In *Medical Science Monitor*, 7(Suppl. 1):327-332, May 2001

Wasyluk H.[1], Onisko A.[2], Druzdzel MJ.[3]
[1] Medical Center of Postgraduate Education, Marymoncka 99, 01-813 Warszawa, Institute of Biocybernetics and Medical Engineering PAS, Email: hwasyluk@cmkp.edu.pl
[2] Institute of Computer Science, Bialystok University of Technology, Wiejska 45A, 15-351 Bialystok, Email: aonisko@ii.pb.bialystok.pl
[3] Decision Systems Laboratory, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A., Email: marek@sis.pitt.edu

SUPPORT OF DIAGNOSIS OF LIVER DISORDERS BASED ON A CAUSAL BAYESIAN NETWORK MODEL

Key words: medical decision support systems, Bayesian networks

SUMMARY

We describe our work on HEPAR II, a probabilistic causal model for diagnosis of liver disorders. The model, a Bayesian network capturing the causal interactions among various risk factors, diseases, symptoms, and test results, is based on expert knowledge combined with clinical data captured in medical records. The main applications of HEPAR II are assistance in diagnosis and training of beginning diagnosticians. We outline the principles of the applied approach, present a brief description of the model, and report its diagnostic performance.

INTRODUCTION

The last two decades have brought considerable advances in the field of computer-based medical systems. These advances have resulted in noticible improvements in medical care, starting from ease of storage and access of digital imaging through gathering of computerized medical data, accessing on-line literature, patient monitoring, therapy planning, and support for medical diagnosis. Similarly to other domains, decision-support systems have proven to be valuable tools that help practicioners in facing challenging medical problems, such as diagnosis. Given that stakes are typically high, support of decision making plays a particularly important role in the field of medicine. As an example, in the domain of hepatology, inexperienced clinicians have been found to make a correct diagnosis in jaundiced patients in less than 45% of the cases [1]. Incorrect diagnosis leads to suboptimal treatment that means waste of resources, time, and human life. Automating those processes that can be captured by formal models usually allows the physicians to focus on important things, eliminates human error and leads to overall improvements in the quality of medical care.

We believe that the domain of hepatology is especially suited to application of computer-based methods and there are several reasons for high expectations from computer-aided diagnosis. Firstly, the number of cases of liver disorders is on the rise. In Poland, the number of new acute and chronic hepatitis cases is roughly half a million per year. Secondly, correct diagnosis, especially in early stages of a disease, is difficult. There is variety of diseases that manifests with similar symptoms. Finally, early diagnosis is critical, as in some cases damage to the liver caused by an untreated disorder may be irreversible.

Typically, a patient suffering from symptoms suggestive of abdominal disorders seeks help at a

primary health clinic. Primary care physicians face the daunting task of determining the source of discomfort based on patient-reported data and physical examination, possibly enhanced with the results of basic medical tests. Correct diagnosis, under these circumstances is difficult and accuracy can be low. Based on our observations, we estimate that at this stage only 40-60% of the cases are diagnosed correctly. This rather low diagnostic performance is caused by several etiophysiological and organizational factors. These include the nature of the liver, e.g., its high productive reserves that often make the abnormalities noticeable only when the disease reaches an advanced stage. Development of liver diseases is often slow and tracherous. Symptoms may be hardly noticeable. Liver disorders are influenced by environmental factors, such as alcohol intake, medications, and diet. There is still insufficient knowledge of immunological factors that cause certain pathologies of the liver. Undiagnosed or misdiagnosed viral hepatitis often leads to irreversible defects of the liver that may have bearing on the manifectation of the possible later disorders. Reliable diagnosis of a liver disorder can be often established only based on the results of liver biopsy. However, biopsy is an invasive examination that is performed only in specialized clinics and may not be available to primary care physicians.

The factors listed above have encouraged us to explore the use of approaches that are applicable in computer systems. Our main goal was aiding and improving the quality of a diagnosis in non-specialist medical centers and clinics. Our initial work focused on the HEPAR system [2-3], designed and built in collaboration beteween the Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences and the Medical Center of Postgraduate Education in Warsaw. The HEPAR system, used for the last ten years in the Gastroentorogical Clinic of the Institute of Food and Feeding in Warsaw, allows for systematic collection and processing of clinical data of hepatological patients diagnosed and treated in the clinic. The system is equipped with a diagnostic module built around several statistical methods that have been found useful in diagnosis. An integral part of the HEPAR system is its database, created in 1990 and thoroughly maintained since then. The current database contains roughly 860 patient records with the ultimate diagnosis verified by means of biopsy, laparoscopy, and often longitudinal follow-up. Each hepatological case is described by over 150 different medical findings, such as patient self-reported data, results of physical examination, laboratory tests, and finally a histopathologically verified diagnosis. The system includes the module supporting the physicians in making a diagnosis. The HEPAR system has been found useful in practice [3].

We believe that the development of computer-based diagnostic systems in the domain of hepatology is important and should lead to a significant improvement in the timeliness and quality of diagnosis. We believe especially that computer-based systems that are capable of utilizing existing clinical databases of patient cases and combining these with expert opinion can be particularly useful. The work on a decision-analytic version of the HEPAR system (the HEPAR II system) reviewed in this paper allows for a smooth combination of expert knowledge with available statistics. We expect that this approach will lead to a further advances in computer-based decision support in medicine.

BAYESIAN NETWORKS

Probabilistic graphical models, such as Bayesian networks [4] (also called *belief networks* or *causal networks*) offer a coherent and intuitive representation of uncertain domain knowledge.

Mathematically, Bayesian networks are acyclic directed graphs modeling probabilistic dependencies among variables. The graphical part of a Bayesian network reflects the structure of a problem (usually a directed graph of causal dependences in the modeled domain), while local interactions among neighboring variables are quantified by conditional probability distributions.

Figure 1 shows a simple Bayesian network (a simplified fragment of the HEPAR II model). The network models 18 variables related to diagnosis of a small set of hepatic disorders: three risk factors (*Age*, *Sex*, *History of alcohol abuse*), 12 symptoms and test results (e.g., *Jaundice*, *Itching*, *Hepatomegaly*, *GGTP*, *Total bilirubin*), and three disorder nodes (*Primary biliary cirrhosis -- PBC*, *Steatosis hepatis*, and *Functional hyperbilirubinemia*). Arcs denote direct probabilistic relationships between pairs of nodes. And so, the arc between *History of alcohol abuse* and *Steatosis hepatis* represents the fact that alcohol abuse increases the likelihood of steatosis. Relations like this are quantified numerically by means of conditional probability distributions (in this case, the probability distribution over steatosis hepatis given abuse or no abuse of alcohol).

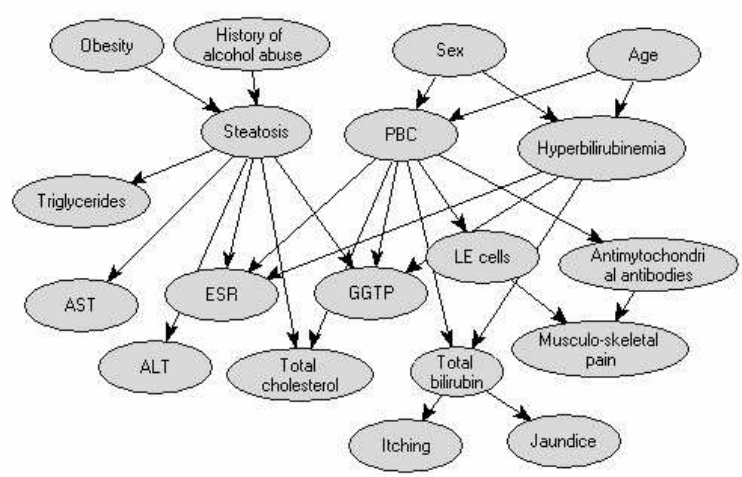

Figure 1: An example Bayesian network model

One of the main advantages of Bayesian networks over other artificial intelligence schemes for reasoning under uncertainty is that they readily combine existing frequency data with expert judgment within the probabilistic framework. Often, for example, hospitals and clinics collect patient data, which over time allow for discovering statistical dependencies and potentially improving the overall quality of diagnosis. When incorporated into a model, they can provide a valuable enhancement to the subjective knowledge obtained from an expert.

Reasoning in Bayesian networks essentially comes down to updating of probabilities given observed variables of a model. For example, the model without any observations will allow for deriving the prevalence rate of each of the disorders. Once specific characteristics of a patient are entered, however, the model will produce the prevalence among the population group that the patient is coming from. Entering risk factors, symptoms, and test results in the course of the interaction with the system will further allow to compute the probability distribution of each of the disordes for this particular patient case. This distribution can be directly used in assisting of medical decisions, for example, by allowing the physician to focus on those

disorders that are most likely. When a model is enhanced with measures expressing the benefits and costs of correct diagnosis and misdiagnosis, respectively, the system can further suggest the diagnosis that is optimal given the circumstances. The probabilistic approach allows for application of such methods as value of information (VOI) computation that essentially computes the expected gain from performing various medical tests and allows for prioritizng various steps of the diagnostic procedure.

Another major advantage of Bayesian networks, compared to other modeling tools, is that they readily model simultaneous presence of multiple disorders. Many approaches, such as those based on classification methods, assume that in each diagnostic case only one disorder is possible, i.e., various disorders are mutually exclusive. This is often an unnecessarily restrictive assumption. It happens fairly often that a patient suffers from multiple disorders and a single disorder may not account for all observed symptoms. Worse even, a situation can arise that a single disorder offers a better explanation for all observations than any other single disorder, while the true diagnosis consists of, for example, two other disorders appearing simultaneously.

Bayesian networks are successfully applied to a variety of problems, including machine diagnosis, user interfaces, natural language interpretation, planning, data mining, and many others (for examples of successful real world applications of Bayesian networks, see March 1995 special issue of the *Communications of ACM*). There have been also successful applications in medicine, for example in medical diagnosis [5-8].

THE HEPAR II MODEL

Using decision-analytic techniques as described above, and particularly Bayesian networks, we have built a model that aims at assisting medical decisions in diagnosis of liver disorders. The technical components of our work are reported elsewhere [9-10]. In this paper, we review the most important elements of the system from the point of view of practical application in medical diagnosis.

The structure of the model is a representation of the relationships between various components of the diagnosis process, i.e., disorders, demographic and risk factors, symptoms, and test results. We constructed the structure based on medical literature and conversations with our domain expert, a hepatologist Dr. Hanna Wasyluk (first author) and two American experts, a pathologist, Dr. Daniel Schwartz, and a specialist in infections diseases, Dr. John N. Dowling from the University of Pittsburgh. We estimate that elicitation of the structure took approximately 40 hours of interviews with the experts, of which roughly 30 hours were spent with Dr. Wasyluk and roughly 10 hours spent with Drs. Schwartz and Dowling. This includes model refinement sessions, where previously elicited structure was reevaluated in a group setting. The HEPAR II model consists of over 70 nodes representing risk factors, symptoms, laboratory tests results, and nine liver disorders.

The numerical parameters of the model, i.e., the prior and conditional probability distributions, were extracted from the HEPAR database. Prior probability distributions are simply relative counts of various outcomes for each of the variables in question. Conditional probability distributions are relative counts of various outcomes in those data records that fulfill the conditions described by every combination of the outcomes of the predecessors.

The data used to extract the numerical parameters contained 505 patient records. All continuous variables were discretized by our expert. One of the assumptions that we used in learning the model parameters was that missing values for discrete finding variables corresponded to state *absent* (e.g., a missing value for *Jaundice* was interpreted as *absent*. In case of continuous variables, a missing value corresponded to a normal value, elicited from the expert (e.g., a missing value for *Total bilirubin* was interpreted as being in the range of 0—1 mg/dl) as the typical value for a healthy patient. We followed here the observation reported by Peot and Shachter [11] that missing values in medical data sets are not missing at random and are either indications of normal or less severe symptoms.

Given a patient case, i.e., values of some of the modeled variables, such as symptoms or the laboratory tests results, the HEPAR II system calculates the posterior probability distribution over the possible liver disorders. We would like to stress here that it is not necessary to enter all possible observations – the system will compute the probability of the disorders for any number of observations available, even if these are just a few symptoms. Obviously, the quality of the system's suggestions improve with the number of observations entered and with their diagnostic value. We present a list of suggested diagnoses rank-ordered by their probability to the user, who is the ultimate decision maker.

HEPAR II as a computer-based diagnostic tool

To test how intuitive and how useful the model is in practice, we have built a simple user interface to our model that lists all observable features, allows the user to set the values of any of them, updates the probability distribution over different disorders and presents an ordered list of possible diagnoses with their probabilities.

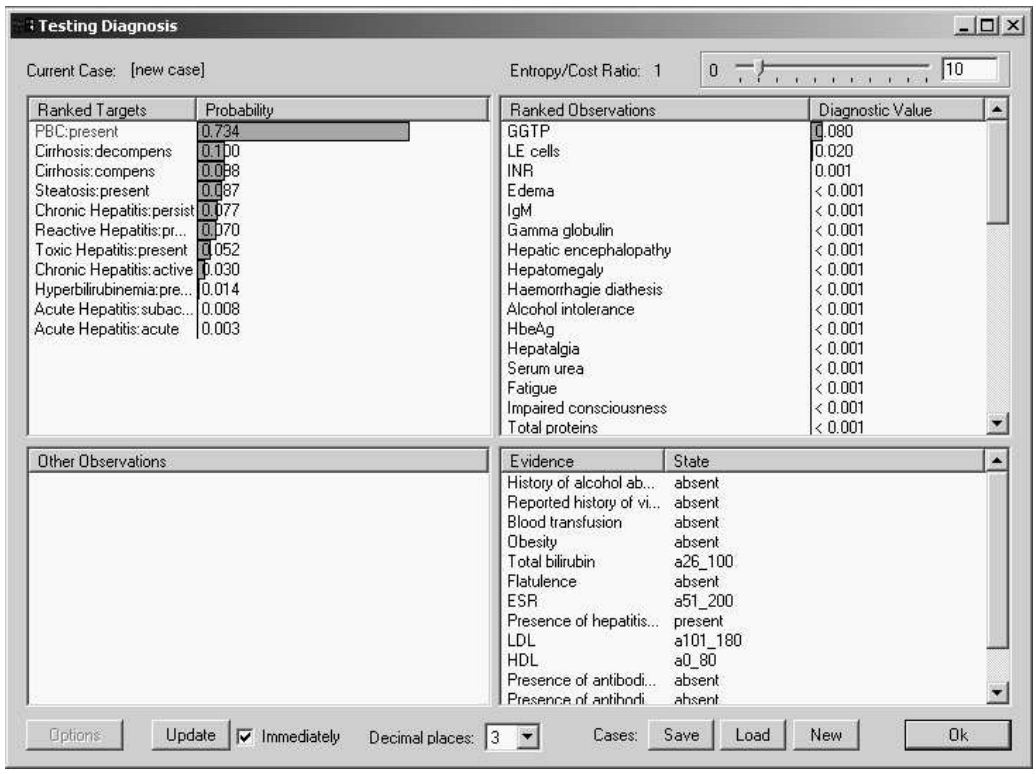

Figure 2: An interactive diagnostic interface to HEPAR II

Figure 2 shows a screen shot of this user interface. The right-hand side of the window contains a complete list of all possible symptoms and observations included in the model. The top right part of the winow contains a list of those possible observations that have not yet been made along with an indiation of their diagnostic value for the pursued disorder (PBC in this case). Those features that have been observed are brought over to the bottom part of the winow. Right-clicking on any of the features brings up a pop-up menu that lists all possible values of the selected variable. By choosing one of these values the user can enter a finding. The top left column presents an ordered list of the possible diagnoses along with their associated probabilities, the latter being presented graphically. The probabilities are updated immediately after entering each finding. Updating the probabilities and presenting a newly ordered list of possible disorders takes in the current version of the model a fraction of a second and is from the point of view of the user instantenous. Our interface allows further to save a patient case in a repository of cases and to return to it at a later time.

Diagnostic accuracy of the HEPAR II model

One question that a critical reader may ask is whether a model like HEPAR II is suffciently accurate to be useful in diagnostic practice. In the course of the project, we performed several tests of diagnostic accuracy of the system [9-10]. In our experiments, we typically took individual patient cases from the HEPAR database and compared the system suggestions to the "gold standard' diagnosis captured in the database. We were interested in both (1) whether the most probable diagnosis indicated by the system is indeed the correct diagnosis, and (2)

whether the set of *k* most probable diagnoses contains the correct diagnosis for small values of *k* (we chose a "window" of *k*=1, 2, 3, and 4). The latter focus is of interest in diagnostic settings like ours, where a decision support system only suggest possible diagnoses to a physician. The physician, who is the ultimate decision maker, may want to see several alternative most likely diagnoses before focusing on one.

Performance for window size of four for each of the 9 disorders individually is pictured graphically in Figure 3.

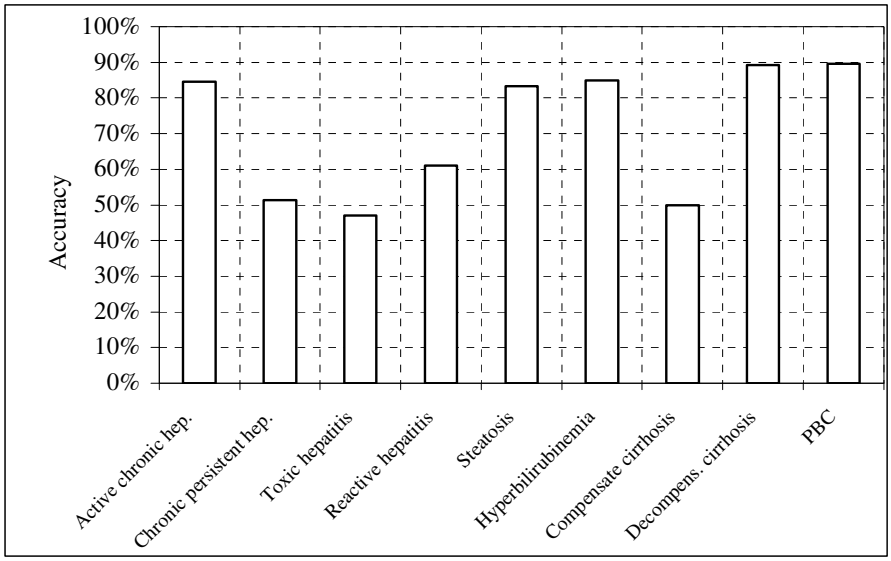

Figure 3: Diagnostic performance of the HEPAR II system

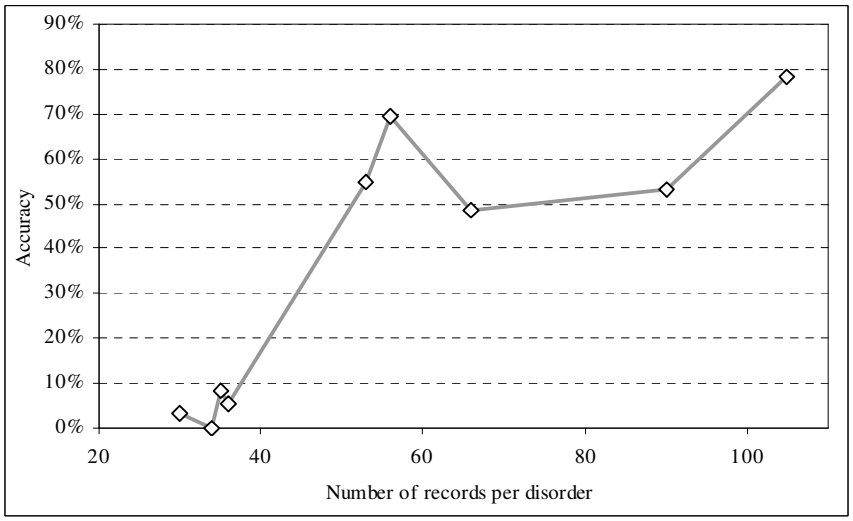

Figure 4: Relationship between the number of patient cases in the database and Hepar II's diagnostic accuracy

Figure 4 shows the relationship between the number of records in the HEPAR database for each disorder and the system accuracy in diagnosing this disorder (window size equals one,

i.e., the most likely disorder). Disorders with more than 50 patient cases captured in the database showed quite high diagnostic accuracy. This promises a higher diagnostic value of our approach when the available data set of patient cases is sufficiently large. This reflects the reliance of the numerical parameters of the model on clinical data in our approach. In those cases where there are no clinical data available, the numerical parameters can be elicited directly from experts. There are quite a number of numerical parameters needed in a model of Hepar II's size (almost 3,000 in the current version of the model) and elicitation of these parameters from experts may be daunting. From the practical point of view, thorough collection an maintanance of clinical data is extremely important. Fortunately, this need has been appreciated and facilitated along with the development of computer-based clinical information systems and most hospital do collect data.

HEPAR II in training beginning dignosticians

In addition to testing the diagnostic accuracy of the system on past patient cases, we were interested in comparing how the system is doing compared to human diagnosticians and whether its impact on the ultimate diagnosis as peformed by its user is beneficial. We applied HEPAR II in training of beginning practitioners (internists and pediatricians) doing a fellowship in general medicine (The results of this study are reported in [12]). The system's accuracy was higher than that of the physicians. The reaction of the users to the system was very favorable and several of them said that the system had been very understandable and that they had learned a lot from it. A quantitative measure of the system impact on the physicians showed that working with the system was beneficial for them. The diagnostic accuracy of physicians increased significantly after they had seen HEPAR II answers. Additionally, there was no negative impact of the system on the users, i.e., none of the correct decisions made by the physicians were changed, even if the system provided an incorrect answer. It has led us to the conclusion that the system could be useful in training of physicians with limited clinical practice.

CONCLUSIONS

The HEPAR II project aims at applying decision-theoretic methods to the problem of diagnosis of liver disorders. It's domain model, a Bayesian network modeling various disorders, risk factors, symptoms, and test results, is built based on a combination of expert knowledge and a database of patient cases. It allows for computing the impact of various observations on the probability of various disorders and aids directly the process of diagnosis. The program has been welcomed as a useful interactive diagnostic and training tool by our colleague physicians. One obvious application of the system is in aiding diagnosis. The diagnostic accuracy of the model seems to be reasonable and for those disorders that are well represented in the database of patient cases reaches almost 80%. The overall model accuracy seems to be better than that of beginning diagnosticians.

The second application of HEPAR II is in teaching. A niche that can be currently filled by HEPAR II is training family doctors, a new specialization in Poland created by the recent health reform. The new family doctors participate in courses at the Medical Center of Postgraduate Education in Warsaw. We believe that the system is particularly useful in teaching the new family doctors the strategy and optimization of the process of diagnosis by

allowing a physician to verify and improve his or her skills. A physician using HEPAR II can by means of simulation assess or verify the diagnostic value of medical findings, such as results of physical examination, self-reported data, or tests in a given situation involving differential diagnosis. A thorough training in differential diagnosis allows a physician to recognize and choose an optimal diagnostic strategy, which is practice comes down to ordering the right tests for the right patients. Selection of those laboratory tests that have the highest diagnostic value can help in reducing the costs of health care. It also has a positive social impact by increasing the accuracy of diagnosis and correctness of therapy.

ACKNOWLEDGMENTS

Support for this research was provided by the United States Air Force Office of Scientific Research, grant F49620-00-1-0112, by the National Science Foundation under Faculty Early Career Development (CAREER) Program, grant IRI-9624629, by the Polish Committee for Scientific Research, grant 8T11E02917, by the Medical Center of Postgraduate Education, Poland, grant 501-2-1-02-18/00, and by the Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, grant 16/ST/00 and by Bialystok University of Technology grant W/II/1/00. Our collaboration was enhanced by travel support from a NATO Collaborative Linkage Grant PST.CLG.976167.
The HEPAR II model was created and tested using SMILE, an inference engine, and GeNIe, a development environment for reasoning in graphical probabilistic models, both developed at the Decision Systems Laboratory, University of Pittsburgh, and available at http://www2.sis.pitt.edu/GeNIe.

REFERENCES:

1. Theodossi, A., Spiegelhalter, D. J. and Portman, B.: The value of clinical, biochemical, ultrasound and liver biopsy data assessing patients with liver disease. Liver, 1983: 3: 315-326.
2. Bobrowski L., Wasyluk H., Dmowski B.; Analysis of Medical Data in Computer system HEPAR. Computers in Medicine 4; 66-71, 1997.
3. Wasyluk H.; The Four Year Experience with HEPAR – Computer Assisted Diagnostic Program. MEDINFO-95, ed. R.A.Greenes, IMIA, Vancouver 1995.
4. Pearl J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, Inc., San Mateo, 1988.
5. Diez, Francisco J., Mira, J., Iturralde, E. and Zubillaga S.: DIAVAL, a Bayesian expert system for echocardiography, Artificial Intelligence in Medicine, 1997: 10: 59—73.
6. Heckerman D., Horvitz E., Nathwani B.; Toward normative expert systems. 1 The PATHFINDER project. Methods of Information in Medicine, 1992:31,90-105.
7. Middleton B, Shwe MA, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, Cooper GF: Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base: II. Evaluation of Diagnostic Performance. *Methods of Information in Medicine*, 1991:30(4): 256-267.
8. Andreassen, S., Woldbye, M., Falck, B. and Andersen, S. K.: MUNIN -- A causal probabilistic network for interpretation of electromyographic findings, In Proceedings of the Tenth International Joint Conference on Artificial Intelligence", McDermott, J., Morgan Kaufmann Publishers, Los Altos, CA, 1987, pages: 366—372.

9. Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk. Extension of the HEPAR II Model to Multiple-Disorder Diagnosis. In Advances in Soft Computing, pages 303-313, Physica-Verlag, Heidelberg, New York, 2000.
10. Agnieszka Onisko, Marek J. Druzdzel and Hanna Wasyluk. Learning Bayesian Network Parameters from Small Data Sets:Application of Noisy-OR Gates. In Notes of the ECAI workshop: "*Bayesian and Causal Networks: From Inference to Data Mining*", Berlin, Germany, August, 22, 2000.
11. Peot, Mark and Shachter, Ross. Learning From What You Don't Observe. In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI--98), Morgan Kaufmann Publishers, San Francisco, CA, 1998: 439—446.
12. Agnieszka Onisko. Evaluation of the HEPAR II system for a diagnosis of liver disorders. To appear in *Proceedings of the Tenth Symposium on Intelligent Information Systems (IIS-2001)*, Poland, June 18-22, 2001.