Predicting the Risk of Invasive Versus In Situ Breast Cancer to Aid Clinical Management

Jagpreet Chhatwal, PhD; Oguzhan Alagoz, PhD; Houssam Nassif; Edward Sickles, MD; Christopher Jovais; Elizabeth S Burnside, MD, MPh, MS

University of Wisconsin
SCHOOL OF MEDICINE AND PUBLIC HEALTH
Department of Radiology

INTRODUCTION

- Since the adoption of mammography, the incidence of ductal carcinoma in situ (DCIS) has increased significantly, with the same predominance in older women as invasive breast cancer.
- DCIS is a non-invasive malignant condition with a very favorable prognosis. Depending on the grade of the DCIS and the expected life span of older women, DCIS often will not cause morbidity or mortality for many years, if ever.
- Invasive breast cancer has an increased risk of axillary node metastasis or distant disease, which will more quickly result in morbidity and mortality in older women.
- The ability to accurately predict the probability of DCIS versus invasive disease would enable older women and their referring physicians to make more informed decisions about managing their breast health in the context of their expected life span and co-morbidities.

We built a quantitative model using ***Bayesian Networks*** to predict the risk of DCIS and invasive breast cancer using patient demographic factors and mammography findings.

Data

To build our Mammography Bayesian Network for DCIS versus invasive cancer (**MBN*i***), we used 146,972 mammograms interpreted at the University of California San Francisco Medical Center between 1/6/1997 to 6/29/2007. Total 6,505 biopsies were performed between 1/6/1997 to 6/29/2007. We used a combination of structured data and natural language processing on dictated reports to extract variables including:

1. Patient demographic factors
2. Imaging features according to the standardized Breast Imaging Reporting and Data Systems (BI-RADS) lexicon.

After matching biopsies to the text reports, we were left with 5,926 biopsy records out of which 2,211 biopsy records turned out to be malignant with 1,544 biopsies diagnosed as having invasive cancer and 667 biopsies as DCIS. **We used 2,211 malignancies for our analysis.**

THE MBN*i*

We constructed our MBN*i* by learning the probabilities as well as the structure of the network from the available mammography data using the open source software WEKA (Waikato Environment for Knowledge Analysis). We trained the network using Tree Augmented Naive Bayes (TAN) algorithm.

We used 10-fold cross-validation to test the performance of our model in discriminating between DCIS and invasive cancer (Figure 1). We used receiver operating characteristic (ROC) curve and precision-recall (PR) curve to measure the performance of our MBN*i*.

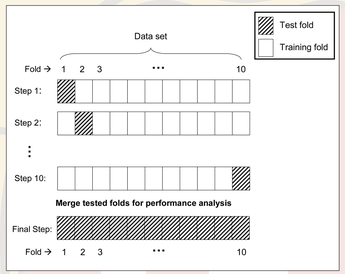


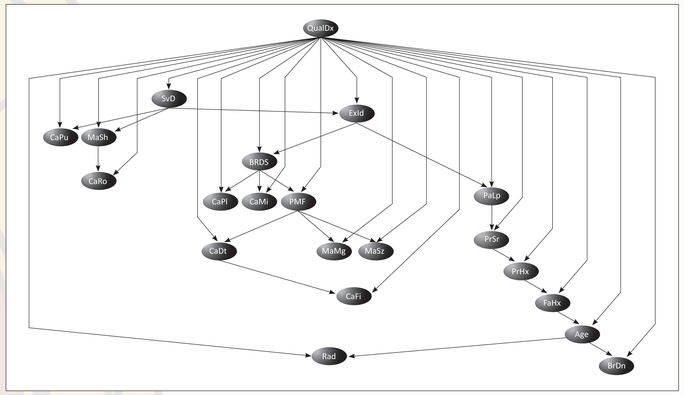
Figure 1
Ten-fold Cross-validation
Step-wise procedure for ten-fold cross-validation to train and test the model on independent dataset

Results

Figure 2 illustrates the structure of the MBN*i* showing dependency relationship between various mammography features and patient demographic factors. The structure demonstrates that all the demographic factors are clustered together showing less interdependence with the mammographic features. The following list defines the variables used in the MBN*i*.

1. Family history (FaHx)
2. Personal history (PrHx)
3. Prior surgery (PrSr)
4. Palpable lump (PaLp)
5. Screening versus diagnostic examination (SvD)
6. Indication for exam, if diagnostic examination (ExId)
7. Interpreting radiologist (Rad)
8. Breast density (BrDn)
9. If abnormal, principal mammographic finding (PMF)
10. Calcification distribution (CaDt)
11. Calcification pleomorphic (CaPl)
12. Calcification punctate (CaPu)
13. Calcification finelinear (CaFi)
14. Calcification round (CaRo)
15. Calcification milk (CaMi)
16. Mass shape (MaSh)
17. Mass margins (MaMg)
18. Mass size (MaSz)
19. BI-RADS codes (BRDS)

Figure 2
Structure of the MBN*i* using TAN algorithm



The outcome of the MBN*i* was the probability of invasive cancer. The area under the ROC curve was 0.853 (Figure 3) and PR curve was 0.928 (Figure 4), showing a high discriminative power of our model in predicting DCIS versus invasive cancer.

Figure 3
ROC Curve of MBN*i* to predict DCIS versus Invasive Cancer

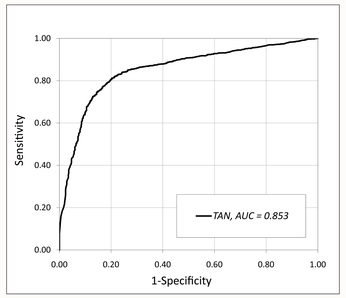
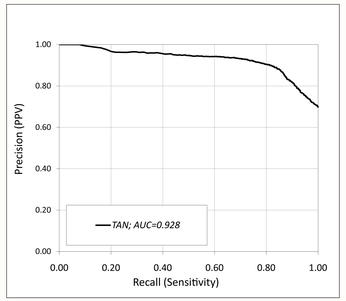


Figure 4
PR Curve of MBN*i* to predict DCIS versus Invasive Cancer



RISK PREDICTION IN OLDER VERSUS YOUNGER WOMEN

Mammography is known to perform better in older women. In addition, mammography performance has been shown to vary with the breast density with lower sensitivity in dense breasts as compared to non-dense breasts.

We stratified our data set in two parts as follows:

1. Mammography data of women less than age 50 (177 DCIS and 361 invasive cancers),
2. Mammography data of women above the age 65 (219 DCIS and 600 invasive cancers).

We trained and tested two different Bayesian networks ($BN_{\leq 50}$ and $BN_{\geq 65}$) using 10-fold cross-validation technique on these two data sets. Again, we trained the structure of the networks using TAN algorithm.

Results

The area under the ROC curve (Figure 5) of $BN_{\geq 65}$ (0.856) was significantly higher (P=0.039) than that of $BN_{\leq 50}$ (0.806). Similarly, area under the PR curve (Figure 6) of $BN_{\geq 65}$ (0.935) was significantly higher (P=0.038) than that of $BN_{\leq 50}$ (0.896).

Figure 5
ROC Curve of MBN*i* with Age Stratification

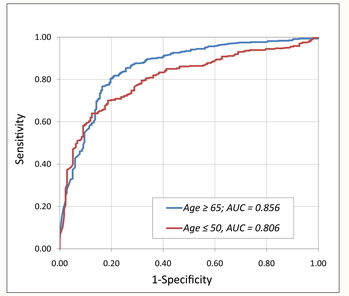
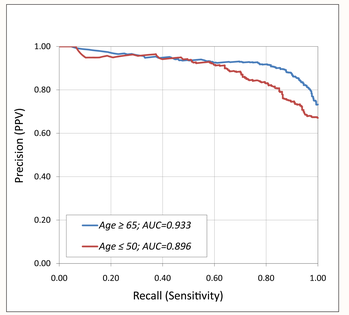


Figure 6
PR Curve of MBN*i* with Age Stratification



CONCLUSIONS

- Our MBN*i* model, which is constructed from the variables observed by radiologists during their daily clinical practice, successfully quantifies the risk of invasive versus in situ breast cancer.
- Our MBN*i* can predict the risk of DCIS versus invasive cancer and may be superior in older women because of the higher accuracy of detection of mammography findings in older women. The lower breast density in older age groups may also contribute to the improved performance in this population.
- Our MBN*i* has the potential to aid in the clinical management decisions such as the need for increased sampling at biopsy and the appropriate selection of surgical interventions.
- Our MBN*i* can accurately predict the presence and type of breast cancer which in the future may empower older women to better manage their breast health in the context of their co-morbidities and life expectancy.

Corresponding Author
Jagpreet Chhatwal, PhD
University of Wisconsin—Madison
1513 University Ave, Room #3233
Madison, WI 53705
Ph: 608.334.3292
Email: jchhatwal@wisc.edu