DATABASE FOR THYROID GLAND DISEASES RESEARCH

Pawe³ Kostka, Sylwia Poœpiech, Ewa Straszecka, Arkadiusz Gertych*
Div. of Biomedical Engineering, Institute of Electronics, Silesian Technical University
16 Akademicka St. 44-100 Gliwice, Poland
phone +48 (32) 2371725, fax +48 (32) 2372235, e-mail ewa@biomed.iele.polsl.gliwice.pl

ABSTRACT: The paper deals with small database design. Customization, flexibility, moderate hardware requirements are considered to most important features for medical research applications. We tried to take advantages of a new programming tool - Delphi system, which offers ready objects for both designing the user interface and creating as well as managing databases. Typical problems arising during data analysis are discussed on example of the thyroid gland diseases database. Dempster-Shafer theory of evidence is used for interpretation of data. It meets the requirements of incomplete data, dependence of parameters and small number of cases.

1. INTRODUCTION.

Database applications enable users to interact with stored information. Databases provide structure for the information and allow it to be gathered and managed. Present hospital information systems are usually huge and expensive programs which strictly prevent anyone from inventing his or her knowledge into them.
Our aim is to provide a physician with a convenient tool to collect patient data and to transfer the stored information into a knowledge base. To improve the reliability of reasoning it should be possible for a user to join different sources of information from medical handbooks and his or her own database. The reasoning process should be clear and calculation of certainty factors simple. An exchange of data is the basic condition for cooperation between physicians and statisticians as well as among research centers.
Tasks which such a database system must face are:

- flexibility from point of view of the design process
- no great hardware requirements; a possibility to use with PC
- a possibility of information exchanging in any format
- guarantee for easy access e.g. through telephone line

Appearance of object-oriented, visual programming environment for rapid application development (RAD) make it possible for a programmer to create in highly efficient way all kinds of applications from general-purpose utilities to sophisticated data access program. We used Delphi 4 programming system to build application for acquisition and analysis data of thyroid gland diseases patients.

2. PROGRAMMING TOOLS USED FOR THYROID GLAND DISEASES DATABASE.

New visual programming environments for rapid application development (RAD) like e.g. Borland Delphi or Borland C++ Builder give opportunities to relatively fast creating very narrow specialized databases. It's the result of providing by these tools ready objects, commonly used fields in user interface part of application like labels, edit fields, notebooks, check boxes, radio buttons groups, list and combo boxes, etc. Therefore, the design of the entering data form can be assign to a specific problem. As an example, we present (fig. 1) fragment of our database form designed for patients suffering from thyroid gland diseases.
At the beginning of this "electronic questionnaire" standard fields – name, sex, age, education etc. describing the examined patient are placed. Next there are fields containing values of such parameters like weight, height, Achilles tendon reflex, and T3, T4, FT3, FT4, TSH with units and norms. Consecutive items in the form are dedicated to particular symptoms. Some of them are grouped in Crooks index. Last field filled in by a physician contains diagnosis. Such a way of designing the application form makes it very flexible and easy to change according to suggestions of physicians. For instance a group of seven additional symptoms elaborated by doctors was placed next to Crooks index. We assumed display of entering data form similar to traditional, paper

* The study is a part of the grant No.8T11E01215 of the State Committee for Scientific Research

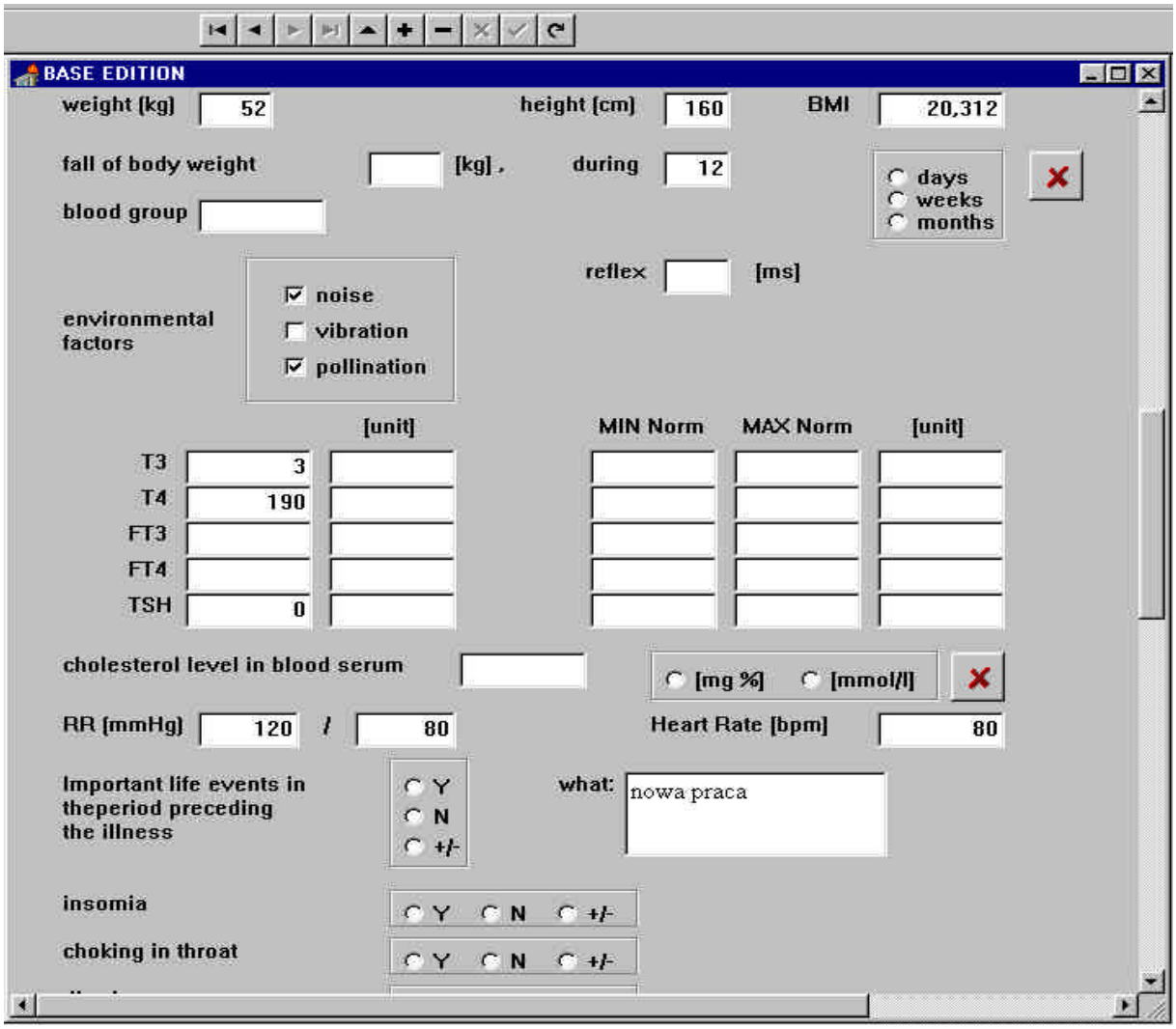

Fig. 1. Fragment of entering data form.

questionnaire used by physicians during talk or interview with a patient. Instead of a pen, computer mouse and keyboard are used. To improve the way of entering data into the form we tried to limit the number of edit fields, by replacing them with mark or multi-choice boxes, wherever it was possible. We introduced logical interconnections between some fields, so their contents depend on the state of other items – some fields can be activated only after choosing specific answer in previous ones. Some of them have assigned mathematical function, which arguments are contents of other fields. An example of such auto-evaluated field is BMI index, which value is determined after filling in weight and height fields. Similarly, SUM in Crooks index is evaluated automatically after marking all of its items. To make data form more flexible, some entered parameters can be scaled in different units – only marking the appropriate unit from a list is required. After that, if necessary, program will convert the value to perform the analysis on data scaled in standard unit.

Apart from objects, which facilitate design of the user interface of the application, RAD programming environments posses tools to create and access databases. They include drivers, which allow to connect to different type databases e.g. Paradox, dBase, FoxPro and Access. While connected, it enables sharing data among users or applications. In our database, we placed the navigation toolbar, which make it possible to look through its contents and edit its records. So we can jump to previous, next, first, last record, and add, remove or modify it. Data records can be presented in the form of original questionnaire or can be placed in a table. Information gathered in the base can also be sorted according to different criterions. The field, which will be the sorting key (e.g. record number or name) should be chosen and marked during the design process. Described above environment tools for database management built in RAD systems, guarantee fast and transparent access to collected data.

Work of the database under Windows platform makes it possible to use all the advantages of this operational system. Records, which contain patients' data can be easily exchanged with other applications; there

is also the possibility to export them to the common accepted ASCII format. Due to included drivers, data forms can be sent to different kind of printers. Fax-modem PC card and an ordinary telephone line makes it possible to transfer database records between remote users. After creating the connection by dialing a number on a computer, it is possible to exchange bi-directional information both among different institution and on the local level. The use of a modem can be alternative for expensive networks systems.

3. DATA ANALYSIS

Database provides fields for storing parameters that are considered of medical interest. Usually not all of them are known for every patient. It results is high percentage of missing information. For most symptoms contributing to Crooks index and other easy to state parameters e.g.: *insomnia, depression, increased thirst,* where possible answers are: yes (*Y*), no (*N*) or don't know (+/-), from 86% to 100% fields are filled in. Hormone levels are known for 25% (*FT3*) – 75% (*TSH*) patients.

In medical databases most of the parameters are not independent. For some symptoms the dependence is obvious even from their definition, e.g. *increased sensitivity to heat* and *increased sensitivity to chilliness*, or *Crooks index* (weighted sum of other symptoms). Other dependencies come up during statistical analysis. Moreover, in medical reasoning practice single symptoms as *T3* and *FT3*, *T4* and *FT4* may be combined to complex symptoms *T3ÙFT3*, *T4ÙFT4*.

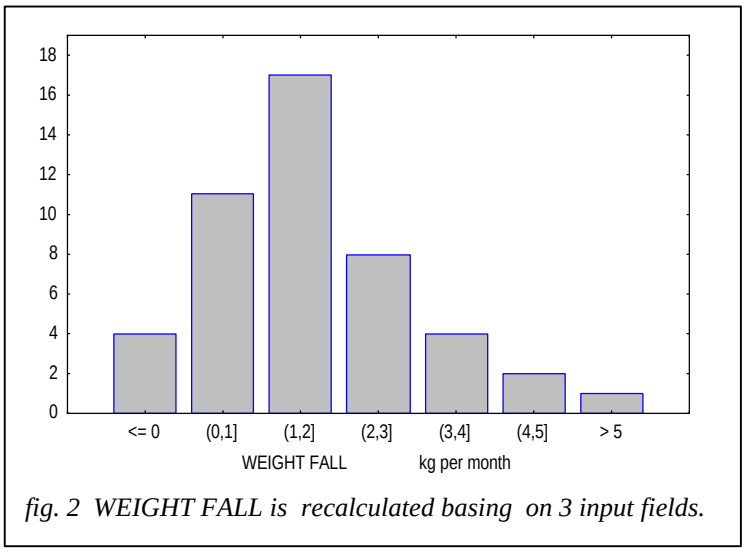

fig. 2 WEIGHT FALL is recalculated basing on 3 input fields.

Many symptoms respond to single fields in the database, but for others 2 to 5 input fields are necessary for proper description. Elementary data processing is built in the database in the form of auto-evaluating fields. Information from input fields is processed from the easy-to-enter form to the form of the symptom that is suitable for statistical analysis (fig.2). This facility allows to get immediately views of the parameters actual values and distributions, which is of particular use whenever instant checking of symptom significance is needed.

For a more advanced interpretation of the database an easy to use, clear and reliable theoretical framework is needed. Classical probability theory cannot be used because of the following, typical in medical applications, limitations:

- Small population does not ensure data for proper statistical descriptions.
- Incomplete data.
- Dependency of symptoms.

An approach with the use of the Dempster –Shafer theory of evidence (DST) may be the solution [3]. DST is similar to probability but has less rigid requirements. In DST the knowledge is represented by a link from a focal element - predicate (symptoms of the disease) to a conclusion (diagnosis), which strength is evaluated by so called measure of belief, calculated on basic probability assignment (*bpa*). *Bpa* describes the information about the predicates. Only two basic assumptions must be satisfied for *bpa* [1]:

$$m(f) = 0$$

$$\sum_{q=T} m(q) = 1$$

Where f - false predicate; T – set of focal elements.

Bpa can be determined in following ways:

- flat *bpa*, every symptom is assigned the same value,
- according to expert's knowledge of the importance of a symptom in the diagnosis,
- using probability (frequency of occurrence).

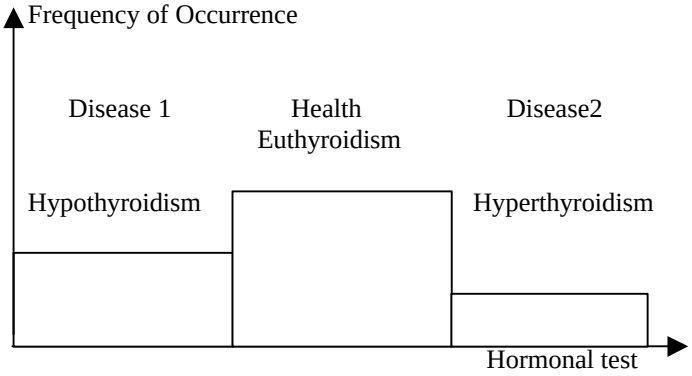

Fig.3 When the same parameter is a symptom of disease or health bpa is determined on base of difference between frequencies of occurrence.

The main advantage of the third method is a possibility to represent empirical distribution. The *bpa* can express which symptoms imply a disease most often in a population, and thus allow to distinguish symptoms which are most frequent. The same parameter in the medicine can be either a symptom of a disease or a symptom of health depending on its value (fig.3). In this case *bpa* can be defined as a difference between symptom occurrence observed for ill and healthy individuals.

Another advantage of DST is the possibility to join two sources of information. The property is particularly useful in case of medical reasoning [2] as it makes it possible to interpret general knowledge in terms of local population. In our application it is used for *Crooks index:*

m_1 – *bpa* according to weights of symptoms in *Crooks index* (general knowledge of the importance of symptoms)

m_2 – *bpa* calculated using frequency of occurrence (local population)

Bpas are combined in the following way [4]:

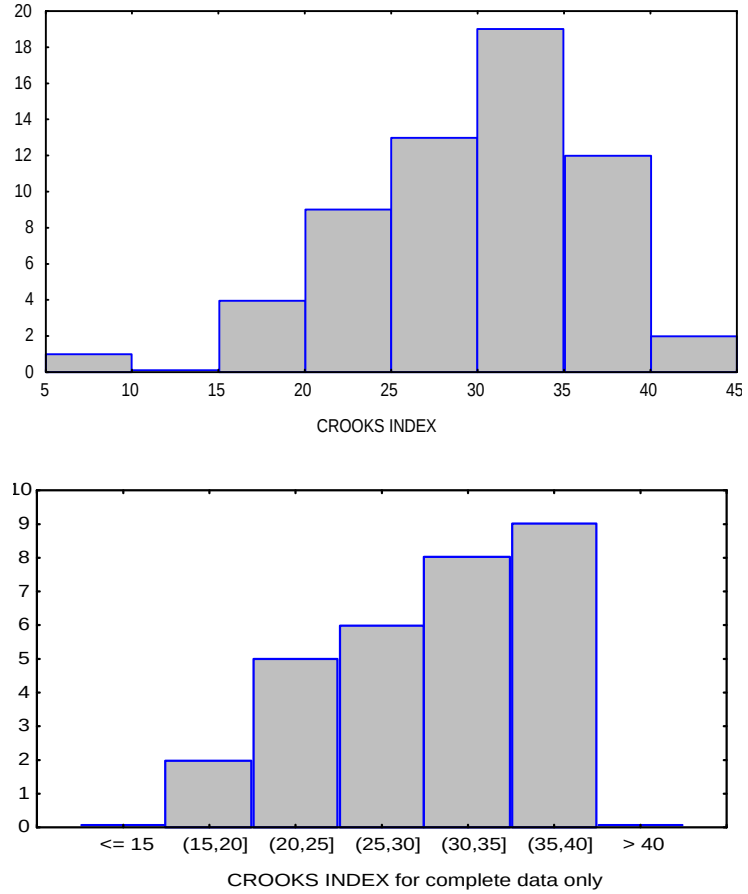

Fig.4 Comparison of the histograms of Crooks index: bottom only cases with complete symptom set; top – all cases.

$$\underset{a\neq f}{\forall} m(a) = \frac{\sum_{b\wedge c=a} m_1(b)\cdot m_2(c)}{\sum_{b\wedge c\neq f} m_1(b)\cdot m_2(c)}$$

Where f – stands for false, m_1 , m_2 – are basic probability assignments of first and second source.

DST gives also a chance to estimate the influence of unknown symptoms on certainty of the diagnosis. For instance complete data to calculate *Crooks index* are provided in 48% of the cases. For the rest 49% 1-3 of the 22 parameters are unknown. The missing information does not affect the final sum as much as to make it out of use (fig.4). It can be noticed unjustified removal of cases with unknown parameters would be an error.

Flexibility of approach is one of the most important requirements during interpreting incomplete information. It is another valuable feature of DST and definition of combined *bpa* that it allows to treat every symptom in the database individually.

Frequency of occurrence may be used to define a *bpa* for well represented parameters, and for parameters with few data we may refer to general medical knowledge.

4. DISCUSSION

The described database is dedicated to the specific, narrow specialized applications. Our database was designed to collect and analyze data of patients suffering of the thyroid gland diseases. Such type of database is convenient at the stage of its creating. Flexible structure of the database makes it possible to design and to modify it easily according to consultations and suggestions of the physicians. It has some specific features, which distinguish it from standard databases. Hormonal tests in diagnostic process are not interpreted as raw values but are considered in reference to their norms. Norms depend on laboratory, where the test was performed. It is the reason for introducing additional fields for inserting norms. It makes it possible to gather data from different sources. During preliminary tests it came up that after accidentally marking one of radio buttons user cannot clear content of the field. Such situation is very important for diagnosis, because selecting "+/-" answer is not equivalent to not choosing any of available options, what means that value of this symptom is unknown.

The analyze of collected data shows that fields corresponding to easy to obtain parameters (e.g. insomia) are filled in almost all cases whereas for parameters requiring more advanced examinations (e.g. laboratory tests, ECG tests) even majority of fields may be empty. Because of small population we have to take under consideration also cases with unknown symptoms. We propose to apply Dempster - Shafer theory of evidence for further investigations of the data.

REFERENCES:

[1] J. Gordon, E.H. Shortliffe –The Dempster-Shafer Theory of Evidence, Rule-Based Expert Systems, Buchanan B.G., Shortliffe E..H. edts, Addison Wesley 1984, pp. 272-292

[2] E.H. Shortliffe, B.G. Buchanan A model of Inexact Reasoning in Medicine, Mathematical Biosciences 23,1975, pp.351-379

[3] Straszecka E., Straszecka J. - Dempster-Shafer Theory of Evidence in Application to Medical Diagnosis Support, Proc.In.Workshop BEMI'97 Sept. 97 Gliwice, Poland, pp.40-44

[4] R.R.Yager, M.Fedrizzi, J.Kacprzyk – Advances in the Dempster-Shafer Theory of Evidence, John Wiley & Sons Inc., 1994