

RESEARCH ARTICLE

Assessment of Breast Cancer Risk in an Iranian Female Population Using Bayesian Networks with Varying Node Number

Abbas Rezaianzadeh¹, Mojtaba Sepandi^{2*}, Salar Rahimikazerooni¹

Abstract

Objective: As a source of information, medical data can feature hidden relationships. However, the high volume of datasets and complexity of decision-making in medicine introduce difficulties for analysis and interpretation and processing steps may be needed before the data can be used by clinicians in their work. This study focused on the use of Bayesian models with different numbers of nodes to aid clinicians in breast cancer risk estimation. **Methods:** Bayesian networks (BNs) with a retrospectively collected dataset including mammographic details, risk factor exposure, and clinical findings was assessed for prediction of the probability of breast cancer in individual patients. Area under the receiver-operating characteristic curve (AUC), accuracy, sensitivity, specificity, and positive and negative predictive values were used to evaluate discriminative performance. **Result:** A network incorporating selected features performed better (AUC = 0.94) than that incorporating all the features (AUC = 0.93). The results revealed no significant difference among 3 models regarding performance indices at the 5% significance level. **Conclusion:** BNs could effectively discriminate malignant from benign abnormalities and accurately predict the risk of breast cancer in individuals. Moreover, the overall performance of the 9-node BN was better, and due to the lower number of nodes it might be more readily be applied in clinical settings.

Keywords: Breast cancer- Bayesian networks- risk assessment

Asian Pac J Cancer Prev, 17 (11), 4913-4916

Introduction

Breast cancer is the leading type of cancer in females and is the fifth most common cause of cancer death (American Cancer Society, 2013). The burden of breast cancer in Iranian women is still increasing although it is still low compared with developing countries (Sharifian et al., 2015). Mammography is a traditional way to detect breast cancer (Gotsche and Nielsen, 2011). However, unfortunately, radiologists may report several interpretations for the same mammography. Fine Needle Aspiration Cytology (FNAC) is a widely used diagnostic tool for breast cancer, with correct classification rate of about 90%. On the other hand, medical data as a source of information may contain many records including valuable hidden relationships. However, the volume of the dataset and complexity of decision making in medicine make it difficult for clinicians to analyze and interpret the data and some processing steps may be needed before the data can be used by clinicians in their work. Hence, several algorithms have been developed to mine patterns through data. Recently, artificial intelligence and machine learning techniques have been widely applied in detection/recognition of breast cancer (Lo et al., 1999).

Distinguishing malignant breast lesions from benign ones and accurately predicting the risk of breast cancer for individual patients are essential for effective clinical decisions (Chan et al., 1999). Thus, mathematical models have been developed for breast cancer risk prediction. The predictive value of mathematical models in risk estimation varies according to age, menopausal status, race/ethnicity, and family history. Current risk prediction models estimate population, not individual, levels of breast cancer risk. Hence, individualized risk prediction models are needed to identify at-risk women who could benefit from timely risk reduction interventions (Jacobi et al., 2009). Up to now, considerable research has been devoted to development of computerized schemes for detection and classification of mammographic abnormalities. However, no perfect research has been conducted in Iran, as an Asian population. It is, therefore, necessary to evaluate a Bayesian Network (BN) model for detection and diagnosis of breast lesions. Clinical data collected as a part of breast cancer screening studies may be modeled using Bayesian classification. The present study aims to determine whether a BN trained on a retrospectively collected dataset including mammographic, risk factors, and clinical findings can accurately predict the probability

¹Colorectal Research Center, Shiraz University of Medical Sciences, Shiraz, ²Faculty of Health, Baqiyatallah University of Medical Sciences, Tehran, Iran. *For Correspondence: msepandi@gmail.com

of breast cancer for individual patients. This study focuses on the use of a computer-aided diagnosis model, which can quantify the risk of cancer using demographic variables and mammography features, to aid clinicians in breast cancer diagnosis and risk estimation.

Materials and Methods

All mammography observations were made by radiologists and all demographic and clinical factors were recorded by trained health workers. The clinical practice we studied routinely converts screening examinations to diagnostic mammography examinations when an abnormality is identified. In our BN, variables affecting the probability of disease are represented as “nodes”. Nodes are data structures that contain an enumeration of possible values they can assume (“states”). They also store probabilities associated with each state. In our system, the “disease” node has two states that represent benign and malignant masses. The structure of the model is also composed of directed arcs that encode the conditional dependence relationships among the variables. Absence of an arc represents conditional independence. Each arc implies a state of conditional dependence between the nodes joined by that arc. Parent-child relationships are defined by the direction of the arcs between the related nodes. A parent node points to a child node. Moreover, each of the finding nodes is associated with a probability table that quantifies the probability of each state of the node depending on the values of incoming nodes. BN is able to calculate a posttest probability of malignancy by using the structure of the model and the probabilities in the conditional probability tables. To implement our BN and perform inference, we used the Weka software. Our dataset consisted of 640 women (196 malignant and 459 benign) retrieved from 11850 screened cases referred to Shahid Motahhari breast clinic affiliated to Shiraz University of Medical Sciences between 2004 and 2012. Malignant and benign cases had been confirmed by pathologic reports. A network incorporating all the features and a hybrid combining the selected image and non-image features were created and compared. Overall, two BNs were built; one with 21 nodes (breast density, mass shape, mass margin, Micro calcification, dimpling, nipple retraction, mass size, asymmetric density, mass density, history of breast surgery, age group, menopause, marital status, body mass index, history of contraceptive use, family history of breast cancer, age at first pregnancy, occupation, education level, parity, and age at menarche) and one with 9 selected nodes. These 9 nodes (breast density, mass shape, mass margin, micro calcification, dimpling, nipple retraction, age, menopause, and marital status) were selected using multiple logistic regression analysis, as a feature selection process, to find the significant variables in the logistic regression model. Performance of the proposed method was evaluated based on cross-validation approach. In addition, BN was trained and tested using 10-fold cross-validation to predict the risk of breast cancer. The 10-fold cross-validation test is used in all experiments. The dataset is firstly divided into 10 subsets randomly and each time, one of the 10 subsets is used as the test set and the other

9 subsets are used in the training set. Every data point appears in the test set only once and in the training set for 9 times. Performance of the BN classification method is evaluated through sensitivity, specificity, and accuracy tests. Sensitivity, specificity, and accuracy are commonly used statistics, using True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) terms. TP is the number of true positives, which means that some cases with ‘positive’ class are correctly classified as positive. FP refers to the number of false positives, which means that some cases with ‘positive’ class are incorrectly classified as positive and should be in the negative class. TN is the number of true negatives, which means that some cases with ‘negative’ class are correctly classified as negative. Finally, FN refers to the number of false negatives, which means that some cases with ‘negative’ class should be classified as positive. In this study, area under the receiver-operating characteristic curve (AUC), accuracy, sensitivity, specificity, and positive and negative predictive values were used to evaluate discriminative performance of the models.

Results

The results indicated that the network incorporating the selected features performed better (AUC = 0.94) compared to that incorporating all the features (AUC = 0.93). The accuracy rate of the 9-node BN was also higher (0.89 and 0.88). Additionally, sensitivity and specificity of the 21-node BN were respectively calculated as 0.8 and 0.9 compared to 0.77 and 0.95 for the 9-node BN. Besides, negative and positive predictive values of the 21-node BN were respectively computed as 0.9 and 0.8 compared to 0.90 and 0.86 for the 9-node BN. Furthermore, performance of the two models was compared to that of the logistic model (Figure 1). The results revealed no significant difference among the 3 models regarding performance indices at 5% significance level.

We now illustrate the use of the 9-node BN model to estimate the probability of cancer using two cases:

Case 1- A 55-year-old, married, menopausal woman presented with a micro lobulated lobular mass, with micro calcification, without density, dimpling, and nipple retraction. BN estimated her probability of cancer to be equal to 0.99. Biopsy of this case was malignant.

Case 2- A 40-year-old single premenopausal woman had a mammogram that showed a circumscribed mass without density, dimpling, nipple retraction, and micro

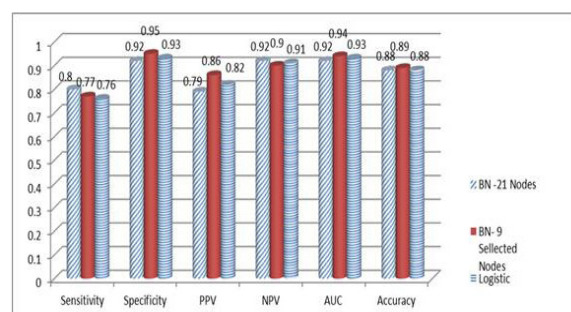


Figure 1. Performance Indices of BN and Logistic Models

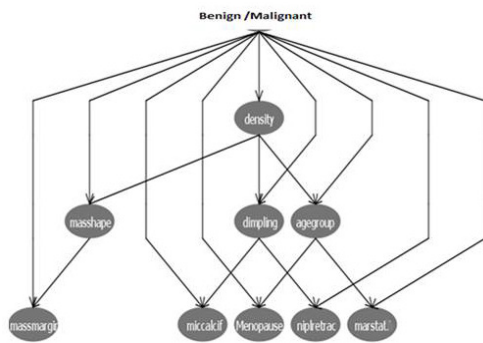


Figure 2. Bayesian Network with 9 Nodes Selected Using Multiple Logistic Regression Analysis

calcification. The probability of malignancy for this finding using BN was 13%. Biopsy of this case was benign.

Discussion

The breast cancer risk assessment may help in the clinical management of patient seeking advice concerning screening and prevention (Erbil et al., 2014). Our results reflect a single practice and must be viewed with some caution with respect to their generalizability because significant variability has been observed in the interpretive performance of screening and diagnostic mammography (Miglioretti et al., 2007; Taplin et al., 2008). In fact, we could not directly compare practice parameters to the literature because screening and diagnostic examinations could not be separated for this database. The model's performance may differ when built separately on screening and diagnostic mammograms. For screening mammograms, the incidence is low and descriptors are less exact because of general imaging protocols, which may result in less accurate model parameters. In contrast, for diagnostic mammograms, the model parameters may be more accurate because more descriptors can be observed as a result of additional specialized views. In addition, the performance of our existing model may differ when tested on screening and diagnostic mammograms separately. The model may perform better when tested on diagnostic examinations, but worse when tested on screening examinations (Sickles et al., 2002). Furthermore, our model differs from some risk prediction models by estimating the risk of cancer at a single time point (i.e., at the time of mammography) instead of over an interval in the future (e.g., over the next 5 years).

In the present study, two BN breast cancer risk estimation models were constructed based on 11850 screened cases referred to Shahid Motahhari breast clinic affiliated to Shiraz University of Medical Sciences between 2004 and 2012 to aid physicians in breast cancer diagnosis. The results demonstrated that the BN could perform as well as the logistic regression in estimating the probability of malignancy, and improve the positive predictive value (PPV) of the decision to perform biopsy. Using a computer-assisted detection program as a second reader has been shown to improve sensitivity in the screening setting (Freer and Ulisse, 2001; Warren

Burhenne et al., 2000). Several systems have also been used to help radiologists improve their decision to sample breast imaging findings for biopsy (Hadjiiski et al., 2004). However, given our present results, our BN had the potential to be used as a decision-support tool that could help underperforming practitioners improve the PPV of biopsy recommendations. Our model reinforced the previously known mammography predictors of breast cancer; i.e., irregular mass shape, speculated mass margins, micro calcifications, and breast density (Chhatwal et al., 2009; Liberman et al., 1998). Besides, the network incorporating the selected features performed better (AUC = 0.94) in comparison to that incorporating all the features (AUC = 0.93). Another study also developed linear discriminant analysis and artificial neural network models using a combination of mammographic and sonographic features and revealed an AUC of 0.92 (Jesneck et al., 2007).

The current study had some limitations. First, this study was a retrospective analysis using registered data about the risk factors and clinical data and existing mammographic reports of a group of patients selected because they had suspicious findings. This may limit the generalizability of the results and, therefore, prospective testing in a larger patient population is needed. Second, all the missing data were labeled as "not present" in our study. Our approach to handling missing data is appropriate for mammography data where radiologists often leave the descriptors blank if nothing is observed on the mammogram. Although complete data are ideally better, it is common for clinical datasets to contain a number of missing data. Third, Breast Imaging-Reporting and Data System (BI-RADS) categories were not incorporated in our study, because of lack of the technology in our center. Adding BI-RADS, as radiologist's integration of the imaging findings, can improve the model's performance and does augment predictions based on the observed mammographic features (Lo et al., 2002).

However, we are encouraged by our promising preliminary results. We demonstrated that our BN model could produce an accurate probability estimate of the risk of malignancy. By generating an accurate estimate of the posttest probability of malignancy, a BN for mammography may provide the opportunity for intuitive and collaborative decision making between patients and physicians in the future. Ultimately, we hope that, with further testing and use, probabilistic models will aid decision making in mammography practice.

The authors' both BNs could effectively discriminate malignant abnormalities from benign ones and accurately predict the risk of breast cancer for individual abnormalities. Moreover, the overall performance of the 9-node BN was better and due to the lower number of nodes, it can be applied more in clinical settings.

Funding Statement

The present manuscript was extracted from a thesis written by Mojtaba Sepandi and financially supported by Shiraz University of Medical Sciences.

Statement conflict of Interest

There are no potential conflict of interest relevant to this study

References

- American Cancer Society (2015). Breast Cancer Facts and Figures. Atlanta: American Cancer Society, Inc. 2015.
- Chan HP, Sahiner B, Helvie MA, et al (1999). Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An ROC study. *Radiology*, **212**, 817-27.
- Chhatwal J, Alagoz O, Lindstrom M J, et al (2009). A logistic regression model based on the national mammography database format to aid breast cancer diagnosis. *AJR Am J Roentgenol*, **192**, 1117-27.
- Erbil N, Dundar N, Inan C, Bolukbas N (2014). Breast cancer risk assessment using the Gail model: a Turkish study. *Asian Pac J Cancer prev*, **16**, 303-6.
- Freer TW, Ullissey M J (2001). Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center 1. *Radiology*, **220**, 781-86.
- Gøtzsche PC, Nielsen M (2011). Screening for breast cancer with mammography. The cochrane library.
- Hadjiiski L, Chan HP, Sahiner B, et al (2004). Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: An ROC study 1. *Radiology*, **233**, 255-65.
- Jacobi CE, de Bock GH, Siegerink B, et al (2009). Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose? *Breast Cancer Res Treat*, **115**, 381-90.
- Jesneck JL, Lo JY, Baker JA (2007). Breast mass lesions: computer-aided diagnosis models with mammographic and sonographic descriptors 1. *Radiology*, **244**, 390-98.
- Lieberman L, Abramson AF, Squires FB, et al (1998). The breast imaging reporting and data system: positive predictive value of mammographic features and final assessment categories. *AJR Am J Roentgenol*, **171**, 35-40.
- Lo JY, Baker JA, Kornguth PJ, Floyd CE (1999). Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks. *Academic Radiol*, **6**, 10-15.
- Lo JY, Markey MK, Baker JA, Floyd Jr, Carey E (2002). evaluation of BI-RADS predictive model for mammographic diagnosis of breast cancer. *AJR Am J Roentgenol*, **178**, 457-63.
- Miglioretti DL, Smith-Bindman R, Abraham L , et al (2007). Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst*, **99**, 1854-63.
- Sharifian A , Pourhoseingholi MA, Emadedin M, et al (2015). burden of breast cancer in Iranian women is increasing. *Asian Pac J Cancer Prev*, **16**, 5049-52.
- Sickles E A, Wolverton DE, Dee KE (2002). Performance parameters for screening and diagnostic mammography: specialist and general radiologists 1. *Radiology*, **224**, 861-69.
- Taplin S, Abraham L, Barlow WE, et al (2008). Mammography facility characteristics associated with interpretive accuracy of screening mammography. *J Natl Cancer Inst*, **100**, 876-887.
- Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al (2000). Potential contribution of computer-aided detection to the sensitivity of screening mammography 1. *Radiology*, **215**, 554-62.