# A Bayesian Network for Mammography

Elizabeth Burnside MD, MPH[1], Daniel Rubin MD[1,2], Ross Shachter PhD[3]
Stanford Medical Informatics, Stanford University, Stanford, CA[1],
Center for Health Care Evaluation, VA Palo Alto Health Care System, Palo Alto, CA[2]
Management Science and Engineering Stanford University, Stanford CA[3]

*The interpretation of a mammogram and decisions based on it involve reasoning and management of uncertainty. The wide variation of training and practice among radiologists results in significant variability in screening performance with attendant cost and efficacy consequences. We have created a Bayesian belief network to integrate the findings on a mammogram, based on the standardized lexicon developed for mammography, the Breast Imaging Reporting And Data System (BI-RADS). Our goal in creating this network is to explore the probabilistic underpinnings of this lexicon as well as standardize mammographic decision-making to the level of expert knowledge.*

## INTRODUCTION

Breast cancer is the most common non-skin cancer affecting women in the U.S.[1] Early diagnosis through screening mammography is the most effective means of decreasing the death rate from this disease. At this time, approximately 61% of women over 50 have had a mammogram in the last 2 years.[2] Based on incidence and population estimates in the year 2000, this statistic translates into 20 million mammograms per year in the U.S.[3,4]

Radiologists oversee all mammograms. Some radiologists have subspecialty training in mammography and read these studies exclusively. These individuals are the experts in the field. Community radiologists read the majority of mammograms in the context of a diverse general practice. Community radiologists have higher biopsy rates and thus a lower positive predictive value of malignant disease. [5]

The American College of Radiology (ACR) is addressing this problem by working to standardize decision-making in mammography screening. For example, the ACR has developed a lexicon, BI-RADS, which standardizes mammogram feature distinctions and the terminology used to describe them.[6] BI-RADS arose in part from a study of the common terms used to describe mammography abnormalities. That study used a linear discriminant analysis to analyze these descriptors. The descriptors most highly associated with a benign or malignant diagnosis were considered the most predictive.[7] Subsequently, they were incorporated in the BI-RADS lexicon. Therefore, the BI-RADS lexicon represents a good foundation on which to build an expert system.

Although the mammography community does not yet uniformly use either the BI-RADS lexicon or formal probability calculations, radiologists have recently begun to use probabilistic information in one context. Sickles[8] showed that several mammographic scenarios have a small probability of malignancy. For example, a circumscribed mass, which is not a new finding as compared to prior films, has a 2% chance of malignancy. These types of abnormalities are considered "probably benign". The standard management of a "probably benign" finding is short-term follow-up (in 6 months) to reevaluate the area of concern. The ability to express findings in probabilistic terms has allowed management decisions less invasive than biopsy but more aggressive than routine follow-up to become the standard of care. Our hope is to facilitate increased management based on probabilistic interpretation.

## MODEL

Our first task in building our model was to construct a belief network. Another Bayesian network has been developed previously in this domain and we will examine the differences between it and our model.[9] From the literature, we identified 22 diseases of the breast. Ten of these diseases are malignant and ten are benign. Two of the diseases are pathologically benign but have a high association with or tendency to progress to malignancy. The standard of care for these two diseases is surgical excision. We consider these two diseases, papilloma and radial scar, to be "premalignant."(Table 1)

BI-RADS consists of 43 descriptors organized in a hierarchy. (Figure 1) We used 38 of the descriptors. We excluded five: skin thickening, trabecular thickening, nipple retraction, skin retraction, and

1067—5027/00/$5.00 

asymmetric breast tissue because they are rare, late, or non-contributory findings on screening mammography.

| *Malignant* | symbol | *Benign* | symbol |
|---|---|---|---|
| Invasive Ductal Carcinoma | DC-NOS | Cyst | Cy |
| Ductal Carcinoma in situ | DCIS | Fibroadenoma | FA |
| Lobular Carcinoma | LC | Fibrocystic Change | FC |
| Lobular Carcinoma in situ | LCIS | Hamartoma | Ham |
| Tubular Carcinoma | TubCA | Focal Fibrosis | FF |
| Papillary Carcinoma | PapCA | Fat Necrosis | FN |
| Medullary Carcinoma | MedCA | Secretory Disease | SecDis |
| Colloid Carcinoma | ColCA | Post-operative change | POC |
| Phylloides Tumor | Phy | Skin Lesion | SL |
| Metastasis | Mets | Lymph node | LN |
| *Radial Scar | RS | *Papilloma | Pap |

Table 1: Breast Diseases
*Radial Scar and Papilloma are considered premalignant

To construct our belief net and perform inference we used the GeNIe modeling environment developed by the Decision Systems Laboratory of the University of Pittsburgh (http://www.sis.pitt.edu/~dsl). We began construction of the global belief network assuming that all of the BI-RADS descriptors except breast density would be children of the disease node. (Figure 2) We modeled the calcification descriptors as conditionally independent manifestations of disease. The distribution, or spatial orientation, descriptors of each type of calcifications are the mutually exclusive states of the corresponding calcification nodes when appropriate. For example, punctate calcifications can be clustered, linear, segmental, regional, diffuse/scattered, or absent. We modeled special cases and associated findings as conditionally independent expressions of disease. The deterministic (double bordered) node in the belief network (Figure 2) has three states, "benign," "malignant" and "premalignant." The final decision to biopsy is based on the value of this node which is a deterministic function of the disease node.

The hierarchical structure of the BI-RADS lexicon elicits progressively more detailed descriptors for identified findings. For example, once a mass is identified, the user can describe the margins or shape of that mass. We incorporate the hierarchical structure of the lexicon into the belief network. This also models the conditional dependence among different mass-related findings.

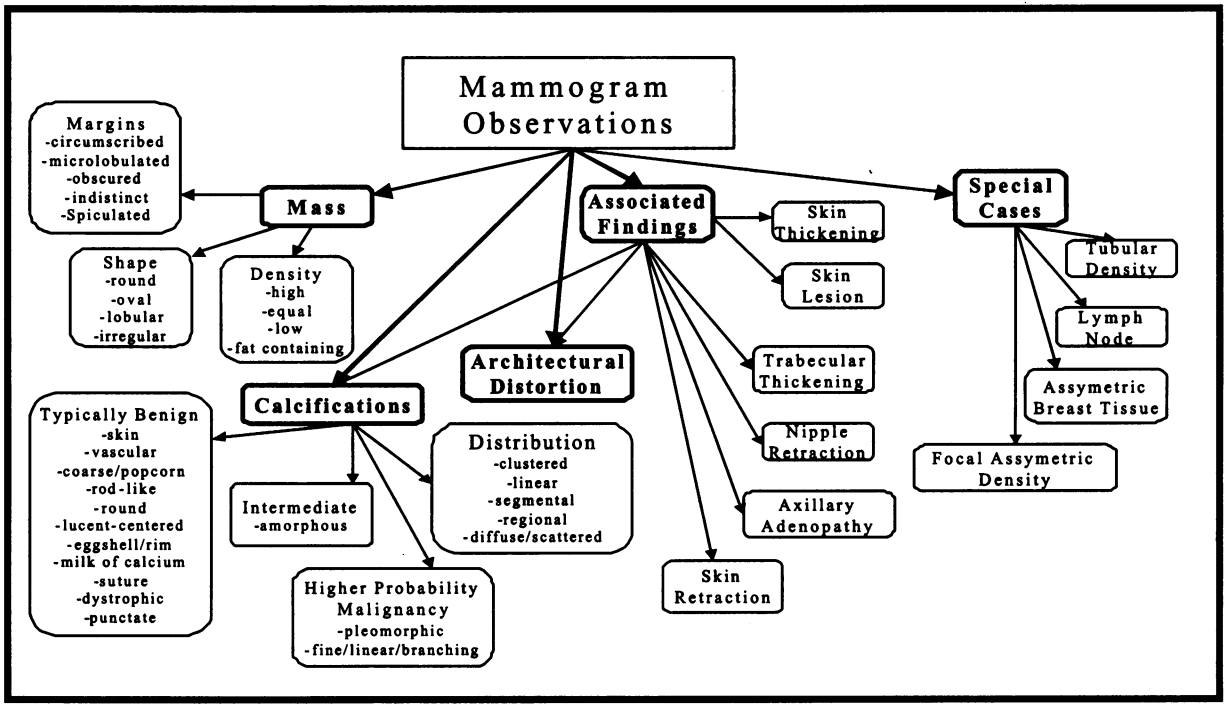


Figure 1: BI-RADS lexicon
(general categories are in bold)

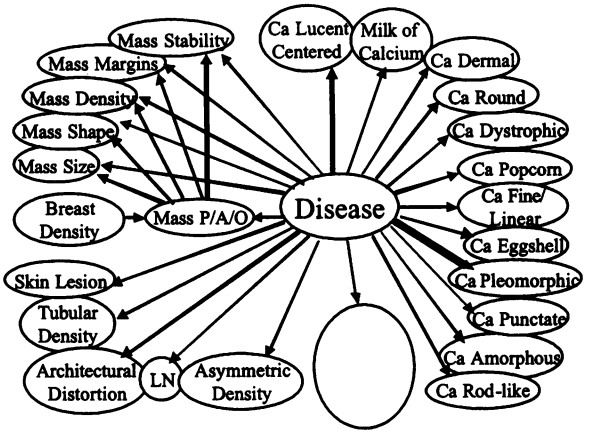


Figure 2: Bayesian Network
Ca=calcifications, LN=lymph node,
P/A/O=present, absent, obscured

The pathophysiology of breast disease presented a challenge in building the model. The transformation of benign cells to atypical cells to malignant cells challenges the mutual exclusivity assumption in our model. For example, the most common breast malignancy, ductal carcinoma not otherwise specified (DCNOS), is generally thought to develop from non-invasive but neoplastic cells termed ductal carcinoma *in situ* (DCIS). Though the rate of transformation is not well known, the causal relationship between these entities is accepted. We therefore represent these two diseases in our model as three mutually exclusive states in the disease node: DCIS, DCNOS, and DCNOS/DCIS. The third state represents a case in which DCNOS and DCIS are both present in the lesion seen on mammography. Our canonical case of DCNOS/DCIS is based on the pathology literature, which describes an entity termed "ductal carcinoma with a predominant intraductal component."[10] Similarly, lobular carcinoma and its noninvasive counterpart lobular carcinoma *in situ* (LCIS) exibit the same pathophysiology. In this way, we represent the spectrum of breast neoplasm in our model. (Table 2) We assume that no other violations of mutual exclusivity between diseases will exist in a single area of the breast. We consider these 24 diseases (and the "normal" state) conditionally exhaustive.

| ***Malignant (Mixed)*** | symbol |
|---|---|
| Ductal Carcinoma in situ/ Invasive Ductal Carcinoma | DCNOS/ DCIS |
| Lobular Carcinoma in situ/ Invasive Lobular Carcinoma | LC/ LCIS |

Table 2: Additional mixed diseases

We made probability assessments from the medical literature and expert opinion. We obtained pretest probabilities, the age specific and risk factor specific distributions of diseases from census data and large randomized trials. We derived many of the joint probabilities from studies of the radiologic/pathologic correlation of individual breast diseases.

## RESULTS

We have tested several standard cases to evaluate the behavior of the model. Table 3 shows the entire probability distribution for the following cases as well as the summation of benign, malignant, and premalignant diseases. Boldface type indicates the most likely diagnoses. No probability is truly zero but many are rounded to zero when we only display four decimal places.

**Case 1:** A 40 year old female with no family history or hormone use has a mammogram which demonstrates a spiculated mass with associated linear and branching calcifications. According to literature and expert opinion, a spiculated mass is typical for ductal carcinoma. The branching calcifications suggest an intraductal component. In this case our model generates the following probabilities: DCNOS/DCIS diagnosis in most likely with a 95% post-test probability. DCNOS alone has a post-test probability of 4.5%, and DCIS alone is unlikely. Variations of this scenario illustrate how the probabilities change as features differ.

**Case 2** A patient with similar demographic characteristics may have a spiculated mass without calcifications detected on her mammogram. This finding elicits an increased post-test probability of DCNOS to 88%. DCNOS/DCIS decreases to 2.9%, and again DCIS is unlikely.

**Case 3:** If the only finding, in a similar patient, is linear calcifications in a clustered distribution the post-test probability for DCIS increases to 70%. DCNOS/DCIS has a post-test probability of 8.9% and DCNOS alone is .001%. These probabilities are consistent with the pathophysiology of the disease as described above.

**Case 4:** A 50 year old patient has a mammogram which demonstrates a round, circumscribed mass. This is our example of a "probably benign" finding. Our system reveals that with these findings there is a 1.3% chance of malignancy in this setting. This is consistent with the radiology literature.[8]

| Disease | Pre-test | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|---|
| *DC-NOS* | 0.0090 | **0.0451** | **0.8851** | 0.0011 | 0.0047 |
| *DCIS* | 0.0019 | 0.0003 | 0.0000 | **0.7053** | 0.0000 |
| *DCNOS/ DCIS* | 0.0012 | **0.9502** | **0.0285** | **0.0892** | 0.0001 |
| *LC* | 0.0009 | 0.0000 | 0.0000 | 0.0007 | 0.0000 |
| *LCIS* | 0.0008 | 0.0000 | 0.0000 | 0.0007 | 0.0000 |
| *LC/ LCIS* | 0.0001 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| *TubCA* | 0.0001 | 0.0005 | 0.0093 | 0.0000 | 0.0000 |
| *PapCA* | 0.0002 | 0.0000 | 0.0009 | 0.0014 | 0.0004 |
| *MedCA* | 0.0001 | 0.0000 | 0.0002 | 0.0000 | 0.0040 |
| *CollCA* | 0.0001 | 0.0000 | 0.0002 | 0.0000 | 0.0040 |
| *Phy* | 0.0010 | 0.0000 | 0.0000 | 0.0000 | 0.0005 |
| *Mets* | 0.0010 | 0.0024 | 0.0474 | 0.0001 | 0.0001 |
| *RS* | 0.0010 | 0.0000 | 0.0000 | 0.0008 | 0.0000 |
| *Cy* | 0.0700 | 0.0000 | 0.0002 | 0.0122 | **0.7749** |
| *FA* | **0.1200** | 0.0014 | 0.0274 | 0.0218 | **0.1815** |
| *FC* | **0.1300** | 0.0000 | 0.0003 | 0.1098 | 0.0000 |
| *Ham* | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| *FF* | 0.0050 | 0.0000 | 0.0000 | 0.0014 | 0.0137 |
| *FN* | 0.0050 | 0.0000 | 0.0000 | 0.0042 | 0.0000 |
| *SecDis* | 0.0010 | 0.0000 | 0.0000 | 0.0008 | 0.0000 |
| *POC* | 0.0010 | 0.0000 | 0.0000 | 0.0008 | 0.0000 |
| *SL* | 0.0230 | 0.0000 | 0.0000 | 0.0181 | 0.0130 |
| *LN* | 0.0300 | 0.0000 | 0.0000 | 0.0253 | 0.0026 |
| *Pap* | 0.0020 | 0.0000 | 0.0002 | 0.0000 | 0.0004 |
| *Normal* | **0.5945** | 0.0000 | 0.0002 | 0.0051 | 0.0000 |
| *Benign* | **0.9708** | **0.0014** | **0.0282** | **0.1996** | **0.9864** |
| *Malignant* | **0.0262** | **0.9986** | **0.9716** | **0.7973** | **0.0133** |
| *PreMalignant* | **0.0030** | **0.0000** | **0.0002** | **0.0031** | **0.0004** |

Table 3: Differential diagnosis as well as summation into management categories with associated post-test probabilities for example cases

For a more systematic evaluation, we used a teaching atlas[11], which contains sufficient clinical information and mammographic descriptors to enter into our Bayesian network. This evaluation methodology is similar to that used in a Bayesian network developed in the same domain.[9] We did this order to facilitate the comparison of our modeling decisions to those of the previous model. We entered 105 cases from the mammography atlas into our model and constructed an ROC curve from the resulting posterior probabilities. To construct the ROC curve we used the ROCKIT 0.9B program (http://www-radiology.uchicago.edu/krl/toppage11.htm). This analysis generated an area under the curve (Az ± SD) of .953 ± .0409. (Figure 3)

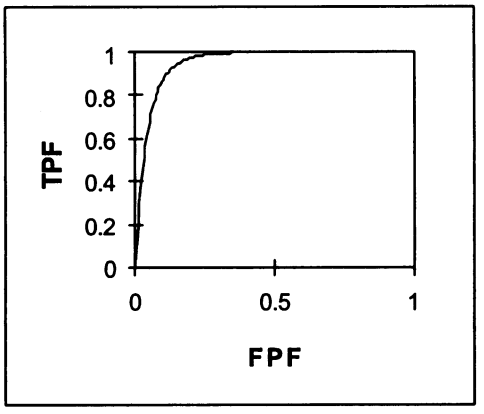


Figure 3: ROC analysis of 105 teaching cases

## DISCUSSION

Our Bayesian network uses a Web-based interface to elicit mammography findings. The structured entry system mandates the use of BI-RADS descriptors. Given mammography findings, our system provides post-test probabilities formulated as a differential diagnosis. In order to facilitate management decisions, it also provides the probabilities associated with benign, pre-malignant, and malignant disease.

A previous attempt to model the mammography screening process with a Bayesian network used a different approach.[9] Disease entities were not used in this model. The ultimate disease state, breast cancer (present or absent), also served as the management decision. Our more granular approach using the individual diseases commonly encountered in the breast helped us assess the conditional probabilities in the causal and pathophysiologic direction. We base our management decision on the summation of the disease probabilities. We believe that our more granular and causal approach will output more accurate post-test probabilities. Our preliminary results are promising.

The Az value of our model, .953, exceeds that of the earlier model, .881.[9] In fact, our system compares favorably with several computer diagnostic aids developed in the domain of screening mammography. A similar area under the ROC curve methodology has been used to evaluate many of these systems. We realize that the composition of the test set is important in the Az value and have included this parameter to facilitate comparison. Two neural nets (NN) have been developed to aid in the diagnosis of breast malignancy. One used BI-RADS descriptors and eight variables from the patient's medical history as inputs into the NN. Their test set contained 56%

malignant cases. The Az value of their ROC curve was .85.[12] The second NN used eight features of calcifications as inputs. The evaluation, done on 104 cases of which 46 were malignant (44% malignant), revealed an Az value of .76.[13] Another group used computer assisted diagnosis for masses. They used 253 mammograms (103 patients) of which 127 were malignant (50% malignant). The Az values were .92 and .96 for one or two views respectively.[14] Finally, the survey of US radiologists evaluating performance used a test set containing 79 cases of which 45 were malignant (56% malignant). The average Az value for these radiologists was .85.[15] Despite possible differences in these studies, we believe this comparison is encouraging. Our Bayesian approach appears to be powerful in isolation, and we might be able to enhance our diagnostic power by incorporating some of these other tools in our model.

The radiology community has only incorporated a small portion of the BI-RADS descriptors into the decision-making process in this field. The entire lexicon, when coupled with our Bayesian model, has great potential to communicate quantitative probabilistic information that will aid management decisions. Our model relates the benign and malignant breast diseases to BI-RADS descriptors and allows us to integrate radiological observations in a principled fashion. We hope that with further testing and use our model will help to elevate the standard of all mammography practice to the level of the expert.

We are looking forward to evaluating and refining our model in the future. We plan a formal evaluation comparing the system to experts as well as general practice radiologists. We also plan to evaluate the software design aspects of our web-based approach as we test the system with clinicians to gauge the usability of our interface. We will also incorporate value of information capabilities to guide the collection of observations and enhance the benefits of this system. In the future, we hope that radiologists of all levels will use this system not only as a decision tool but also as an educational aid for those interested in learning the BI-RADS lexicon from a probabilistic perspective. Our system has the potential to aid in both normative decision-making and education.

Acknowlegements: We thank Mark Peot for key suggestions. This work was funded by Grant LM 07033 from the National Library of Medicine.

1. Cancer Surveillance Section Annual Report: California Department of Health Services, 1999.
2. E. Pamuk, D. Makuc, K. Heck, C. Reuben and K. Lochner. Socioeconomic Status and Health Chartbook. Hyattsvill, Maryland: National Center for Health Statistics 1998.
3. Resident Population of the United States: Middle Series Projections, 1996-2000, by Age and Sex: U.S. Bureau of the Census, 1996.
4. SEER Cancer Statistics Review 1973-1996: National Cancer Institute, 1998.
5. ML Brown, F. Houn and E. Sickles. Screening mammography in community practice- positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. AJR 1995; 165:1373-1377.
6. Breast Imaging Reporting And Data System (BI-RADS). Reston VA: American College of Radiology, 1998.
7. J Swets, D. Getty, R. Pickett *et al.*. Enhancing and Evaluating Diagnostic Accuracy. Medical Decision Making 1991; 11:9-16.
8. EA Sickles. Management of probably benign breast lesions. Radiologic Clinics of North America 1995; 33:1123-30.
9. C. Kahn, L. Roberts, K. Wang, D. Jenks, P. Haddawy. Preliminary investigation of a Bayesian network for mammographic diagnosis of breast cancer, Proceedings of th 19th Annual Symp. on Computer Applications in Medical Care, Philadelphia, 1995. Hanley & Belfus.
10. C. Sewell. Pathology of benign and malignant breast disorders. In: V. Jackson, ed. Radiologic Clinics of North America. Vol. 33:6. Philadelphia: W.B. Saunders Co., 1995:1067-80.
11. L. Tabar and P. Dean. Teaching Atlas of Mammography. NY,NY: Thieme-Stratton, 1985.
12. J. Baker, P. Kornguth, J. Lo, M. Williford and C. Floyd. Breast cancer: prediction with artificial neural networks based on BI-RADS standardized lexicon. Radiology 1995; 196:817-822.
13. Y. Jiang, R. Nishikawa, R. Schmidt, C. *et al.* Improving breast cancer diagnosis with computer assisted diagnosis. Acad Radiol 1999; 6:22-30.
14. H. Chan, S. Merkman. *et al.* Improvement of radiologists' characterization of mammographic masses by using computer aided diagnosis: an ROC study. Radiology 1999; 212:817-27.
15. C. Beam, P. Ladge and D. Sullivan. Variability in the interpretation given to screening mammograms by U.S. radiologists: findings from a national sample. *Arch Int Med* 1999;156, 209-213.