# A Bayesian Network Model for Diagnosis of Liver Disorders

Agnieszka Onisko, M.S.,[1,2] Marek J. Druzdzel, Ph.D.,[1] and Hanna Wasyluk, M.D.,Ph.D.[3]

[1] Decision Systems Laboratory, School of Information Sciences, Intelligent Systems Program, and Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A., *marek@sis.pitt.edu, aonisko@sis.pitt.edu*

[2] Institute of Computer Science, Bialystok University of Technology, Bialystok, 15-351, Poland, *aonisko@ii.pb.bialystok.pl*

[3] Medical Center of Postgraduate Education, Warsaw, 01-813, Marymoncka 99, Poland, *hwasyluk@cmkp.edu.pl*

*Probabilistic graphical models, such as Bayesian networks and influence diagrams, offer coherent representation of domain knowledge under uncertainty. They are based on the sound foundations of probability theory and they readily combine available statistics with expert judgment. This paper describes our work in progress on a probabilistic causal model for diagnosis of liver disorders that we plan to apply in both clinical practice and medical training. The network, and especially its numerical parameters, is based on data from a clinical database. We present the Bayesian model and report initial results of our diagnostic performance tests.*

## INTRODUCTION

Some of the earliest Artificial Intelligence (AI) approaches to medical diagnosis were based on Bayesian and decision-theoretic schemes. Difficulties in obtaining and representing quantities of numbers and both the computational and representational complexity of probabilistic schemes caused a long-lasting departure from these approaches. Only recently, development of probabilistic graphical models, such as Bayesian networks[2] and closely related influence diagrams, has caused a renewed interest in applying probability theory in intelligent systems (see [4] for an accessible overview of decision-analytic methods in intelligent systems).

Today, Bayesian networks are successfully applied to a variety of problems, including machine diagnosis, user interfaces, natural language interpretation, planning, vision, robotics, 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.[8,9].

In this paper, we describe our work in progress on a probabilistic causal model for diagnosis of liver disorders. Our work is continuation of the HEPAR project,[1] conducted in the Institute of Biocybernetics and

Biomedical Engineering of the Polish Academy of Sciences in co-operation with physicians at the Medical Center of Postgraduate Education. The HEPAR system contains a database of patient records of the Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw. This database is thoroughly maintained and enlarged with new cases. The system is currently used in the clinic as a diagnostic and training aid. Our model is essentially a Bayesian network modeling causal relations among its variables with its numerical parameters extracted from the HEPAR database. One application of our model, in addition to its diagnostic value, is in training physicians. We present the model and report the initial results of our diagnostic performance tests.

## BAYESIAN NETWORKS

A Bayesian network[2] (also referred to as *Bayesian belief network*, *belief network*, *probabilistic network*, or *causal network*) consists of a qualitative part, encoding existence of probabilistic influences among a domain's variables in a directed graph, and a quantitative part, encoding the joint probability distribution over these variables. Each node of the graph represents a random variable and each arc represents a direct dependence between two variables. Formally, the structure of the directed graph is a representation of a factorization of the joint probability distribution. As many factorizations are possible, there are many graphs that are capable of encoding the same joint probability distribution. Of these, those that minimize the number of arcs are preferred. From the point of view of knowledge engineering, graphs that reflect the causal structure of the domain are especially convenient - they normally reflect expert's understanding of the domain, enhance interaction with a human expert at the model building stage and are readily extendible with new information. Finally, causal models facilitate user insight once a model is employed. This is important in all those systems that aid decisions and fulfill in part a training role, like most
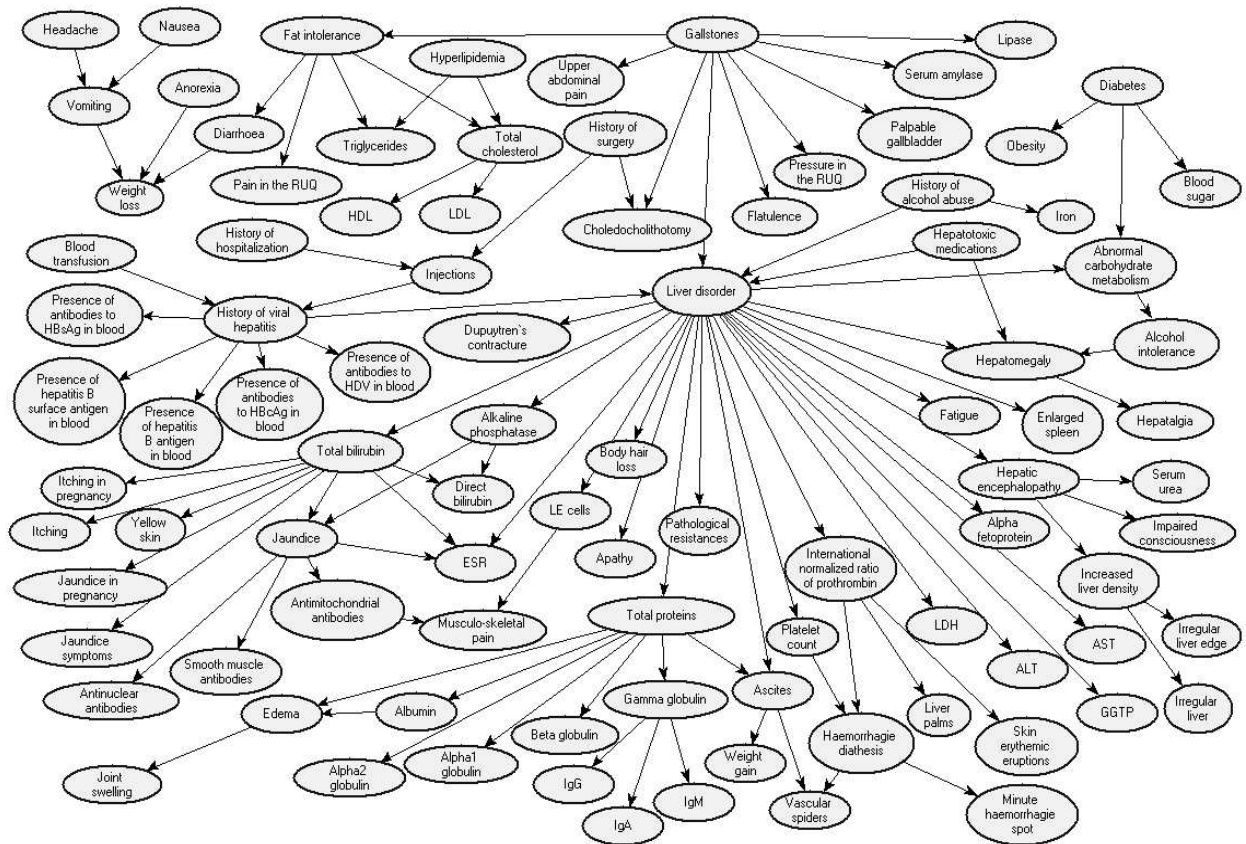
Figure 1: The structure of the model.

diagnostic systems. Quantification of a Bayesian network consists of prior probability distributions over those variables that have no predecessors in the network and conditional probability distributions over those variables that have predecessors. These probabilities can easily incorporate available statistics and, where no data are available, expert judgment. A probabilistic graph represents explicitly independencies among model variables and allows for representing a full joint probability distribution by a fraction of numbers that would be required if no independencies were known. Every independence leads to omitting an arc from the graph and leads to significant reductions of the numbers needed to fully quantify the domain. It should be stressed here that Bayesian networks are capable of representing any independencies, not only those assumed to exist in early Bayesian systems. In particular, a domain where no independencies exist will be represented correctly by a Bayesian network that is a complete graph. The most important type of reasoning in Bayesian networks is known as *belief updating*, and amounts to computing the probability distribution over variables of interest conditional on others, observed variables. In other words, the probability distribution over the model variables is adjusted for a particular case, in which some of the model variables assume given values.

While belief updating in Bayesian networks is in the worst case NP-hard,[3] there are several very efficient algorithms capable of updating beliefs in networks on the order of hundreds of variables within seconds (this depends strongly on the topology of the network - the sparser a network, the shorter it takes to update).

## DIAGNOSTIC MODEL

The starting point for building our model has been HEPAR's database of patient cases. The database available to us included about 600 patient records, each of these records was described by 119 features (binary, denoting presence or absence of a feature or continuous, expressing the value of a feature) and each record belonged to one of 16 liver disorders. One limitation of the HEPAR database is the assumption that a patient appearing in the clinic has at most one disorder, i.e., disorders are mutually exclusive. The features can be divided conceptually into three groups: symptoms and findings volunteered by the patient, objective evidence observed by the physician, and results of laboratory tests.

**Model Structure**
We elicited the structure of dependencies among the variables from our domain experts: Dr. Hanna Wasyluk (third author) from the Medical Center of Postgraduate Education, and two American experts, a pa-

thologist, Dr. Daniel Schwartz, and an epidemiologist, Dr. John N. Dowling, both at the University of Pittsburgh. We estimate that elicitation of the structure took about 40 hours with the experts, 30 hours with Dr. Wasyluk and 10 hours with Drs. Schwartz and Dowling. This includes model refinement sessions, where previously elicited structure was re-evaluated in a group setting. We started with an initial model comprising 40 variables of the highest diagnostic value (according to the expert)[7] and gradually extended it by adding variables one at a time.

Our current network is comprised of 94 variables: the disorder variable with 16 outcomes and 93 feature variables. The structure of our current model is shown above (Figure 1). We believe that it models reasonably causal interactions among the selected variables. We have also created a hierarchical version of the model, where groups of nodes were clustered into such submodels as *Hemotologic and Vascular Changes, Hepatocellular markers, Late stage diseases* or *Viral Hepatitis.*

**Model Parameters**

While the underlying formalism of Bayesian networks allows both discrete and continuous variables, all general purpose exact algorithms for Bayesian networks deal with models containing only discrete variables. In order to take advantage of these algorithms, we decided to discretize continuous variables. Our discretization is based on expert opinion that variables such as urea, bilirubin, or blood sugar have essentially *low*, *normal*, *high*, and *very high* values. The numerical boundaries of these intervals are based on expert judgment. Given a structure of the model, the specification of the desired discretization, and the HEPAR database, our program learns the parameters of the network, i.e., prior probabilities over all nodes without predecessors and conditional probabilities over all nodes with predecessors, conditional on these predecessors. 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. We would like to make two remarks here. The first is that the HEPAR database contains many missing measurements. We interpreted the missing measurements as possible values of the variables in question. This interpretation requires some care when using our system. We assume namely that the fact that a measurement was not taken is meaningful - the physician did not find taking the measurement appropriate. The meaning of the thus construed outcome *unmeasured* is in this way equivalent to a measured value of the variable. The second remark concerns

the accuracy of the learned parameters. While prior probabilities can be learned reasonably accurately from a database of hundreds records, conditional probabilities present more of a challenge. In cases where there are several variables directly preceding a variable in question, individual combinations of their values may be very unlikely to the point of being absent from the data file. In such cases, we assumed that the distribution is uniform. In all cases where the counts were zero, and naively interpreted would suggest a zero probability, we inserted a small probability reflecting the fact that almost no probabilities in the domain of medicine are zero or one. We found empirically that a value around 0.1 led to the best diagnostic performance. Generally, conditional probabilities learned from a data file of this size are not very reliable and need to be verified by an expert. There is much anecdotal and some empirical evidence that imprecision in probabilities has only small impact on the diagnostic accuracy of a system based on a Bayesian network.[6] This remains to be tested in our system.

**Diagnostic performance**

We tested the model in a variety of ways to verify its diagnostic value. Our first test involved testing the overall performance of the model in terms of classification accuracy (each of the diseases was viewed as a separate class that the program predicted based on the values of all the other variables). We applied a standard leave-one-out approach[5] (i.e., using repeatedly all but one record in the database to learn the parameters and then using the remaining record to test the prediction). We were interested in both (1) whether the most probable diagnosis indicated by the model 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). Results were approximately 34%, 47%, 56%, and 67% for $k$=1, 2, 3, and 4 respectively. In other words, the most likely diagnosis indicated by the model was the correct diagnosis in 34% of the cases. The correct diagnosis was among the four most probable diagnoses as indicated by the model in 67% of the cases. We consider this performance to be quite good given the difficulty of the problem, small size of the data set and many missing values. Please note that given 16 diseases, mean performance based on random guessing would barely exceed 6%.

Our second test focused on studying the relationship between the number of records in the database and the accuracy within a class. Some of the diseases has single records in the database, others have as many as a hundred of records. Figure 2 shows the relationship between the number of records for a particular dis-

ease and the system accuracy in diagnosing this disease. It is clear that accuracy increases significantly with the number of data records. In our model, diseases with more than 80 records present in the database showed high diagnostic accuracy. This raises high hopes for the diagnostic value of Bayesian network approach when the available data set is sufficiently large.
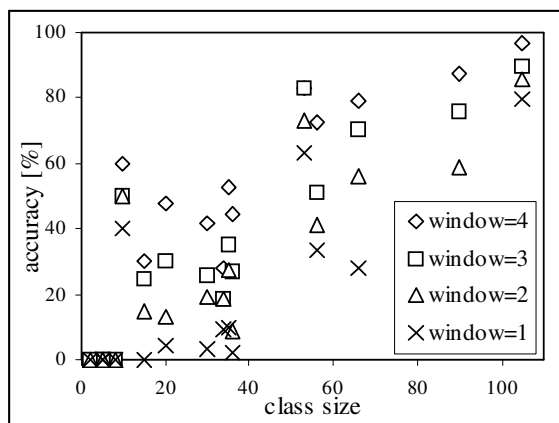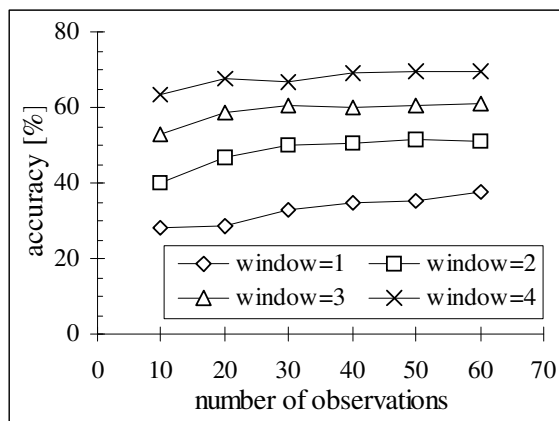


Figure 2: Influence of the number of records for a disease (class size) and accuracy in predicting this disease.

Figure 3: Influence of the number of observations on



the diagnostic accuracy.

Finally, we studied the diagnostic accuracy of our model when only a subset of the possible observations is entered. Bayesian networks are quite unique among other classification approaches in that they support classification based on any subset of the possible features. We started our test by entering a set of 10 randomly chosen findings and then added additional findings in batches of 10 until 60 findings were entered. The results of this test are presented on Figure 3. As expected, the diagnostic accuracy increases with the number of findings, although minimally. The additional gain of entering 50 more findings in addi-

tion to the first 10 is around 10% additional accuracy. This we found somewhat surprising. One explanation of this phenomenon is that our model is quite sparse and additional findings are often screened off from the disease node by the previous findings and have little additional diagnostic value.

We believe that pure diagnostic performance, in terms of the percentage of correct diagnoses, is in itself not an adequate measure of quality of a medical decision support system. In the domain of medicine, the physician user carries the ultimate responsibility for the patient and he or she will be unwilling to accept a system's advice without understanding it. While a causal model may perform worse in numerical terms than a regression-based model[*], it offers three important advantages: (1) its intuitive and meaningful graphical structure can be examined by the user, (2) the system can automatically generate explanations of its advice that will follow the model structure and will be reasonably understandable, and (3) the model can be easily enhanced with expert opinion; interactions absent from the database can be added based on knowledge of local causal interactions with the existing parts and can be parameterized by expert judgment.

## A DIAGNOSTIC TOOL BASED ON OUR MODEL

The probabilistic approach based on Bayesian networks allows to query the system with partial observations, something that is not natural for classification systems. To test how intuitive and how useful the model is in practice, we have built a simple user interface to our model (Figure 4). It lists all observable features we have been considered in our Bayesian network, 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. The left column contains a complete list of all variables included in the model. 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. Please note, that each variable has a possible value defined as *unknown*. We use this value to model the situation common in the data file that there is no information available about the value of that feature (a missing

---

[*] We would like to point out that this parallels the historical fact that Copernican theory of the structure of the Solar System, when introduced, predicted the positions of planets with less precision than the existing Sun-centered theory due to Ptolemy. Still, Copernican theory is preferable because it reflects better the physical structure of the system and explains the movements of planets with fewer free parameters.

value). *Not observed* denotes simply that the feature has not yet been observed. The right column presents an ordered list of the possible diagnoses along with their associated probabilities, the latter being presented graphically. The probabilities can be updated immediately after entering each finding (the default) or only after the *Update* button is pressed. Belief updating 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 instantaneous. Our colleague physicians have welcomed our program as a useful interactive diagnostic and training tool.
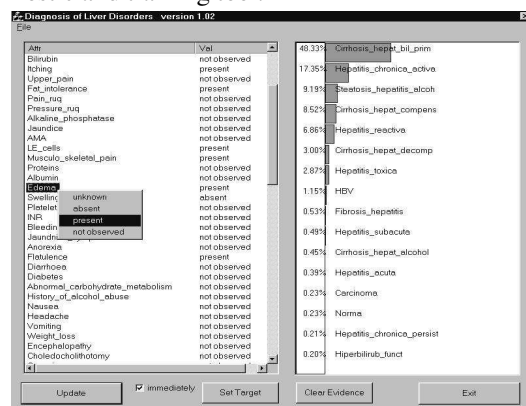


Figure 4: A dedicated interface to our model.

## CONCLUSIONS

We described our work on a probabilistic causal model for diagnosis of liver disorders. The model includes 16 liver disorders and 93 features, such as important symptoms, signs and risk factors. Given a patient's case, i.e., observation of values of any subset of the 93 features, the model computes the posterior probability distribution over the possible 16 liver disorders. This probability can be directly used in diagnostic decisions. We would like to remark that the model output, probability distribution over the possible disorders, is something that internists are used to and know how to interpret. Since our model follows reasonably the causal structure of the domain, and its output has a sound and unambiguous meaning, we hope that in addition to its value as a diagnostic aid, it will be useful in training beginning diagnosticians.

So far, the structure of the network was elicited from human experts and the numerical parameters were learned from a database of cases. One of our next steps will include combining expert judgment with the database to extract the numerical parameters of the network. In the long run, we plan to enhance our model with an explicit representation of diagnostic decisions and utilities of correct and incorrect diagnoses.

## References

1.  Bobrowski L. HEPAR: Computer system for diagnosis support and data analysis. Prace IBIB 31, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland, 1992.
2.  Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.
3.  Cooper GF. The computational complexity of probabilistic inference using Bayesian belief networks. AI, 1990:42(2-3):393-405.
4.  Henrion M, Breese JS, Horwitz EJ. Decision Analysis and Expert Systems. AI Magazine, Winter 1991: 12(4):64-91.
5.  Moore W, Lee MS. Efficient algorithms for minimizing cross validation error. In Proceedings of the 11th International Conference on Machine Learning, San Francisco, Morgan Kaufmann, 1994.
6.  Pradhan M, Henrion M, Provan G, B. Del Favero, Huang K. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. AI, 1996:85(1-2):363-397.
7.  Onisko A, Druzdzel MJ, Wasyluk H. A probabilistic causal model for diagnosis of liver disorders. In Proceedings of the 7th Symposium on Intelligent Information Systems, pages 379-387, Malbork, Poland, June 1998.
8.  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 Methods of Information in Medicine, 1991:30(4): 256-267.

9.  Heckerman D, Horvitz E, Nathwani B. Toward normative expert systems. 1: The PATHFINDER project. Methods of Information in Medicine, 1992:31:90-105.