

Improving Bayesian Networks Breast Mass Diagnosis by Using Clinical Data

Verónica Rodríguez-López^(✉) and Raúl Cruz-Barbosa

Universidad Tecnológica de la Mixteca,
69000 Huajuapán de León, Oaxaca, Mexico
{veromix,rcruz}@mixteco.utm.mx

Abstract. Nowadays, breast cancer is considered a significant health problem in Mexico. Mammogram is an effective study for early detecting signs of this disease. One of the most important findings in this study is a mass, which is the main indicator of malignancy. However, mass detection and diagnosis are difficult. In this study, the impact of the inclusion of seven clinical features on the performance of Bayesian Networks models for mass diagnosis is presented. Here, Naïve Bayes, Tree Augmented Naïve Bayes, K-dependence Bayesian classifier, and Forest Augmented Naïve Bayes models with eight image features nodes were augmented with several clinical features subsets. These models were trained with a data set extracted from the public BCDR-F01 database. The experimental results have shown that the Bayesian networks models augmented with a subset of three clinical features have improved their performance up to 0.82 in accuracy, 0.80 in sensitivity, and 0.83 in specificity. Therefore, these augmented models are considered as suitable and promising methods for mass classification.

Keywords: Breast cancer · Bayesian networks · Clinical data · Mass diagnosis

1 Introduction

Breast cancer is a disease in which malignant cells are formed in the breast tissues [17]. Nowadays, it is the most common cancer among women in the world. Although the highest rates of incidence for this disease are found in developed countries, the lowest survival rates occur in less developed countries [26]. As an example of this situation, Mexico is found at an intermediate level, with incidence rates four times lower than the highest ones, however, since 2006 breast cancer is the primary cause of death from malignant tumors among women [4].

A mammogram is an X-ray film of the breast that has shown to be an effective tool for early detection of breast cancer. One of the most important findings in this study is a mass, which is one of the main signs for detection of this disease. The American College of Radiology (ACR) defines a mass as a three-dimensional structure demonstrating convex outward borders, usually evident on two orthogonal views. These lesions are diagnosed from their characteristics of size,

shape, margins and density [1]. Margin refers to the mass border, and density is the amount of fat tissue in the mass compared with the surrounding breast tissue. Fatty masses with round shaped and well defined margins usually indicate benign cellular changes, while irregular shaped masses with high density and ill-defined margins, have a high probability of malignancy [8].

The diagnostic of masses is difficult, because they often are hidden by complex or similar breast tissue, which make them hard to distinguish from their characteristics. Moreover, the wide range of mass characteristics add more complexity to this, which becomes a more difficult, tedious and time-consuming task. For this, Computer Aided Diagnosis Systems (CADx) have been developed in order to help in the mammogram analysis. These systems have shown to improve the performance in mass diagnosis from 75 % to 82 % [6, 24, 25].

Bayesian networks are probabilistic models that have been applied on breast diagnosis. There are several reports about the performance of Bayesian networks for breast diagnosis when they are trained with clinical data and mammogram findings provided by radiologists [2, 10, 13]. However, few works have been discussed about their performance when they are trained from automatically obtained features [18, 25]. In this study, the performance comparison of Bayesian networks models for breast mass diagnosis is presented. The main aim of this paper is to analyze the impact of the inclusion of seven clinical features on the performance of Bayesian Networks models as Naïve Bayes, Tree Augmented Naïve Bayes, K-dependence Bayesian classifier, and Forest Augmented Naïve Bayes. These models are trained with automatically obtained image features and are augmented with several clinical features subsets.

The paper is organized as follows. In Sect. 2, we explain the image features used for mass description. A brief review of Bayesian networks classifiers is presented in Sect. 3. In Sect. 4, the data sets used in our experiments is described, and the experimental results are presented in Sect. 5. Finally, conclusions and future work are given in Sect. 6.

2 Image Features

In this study, eight image features extracted from mammogram images are used to characterize the shape, margins and density of masses. This set of image descriptors were selected as in [22].

The mass shape can be round, oval, lobular, or irregular [1]. Round and convex masses are probably benign, whereas the malignant ones have irregular shape [8]. The compactness feature is used to describe the mass shape.

The most important predictor of malignancy is the mass margin. A mass can appear with circumscribed (well-defined), microlobulated, obscured (by density of adjacent tissue), indistinct (ill-defined), or spiculated margins. Benign masses usually have very well defined margins, while the poorly defined are associated with malignant ones [8]. The mean, standard deviation and entropy of the normalized radial length (NRL) are calculated and used as margins descriptors. The NRL is defined as the normalized Euclidean distance from a point on the mass boundary to the mass centroid [9].

The mass density can be fatty, or also be lower, higher or isodense than the surrounding glandular breast tissue. Fatty or isodense masses have less probability of malignancy than the ones with higher density [8]. To describe this density, three intensity features and energy of gray-level co-occurrence matrices (GLCM) are used. From the intensity of pixels belonging to the mass region, three basic statistics are calculated: median, skewness, and kurtosis [6, 9]. The energy is obtained from a sample of 32×32 pixels of the mass center for the four directions $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, and $d = 1$. The range, in the four directions, is taken as density descriptor.

3 Bayesian Networks Classifiers

A Bayesian Network (BN) is a probabilistic graphical model where the nodes represent variables, and the arcs, dependence among variables. BN are recognized as powerful tools for knowledge representation and inference under conditions of uncertainty [7].

A formal definition of BN is as follows: A Bayesian network is a pair (D, P) , where D is a directed acyclic graph, and $P = \{p(x_1|\pi_1), p(x_2|\pi_2), \dots, p(x_n|\pi_n)\}$ is a set of n conditional probability distributions, one for each variable, and Π_i is the parent set of node X_i in D . The set P defines the associated joint probability distributions as [5],

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i|\pi_i) \quad (1)$$

Several types of BN models have been proposed as classifiers, some of them are: Naïve Bayes (NB), Tree Augmented Naïve Bayes (TAN), K-dependence Bayesian classifier (KDB), and Forest Augmented Naïve Bayes (FAN).

3.1 Naïve Bayes

Naïve Bayes model is the simplest form of a Bayesian network, in which the root node of a tree-like structure corresponds to a class variable. Also, the class node is the only parent for each attribute variable. The key assumption of a NB model is that all attributes are independent given the value of the class variable [7].

3.2 Tree Augmented Naïve Bayes

A TAN classifier is an extension of the Naïve-Bayes model, and also has a tree-like structure. In this model is allowed that each variable has at most two parents: the class variable, and other attribute. The Friedman method [11] can be applied to learn the structure of a TAN model.

3.3 K-Dependence Bayesian Classifier

A KDB classifier is a Bayesian Network where each attribute variable has at most k parents. This BN model is also considered as an extension of Naïve-Bayes, and its structure can be learned with the algorithm proposed by Sahami [23].

Table 1. Clinical features included in the BCDR-F01 database.

Feature	Description	Type
Age	The age of the patient at the time of the study	Nominal (range [28,82])
Breast density	The density of the breast at the time of the study according to the BI-RADS standard	Ordinal (range [1–4])
Mammography nodule	The lesion contains a mass	Binary
Mammography calcification	Calcifications were detected in the lesion	Binary
Mammography microcalcification	Microcalcifications were detected in the lesion	Binary
Mammography axillary adenopathy	Axillary adenopathy detected	Binary
Mammography architectural distortion	Signs of architectural distortion	Binary
Mammography stroma distortion	Signs of stroma distortion	Binary

3.4 Forest Augmented Naïve Bayes

This BN classifier is a variant of TAN where the attribute variables form a forest graph. One of the advantages of this classifier is that eliminates unnecessary relations among attributes. A method to learn the structure of FAN classifiers is proposed by Lucas [15].

4 Datasets

In this paper, a dataset extracted from the public BCDR-F01 (Film Mammography dataset number 1) database [20] was used. This database is the first public released dataset of the Breast Cancer Digital Repository (BCDR) which contains craniocaudal and mediolateral oblique mammograms of 190 patients. The mammograms were digitized with a resolution of 720×1167 pixels, using 256 grey levels. For each mammography, the coordinates for the lesion contours, and numerical anonymous identifiers for linking instances and lesions are provided. In addition to this, information including clinical and image-based descriptors of each lesion, is also available. The summary of the clinical descriptors available for this dataset is shown in Table 1.

In order to form the dataset used in our experiments, the 224 mammogram images with mass lesion that include all clinical descriptors were selected from the BCDR-F01 database. Next, the eight image features explained in Sect. 2 were extracted from each smallest bounding box containing a mass (region of interest-ROI, see Fig. 1). The ROIs were obtained with the help of ImageJ program [21]. In addition to the image features, for each mass the clinical information about the

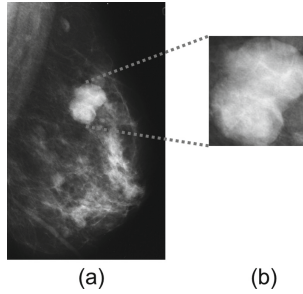


Fig. 1. An example of a BCDR image used in this study: (a) original image and (b) the corresponding ROI containing a mass.

age of patient, breast density, calcification, microcalcification, axillary adenopathy, architectural distortion, and stroma distortion were included. In summary, this dataset includes 224 mass cases: 112 benigns and 112 malignants where each mass is represented by eight image descriptors and seven clinical features.

5 Experiments and Results

The goal of the experiments is to analyze whether the Bayesian networks performance on mass diagnosis can be improved with a combination of image and clinical features. The performance of the BN models is evaluated with the Leave-one-out cross validation technique. The performance measurements used to report the results are accuracy, sensitivity, and specificity. Classification accuracy is the proportion of masses that are correctly classified by the model. The ratio of malignant masses that are correctly identified is the sensitivity; and specificity is the ratio of benign masses that are correctly identified [24]. All Bayesian networks models are trained and tested using the Matlab® software with help of the Bayes Net toolbox [16] and the BNT Structure Learning Package [14].

We start our experiments with four types of Bayesian networks models: Naïve Bayes, Tree Augmented Naïve Bayes, K-dependence Bayesian classifier (for $K = 2$), and Forest Augmented Naïve Bayes. All models were trained with the values of the eighth image features explained in Sect. 2, which were selected as in [22]. The corresponding network topology for these models are presented in Fig. 2.

5.1 Bayesian Networks Models Using the Complete Set of Clinical Features

In our first experiment, the impact of inclusion of the complete set of clinical features on the initial Bayesian networks models (shown in Fig. 2) is evaluated.

The seven clinical features were added to each Bayesian network model: NB, TAN, KDB, and FAN. Each clinical feature was added taking into account its

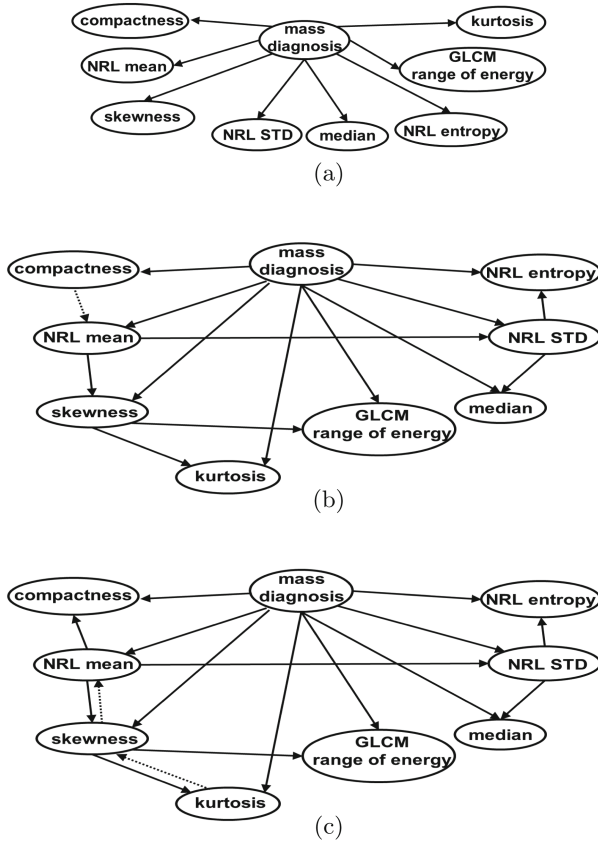


Fig. 2. Topology for the initial Bayesian network models trained with the image features: (a) NB, (b) TAN, (c) KDB and FAN, the former is represented with solid arc lines and the latter is the same structure except where the dotted arcs appear.

causal relationship with the diagnosis node or with the mass descriptor nodes. According to medical literature, the age of patient impact on breast diagnosis; breast density impact on shape, margins, and density of masses; presence of calcification, microcalcification, axillary adenopathy, architectural distortion, and stroma distortion, are associated with breast disease [3, 8].

The classification results for this experiment are presented in Table 2. From this table, it can be seen that the addition of all clinical features do not help to improve the previous performance on the initial BN models. Only, the extended NB model showed a significant improvement that can be explained by its simple topology. A decrease in sensitivity with a important improvement for the specificity are observed in the other models.

Table 2. Performance results of mass diagnosis for the Bayesian networks models using the complete set of clinical features.

Model	Accuracy		Sensitivity		Specificity	
	Image features	Clinical and image features	Image features	Clinical and image features	Image features	Clinical and image features
NB	0.73	0.78	0.65	0.69	0.81	0.88
TAN	0.79	0.78	0.80	0.70	0.77	0.85
KDB	0.79	0.78	0.80	0.70	0.77	0.85
FAN	0.79	0.78	0.80	0.70	0.77	0.85

5.2 Bayesian Networks Models Using Subsets of Clinical Features

In our second experiment, the impact of inclusion of different subset of clinical features on the initial Bayesian Networks models are evaluated.

In order to find the best subset of clinical features, the sequential forward selection method over all four types of Bayesian Network models is applied [19]. Following the same convention as the previous experiment, each clinical feature is added in the Bayesian network models taking into account its causal relationship with the other nodes. According to this procedure, the best subset is formed by the calcification, axillary adenopathy, and architectural distortion features. Other possible topologies for TAN, KDB and FAN models with this best combination of clinical and image features were searched by using structure learning algorithms, but was not observed an improvement on the performance results.

The performance results of the Bayesian networks models trained with the best combination of clinical and image features are summarized in Table 3. From this table, it can be seen that all models trained with the best combination of clinical and image features outperformed those obtained with only image features. For all models, this combination significantly improves the specificity. That is, the results show that the inclusion of clinical features helps to improve particularly the benign mass diagnosis of our dataset.

Table 3. Performance results of mass diagnosis for the Bayesian networks models using the best combination of clinical and image features.

Model	Accuracy		Sensitivity		Specificity	
	Image features	Clinical and image features	Image features	Clinical and image features	Image features	Clinical and image features
NB	0.73	0.76	0.65	0.67	0.81	0.85
TAN	0.79	0.82	0.80	0.80	0.77	0.83
KDB	0.79	0.82	0.80	0.80	0.77	0.83
FAN	0.79	0.82	0.80	0.80	0.77	0.83

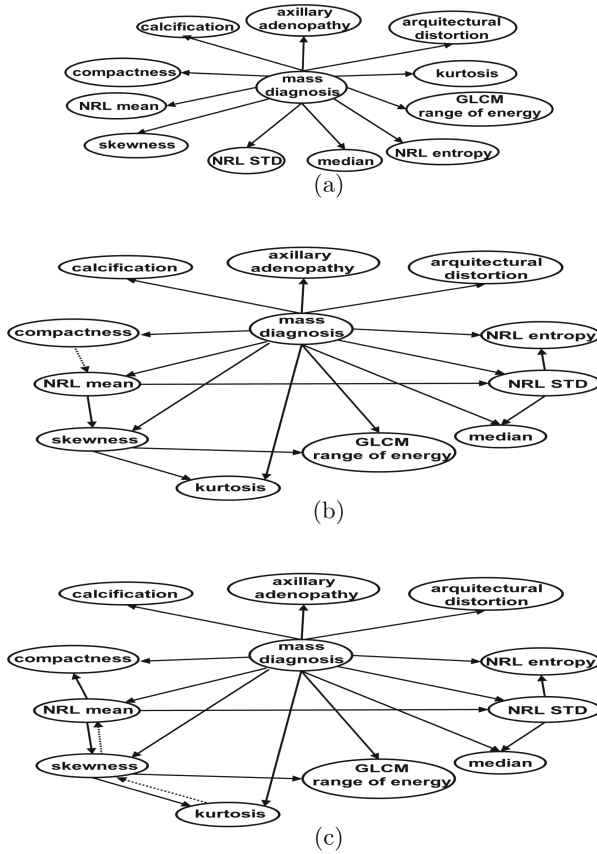


Fig. 3. Topology for the final Bayesian network models using the best combination of clinical features: (a) NB, (b) TAN, (c) KDB and FAN, the former is represented with solid arc lines and the latter is the same structure except where the dotted arcs appear.

Also from Table 3, it can be seen that the best obtained models are TAN, KDB, and FAN. These models, which include dependence among image features, show a good accuracy performance that is comparable with the average performance of radiologists [6] and with the average performance (0.82) of experimental CADx [24]. In summary, these results indicate that they are more suitable for identification of both mass type than the ones only using image features.

The topologies presented in Fig. 3 for TAN, KDB, and FAN using the best combination of clinical and image features, show a structure that is easy to interpret and validate by experts. They suggest that the clinical factors like the presence of calcification, axillary adenopathy, and architectural distortion; as well as, shape, margins, and density of masses (see Sect. 2 for image descriptors) are important factors for mass diagnosis. Moreover, the dependences among image features captured by the extended models, reveal that the shape of a mass has influence on margins, and this impact on the density attributes. These

findings, factors and their relationships, are consistent with both the analysis followed by experts and medical literature for mass diagnosis [8, 12].

6 Conclusions

The impact of a combination of clinical and image features on the performance of Bayesian Networks models for mass diagnosis was presented in this work. Seven clinical features about age of patient, breast density, calcification, microcalcification, axillary adenopathy, architectural distortion, and stroma distortion were analyzed. Several subsets of them were included on initial Bayesian Networks classifiers combined with eight image features nodes. The experimental results have shown that Bayesian networks models trained with a combination of three clinical features outperformed with 0.82 in accuracy, 0.80 in sensitivity, and 0.83 in specificity, those obtained with only image features. This improvement, particularly has more impact on the diagnosis of benign masses than on the malignant ones. Furthermore, the results have also shown that the augmented models: TAN, KDB, and FAN, have a structure that is easy to interpret and validate by experts. For this, it can be said that Bayesian networks with a combination of clinical and automatically obtained image features are promising models for mass classification.

As future work, interpretation and validation by experts of the best found models and the comparison of them with other type of classifiers are considered.

References

1. Balleyguier, C., Ayadi, S., Van Nguyen, K., Vanel, D., Dromain, C., Sigal, R.: BIRADSTM classification in mammography. *Eur. J. Radiol.* **61**(2), 192–194 (2007)
2. Burnside, E.S., Davis, J., Chhatwal, J., Alagoz, O., Lindstrom, M.J., Geller, B.M., Littenberg, B., Shaffer, K.A., Kahn Jr, C.E., Page, C.D.: Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings 1. *Radiology* **251**(3), 663–672 (2009)
3. Burnside, E.S., Rubin, D.L., Fine, J.P., Shachter, R.D., Sisney, G.A., Leung, W.K.: Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience 1. *Radiology* **240**(3), 666–673 (2006)
4. Cárdenas, J.S., Bargalló, E.R., Erazo, A.V., Maafs, E.M., Poitevin, A.C.: Consenso mexicano sobre diagnóstico y tratamiento del cáncer mamario: quinta revisión. Elsevier Masson Doyma México (2013)
5. Castillo, E.: Expert Systems and Probabilistic Network Models. Monographs in Computer Science. Springer-Verlag, New York (1997)
6. Cheng, H., Shi, X., Min, R., Hu, L., Cai, X., Du, H.: Approaches for automated detection and classification of masses in mammograms. *Pattern Recogn.* **39**(4), 646–668 (2006)
7. Cheng, J., Greiner, R.: Learning Bayesian belief network classifiers: algorithms and system. In: Stroulia, E., Matwin, S. (eds.) *Canadian AI 2001. LNCS (LNAI)*, vol. 2056, p. 141. Springer, Heidelberg (2001)
8. De Paredes, E.S.: Atlas of Mammography. Lippincott Williams & Wilkins, Philadelphia (2007)

9. Delogu, P., Evelina Fantacci, M., Kasae, P., Retico, A.: Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier. *Comput. Biol. Med.* **37**(10), 1479–1491 (2007)
10. Fischer, E., Lo, J., Markey, M.: Bayesian networks of BI-RADS descriptors for breast lesion classification. In: *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS 2004)*. vol. 2, pp. 3031–3034 (2004)
11. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. *Mach. Learn.* **29**(2–3), 131–163 (1997)
12. Jackson, V., Dines, K., Bassett, L., Gold, R., Reynolds, H.: Diagnostic importance of the radiographic density of noncalcified breast masses: analysis of 91 lesions. *Am. J. Roentgenol.* **157**(1), 25–28 (1991)
13. Kahn Jr, C.E., Roberts, L.M., Shaffer, K.A., Haddawy, P.: Construction of a bayesian network for mammographic diagnosis of breast cancer. *Comput. Biol. Med.* **27**(1), 19–29 (1997)
14. Leray, P., Francois, O.: BNTStructure Learning Package. Technical report FRE CNRS 2645, Laboratoire PSI-INSA Rouen-FRE CNRS, France (2004)
15. Lucas, P.J.S.: Restricted bayesian network structure learning. In: Gamez, J.A., Moral, S., Salmeron, A. (eds.) *Advances in Bayesian Networks*. STUDEFUZZ, vol. 146, pp. 217–234. Springer, Heidelberg (2004)
16. Murphy, K.: How to use Bayes net toolbox (2004). <http://www.ai.mit.edu/murphyk/Software/BNT/bnt.html>
17. National Cancer Institute, NCI: General information about breast cancer. <http://www.cancer.gov/cancertopics/pdq/treatment/breast/Patient/page1>. Accessed Jan 2015
18. Patrocínio, A.C., Schiabel, H., Romero, R.A.: Evaluation of Bayesian network to classify clustered microcalcifications. In: *Proceedings of the SPIE Medical Imaging 2004*. pp. 1026–1033. International Society for Optics and Photonics (2004)
19. Pazzani, M.: Searching for dependencies in Bayesian classifiers. In: Fisher, D., Lenz, H.-J. (eds.) *Learning from Data*. LNS, pp. 239–248. Springer, New York (1996)
20. Ramos-Pollán, R., Guevara-López, M.A., Suárez-Ortega, C., Díaz-Herrero, G., Franco-Valiente, J.M., Rubio-del Solar, M., González-de Posada, N., Vaz, M.A.P., Loureiro, J., Ramos, I.: Discovering mammography-based machine learning classifiers for breast cancer diagnosis. *J. Med. Syst.* **36**(4), 2259–2269 (2012)
21. Rasband, W.: ImageJ: Image processing and analysis in Java. *Astrophysics Source Code Library* (2012)
22. Rodríguez-López, V., Cruz-Barbosa, R.: On the breast mass diagnosis using Bayesian networks. In: Gelbukh, A., Espinoza, F., Galicia-Haro, S. (eds.) *MICAI 2014, Part II*. LNCS, vol. 8857, pp. 474–485. Springer, Heidelberg (2014)
23. Sahami, M.: Learning limited dependence Bayesian classifiers. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. vol. 96, pp. 335–338. AAAI Press (1996)
24. Sampat, M., Markey, M., Bovik, A.: Computer-aided detection and diagnosis in mammography. In: *Handbook of Image and Video Processing*. ch. 10.4, pp. 1195–1217. Elsevier Academic Press (2005)
25. Velikova, M., Lucas, P.J., Samulski, M., Karssemeijer, N.: A probabilistic framework for image information fusion with an application to mammographic analysis. *Med. Image Anal.* **16**(4), 865–875 (2012)
26. World Health Organization: Breast cancer prevention and control. <http://www.who.int/cancer/detection/breastcancer/en/index1.html>. Accessed Jan 2015