Comparison of naive Bayes and logistic regression for computer-aided diagnosis of breast masses using ultrasound imaging

Cary, Theodore, Cwanger, Alyssa, Venkatesh, Santosh, Conant, Emily, Sehgal, Chandra

Theodore W. Cary, Alyssa Cwanger, Santosh S. Venkatesh, Emily F. Conant, Chandra M. Sehgal, "Comparison of naive Bayes and logistic regression for computer-aided diagnosis of breast masses using ultrasound imaging," Proc. SPIE 8320, Medical Imaging 2012: Ultrasonic Imaging, Tomography, and Therapy, 83200M (24 February 2012); doi: 10.1117/12.911916

SPIE.

Event: SPIE Medical Imaging, 2012, San Diego, California, United States

Comparison of Naïve Bayes and logistic regression for computer-aided diagnosis of breast masses using ultrasound imaging

Theodore W. Cary [a], Alyssa Cwanger [a], Santosh S. Venkatesh [b], Emily F. Conant [a], Chandra M. Sehgal [a]
[a]Department of Radiology, 3400 Spruce Street, University of Pennsylvania, Philadelphia PA, USA 19104; [b]Department of Electrical and Systems Engineering, 200 south 33rd Street, University of Pennsylvania, Philadelphia, PA USA 19104

ABSTRACT

This study compares the performance of two proven but very different machine learners, Naïve Bayes and logistic regression, for differentiating malignant and benign breast masses using ultrasound imaging.

Ultrasound images of 266 masses were analyzed quantitatively for shape, echogenicity, margin characteristics, and texture features. These features along with patient age, race, and mammographic BI-RADS category were used to train Naïve Bayes and logistic regression classifiers to diagnose lesions as malignant or benign. ROC analysis was performed using all of the features and using only a subset that maximized information gain. Performance was determined by the area under the ROC curve, Az, obtained from leave-one-out cross validation.

Naïve Bayes showed significant variation (Az 0.733 ± 0.035 to 0.840 ± 0.029, $P < 0.002$) with the choice of features, but the performance of logistic regression was relatively unchanged under feature selection (Az 0.839 ± 0.029 to 0.859 ± 0.028, $P = 0.605$). Out of 34 features, a subset of 6 gave the highest information gain: brightness difference, margin sharpness, depth-to-width, mammographic BI-RADs, age, and race. The probabilities of malignancy determined by Naïve Bayes and logistic regression after feature selection showed significant correlation ($R^2 = 0.87$, $P < 0.0001$).

The diagnostic performance of Naïve Bayes and logistic regression can be comparable, but logistic regression is more robust. Since probability of malignancy cannot be measured directly, high correlation between the probabilities derived from two basic but dissimilar models increases confidence in the predictive power of machine learning models for characterizing solid breast masses on ultrasound.

Keywords: breast ultrasound, computer-aided diagnosis, ultrasound image analysis, logistic regression, Naïve Bayes, breast cancer; quantitative diagnostic ultrasound.

1. INTRODUCTION

Breast cancer is the second leading cause of cancer death after lung cancer in women. Approximately one of every eight women in the United States and one of every ten women in Europe will develop invasive breast cancer. In 2010, an estimated 200,000 new cases of invasive breast cancer were diagnosed and approximately 40,000 women died from this disease [1]. Mammography and sonography are the two imaging methods most commonly used to detect and evaluate solid breast lesions. Mammography is currently accepted as the most effective imaging technique for breast cancer detection. This imaging modality alone, however, is not sufficient for the complete differentiation of benign and malignant breast masses. A biopsy, which consists of minor surgery, is necessary for a final diagnosis. There are many false positive screenings of benign mammographic images, which result in a low biopsy yield (21% to 34%) [2]. Although sonography can reliably identify some simple benign cases, it is not sufficient to distinguish benign from malignant solid breast masses. Therefore, a complete diagnostic work-up including ultrasonography is often recommended to improve the positive predictive value of the biopsy yield, and computer-based analysis of ultrasound images continues to be an area of active research.

Medical Imaging 2012: Ultrasonic Imaging, Tomography, and Therapy, edited by Johan G. Bosch, Marvin M. Doyley, Proc. of SPIE Vol. 8320, 83200M · © 2012 SPIE
CCC code: 1605-7422/12/$18 · doi: 10.1117/12.911916

Several quantitative approaches have been proposed to improve the accuracy of classification schemes [3-16]. Although the results on computer models used to classify masses are encouraging, the analysis is often data specific, depending on derived image features, so the relative performance of classification schemes is difficult to assess. The motivation of this study is to evaluate two classification approaches on the same dataset using the same image features. We compare the performance of two probability estimating classifiers, Naïve Bayes and logistic regression, for the diagnosis of breast masses on grayscale ultrasound images. Although the learning strategies are fundamentally different, both methods naturally estimate probabilities of malignancy under the assumption that individual mass attributes, or features, are conditionally independent. The resulting models are easy to understand and to compare, and the instance probabilities can be ranked and correlated case by case for analysis.

2. METHODS

2.1 Classification schemes

Naïve Bayes

Consider binary classification of breast masses as malignant (M) or benign (B). According to Bayes' theorem, the conditional probability $P(M|F_1, \ldots F_n)$ that a mass is malignant given observed features $(F_1, F_{21}, \ldots . F_{n1})$ is equal to the normalized product of the likelihood of the features in the malignant masses $P(F_1, \ldots F_n|M)$ and the probability of the mass being malignant prior to observing any features, $P(M)$:

$$P(M|F_1, \ldots F_n) = \frac{P(F_1,\ldots F_n|M)\, P(M)}{P(F_1,\ldots F_n)}, \tag{1}$$

where the normalizing factor $P(F_1, \ldots F_n)$ is the probability of all the evidence. Under the approximation that each feature is conditionally independent, Equation 1 simplifies to Naïve Bayes:

$$P(M|F_1, \ldots F_n) = \frac{P(M)\prod_{i=1}^{n} P(F_i|M)}{\prod_{i=1}^{n} P(F_i|M) + \prod_{i=1}^{n} P(F_i|B)}\,. \tag{2}$$

Continuous feature distributions $P(F_i|M)$ and $P(F_i|B)$ were determined assuming the features are normally distributed. For nominal features the probabilities were determined by counting the number of cases with features that were malignant or benign.

Logistic Regression

In logistic regression the class $Y = \{0,1\} = \{B, M\}$ is the linear combination of input features $(F_1, F_2, \ldots . F_n)$:

$$Y = \alpha + \beta_1 F_1 + \beta_2 F_2 + \ \ldots . \beta_n F_n = \ \alpha + \sum_{i=1}^{n} \beta_i F_i. \tag{3}$$

The coefficients α and β_i were determined by fitting the training data to Equation 3 using the maximum likelihood function. The estimated α and β_i were used to determine the class conditional probabilities:

$$P(M|F_1, \ldots F_n) = \frac{1}{1+exp\left[-\left(\alpha+\sum_{i=1}^{n} \beta_i F_i\right)\right]}, \tag{4}$$

$$P(B|F_1, \ldots F_n) = \frac{exp[-(\alpha+\sum_{i=1}^{n}\beta_i F_i)]}{1+exp[-(\alpha+\sum_{i=1}^{n}\beta_i F_i)]}. \quad (5)$$

2.2 Features

Feature extraction is a critical step in lesion classification. A compact vector of numerical features describing the lesion is desired. Good features are characterized by large interclass mean distance, small intraclass variance, weak inter-feature correlation, and low sensitivity to noise. It is useful if the features are explainable in physical terms [17]**.** In practice individual features have significant overlap between classes, so many features are extracted and used together to improve class discrimination. In this study, 31 features characterizing morphology and texture were extracted from the ultrasound images (Table 1). These features were used with mammography and patient characteristics (Table 1) to discriminate malignant and benign masses.

Table 1. Features used for classifying malignant and benign breast masses.

Category	**Features**			
Morphological	Angular variation interior Angular variation margin Brightness difference Margin sharpness Axis ratio Depth-to-width ratio Radius variation Skeleton norm Tortuosity			
Textural [Chu et al] [Haralick]	1st order statistics	Runlength texture	Gray-level cooccurrence (GLCM) texture	Gray-level connectivity texture
	Mean Standard deviation Kurtosis Skewness	Short-run emphasis Long-run emphasis Gray-level nonuniformity Runlength nonuniformity Run percentage	ASM Contrast Dissimilarity Entropy GLCM mean GLCM variance GLCM correlation Homogeneity	Fragmentation Frag area mean Frag area range Spread
Mammo	BI-RADS categories 1 to 5			
Patient	Age Race			

2.3 Dataset and diagnostic performance

266 cases with biopsy-proven diagnoses, mammographic BI-RADS categories, and ultrasound images in radial and anti-radial planes were used for the study. B-mode sonographic imaging was performed using a broadband 12-5 MHz transducer and a Philips ATL 5000 scanner. Each mass was outlined in 2 to 3 images per patient, and the ultrasound features characterizing shape, margin, echogenicity and texture were extracted from each image. These features along with patient age, race, and mammographic BI-RADS category were used with Naïve Bayes or logistic regression for supervised learning using leave-one-out cross validation. With each classifier, two approaches were used. The first approach was to train on all the features without any feature selection. In the second approach, features were selected

based on information gain. For each model, receiver operating characteristic (ROC) analysis was performed to assess diagnostic performance. Estimates of probability of malignancy provided by Naïve Bayes and by logistic regression were correlated using linear regression.

3. RESULTS

Of the 266 cases, 180 (67.7%) were benign and 86 (32.3%) were malignant. Patients with malignant masses were significantly older (58.7 ± 12.1) than the patients with benign masses (48.1 ± 14.5).The age difference between the two groups was significant (P <0.0001).

Figure 1 shows that when all the features were used, Naïve Bayes performed worse than logistic regression: the area under the ROC curve Az was 0.733 ± 0.035 for Naïve Bayes versus 0.839 ± 0.029 for logistic regression. The difference was statistically significant (P = 0.002, Table 2).

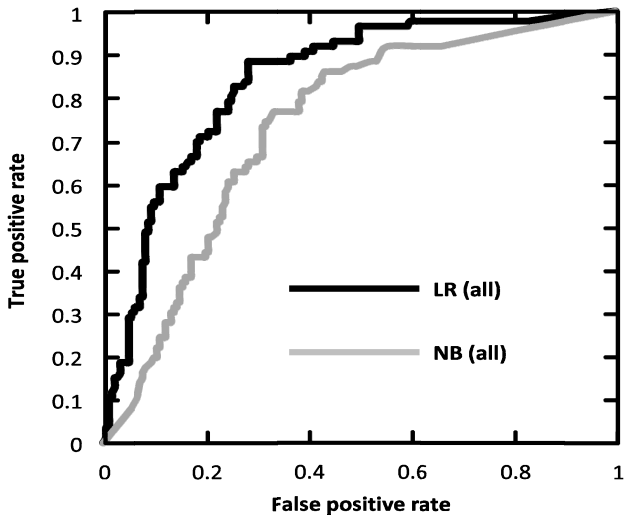

Figure 1. ROC curves for logistic regression, LR(all)), and Naïve Bayes NB(all) using all 34 features.

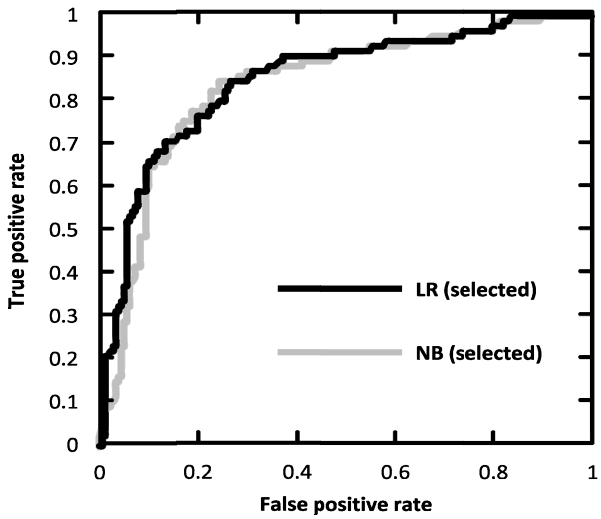

Figure 2. ROC curves for logistic regression, LR(selected) and Naïve Bayes, NB (selected), using 6 of 34 features, selected based on information gain.

Of the 34 features tested, brightness difference, margin sharpness, depth-to-width ratio, mammographic BI-RADS, age, and race showed information gain (Table 3). When these six features were used, the ROC performance for both the classifiers improved (Figure 2, Table 2). The improvement was statistically significant for Naïve Bayes but not for logistic regression. With selected features, Naïve Bayes performance approached that of logistic regression (Table 2).

Table 2. Area under the ROC curves (Az) for Naïve Bayes and logistic regression using all the features (Row 2) and the features selected on the basis of information gain (Row 3). The P values in Column 4 compare Az values in Columns 2 and 3. The P values in Row 4 compare Az values in Rows 2 and 3.

Classifier / Features	Naïve Bayes	Logistic regression	P
All	0.733 ± 0.035	0.839 ± 0.029	0.002
Selected	0.840 ± 0.029	0.851 ± 0.028	0.501
P	< 0.001	0.605	

Figure 3 shows correlation between probability of malignancy, P(malig), determined by Naïve Bayes and by logistic regression. The data fit a linear model with 1.1 slope and near-zero (0.03) intercept. The regression coefficient (R^2) of 0.87 is highly significant ($P < 0.0001$).

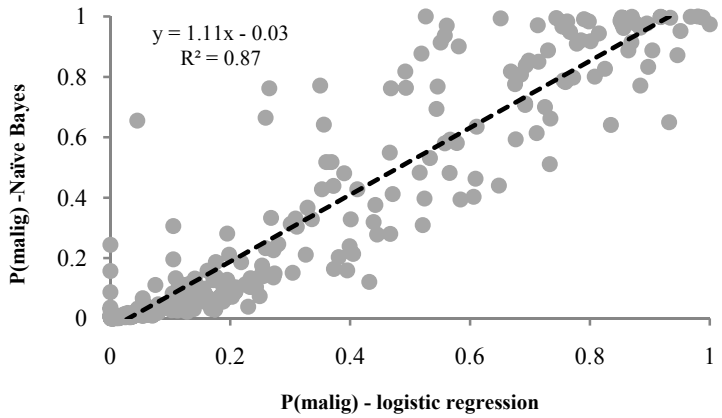

Figure 3. Correlation between probability of malignancy P(malig) estimated by Naïve Bayes and logistic regression.

4. DISCUSSION

Despite considerable technical improvements in mammography and in ultrasound imaging, characterization of solid breast masses as malignant and benign continues to be difficult without biopsy. There is growing interest in using computer image analysis to improve diagnostic confidence. In our previous work [3, 5, 6, 7, 8] we identified sonographic margin features and patient age as statistically discriminating features for benign and malignant breast masses. In this study we expanded the feature domain to include morphological and textural features as well as mammographic BI-RADS confidence levels. These features were used with Naïve Bayes and logistic regression classifiers to assess the probability of malignancy. The two classifiers are fundamentally different. Naïve Bayes is a generative classifier that learns the joint probability of the input features (F) and the label (malignant or benign) from the known examples, and then predicts class probabilities for the unknown cases using the Bayes' rule under the assumption that the features are independent. Logistic regression, on the other hand, is a discriminative classifier that models the posterior probability by directly mapping the input features to the class labels. The learned weights of the logistic regression equation are used to predict the probability of malignancy.

The results of the study show that when all the morphological, textural and mammographic features were used, Naïve Bayes was outperformed by logistic regression by a significant margin: Az was 0.733 for Naïve Bayes versus 0.839 for logistic regression. The reason for the difference in performances could be several factors, including the number of features used, the sample size, and the interdependence of features. In particular, since most of the features are extracted from the images, the assumption that they are independent is likely to be violated. To obtain better insight into some of these issues, a subset of features was selected based on the information gain. Six of the 34 features provided information gain. Of these six features, only three were derived from ultrasound images. When only the selected six features are used, the diagnostic performance of Naïve Bayes increases significantly from Az of 0.733 to 0.839 ($P = 0.002$). In contrast, the performance of logistic regression increases modestly from Az of 0.840 to 0.851 ($P = 0.5$). These results suggest that the image features may not be independent. This feature interdependence reduces the performance of Naïve Bayes, but logistic regression compensates by weighting correlated features lower in the learned model. Elimination of dependent features by information gain selection improves the performance of Naïve Bayes and makes it comparable to that of logistic regression, and the individual case-by-case class probability estimates from Naïve Bayes using six features correlate closely (R^2=0.87) with the probability estimates from logistic regression.

5. CONCLUSION

This study shows that two classifiers employing very different learning strategies can provide highly correlated estimates of probability of malignancy for breast masses on ultrasound. Naïve Bayes is more sensitive to feature selection than logistic regression, but selection by entropy-based information gain can improve performance so that the two classifiers are comparable. Since probability of malignancy cannot be measured directly, a high degree of agreement (R^2=0.87) between two models provides confidence that machine learning probability estimates can be used to differentiate malignant and benign breast masses. Since these two dissimilar approaches agree with high diagnostic performance (Az ~ 0.85) under naïve independence assumptions, other probability estimating models that can represent more complex distributions, such as Bayesian networks, may improve diagnostic performance and facilitate exploratory data analysis by identifying and modeling the interdependence of individual features.

ACKNOWLEDGEMENT

We thank Susan Schultz RDMS for help with patient studies. This work was supported in part by NIH grant CA130946

REFERENCES

[1] American Cancer Society, "Cancer Facts and Figures – 2011", American Cancer Society, NY, 1 (2011).

[2] Mainiero, M. B., Goldkamp, A., Lazarus, E., Livingston, L., Koelliker, S. L., Schepps B., and Mayo-Smith, W. W., "Characterization of breast masses with sonography: can biopsy of some solid masses be deferred?" Journal of Ultrasound in Medicine, 24(2), 161-167 (2005).

[3] Sehgal, C. M., Kangas, S.A., Cary, T.W., Weinstein, S.P., Schultz, S.M., Arger, P.H., and Conant, E.F. "Quantitative description of solid breast nodules by ultrasound imaging," Proc. SPIE 5373, 324-330 (2004).

[4] Sehgal, C. M., Arger, P. H., Rowling S. E., Conant E. F., Reynolds, C., and Patton J. A., "Quantitative vascularity of breast masses by Doppler imaging: regional variations and diagnostic implications," J Ultrasound Med. 19, 427-40 (2000).

[5] Sehgal, C. M., Cary, T. W., Kangas, S. A., Weinstein, S. P., Schultz, S. M., Arger, P. H., and Conant, E. F., "Computer-based margin analysis of breast ultrasound for differentiating malignant and benign masses," J Ultrasound in Medicine, 23, 1201-1209 (2004).

[6] Song, J. H., Venkatesh, S. S., Conant, E. F., Arger, P. H., Sehgal, C. M., "Comparative Analysis of Logistic Regression and Artificial Neural Network for Computer-Aided Diagnosis of Breast Masses," Academic Radiology, 4, 487-495 (2005).

[7] Sehgal, C. M., Weinstein, S. P., Arger, P. H., and Conant, E. F., "Review of Breast Ultrasound," J Mammary Gland. Biol. Neoplasia, 11(2), 113-23 (2006).

[8] Harvey, P., Arger, P. H., Conant, E. F., and Sehgal, C. M., "Differentiation of the solid benign and malignant breast masses by quantitative analysis of the ultrasound images," IEEE International Ultrasonics Symposium Proceedings, 2009, 530 – 533 (2009).

[9] Chen, D. R., Chang, R. F., and Kuo, W. J., "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks," Ultrasound Med. Biol. 28, 1301–1310 (2002)

[10] Drukker, K., Gruszauskas, N. P., Sennett, C. A., and Giger, M. L., "Breast US computer-aided diagnosis workstation: performance with a large clinical diagnostic population," Radiology, 248(2), 392-397 (2008).

[11] Gruszauskas, N. P., Drukker, K., Giger, M. L., Chang, R. F., Sennett, C. A., Moon, W. K., and Pesce, L. L., "Breast US computer-aided diagnosis system: robustness across urban populations in South Korea and the United States," Radiology, 253(3), 661-71 (2009).

[12] Joo, S., Yang, Y. S., and Moon, W. K., "Computer- aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features," IEEE Trans. Med. Imaging 23, 1292–1300 (2004).
[13] Rodrigues, P. S., Giraldi, G. A., Provenzano, M., Faria, M. D., Chang, R. F. and Suri J. S., "A new methodology based on q-entropy for breast lesion classification in 3-D ultrasound images," Conf Proc IEEE Eng Med Biol Soc., 1048–1051 (2006).
[14] Sahiner, B., Chan, H. P., Roubidoux, M. A., Hadjiiski, L. M., Helvie, M. A., Paramagul, C., Bailey, J, Nees, A. V. and Blane C., "Malignant and benign breast masses on 3D US volumetric images: Effect of computer-aided diagnosis on radiologist accuracy," Radiology 242, 716–724 (2007).
[15] Shen, W. C., Chang, R. F., Moon, W. K., Chou, Y. H. and Huang, C. S., "Breast ultrasound computer-aided diagnosis using BI-RADS features," Academic Radiology, 14, 928–939 (2007).
[16] Wang, Y., Jiang, S., Wang, H., Guo, Y. H., Liu, B., Hou, Y., Cheng, H. and Tian J. "CAD algorithms for solid breast masses discrimination: evaluation of the accuracy and interobserver variability," Ultrasound Med Biol., 36(8), 1273-1281 (2010).
[17] Kil, D. H. and Shin, F. B., "Pattern recognition and prediction with applications to signal characterization," AIP Press, New York, 73-109 (1996).