

A Bayesian Network Model for Diagnosis of Liver Disorders

Agnieszka Onisko,^{1,2} Marek J. Druzdzel,¹ and Hanna Wasyluk³

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*

Abstract

This paper describes our work on a probabilistic causal model for diagnosis of liver disorders that we plan to apply in both clinical practice and medical training. The Bayesian network, and especially its numerical parameters, is based on data from a clinical database. We present the model and report the results of our diagnostic performance tests.

1. INTRODUCTION

Probabilistic graphical models, such as Bayesian networks [2] 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. 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 [6,7].

We describe our work on a probabilistic model for diagnosis of liver disorders. Our model is essentially a Bayesian network modeling causal relations among its variables as elicited from a domain expert and its numerical parameters extracted from a database. The 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.

2. DIAGNOSTIC MODEL

The starting point for building our model has been HEPAR's database of patient cases collected in the Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw. The data 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. 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.

2.1 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 pathologist, Dr. Daniel Schwartz, and an epidemiologist, Dr. John N. Dowling, both at the University of Pittsburgh. We estimate that elicitation of the structure

took approximately 40 hours with the experts, of which 30 hours spent with Dr. Wasyluk and 10 hours spent with Drs. Schwartz and Dowling. This includes model refinement sessions, where previously elicited structure was reevaluated in a group setting. We started with an initial model comprising 40 variables of the highest diagnostic value (according to the expert) [5] 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 in 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 submodels: *Elevated bilirubin*, *Gallstones*, *Hematologic and vascular changes*, *Hepatocellular markers*, *Late stage diseases* or *Viral hepatitis*, *Serum proteins abnormalities* and *Systemic illness*.

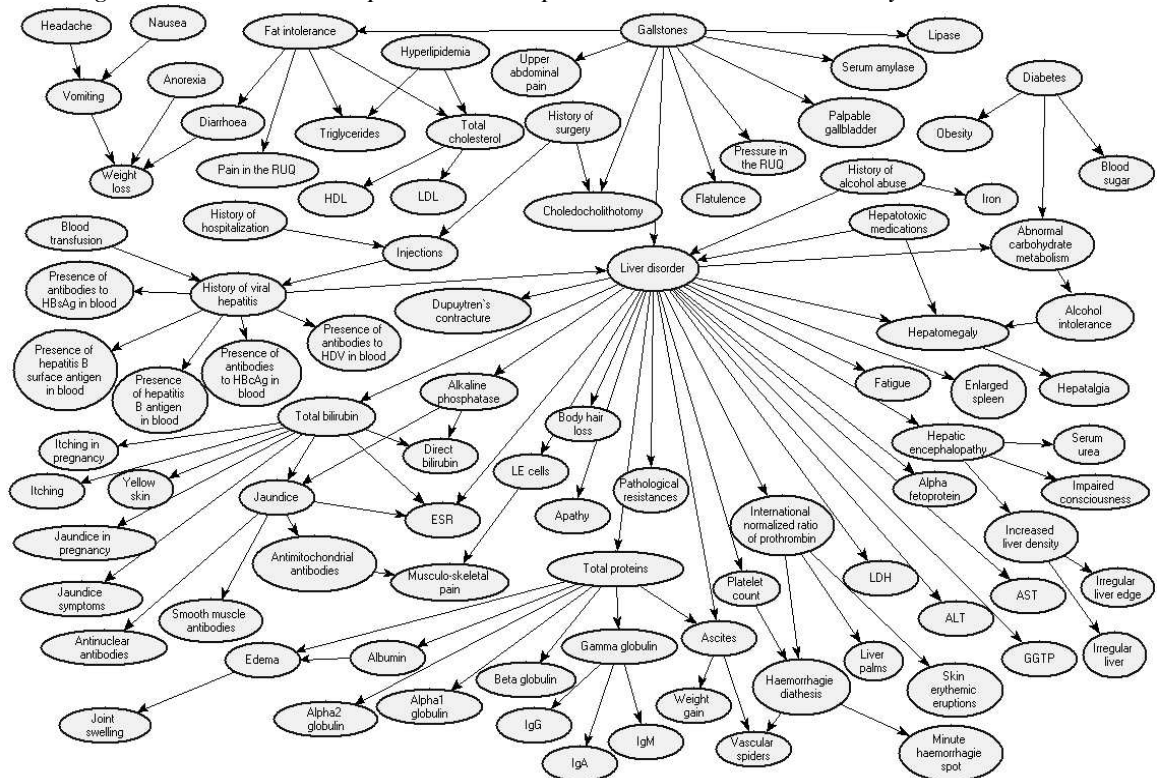


Figure 1. The structure of the model.

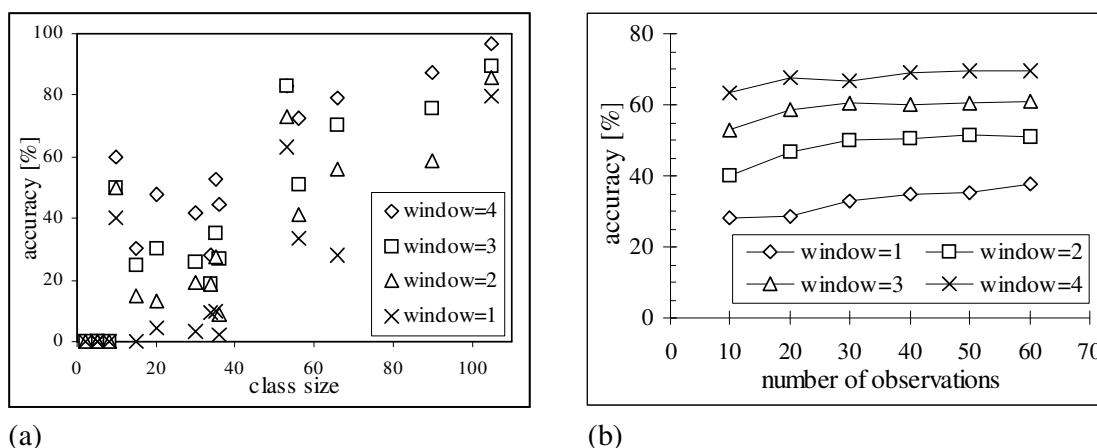
2.2 Model Parameters

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. While prior probabilities can be learned reasonably accurately from a database of consisting of a few hundred records, condi-

tional 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 [4]. This remains to be tested in our system.

2.3 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 the standard leave-one-out approach [3] (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



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

(b) Figure 2b: Influence of the number of observations on the diagnostic accuracy.

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 have single records in the database, others have as many as a hundred of records. Figure 2a shows the relationship between the number of records for a particular disease 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 very high diagnostic accuracy. This raises high hopes for the diagnostic value of Bayesian network approach when the available data set is sufficiently large.

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 2b. As expected, the diagnostic accuracy increases with the number of findings, although not rapidly. The additional gain of entering 50 more findings in addition 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 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.

3. 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 laboratory tests. 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 and explanation of the model.

Acknowledgments

This research was supported by the Air Force Office of Scientific Research, grant F49620-97-1-0225, by the Polish Committee for Scientific Research, grant 8T11E02917, by the Bialystok University of Technology, grant W/II/1/99. In addition Dr. Wasyluk was supported by the Medical Center of Postgraduate Education of Poland grant 501-2-1-02-14/98, and by the

Institute of Biocybernetics and Biomedical Engineering Polish Academy of Sciences, grant 16/ST/99. Dr. Druzdzal was also supported by the National Science Foundation under Faculty Early Career Development (CAREER) Program, grant IRI-9624629. We thank Drs. Schwartz and Dowling for their help in building and refining the structure of the model. Our model was created using **GeNIe**, a development environment for graphical probabilistic models developed at the Decision Systems Laboratory and available at <http://www2.sis.pitt.edu/~genie>.

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] 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.
- [4] Pradhan M, Henrion M, Provan G, B. Del Favero, Huang K. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence*, 1996;85(1-2):363-397.
- [5] Onisko A, Druzdzal 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.
- [6] 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.
- [7] Heckerman D, Horvitz E, Nathwani B. Toward normative expert systems. 1: The PATHFINDER project. *Methods of Information in Medicine*, 1992;31:90-105.