Preliminary Investigation of a Bayesian Network for Mammographic Diagnosis of Breast Cancer

Charles E. Kahn, Jr., M.D., Linda M. Roberts, M.S., Kun Wang, B.S., Deb Jenks, M.S.N., Peter Haddawy, Ph.D.

The Medical Informatics and Decision Science (MIDAS) Consortium, Milwaukee, Wisconsin*

Bayesian networks use the techniques of probability theory to reason under conditions of uncertainty. We investigated the use of Bayesian networks for radiological decision support. A Bayesian network for the interpretation of mammograms (MammoNet) was developed based on five patient-history features, two physical findings, and 15 mammographic features extracted by experienced radiologists. Conditional-probability data, such as sensitivity and specificity, were derived from peer-reviewed journal articles and from expert opinion. In testing with a set of 77 cases from a mammography atlas and a clinical teaching file, MammoNet performed well in distinguishing between benign and malignant lesions, and yielded a value of 0.881 (± 0.045) for the area under the receiver operating characteristic curve. We conclude that Bayesian networks provide a potentially useful tool for mammographic decision support.

INTRODUCTION

In 1995, an estimated 183,400 women in the United States will be newly diagnosed with breast cancer, and 46,240 will die of the disease [1]. Screening mammography effectively detects early breast cancers and can increase the likelihood of cure and long-term survival [2]. Differentiating between benign and malignant mammographic findings, however, is difficult. Only 15%-30% of biopsies performed on nonpalpable but mammographically suspicious lesions prove malignant [3]. Automated classification of mammographic findings using discriminant analysis and artificial neural networks has indicated the potential usefulness of computer-aided diagnosis [4,5].

We explored the use of Bayesian networks as a diagnostic decision aid in mammography. Bayesian networks—also called belief networks or causal probabilistic networks—use probability theory as a formalism for reasoning under conditions of uncertainty [6,7]. Bayesian networks can express the relationships between diagnoses, physical findings, laboratory test results, and imaging study findings. In radiology, Bayesian networks have been applied to the diagnosis of liver lesions on MR images [8] and to the selection of imaging procedures for patients with suspected gallbladder disease [9].

Bayesian networks are directed acyclic graphs: they are "directed" in that the connections between nodes are "one-way," and they are "acyclic" because they cannot include loops. Each node represents a variable and has two or more possible states. For example, the variable "Breast Cancer" has two states: "present" and "absent." For each node, the probability values associated with the states sum to 1. The connections between variables represent direct influences, expressed as conditional probabilities such as sensitivity and specificity.

METHODS

We created a Bayesian network model of breast cancer diagnosis, called MammoNet, that incorporates five patient-history features, two physical findings, and 15 mammographic findings. The model assumes that all of the evidence pertains to one particular site identified by mammography. MammoNet infers the posterior probability of breast cancer at that site based on the available evidence.

Four of the patient-history features influence the presence of breast cancer, which in turn influences the presence of the physical findings and mammographic findings (Figure 1). The mammographic findings are divided into direct manifestations of malignancy, such as mass or calcification, and indirect signs, such as architectural distortion. One of the patient-history factors, that of a prior biopsy, serves as a competing cause of the mammographic finding of architectural distortion.

*Section of Information and Decision Sciences, Department of Radiology, Medical College of Wisconsin, Milwaukee, Wisconsin 53226-0509; and Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201. E-MAIL: ckahn@mcw.edu

The nodes and their states are described in Table 1. Wherever possible, we used standardized terminology as proposed in the American College of Radiology's Breast Imaging Reporting and Data Systems (BIRADS) lexicon [10]. MammoNet's knowledge base was constructed from the peer-reviewed medical literature, census data, and health statistics reports. When required probability data were unavailable or the sample size too small, we obtained estimates from an expert mammographer.

Mammographically detectable mass and calcification are modeled as conditionally independent manifestations of malignancy. The Mass and Calcification nodes have three states: "malignant," "benign," and "none." If no mass is evident, for example, the Mass Present node is set to "no", which forces the Mass node to the state "none" and nodes such as Mass Margin to the state "not applicable" (NA). The Mass Present node allows one to express uncertainty regarding the presence of mass independently of the descriptive features. The mammographic features of a mass (Mass Margin, Mass Density, etc.)—although conditionally independent of Breast Cancer given Mass—affect the diagnosis by their influence on the Mass node's "malignant" and "benign" states. The model treats Calcification and its related nodes in similar fashion.

We used the Bayesian Network Generator (BNG) system [11,12] to generate a Bayesian network model from a set of probabilistic rules. Inference (i.e., calculation of posterior probabilities) was performed using the public-domain IDEAL system [13] on a DEC 5000/240 computer (Digital Equipment Corp., Maynard, MA).

To test MammoNet, we encoded 67 cases from a mammography atlas [14] and 10 cases from a clinical teaching file. Each case included clinical data, the mammographic findings, the expert mammographer's diagnosis, and the histological diagnosis based on clinical follow-up and/or biopsy results. The atlas provided a set of relatively straightforward cases; the clinical teaching file contained cases considered diagnostically challenging. Of the 77 cases, 25 were positive for breast cancer.

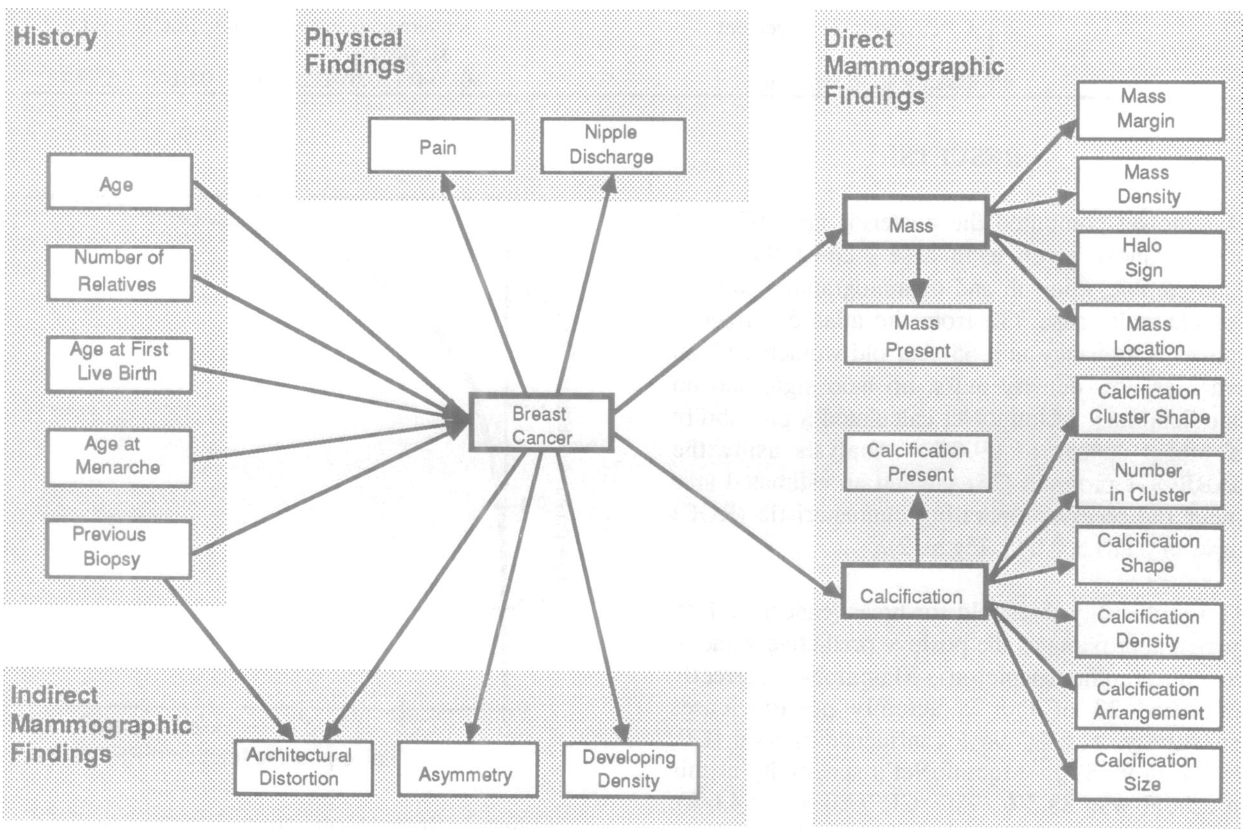

Figure 1. Topology of MammoNet's network model for mammographic diagnosis. The Breast Center, Mass, and Calcification nodes are inferred, and do not receive direct input from the user.

Table 1. Definitions of MammoNet's nodes and their states. NA = not applicable.

Category	*Node*	*States*
DIAGNOSIS	Breast Cancer	present, absent
HISTORY	Age (years)	20-24, 25-29, ..., 75-79
	Age at Menarche (years)	<12, 12-13, ≥14
	Age at First Live Birth (years)	<20, 20-24, 25-29, ≥30
	Number of First-Degree Relatives with Breast Cancer	0, 1, 2
	Previous Biopsy	yes, no
PHYSICAL FINDINGS	Pain	present, absent
	Nipple Discharge	present, absent
INDIRECT MAMMOGRAPHIC FINDINGS	Architectural Distortion	present, absent
	Asymmetry	present, absent
	Developing Density	present, absent
DIRECT MAMMOGRAPHIC FINDINGS	Mass	malignant, benign, none
	Mass Present	yes, no
	Mass Margin	spiculated, irregular, relatively well defined, NA
	Mass Density	high, low, NA
	Halo Sign	present, absent, NA
	Tumor Location	upper outer, upper inner, lower outer, lower inner, retroareolar, NA
	Calcification	malignant, benign, none
	Calcification Present	yes, no
	Calcification Cluster Shape	punctate, round, linear, variable, NA
	Number of Calcifications in Cluster	≤5, 6-10, 11-15, 16-25, 26-50, >50, NA
	Calcification Shape	linear branching, irregular, indeterminate, round, NA
	Calcification Density	1-2, 1-3, 2-3, 3-4, NA
	Calcification Arrangement	scattered, clustered, scattered&clustered, single, NA
	Calcification Size (mm)	0.05-0.1, 0.05-0.2, 0.01-1, 0.01-2, 1-3, NA

RESULTS

MammoNet computed the posterior probability of breast cancer given each case's constellation of demographic, clinical, and mammographic features. For example, case #32 from the atlas described a known malignancy in a 65-year-old woman with an irregular, low-density mass, no halo sign, and no calcifications; MammoNet calculated a probability of breast cancer of 99.6%. Analysis using the LABROC1 program [15] yielded an estimated area under the receiver operating characteristic (ROC) curve of 0.881 ± 0.045 (Figure 2).

At a probability threshold for breast cancer of 15% (which approximates the positive predictive value of mammographic suspicion), MammoNet correctly identified 23 of the 25 actually positive cases (sensitivity, 92.0%; 95% confidence interval [CI], 75.0% to 97.8%). MammoNet's specificity at this threshold was 88.5% (95% CI, 77.0% to 94.6%). Three benign lesions that MammoNet falsely identified as positive were considered suspicious by the mammographers and were referred for biopsy.

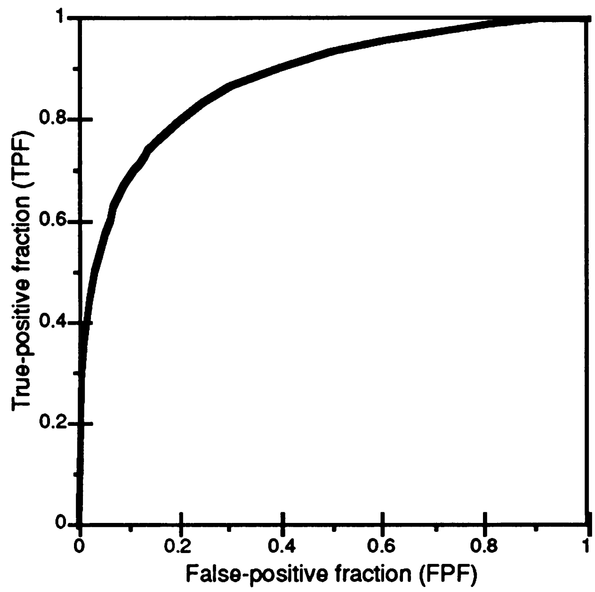

Figure 2. Receiver operating characteristic (ROC) curve for MammoNet.

DISCUSSION

Bayesian networks represent a promising technique for clinical decision support and provide a number of powerful capabilities for representing uncertain knowledge. They provide a flexible representation that allows one to specify dependence and independence of variables in a natural way through the network topology. Because dependencies are expressed qualitatively as links between nodes, one can structure the domain knowledge qualitatively before any numeric probabilities need be assigned. The graphical representation also makes explicit the structure of the domain model: a link indicates a causal relation or known association. The encoding of independencies in the network topology admits the design of efficient procedures for performing computations over the network. A further advantage of the graphical representation is the perspicuity of the resulting domain model. Finally, since Bayesian networks represent uncertainty using standard probability, one can collect the necessary data for the domain model by drawing directly on published statistical studies.

MammoNet's performance—as measured by its area under the ROC curve (A_Z)—compares very favorably with that of artificial neural network (ANN) models and expert mammographers. An ANN with 14 input features achieved an A_Z value of 0.89 (vs. 0.84 for attending radiologists) [5] on cases from the same mammography atlas as used in this study [14]. ANNs learn directly from observations, but cannot meaningfully explain their decisions. Their knowledge consists of an "impenetrable thicket" of numerical connection values. The ability of Bayesian networks to explain their reasoning [16,17] in an important advantage over ANNs; physicians generally will not accept and act on a computer system's advice without knowing the basis for the system's decision [18].

Another computer-assisted decision aid for mammographic interpretation included a checklist of 12 features determined to have particular diagnostic value [19]. Given a quantitative assessment of the 12 features, the decision aid estimated the probability of malignancy using weighting factors obtained from discriminant analysis. ROC analysis of this model showed a gain of about 0.05 in sensitivity or specificity when the other value remained constant at 0.85.

The ongoing refinement of MammoNet includes adding variables and states to the model, acquiring conditional-probability data from large case series, and rigorous testing and evaluation. We are converting the system to run using the Hugin inference system [20]. We are considering the addition of demographic features such as race and geographic location, and patient-history features such as diet, body habitus, history of hormone therapy, and previous cancers. The granularity of the model's variables could be increased by partitioning the Breast Cancer node into more than the current two states to represent the numerous types of cancer and benign conditions. We are developing links between MammoNet and a database to allow collection and analysis of a large set of clinical cases. Our goal is to create a decision support tool to improve the diagnostic accuracy and cost-effectiveness of screening mammography. Such a decision aid must be reliable, integrated with a clinical database and reporting system, and able to generate explanations to the physicians who use it.

ACKNOWLEDGMENTS

We thank Robert A. Krieger, M.S., and Katherine A. Shaffer, M.D. for many helpful discussions and Charles E. Metz, Ph.D., for the use of the LABROC software. This work was supported in part by grants from the National Science Foundation (IRI-9207262 to P.H.), Sun Microsystems (AEG award to P.H.), and the National Library of Medicine (USPHS G08 LM05705 for Integrated Advanced Information Management Systems [IAIMS] planning at the Medical College of Wisconsin).

References

[1] Wingo PA, Tong T, Bolden S. Cancer statistics, 1995. *CA* 1995; 45:8-30.

[2] Feig SA. Decreased cancer mortality through mammographic screening: results of clinical trials. *Radiology* 1988; 167:659-665.

[3] Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology* 1988; 167:353-358.

[4] Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of diagnostic images. *Invest Radiol* 1988; 23:240-252.

[5] Wu Y, Giger ML, Doi K, Vybomy CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* 1993; 187:81-87.

[6] Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers, 1988.

[7] Andreassen S, Jensen FV, Olesen KG. Medical expert systems based on causal probabilistic networks. *Int J Biomed Comput* 1991; 28:1-30.

[8] Tombropoulos R, Shiffman S, Davidson C. A decision aid for diagnosis of liver lesions on MRI. In: Safran C, ed. *Proceedings of the 17th Annual Symposium of Computer Applications in Medical Care*. New York: McGraw-Hill, 1993: 439-443.

[9] Haddawy P, Kahn CE Jr, Butarbutar M. A Bayesian network model for radiological diagnosis and procedure selection: work-up of suspected gallbladder disease. *Med Phys* 1994; 21:1185-1192.

[10] D'Orsi CJ, Kopans DB. Mammographic feature analysis. *Sem Roentgenol* 1993; 28:204-230.

[11] Haddawy P. Generating Bayesian networks from probability logic knowledge bases. In: *Uncertainty in Artificial Intelligence: Proceedings of the Tenth Conference*. San Mateo, CA: Morgan Kaufmann, 1994: 262-269.

[12] Haddawy P, Krieger RA. Principled construction of minimal Bayesian networks from probability logic knowledge bases. *J Artif Intel Res* (under review).

[13] Srinivas S, Breese J. IDEAL: a software package for analysis of influence diagrams. In: *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann, 1990: 212-219.

[14] Tabar L, Dean PB. *Teaching Atlas of Mammography*. (2 ed.) New York, NY: Thieme-Stratton, 1985.

[15] Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986; 21:720-733.

[16] Suermondt HJ, Cooper GF. An evaluation of explanations of probabilistic inference. *Comput Biomed Res* 1993; 26:242-254.

[17] Haddawy P, Jacobson J, Kahn CE Jr. Generating explanations and tutorial problems from Bayesian networks. In: Ozbolt JP, ed. *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*. Philadelphia: Hanley & Belfus, 1994: 770-774.

[18] Teach RL, Shortliffe EH. An analysis of physician attitudes regarding computer-based clinical consultation systems. *Comput Biomed Res* 1981; 14:542-558.

[19] D'Orsi CJ, Getty DJ, Swets JA, Pickett RM, Seltzer SE, McNeil BJ. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. *Radiology* 1992; 184:619-622.

[20] Andersen SK, Olesen KG, Jensen FV, Jensen F. HUGIN — a shell for building Bayesian belief universes for expert sytems. In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press, 1989: 1080-1085.