

Wioletta SZAJNAR, Galina SETLAK  
The State Professional High School, Jarosław, Institute of Technical Engineering,  
Rzeszow University of Technology, Department of Computer Science

## A CONCEPT OF BUILDING AN INTELLIGENT SYSTEM TO SUPPORT DIABETES DIAGNOSTICS

**Summary.** The paper presents an approach to designing an intelligent system to be used in diabetes diagnostics as well as to creating a problem knowledge base for the system. The work also gives a description of constructing a decision tree for diabetes diagnostics and shows how to use it as a basis for generating knowledge base rules. Finally, the article also outlines how an intelligent diagnostic system works.

**Keywords:** neural networks, artificial intelligence, expert systems, hybrid systems, decision tree diabetes diagnostics

## KONCEPCJA BUDOWY INTELIGENTNEGO SYSTEMU WSPOMAGANIA DIAGNOSTYKI CUKRZYCY

**Streszczenie.** Artykuł przedstawia koncepcję budowy inteligentnego systemu wspomagania diagnostyki cukrzycy. Przedstawiony został opis budowy drzewa decyzyjnego, jak również zasady tworzenia bazy wiedzy dla systemu diagnostycznego.

**Słowa kluczowe:** sieci neuronowe, sztuczna inteligencja, systemy ekspertowe, systemy hybrydowe, drzewo decyzyjne, diagnostyka cukrzycy

### 1. Introduction

Diabetes is nowadays considered to be the one of the greatest health hazards of the XXI century. The number of diabetics patients has been increasing to reach 2 million in Poland and more than 200000 million worldwide. Diabetes has become a social disease, associated with the progress of civilization, that has been afflicting more and more people. It needs intensive monitoring and medical care. On account of this, scientific research to find new approaches and tools in order to support diabetes diagnostics and treatment is socially important

and timely. Presently, such research is being conducted at numerous scientific research centres all over the world.

In recent years, teleinformation systems and Internet portals have been developed to both improve diabetes diagnostics and educate diabetics. A special telemedical programme has been created in EU (Telematic Management of Insulin- Dependent Diabetes Mellitus; TIDDM) for patients with type 1 diabetes. There is, however, a shortage of advanced solutions and computer systems grounded on a knowledge base which support and facilitate medical diagnostics. The purpose of the relevant research is to develop methods of designing and creating intelligent diagnostic systems founded on a knowledge base. The research also aims at implementing state-of-the-art methods and artificial intelligence tools to model and analyze knowledge acquired from various sources. The modelling of knowledge and processes will copy the human way of reasoning. The basic task of the research is the development of a knowledge base as the foundation of an intelligent diagnostics supporting system which can be used in a dispersed medium, in teleinformation systems supporting the monitoring and treatment of people who have diabetes.

The paper presents an approach to the designing of an intelligent diabetes diagnostics system as well as to the creation of a problem knowledge base for the system. The work describes the construction of the decision tree for diabetes diagnostics and explains how to use it to generate knowledge base rules. In recent years artificial intelligence tools (such as neural networks, genetic algorithms, fuzzy logic and others) have been very popular and found their application in solving newer, more complicated problems including diagnosing and recognizing diseases. Artificial intelligence tools are the components of the expert systems that are being built to effectively support diagnostics. Many of them have already been developed and applied, among other things, to diagnose neoplastic diseases, heart and cerebrovascular diseases and many others.

The article is an introduction to developing an intelligent diagnostic system whose task is to initially find out whether the patient has diabetes and then to decide whether the illness is of type 1 or type 2. The paper also presents a decision tree which will help to tentatively determine whether a patient with clinical symptoms suggesting diabetes really has the disease or if it is only a group of symptoms indicative of his, e.g. sugar intolerance. In order to decide on the right line of further analysis we have to prepare, in the first instance, a set of data. The system is very helpful in diagnosing and can make a diabetologist's work much easier [7].

## **2. Basic principles of an intelligent diagnostic system designing**

### **2.1. An intelligent system supporting diabetes diagnostics**

The proposed system of diabetes diagnostics should be a computer system of open architecture and modular structure which will enable the application of a few methods of knowledge representation and integration of different knowledge processing schemes during the process of inference.

Intelligent diagnostic systems should integrate chosen contemporary methods and techniques of knowledge and process modelling, namely: artificial neuron networks – the most fascinating tool of artificial intelligence that can model complex functions and imitate (to some extent) the activities of human mind, namely the ability to learn, fuzzy logic-technologies and methods of natural language formalization, linguistic and qualitative knowledge processing and fuzzy inference, genetic algorithms and methods of evolutionary modelling – algorithms that can learn, based on the theoretical achievements of the theory of evolution and enriching the two aforesaid artificial intelligence techniques (as optimization algorithms they can be applied to teach neuron networks and neuro fuzzy systems and/or to optimize the structure of networks and systems) [10].

The combination of these tools with a traditional expert system where knowledge is symbolically represented will allow for the creation of comprehensive programme tools. They will be used to solve complex problems and tasks necessitating the processing of incomplete, fragmentary, fuzzy or contradictory knowledge as well as in all the situations where the knowledge is difficult to formalize.

An expert system is assumed to be a computer system which, as its name indicates, performs its tasks as efficiently as an expert in a particular branch of knowledge or science. The programme, using a given knowledge base as well as a rule base and facts, draws conclusions and makes decisions, just like a human being. Creating a system founded on a knowledge base means acquiring the knowledge of an expert who often finds the solution making use of relevant information and experience. The expert system that applies the written expert knowledge of a particular field can use it many times in an economical and effective way without the expert's presence. This enables the expert to avoid analogous reports and take up more creative tasks. Thus, a particular asset of such systems is the possibility of solving given tasks without the expert's participation, and also the possibility of aggregating knowledge within one system of a numerous expert team [1].

Any expert system should contain a few fundamental components: knowledge base, data base, inference procedures – inferring engine, explanation procedures – explain inference strategy, procedures of dialogue control- input/output procedures enable the user to formulate

tasks and the programme to transmit the solution, procedures that make it possible to enlarge and modify knowledge – its acquisition [2].

Hybrid intelligent systems are a relatively new category of systems based on artificial intelligence. They consist in the combination of the best features of such tools of artificial intelligence as: expert systems, learning systems, neural networks and genetic algorithms. Owing to this, the hybrid system is able to handle more difficult problems which a single system, which is a part of the hybrid system, could not cope with.

## 2.2. Structure of the hybrid expert system under construction

A hybrid system is a combination of an expert system with neural networks. This way, symbolic processing, characteristic of traditional expert systems is complementary in relation to dispersed, parallel processing typical of neural networks. In the simplest case, both ways of processing may occur independently. It is assumed that there should be a kind of superior medium. It distributes tasks between particular systems. Depending on the sort of problem the tasks are allotted to the system which guarantees the best solution. The functioning of this solution can be described briefly as follows. An expert system operates the interface, processes the data that have been entered into appropriate coefficients and searches the data base. It is accepted that in this field there are no clear and explicit rules; that is why an artificial neural network is the appropriate tool for processing this kind of fuzzy knowledge. An expert system forms a file containing the calculated business coefficients as the input data for the neural network. After the calculation has been carried out by the neural network, the results obtained at its output are passed back to the expert system to be interpreted. In the end, the final results can be read in the displayed window [3].

This is usually a long and tiring process since expert knowledge is of intuitive – practical nature, often difficult to verbalize [4].

In the case when various methods can be used for solving the same problem, same features of one of them can be decisive (e.g. it may be the ability of the expert systems to explain and codify knowledge).

The building process of a hybrid system, which is the effect of the neural network and expert system integration, in the PC- Shell system consists of the following stages (the sequence of points 1 and 2 is arbitrary)

Generating neural applications ( $SN_1, \dots, SN_m$ ), for selected subproblems by means of the Neuronix system.

Elaborating knowledge bases (as knowledge sources)  $SE_1, \dots, SE_n$ .

Integration of the knowledge sources elaborated at the previous stages at the level of the knowledge presentation language of the PC- Shell system.

From the point of view of practical realization, the role of the Neuronix system boils down to teaching the network and generating the definition of the neural application as a defining file (files with npr extension) and a file containing weighing kit (files with weight extension) [8].

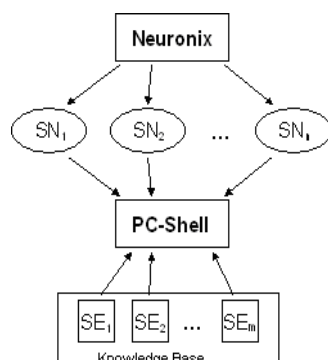


Fig. 1. Structure of hybrid application (SE+SN) in the PC-Shell system. Own preparation  
Rys. 1. Struktura aplikacji hybrydowej (SE+SN) w systemie PC-Shell

### 3. Decision tree as a model of diabetes classification

#### 3.1. Decision tree structure

The decision tree is a flow diagram of a tree-like structure where each top means attribute testing, each edge stands for test exit (value or attribute value set), and each leaf represents a class.

To classify an unknown object the values of its attributes are tested following the information comprised in the decision tree. During the testing process, the path in the tree from the root to one of the leaves is covered – in this way the class the object is going to be classified into will be defined.

Decision trees can easily be transformed into classification rules. The object to be classified is the set of objects,  $S$ , which are described by  $(m+1)$  – the dimensional vector of characteristics  $(x_1, x_2, \dots, x_m, y)$ . Variable  $y$  is the explained variable – i.e. singled out, because of which the classification is performed. What is more, variable  $y$ , depending on the problem being solved, can be a quality variable – (linguistic, nominal or order variable) or continuous and binary variable [7]. The result of the classification in the decision tree induction method is the decision tree.

The main problem to face while building a decision tree is the defining of the criterion enabling the choice of the attribute applied in the development of the tree. In order to choose such an attribute it is necessary to calculate the entropy (entropy is the information contained in the teaching example set). Entropies are calculated following the formula:

$$I(E) = -\sum_{i=1}^{|E|} \frac{|E_i|}{|E|} \cdot \log_2\left(\frac{|E_i|}{|E|}\right) \quad (1)$$

Where:  $E$  – set of teaching examples,  $|E_i|$  – number of examples which describe  $i$ -th object,  $|E|$  – number of examples in the teaching set.

In order to choose the attribute that will be ascribed to the generated node of the decision tree, the criterion of the maximum relative information increase is applied; the increase is caused by the use of a particular attribute (the attribute for which the criterion function value is the highest is applied as the successive attribute):

$$\Delta I(E, a) = I(E) - I(E, a) \quad (2)$$

### 3.2. Diabetes diagnostics

To start a diagnosis it is necessary to gather the proper number of data. The source of the data can be: the patient himself, his record, a primary doctor, a specialist, a biochemical lab and specialist tests. The system gathers data by means of a number of tests: subjective, objective tests, laboratory tests and extra tests.

- subjective tests- general interview: the system takes such data as: name, surname, age, occupation, place of work, living conditions (diet, smoking addiction, nicotine addiction, alcohol), general health condition, eating habits, nutritional status (obesity, emaciation), history of body mass changes, family interview (did the family members suffer/suffered from similar diseases), the beginning of the disease (how many years ago), course of the disease so far, other organ and system complaints, past diseases and operations, present or previous infections, diabetes education, course of treatment (in the case of previously treated diabetes patients) medication applied,
- objective tests: height and body mass measurement (BMI), calculating the due body mass and comparing it with the real mass, personal development phase evaluation (sexual-bodily phase, old age phase), arterial pressure evaluation in the lying and standing position (with the measurement of the orthostatic reaction), ophthalmoscopic tests of fundus (at pupillary dilation), thyroid test, heart test, taking pulse and testing all the peripheral arteries accessible when fingering and auscultating, feet test, neurological test, teeth and gums test, skin and mucosa tests,
- laboratory tests: glycaemia test on an empty stomach and daily glycaemia profile, notation of glycated haemoglobin and fructosamine, notation of the lipid profile on an empty stomach: total cholesterol and cholesterol in lipoproteins of high density (HDL – high density lipoproteins), cholesterol in lipoproteins of low density (LDL – low density lipoproteins) and triglycerides, urine test (apart from glucosuria) to check ketone bodies and protein presence, (macro- and micro-albuminuria) and microscope test of the sediment,

bacteriological test (urine cultures and possibly antibiogram), euthyroidism test and morphological status of the thyroid (concentration evaluation of T3, T4 and TSH, scintigraphy of the thyroid), peripheral arteries tests (potency and the rush of blood), electrocardiogram, echocardiography, ergometer test, neurological tests, especially the electromyography test, ophthalmic review (general test of the organ of sight),

- Additional tests: fundus test [5].

The diagnosis of diabetes is based, most of all, on checking the glucose level. However, many other symptoms that may suggest diabetes are also considered. They include: increased thirst (polidypsia), polyuria, loss of body mass, urinating during the night, increased appetite, general fatigue, dermatomycosis and mucous membrane mycosis, skin infections (mycosis, furuncles) feeling of dryness in the mouth cavity, pruritus of the vulva, cramps, vision disorders [6].

Biochemical criteria:

hyperglycaemia (glucose concentration of above 126mg/dl in venous blood found at least twice), casual blood glucose level in the serum (at any time of the day) of above 200mg/dl, blood glucose level of above 200mg/dl, two hours after a 75g glucose oral load (oral glucose tolerance test OGTT), glucosuria [9].

### 3.3. Principles of constructing a detailed decision tree for diabetes diagnostics

There is a set of three objects:

- diabetes (object  $E_1$ ),
- impaired glucose tolerance (object  $E_2$ ),
- lack of the disease (object  $E_3$ ).

The characteristic features of the objects are: occurrence of diabetes in the family: (yes, no), thirst: (great, normal), urination: (frequent, normal), body mass: (obesity, normal), appetite: (great, normal), fatigue: (great, little), age: (over 45, under 45), hypertension: (yes, no), cholesterol level: (high, normal), triglyceride level: (high, normal), cardiovascular diseases: (yes, no), physical exercise: (little, a lot), dermatomycosis: (yes, no), skin infections: (often, rarely), xerostomia: (yes, no), cramps: (yes, no), vision disorders, glucose level: (above 126mg/dl, below 126mg/dl), occasional glucose content in the serum: (above 200mg/dl, below 200mg/dl), glucose tolerance test: (above 200mg/dl, below 200mg/dl), glucosuria: (yes, no).

Figure 2 shows a decision tree representing the algorithm of type 2 diabetes diagnosis (the most common), considering the factors increasing diabetes hazard.

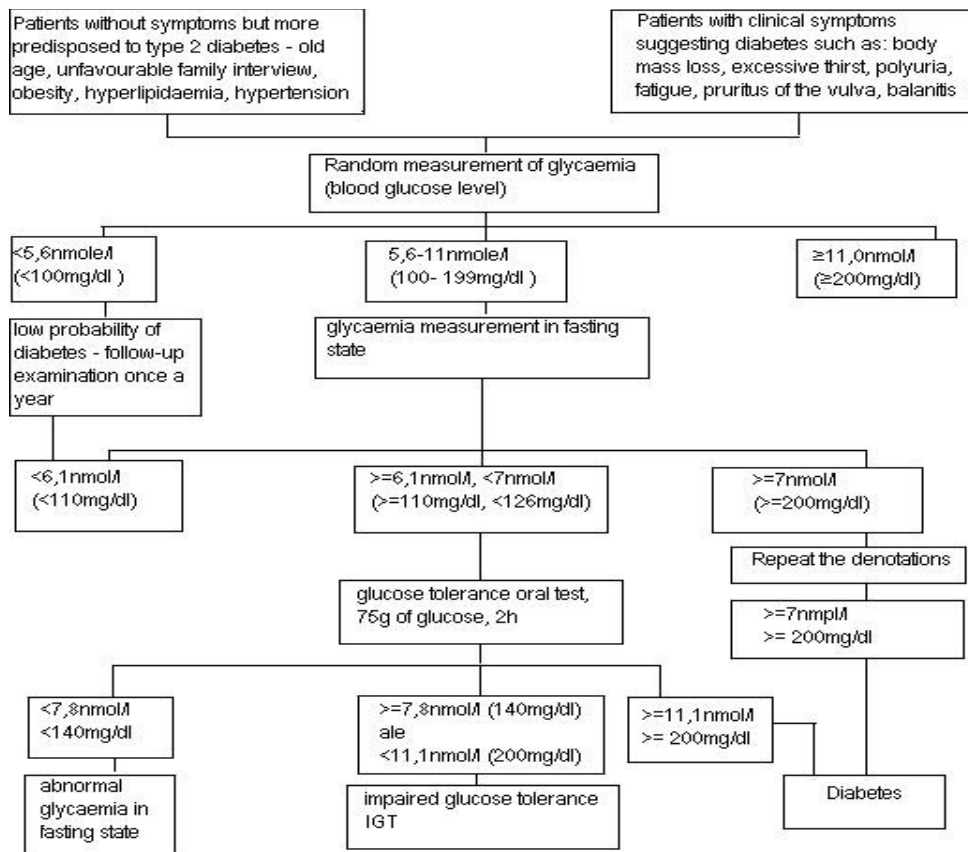


Fig. 2. Diagnostic scheme of diabetes with regard to factors increasing diabetes risk. Own preparation

Rys. 2. Schemat diagnostyczny cukrzycy uwzględniający czynniki zwiększające ryzyko wystąpienia cukrzycy

Due to a multitude of symptoms that might suggest diabetes two groups of patients were selected

- patients without clinical symptoms, but with more diabetes predisposition,
- patients with clinical symptoms suggesting diabetes.

The beginning of the tree consists of the alternatively presented syndromes of the patient's condition. The next component is the examination (diagnostic activity) for which the scheme predicts three possible results: the first suggest completing the procedure, the second – another examination, the third points to the specific condition of the patient (diagnosis). Two results are anticipated for the next examination. One of them- like the result of the first examination – suggests diagnosis, the other, however, insists on another examination whose possible results (given separately) somehow enhance the previously suggested diagnoses. The principle of the practical interpretation of the scheme is like that in the previous case: finding such and such initial state necessitates conducting such and such examinations whose nature and sequence depend on the results obtained and which finally lead to such and such diagnosis [10].



Inference rules can be formulated based on the generated decision tree.

*If (the patient's state without symptoms and more predisposed to diabetes = random glycaemia measurement  $<100\text{mg/dl}$ ), then the probability of diabetes low.*

*If (the patient's condition without clinical symptoms and more predisposed to diabetes = random glycaemia measurement  $>100\text{mg/dl}$  but  $\leq 199\text{mg/dl}$ ) and (glycaemia fasting state measurement  $= <110\text{mg/dl}$ ), then the probability of diabetes low.*

*If (the patient's condition without clinical symptoms, and more predisposed to diabetes = random glycaemia measurement from  $100\text{mg/dl}$  to  $199\text{mg/dl}$ ) and (glycaemia fasting state measurement  $= \geq 110$  but  $<126\text{mg/dl}$  and (oral glucose tolerance test  $= <140\text{mg/dl}$ ), then abnormal fasting state glycaemia.*

*If (the patient's condition without clinical symptoms, and more predisposed to diabetes = random glycaemia measurement from  $100\text{mg/dl}$  to  $199\text{mg/dl}$ ) and (fasting state glycaemia  $= \geq 110$  but  $<126\text{mg/dl}$ ) and (oral glucose tolerance test  $= <200\text{mg/dl}$ ), then impaired glucose tolerance.*

*If (the patient's condition without clinical symptoms, and more predisposed to diabetes = random glycaemia measurement from  $100\text{mg/dl}$  to  $199\text{mg/dl}$ ) and (fasting state glycaemia  $= \geq 110$  but  $<126\text{mg/dl}$ ) and (oral glucose tolerance test  $= \geq 200\text{mg/dl}$ ), then diabetes.*

The other rules are formulated in the same way.

#### 4. Conclusions

The hybrid system we made is to function as a medical expert diagnosing diabetes. It is supposed to replace, in a sense, the doctor in the first phase of the illness diagnostics (it allows for an interview and a detailed blood test).

The knowledge base created so far was confined to the determination of three object groups.

After getting the information about the patient, symptoms and the disease, and after completing the basic medical examinations, the moment is reached when our expert system, based on the above information, draws the relevant conclusions. Diagnosing involves moving in the decision tree. As aforesaid, the system discussed recognizes three classes of the disease: type 1 diabetes, type 2 diabetes and diabetes mellitus in pregnancy. Secondary diabetes is very difficult to recognize and of great diversity. It takes a separate interpretation and analysis. For that reason it was ignored in initial tests.

The scheme of the diagnosis is quite simple and in a way obvious. The final result may be a statement that the patient suffers from a particular kind of diabetes and the recommended remedy is simply an appropriate diet together with pharmacological treatment. It is this way that we reach the stage of diagnosis. If we analysed in detail one of the possible ways, we

would observe that its particular stages do not differ from the proceedings adopted by the doctor. Our expert system is to come to the right conclusion behaving like an expert in a given field, that is, a doctor.

The next stage is the classification of the symptoms and the division into two types of diabetes (objects type 1 and 2), considering differential diagnostics. Training files containing verified patient data will be prepared by an expert. Further analysis can be carried on in different ways, which will be decided after the data set is ready.

It is very important that there should be a close cooperation between the system engineer and the experts in the field in question when constructing the appropriate expert system. Nothing can replace the doctor's knowledge and his long-standing experience.

The system of diabetes diagnostics based on artificial intelligence that we constructed would be the first system of that kind made in Poland.

## BIBLIOGRAPHY

1. Kowalczyk Z., Wiszniewski B.: Inteligentne wydobywanie informacji w celach diagnostycznych. Pomorskie Wydawnictwo Naukowo-Techniczne, Gdańsk 2007.
2. Kwaśnicka H.: Sztuczna inteligencja i systemy ekspertowe. Rozwój i perspektywy, Wydawnictwo Wyższej Szkoły Zarządzania i Finansów, Wrocław 2005.
3. Bubnicki Z.: Wstęp do systemów ekspertowych. PWN, Warszawa 1990.
4. Białko M.: Podstawowe właściwości sieci neuronowych i hybrydowych systemów ekspertowych. Wydawnictwo Uczelniane Politechniki Koszalińskiej, Koszalin 2000.
5. World Health Organization.: Definition, diagnosis and classification of diabetes mellitus and its complications. WHO, Geneva, Switzerland 1999.
6. strona PTD poświęcona cukrzycy <http://www.cukrzyca.info.pl>.
7. Bizoń D.: Analiza klasyfikacji danych medycznych za pomocą wybranych metod sztucznej inteligencji. Technologie Informatyczne w Medycynie, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2008.
8. Mulawka J. J.: Systemy ekspertowe. WNT, Warszawa 1996.
9. Czech A., Tatoń J.: Cukrzyca, podręcznik diagnostyki i terapii. Wydawnictwo ELAMED, Katowice 2009.
10. Gorzałczany M. B.: Neuro-fuzzy classifying system for intelligent decision support, part II. „Applications”. Archives of Control Sciences (Polish Academy of Sciences, Committee of Automatic Control and Robotics), 2000.

Recenzenci: Dr inż. Robert Brzeski  
Prof. dr hab. inż. Andrzej Polański

Wpłynęło do Redakcji 16 stycznia 2011 r.

## Omówienie

Niniejszy artykuł jest poświęcony problematyce tworzenia inteligentnego systemu wspomaganego diagnostyki cukrzycy, opartego na bazie wiedzy. Cukrzyca jest chorobą cywilizacyjną, społeczną, o charakterze postępującym, wymagającą intensywnego monitorowania i leczenia. Badania naukowe wspomagające diagnostykę cukrzycy są bardzo aktualne i ważne społecznie. Obecnie badania w tym zakresie realizowane są w wielu ośrodkach naukowo-badawczych na całym świecie. Brakuje natomiast zaawansowanych rozwiązań i systemów informatycznych, opartych na bazie wiedzy, ułatwiających diagnostykę medyczną w tym zakresie.

Zaproponowano zastosowanie nowoczesnych metod i narzędzi sztucznej inteligencji do modelowania i analizy wiedzy, pozyskiwanej z różnych źródeł, w tym modelowanie wiedzy i procesów na podstawie sposobu rozumowania człowieka. Tworzono hybrydowy system ekspertowy, opierający się na szkieletowym systemie PC Shell 4.5. Jako źródło wiedzy wykorzystano symulator sieci neuronowej NEURONIX, który jest elementem składowym zintegrowanego pakietu sztucznej inteligencji AITECHSPHINX firmy AITECH. Do rozwiązywania takich zadań, jak identyfikacja, klasyfikacja i grupowanie danych diagnostycznych wykorzystywane są zarówno sieci neuronowe, jak i neuronowo-rozmyte systemy. Natomiast zadaniem systemu ekspertowego jest ocena efektywności tych metod, analizy danych diagnostycznych, porównanie wyników i wybór najlepszych rozwiązań.

W artykule przedstawiona została również koncepcja budowy drzewa decyzyjnego, na podstawie którego są tworzone reguły dla bazy wiedzy inteligentnego systemu diagnostycznego.

Opracowane podejście do projektowania inteligentnego systemu diagnostyki cukrzycy umożliwi tworzenie systemów informatycznych o otwartej architekturze, opartych na bazie wiedzy, zawierających kompleksowe narzędzia programowe, pozwalające na praktyczne rozwiązywanie złożonych problemów diagnostycznych. Opracowana baza wiedzy – podstawa inteligentnego systemu wspomaganego diagnostykę – może być wykorzystana w systemach teleinformatycznych wspierających monitorowanie i leczenie chorych na cukrzycę.

**Addresses**

Wioletta SZAJNAR: The State Professional High School, Jarosław, Institute of Technical Engineering, ul. Czarnieckiego 16, 37-500 Jarosław, Polska, wszajnar@wp.pl.

Galina SETLAK: Rzeszow University of Technology, Department of Computer Science, Powstańców Warszawy Str. 8, 35-959 Rzeszów,  
The State Professional High School, Jarosław, Institute of Technical Engineering,  
ul. Czarnieckiego 16, 37-500 Jarosław, Polska, gsetlak@prz.edu.pl.