AND METHODS

Medical expert systems allow the doctor not only to check his own diagnostic assumptions, but also to ask the computer for advice in difficult diagnostic cases [1].

Medical expert systems (MES) are based on the doctor's experience in observing and treating patients with various diseases PUFF, MYCIN, Oncocyn, SPE, ABEL, CADUCEOS, BLUE, FOX, GUIDON SEIZ [2-4]. It should be noted that the SEIZ expert system works according to a logic similar to that used by the doctor in everyday practice. The system checks the presentation of a symptom based on a knowledge base (KB). In addition, it queries a number of other clinical features (parameters), different variants of which carry different weights for different diagnoses (hypotheses), for example, the parameter "aura" carries a certain positive weighting for complex partial seizure and a negative weighting for typical absence[5]. Other researchers at the University of Tennessee (USA, Memphis) conducted a survey on childhood epilepsy and seizures [6]. The aim of this study was to treat epilepsy with the knowledge of experienced experts from different countries.

A comprehensive study of medical and social problems makes it possible to propose a system of measures aimed at preventing and improving the organization of neurological care for children and adolescents with epilepsy.

RESULTS

The aim of the study is to investigate the age-specific features of the course of epilepsy in children and adolescents and to develop a computer consultative expert system for diagnosing epilepsy. The proposed system consists of: a database of patients with all the signs and parameters necessary for Dataset; an algorithm to diagnose the general condition of the patient at his admission to the clinic; a decision tree method to determine epilepsy in children; an algorithm to determine the type of epilepsy; software to implement the algorithms; evaluation and testing the effectiveness of the developed ES.

We chose a decision tree method for predicting epilepsy in children, which works according to a logic similar to that used by the doctor in everyday practice.

Development of a database of EEG analyses in children to determine epilepsy was conducted with the help of a neurophysiologist at the family polyclinic No. 45 in Tashkent, Uzbekistan.

Four forms of epilepsy were chosen to create ES: febrile seizures in children, benign rolandic epilepsy, juvenile myoclonic epilepsy, and Lennox-Gasto syndrome.

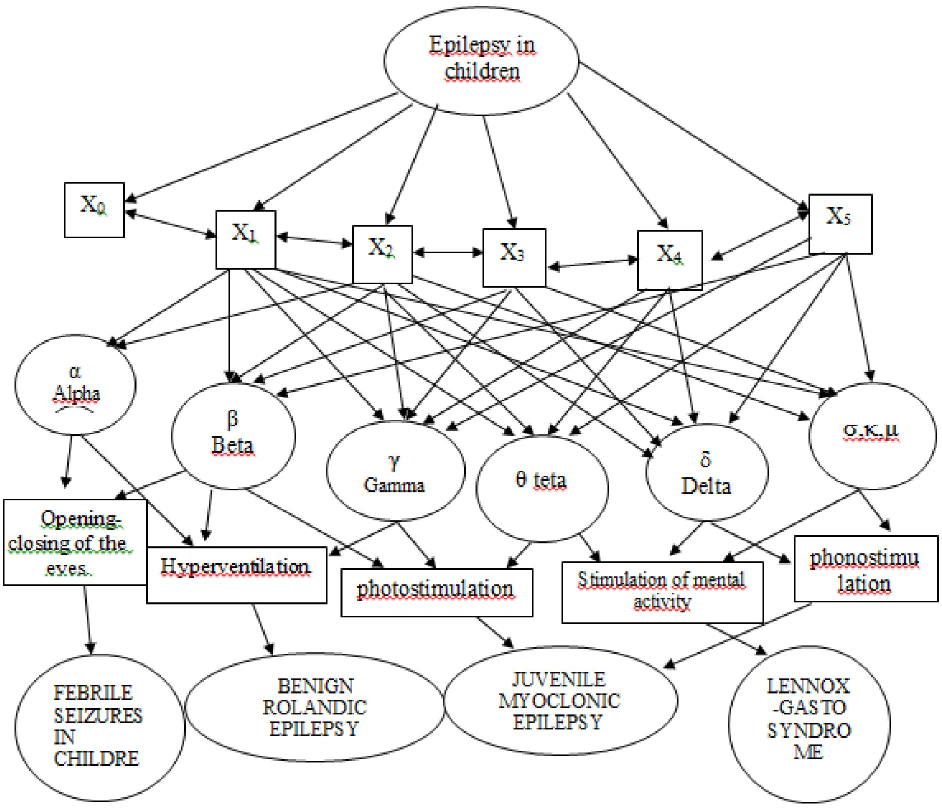
Decision tree algorithm to determine epilepsy in children.

We propose a decision tree algorithm for predicting epilepsy in children (Fig. 1). The decision on such a tree is made as follows:

1. Patient data goes to the root node of the tree.
2. An EEG examination is introduced and the basic EEG rhythms are examined, i.e. alpha, beta, delta, theta rhythms, etc.
3. If pathological activity (adhesions, abrupt changes in the encephalogram) is detected when looking at EEG rhythms, an accurate diagnosis is assigned through provocative tests.
4. Point 2 and point 3 are repeated until the conclusion is reached, which represents the desired solution.
5. After all the stages, the type of epilepsy and the number of patients with epilepsy are determined.

Fig 1.

Decision tree for determining epilepsy in children.



Analysis of the characteristics of the decision tree.

$X_0 = \{ X_{0,1}, X_{0,2}, X_{0,3} \ldots X_{0,13} \}$ - General blood count. $X_{0,1}$ - Saturation (SPO2); $X_{0,8}$ - creatinine; X_{02} - hemoglobin; X_{09} - sed rate; X_{03} - red blood cell; X_{09} - platelets; X_{04} - color.index; X_{010} - hematocrit; X_{05} - leukocyte; X_{011} - lymphocytes%; X_{06} - bilirubin; X_{012} - monocytes%; X_{07} - urea;

$X_1 \{ X_{1\,1}, X X_{1\,2, 1\,3, \ldots} X_{1\,8} \}$ $X_{1\,1}$ - headache; $X_{1\,2}$ - birth trauma; $X_{1\,3}$ - congenital brain abnormalities; X_{14} - chromosomal syndromes; X_{15} - hereditary neurogenic syndromes; X_{16} - infections; X_{17} - brain tumors; X_{18} - hypoxia; X_{18} - cranial nerve pathology;

$X_2 \{ X_{2\,1}, X_{2\,2}, X_{2\,3, \ldots} X_{2\,7} \}$ $X_{2\,1}$ - hereditary metabolic defects; $X_{2\,2}$ - unnatural abrupt extension of the torso; $X_{2\,3}$ - muscle tension throughout the body; $X_{2\,4}$ - abrupt unreasonable shudders; $X_{2\,5}$ - subtle nodding of the head; $X_{2\,6}$ - legs bent, pulled up to the stomach; $X_{2\,7}$ - vascular pathology;

X_3 {$X_{3\,1}$, $X_{3\,2}$, $X_{3\,3}$.... $X_{3\,6}$ } $X_{3\,1}$ - behavioral deficits in the speech and facial areas; $X_{3\,2}$ - cerebral palsy; $X_{3\,3}$ - hydrocephalus (this is an accumulation of excess cerebrospinal fluid in the ventricles (hollow spaces) inside the brain. This leads to an increase in the size of the ventricles and excessive pressure on the brain [7]); $X_{3\,4}$ - encephalitis (then inflammation of the brain. [8]; $X_{3\,5}$ - prolonged somnolence (interventional epileptiform activity in the left paracentral region(cerebral palsy) is spontaneously registered);

X_4 {$X_{4\,1}$, $X_{4\,2}$, $X_{4\,3}$ $X_{4\,4}$ } $X_{4\,1}$ - Quantization Frequency (Hz); $X_{4\,2}$ - sleep deprivation (lack or complete absence of sleep); $X_{4\,3}$ - photostimulation (this is exposure to bright flashes of light directed into the patient's eyes);

X_5 {$X_{5\,1}$, $X_{5\,2}$, $X_{5\,3}$ $X_{5\,4}$ } $X_{5\,1}$ - paroxysmal activity (possible susceptibility to seizures); $X_{5\,2}$ - paroxysmal rapid sleep activity;

Thus, our decision tree for determining epilepsy produces one of the following results: febrile seizures in children, benign rolandic epilepsy, juvenile myoclonic epilepsy, Lennox-Gasto syndrome. math analysis of DTM in epilepsy in children.

We will denote the attributes by *E (epilepsy)* and associate them with each of the patient's analyses *(X_0)* and endings *(X_i)*, EEG rhythms (R), and provocative trials (P).

$E = E(X_0\,, X_i\,, R, P)$ *(1,1)*

$R = \sum(R_\alpha\,, R_\beta\,, R_\gamma, R_\theta\,, R_\delta \ldots R_i)$, here R = EEG rhythms.

1. $X_0 = \{x_{01}, x_{02} \ldots x_{012}\}$ 2. $X_1 = \{x_{1\,1}, x_{1\,2} \ldots x_{1\,9}\}$; $X_{1i} \epsilon X_1$

3. $X_2 = \{x_{2\,1}, x_{2\,2} \ldots x_{2\,7}\}$; $X_{2i} \epsilon X_2$ 4. $X_3 = \{x_{3\,1}, x_{3\,2} \ldots x_{3\,5}\}$; $X_{3i} \epsilon X_3$

5. $X_4 = \{x_{4\,1}, x_{4\,2} \ldots x_{4\,3}\}$; $X_{4i} \epsilon X_4$ 6. $X_5 = \{x_{5\,1}, x_{5\,2} \ldots x_{5\,3}\}$; $X_{5i} \epsilon X_5$;

Let's consider different variants of relations of conceptual elements. And so, the received data on patients in the most general form can be represented as follows:

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5,$$

where X_i - analysis parameters (i=1,2,...,n).

Each of the stages is a structured set of elements. For example, for stage X_1 it may have the form:

$$X_{11} \rightarrow X_{1\,2} \rightarrow (X_{1\,3} \leftrightarrow X_{1\,4}) \rightarrow (X_{1\,5} \uparrow\downarrow X_{1\,5})$$

Step 1. $X_0 \epsilon X_1, X_2, X_3, X_4, X_5$.

Step 2. $R_i \epsilon\ \epsilon X_1, X_2, X_3, X_4, X_5$.

Step 3. $P_i \epsilon$ n, where n is the number of provoking trials (hyperventilation, photostimulation, etc.)

$E_1 = \sum X_I\,(R_i\,, P_i)$, here X_I is the number of analysis parameters.

Decision tree training was based on data from 370 patients with various forms of epilepsy. This system has promising value for use by physicians in the examination of patients under 18 years of age for the diagnosis of 4 forms of epilepsy.

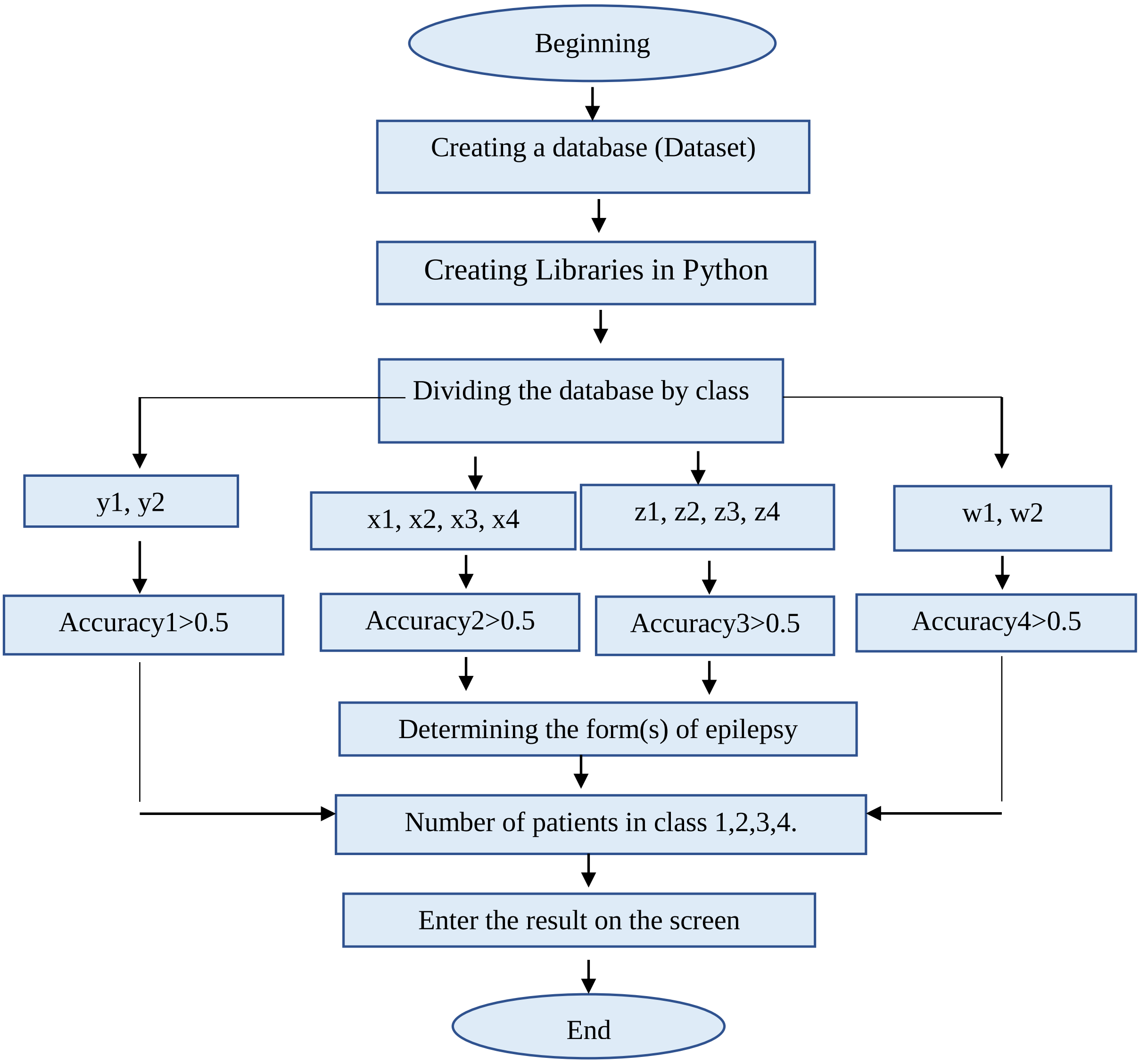
Software organization

To implement the software, we chose Data Science CRISP-DM methodology.The software implementation of the algorithm is shown in Fig. 2.

Fig. 2.

SCIENCE AND INNOVATION
INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 7
UIF-2022: 8.2 | ISSN: 2181-3337

Algorithm of software implementation.



The dialog between the doctor and the expert system is implemented as a thin client (web application). For each visitor at the next visit a unique session number is created, by which the expert system distinguishes users. Using the data entry interface, the doctor compiles a list of symptoms and diagnoses that the patient has or that the doctor assumes (with or without a positive coefficient of confidence). Also, excluded diagnoses and symptoms are added (with a negative confidence coefficient).

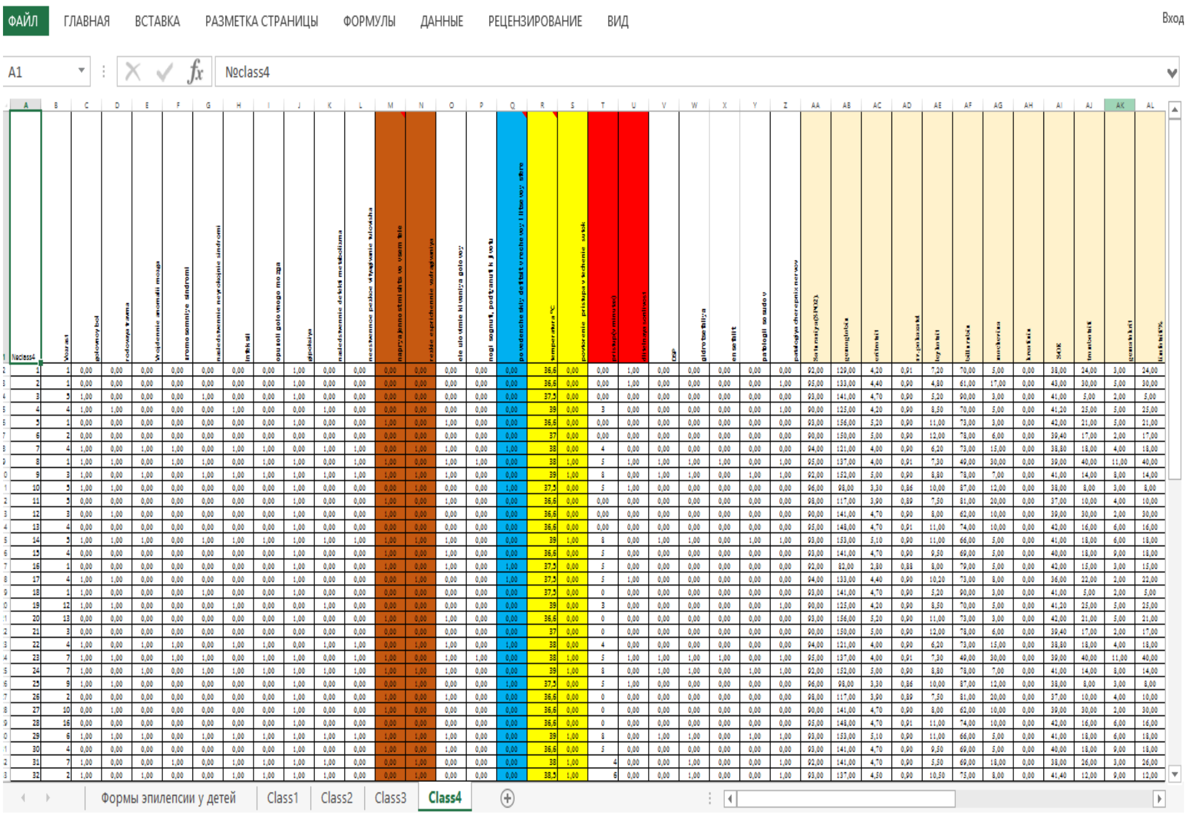
DISCUSSION

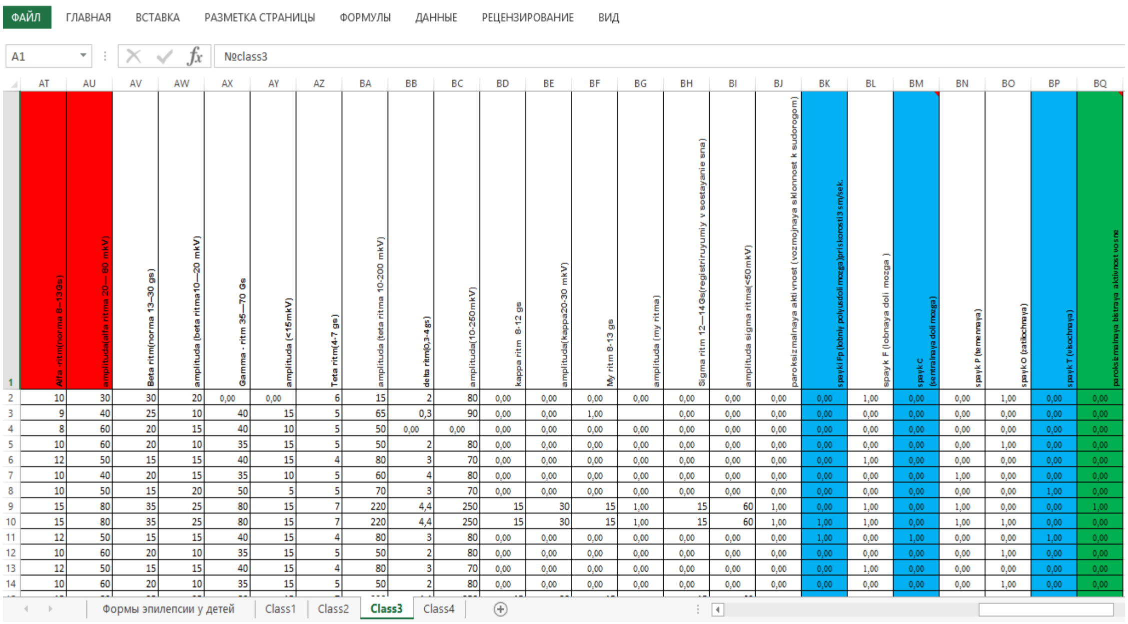
Organization of the database and software implementation of the algorithm.

Fig.2 shows the algorithm for programming. Before we start programming in Python, we collected the analyses of 370 patients by 67 parameters in the form of an Excel spreadsheet (Fig. 3) for our Dataset. This Dataset is divided into 4 classes by parameters, because only 4 forms of epilepsy were chosen.

Fig. 3.

Database (Dataset).





When developing ES I used pandas library (to edit dataset) and numpy (calls library to work with data in the form of a matrix), as well as matplotlib. pyplot (to present data in graphical form). Colab allows you to use all the features of popular Python libraries to analyze and visualize data. For example, the cell below uses the **numpy** library to generate random data and the **matplotlib** library to visualize it.

The implementation of the expert system is developed in the Python programming language with **Google Colaboratory or Colab.** This is another cloud-based service from Google Research. And it is also an IDE that allows anyone to write source code in an editor and run it from a browser. In particular, it supports the Python programming language and is focused on machine learning tasks, data analysis, educational projects, etc. [9].

import pandas as pd # library for editing the dataset

import numpy as np # Calls the library to handle matrix data

df = pd.read_excel('epilepsia_deti.xlsx') # Read the database

df. isnull(). sum() # checks if patients have empty values in the dataset

dfg = df. sort_values(['Vozrast','temperatura ºC'], ascending=False) # checks if the patients have empty values in the dataset

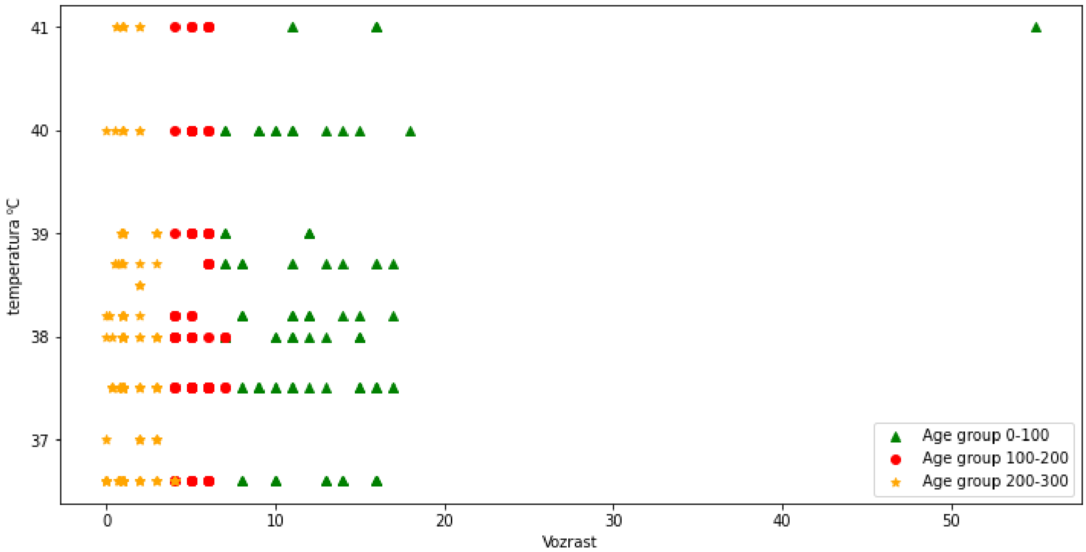
import matplotlib. pyplot as plt # represent data graphically

After editing the library, subtract all the parameters of the patient to the computer and highlight by what parameters you can determine the type of epilepsy.

After the code is executed, a dot plot will appear in the cell above on the page https://colab.research.google.com/drive that displays the age and temperature of the patients (Fig. 4).

Fig.4.

Result, obtained age and temperature of patients in the form of a diagram.



Based on the work of the software, the following results were obtained: 41 children correspond to the first type of epilepsy, 89 children to the second type, 26 children to the third type, and 30 children to the fourth type.

The average value of all results is 92% when working with the necessary medical data of 370 patients. At the same time, 186 patients were obtained in which the diagnosis made by the developed program coincided with the doctor's diagnosis.

CONCLUSION

Algorithms and their software implementation for determining the type of diseases are developed and a comparative assessment of the diagnosis of epilepsy by a physician and the results obtained on the basis of clinical and macroscopic examination data is made. A decision tree method is proposed that provides a therapeutic and diagnostic process for diagnosing epilepsy in children. The developed software of the expert system can be used as intellectual support to the practicing neurologist when interpreting the data of clinical and macroscopic examination of patients and making a preliminary diagnosis. The Data Science CRISP-DM methodology was chosen to implement the ES programs in the Python language. The proposed algorithmic and software for determining epilepsy in children with reliability and efficiency of 92 % determines the type of the disease. In a study of 370 patients, we received 186 patients (the rest are healthy), in which the diagnoses coincided with the doctor's diagnoses. The software was created in the Python language, and a Dataset of patients with their disease parameters was created.

REFERENCES

1. Жариков О.Г., Ковалев А.А., Литвин А.А.,/ Современные возможности использования некоторых экспертных систем в медицине./ Учреждение «Гомельская областная клиническая больница», ГНУ «Объединенный институт проблем информатики Национальной академии наук». – 2008 - №5, стр.-24-29.
2. https://sparm.com/publications/mediczinskie-ekspertnye-sistemy
3. *Андрейчиков А.В., Андрейчикова О.Н.* / Интеллектуальные информационные системы. – М.: Финансы и статистика, 2006.
4. Головин П.А., Нечаев В.А., Нечаев Д.А./ Экспертные системы для классификации болезней в медицинской диагностике.// Научно-технических вестник информационных технологий, механики и оптики. – 2006, №5, стр.-80-84.
5. An expert system for the diagnosis of epilepsy: Results of a clinical trial - SANJEEV V. THOMAS, JAYALEKSHMI R. KURUP, ABRAHAM KURUVILLA, B. NARAYANAN NAIR, K. L. THOMAS, P. SANKARA SARMA/ THE NATIONAL MEDICAL JOURNAL OF INDIA VOL. 14, NO.5, 2001.
6. Wheless, James & Clarke, Dave & Arzimanoglou, Alexis & Carpenter, Daniel (2008). Treatment of pediatric epilepsy: European expert opinion, 2007. Epileptic disorders : international epilepsy journal with videotape. 9. 353-412.
7. https://www.mskcc.org/ru/cancer-care/patient-education/about-hydrocephalus-pediatric-patients
8. https://maremed.ru/article/chto-takoe-jencefalit/
9. https://www.hwlibre.com/ru/google-colaboratory/