Logistic discriminant anlysis of breast cancer using ultrasound measurements

P. Abdolmaleki[*1], M. Mokhtari Dizagi[2], M.R. Vahead[2], M. Gity[3]

[1] *Department of Biophysics, Tarbiat Modarres University, Tehran, Iran*
[2] *Department of Medical Physics, Tarbiat Modares University, Tehran,Iran*
[3] *Tehran, University Medical Sciences, Tehran, Iran*
[4] *Kerman, University Medical Sciences, Kerman, Iran*

ABSTRACT

Background: Logistic discriminant method was applied to differentiate malignant from benign in a group of patients with proved breast lesions on the base of ultrasonic parameters.

Materials and Methods: Our database include 273 patients' ultrasonographic pictures consisting of 14 quantitative variables. The measuered variables were ultrasound propagation velocity, acoustic impedence and attenuation coefficient at 10 MHz in breast lesions at 20, 25, 30 and 35 °C temperature, physicsl density and age. This database was randomly divided into the estimation of 201 and validation of 72 samples. The estimation samples were used to build the logistic discriminant model, and validation samples were used to validate the performance. Finally, important criteria such as sensitivity, specificity, accuracy and area under the receiver operating characteristic curve (ROC) were evaluted.

Results: Our results showed that the logistic discriminant method was able to classify correctly 67 out of 72 cases presented in the validation sample. The results indicate a remarkable diagnostic accuracy of 93%.

Conclusion: A logistic discriminantor approach enable to predict the probability of malignancy of breast cancer using features extracted from ultrasonic measurement on ultrasound imaging is established.

Keywords: Breast cancer, ultrasonic tissue characterization, ultrasonic imaging, logistic discriminant analysis, receiver operating characteristic curve (ROC).

INTRODUCTION

Breast cancer is a serious health problem and the most leading cause of cancer death in the female population. Although conventinal mammography in conjunction with physical examination are the common established modality for early detection of malignancy, approximately 5%-15% of palpable breast cancer are not detected with mammography (Samuels *et al.* 1992). The main reasons of uncapability of mamography include inherent limitations of X-ray mammography, observer limitation, suboptimal mammographic technique and tumor biology (size and nature of the lesion and surrounding breast tissue) (Weinreb *et al.* 1995). These acknowledged limitations of mammography have sparked some interest in using other diagnostic modalities such as ultrasonography. Clinical sonography in adjunct with mammography is therefore being used in breast cancer to reach a high degree of diagnostic accuracy and to decrease the limitation of X-ray mammography (Hata *et al.* 2004, Yang *et al.* 2004).

Studying the ultrasonic tissue characteristics is an important factor in determining diagnostic criteria for breast factor and for developing new

*
Corresponding author:
Dr.Parviz Abdolmaleki, Ph.D., Department of Biophysics, Faculty of Science, Tarbiat Modares University, Tehran. P.O. Box 14115/175, Fax:+98-21-8009730
E-mail: parviz@modares.ac.ir

ultrasonic equipment in clinical breast echography. The most important acoustic parameters of the tissue which has been previousely reported are velocity, attenuation, impedance, backscatter, angular dependence of scattering, frequency dependence of attenuation, and non-linearly parameter (Duck 1980). These quantitative parameters have a potential usefullness to be organized in algorithemic and/or non-algorithmic statistical models to support the radiologist in differentiating the breast lesions (Rosenfield *et al.* 1980). However, each of these measured parameters showed a sort of value in making differentiation between benign and malignant patterns; however, the accurate diagnosis of breast lesions on sonography still re-mains a clinical difficult task. Biopsy is therefore often performed for histo-logic confirmation. Although specific, biopsy is an invasive, costly, and psychologi-cally stressful procedure especially for the women. This is more important if we consider the presence of a large number of indeterminate lesions report by radiologist from different diagnostic modalities. Therefore, it becomes constractive to reduce the number of unnessary biopsies; because, true posi-tive fraction is reported low, due to technical limitation, as well as, inherent biological overlap (Chen *et al.* 2002). Consequently a comput-erized second opinion in the form of an automated disscriminator could be useful, espe-cially in borderline cases, for the differen-tiation of ma-lignat from benign lesions.

Considering the final histologic diagnosis be-ing malignant (with probability of p) or benign (with probability of 1-p) as a binary outcome, the logistic discriminant analysis (Hosmer *et al.* 1989) could be used as a classifier to predict the outcome of biopsy. This is a form of regression which is used when the dependent variable is a dichotomy and the independent variables are continuous, categorical or both. Logistic discriminant analysis has been successfully applied in breast cancer using different diagnostic modalities (Chou *et al.* 2001, Lumachi *et al.* 2001, Ikeda *et al.* 1999).

On the other hand, the study of ultrasonic tissue characteristic is important in determining diagnosic criteria for breast cancer diagnosis. It also can be applied to develop new ultrasonic equipment in clinical breast echography. So far, some reports have suggested the existance of significant differences in calculated ultrasound velosities and attenuation coefficients in malignant and benign lesions (Weiwad et al. 2000, Siva-ramakrishna *et al.* 2002, Mokhtari-Dizaji 2001).

In this study, we intended to establish a logistic regression model to work as a tool for radiologists to predict the outcome of biopsy using data extracted from ultrasonic measurement. The performance of the established model was then evaluated using the common statistical index including accuracy, sensitivity and specificity. Ultimately, ROC analysis was performed to support the obtained statistical results.

MATERIALS AND METHODES

Our goal was to apply the logistic discriminant analysis to the data collected in a study designed to predict the malignancy of breast cancer on the basis of features that had been extracted from measuring the ultrasonic tissue characteristics consisted of ultrasound propagation velocities, acoustic impedance and attenuation coefficients. Our study group consisted of 273 ultrasonography images of breast lesions with histopathlogically proof. This database included 130 fibroadenoma, and 143 ductal carcinoma.

Data acquisition- sonography imaging
To measure the ultrasonic parameters, we used an experimental set consisting of a clamp, an ultrasound A-mode device (Echoscan US-2500 NIDEK, 10 MHz), a video blaster (SE Creative Technology), a personal computer (Pentium 133), a heater (Rena Co.) and a digital thermometer (SIGMA, ± 0.01 °C). The chamber containing the tissues was heated, and the temperature (± 0.01 °C) during the experiment was monitored and controlled by a thermocouple inserted in the

chamber wall. Tissue specimens were collected from Tehran pathologic centers (especially Imam Khomaini Hospital, Tehran, Iran). The specimens were cut approximately with 2 cm thickness by microtom (NVSLM1-Vibroslice, ±0.05mm) and implanted in breast paranchymal tissue of mimicking materials. These materials had acoustic properties similar to real breast tissue (Sivaramakrishna et al. 2002). The specimens were then heated and at 20, 25, 30 and 35 ℃, ultrasound images were taken by A-mode sonography. The images were digitized by the video blaster board and saved by the computer. The ultrasonic images were then processed and the ultrasound velocities and attenuation coefficients in 10 MHz in breast lesions (fibroadenoma and ductal carcinoma) at 20, 25, 30 and 35 ℃ were measured. The details of measurement had been previously reported elsewhere (Mokhtari-Dizaji *et al.* 2001). We measured the mass density for both malignant and benign lesions using Archimedes law at 20, 25, 30 and 35 ℃ in order to estimate the acoustic imped-ance. We also used pulse-echo ultrasound to meas-ure ultrasound velocities according to the analysis of the transit-time at range of 20-35 ℃. Furthermore, we have estimated the acoustic impendence for malignant and benign lesions by simply multiplying the measured mass density with measured ultrasound velocities at 20, 25, 30 and 35 ℃. The attenuation coefficients (db/cm) were measured with processing echo amplitudes from the front and backing interface of samples implanted in tissue mimicking materials at 10 MHz (Mokhtari-Dizaji et al. 2001). To avoid the statistical error, measurements were performed three times for each specimen. Neglecting the fixator effect and death time, conditions for all of the specimens were equal. In all measurements, we assumed that tissue is homogene, isotope and heating is uniform.

Theory of the statistical method

We used logistic regression model as a classifier to predict the outcome of biopsy in breast cancer. The training and validation samples were used to build and validate the logistic regression model, respectively. Briefly, the logistic regression analysis is a statistical technique through which one examines the relationship between a dependent variable (result of biopsy which represented by Y) and a set of independent variables (14 ultrasonic features which represented by X1 to X14).

Then the independent variables, which could provide the best prediction, can be selected. This approach is commonly applied to predict membership in two groups using a set of predictors (n=14). Suppose we have two populations with different prior probabilities. Using the cases presented in the training samples (n= 201) as

$$E\{Y\} = p = \frac{\exp(\alpha + \beta_1 X1 + ... + \beta_{14} X14)}{1 + \exp(\alpha + \beta_1 X1 + ... + \beta_{14} X14)}$$

well as the prior probability the posterior probabilities for each group was obtained. Then, the cases presented in the validation samples (n=72) were separated based on the obtained posterior probability associated with variables. The simplest optimizing method of discrimination is to maximize the posterior probability of correct allocation. To obtain the posterior probability the logit coefficients could be estimated using the Maximum Likelihood Estimation (Anderson 1972, Equihua 1988). Allocation of new cases can be performed using logit function, which could be obtained using the natural logarithm of the ratio of the calculated posterior probabilities (Hosmer *et al.*1989).

If the outcome of the logit function is positive (with the assumption of equal prior probabilities) the individual is allocated to class one (malignant group). On the other hand, if the outcome is negative, the case is allocated to class two (benign group). The featuers that

$$Ln\left(P/_{1-P}\right) = Logit(p) = \alpha + \beta_1 X1 + ... + \beta_{14} X14$$

entered into the allocation rule were selected by Wald statistics. It is the square of the ratio of the

unstandardized logit coefficients to its standard error, which has a chi-square distribution (Hosmer *et al.* 1989). We addressed a breif detail of logistic regression theory in an attached appendix in previous publication (Abdolmaleki *et al.* 2004).

We initially used logistic regression analysis to predict the outcome of biopsy using a data base consisted of 273 images. We randomly selected 74% (n=201) patient's records (including 96 benign and 105 malignant) to compose the estimation samples. To prepare the validation samples the rest of data, 26% patient's records (n=72), (including 34 benign and 38 malignant) were selected. The dependent (criterion) variable was the dichotomized result of biopsy defined as benign (0) or malignant (1). The independent variables entered into the logistic regression equation were ultrasound propagation velocity, acoustic impedance and attenuation coefficient in 10 MHz in breast lesions at 20, 25, 30 and 35 °C, mass density and age. The Variable (location of the lesion) that was not continuous was excluded from the course of the study. Variables were entered conditionally at an alpha significance of 0.05 on the partial *F*-test and on the log likelihood ratio test and then removed at an alpha of 0.10. We computed a covariance matrix containing all continuous variables to fulfill the established guideline for feature selection. The analysis generated Wald statistics, regression coefficients, standard errors, confidence intervals, Nagelkerke R^2, Hosmer-Lemeshow goodness-of-fit chi square, and predicted group membership. The Nagelkerke R^2 attempts to quantify the proportion of explained variance in the logistic regression model, similar to the R^2 in linear regression, although the variation in a logistic regression model must be defined differently. Nagelkerke (Nagelkerke 1991) proposed a modification to the Cox and Snell R^2 so where value of 1 could be achieved. Ultimately, we built logistic regression models using forward stepwise procedure in SPSS statistical package based on MLE method.

In general the following simulations were performed and compared:
1. The logistic regression model was established using ultrasonic measured data through a forward wald stepwise feature selection. This was performed by determination of the logit coefficients on 14 features in estimation samples (201 patient 's records) during the estimation procedure.
2. The established model was then tested on selected features in validation samples (72 patient 's records) during the validation procedure.

Performance Evaluation

Receiver operating characteristic (ROC) analysis is widely used to evaluate diagnostic performance of logistic discriminant. An ROC curve provides a concise description of trade-offs available between sensitivity and specificity. The area under an ROC curve, denoted A_z when the ROC curve was fitted with the conventional binomial model, is often used to summarize the diagnostic performance described by entire ROC curve (Jiang *et al.* 1996).

After logistic discriminator ap-proach had been established perfectly the validation samples were presented to the model giving two posterior probabilities. Taking into consideration the posterior probability of malignancy, the diag-nostic performance of the logistic discriminator ap-proach was estimated. To do so, the true positive and the false-positive fractions was determined . This data was then used to plot the ROC curves. Ultimately, the area under the ROC curve (A_z) was used to compare the performance of the logistic discriminator ap-proach on validation samples (n=72) during the testing procedure.

To evaluate the performance of the logistic discriminator ap-proach, the obtained posterior probability of malignancy was classi-fied into the five following categories;
(1)= (0.0-0.2)= "Benign"
(2)= (0.2- 0.4)= "Probably benign"

(3)= (0.4-0.6)= "Unsure"
(4)= (0.6-0.8)= "Probably malignant"
(5)= (0.8-1)= "Malignant"

RESULTS

We considered 273 tissue images, comprising 130 benign and 143 malignant tumors, in biopsy specimens. Table 1 summaries the means and standard deviations obtained for evaluated features for both malignant and benign tumors at 20, 25, 30 and 35 ℃. The statistical analysis of variables indicate that, although the assessment of the attenuation coefficients, acoustic impedences, mass densities and ages in differentiation of malignant and benign tumors were successful, the ultrasound did not differ between benign and malignant lesions (p-value>0.05) at a tempreture range of 20-30 ℃.

We ran a logistic regression model using a forward stepwise procedure. We used the likelihood-ratio (LR) test to enter variables into the model. Fourteen variables were entered into the

Table 1: The extracted quantitative parameters from ultrasound measurement used as input into the logistic regression model

Indexes	Mean ± S.D.	Range
Age	40.19±14.67	59
Physicsl density	1016.03±30.13	160
Velocity at 20 ℃	1553.19±7.70	41
Velocity at 25 ℃	1556.11±7.79	46
Velocity at 30 ℃	1559.16±7.97	46
Velocity at 35 ℃	1566.25±61.21	1036
Attenuation at 20 ℃	7.79±1.09	3.91
Attenuation at 25 ℃	7.47±1.06	3.73
Attenuation at 30 ℃	7.14±0.98	3.13
Attenuation at 35 ℃	6.68±.89	3.07
Impedence at 20 ℃	1572871.4±16512.7	89657
Impedence at 25 ℃	1575124.5±17362.2	93646
Impedence at 30 ℃	1582409.7±19094.9	250037
Impedence at 35 ℃	1584123.3±12804.8	81734

training model before the forward stepwise procedure was terminated. Variables included in the logistic regression model were: ultrasound propagation velocity, physical impedence and attenuation coefficient in 10 MHz in breast lesions at 20, 25, 30 and 35 ℃ , physicsl density and age. The last step Nagelkerke R^2 for the logistic regression model 0.919 (step 3), suggesting 91% of the variance associated with result of biopsy was accounted for in the model. The Hosmer-Lemeshow chi-square was 4.617 (*df*=8, P=0.798). We optimized the cut-point for logit (p)>0 for the classification of the lesions type based on the prevalance of the data. We tested the performance of model for three different cut-point of 0.4, 0.5 and 0.6 for logit(p)>0. The best performance was obtained for cut-point of 0.4. Table 2 shows the maximum likelihood estimates of the parameters, standard errors, Wald statistic and corresponding p-values of the logistic regression models fitted to the estimation samples (n=201) for the significant features at last step (step 3). The logistic regression equation for the statistically significant predictors was:

Logit(p)= 26.124-9.623*Attenuation at 35 ℃ + 0.015* Velosity at 35 ℃ +0.300* Age

Baesd on the established logit function an individual case presented in the validation samples were allocated to the malignant group if the logit was (p)>0; otherwise to benign group.

The output of the logistic discriminant analysis on validation samples (table 2) showed a high diagnostic accuracy (93%) , a remarkable specificity (91%) and sensitivity (95%).

A receiver operating characteristic (ROC) curve was computed using the predicted probabilities for group membership from the logistic regression model. Each point on the curve represented the true-positive rate (sensitivity) and the false-positive rate (1_specificity) for a single value. The area under the ROC curve based on the logistic regression model was 0.9471 (SE=0.0276) (figure 1).

Table 2: Indicating the maximum likelihood estimates, standard errors, Wald statistic and corresponding p-values of the logistic regression models fitted to the estimation samples (n=201) for the significant parameters.

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCPT	26.124	5.844	19.986	0.001
Age	0.300	0.059	26.013	0.003
Attenuation at 35 ℃	-9.623	1.708	31.755	0.003
Velocity at 35 ℃	0.015	0.004	13.821	0.005

DISCUSSION

Currently there is considerable interest in ultrasonic tissue characterization. The general aim is to isolate and measure those ultrasonic parameters which discriminate the different
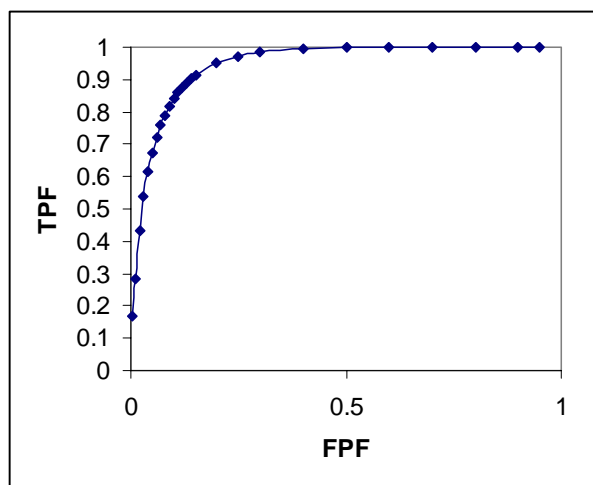

Figure 1: Resulting ROC curves for the logistic discriminant model (LDM) showing the diagnostic performances.

tissues and their pathological states (Rosenfield *et al.* 1980, Edmonds *et al.* 1991, Lefebvre *et al.* 2000). Ultrasound waves are used in breast imaging for the detection of pathological conditions, particularly cystic and cancerous conditions (Weiwad *et al.* 2000, Edmonds *et al.* 1991). Most carcinomas are detected by ultrasound because they are hypoechoic compared to the surrounding dense tissues. Fat and some other

benign lesions may be hypoechoic as well. Thus, carcinomas without shadowing may not be detected by conventional ultrasound. Furthermore, hypoechoic fat lobules, fibroadenomas or inflammation can be mistaken for malignancy, because detection and differentiation of small malignancies, especially in large breasts, is limited and the interpretation of the ultrasound images is only dependent on the experience of radiologist and palpate examinations. Therefore, it would be helpful to obtain additional information from the ultrasound examination to simplify and improve the interpretation.

In the present study we have assumed that applying the objective features extracted from ultrasonic measurement on ultrasound imaging and analyzed by a logistic discriminator approach can possibly reduce the present overlap between the benign and malignant breast tumors. This assumption is justified by the previous reports, suggesting the potential usefulness of the logistic discriminant analysis in making association between many independent continuous and qualitative features (Hosmer 1989, Ikeda *et al.* 1999, Weiwad *et al.* 2000). This took place by establish-ing similarities among evaluated features in the estimation samples during the estimation proc-ess by addressing them as proportional parameters. The estimated parameters were then used during the validation process to evaluate the probability of malignancy for the cases that have not been previously presented to

the model. The average sensitivity specificity and accuracy of the logistic regression model output on validation samples (n=72) were 95%, 91% and 93% respectively. Althoug these results are compareable with the results reported for the artificial neural network (Chen *et al.* 2000) with sensitivity (97.6%), specificity (79.5%) and accuracy (85.6%) some advantages as well as some disadvantages reported for it, which should be addressed elsewhere. These results are valuable because most of the attention in clinical assessment is focused to find a method with higher specificity while keeping high sensitivity. It means that reducing the number of benign cases sent for biopsy, which is an invasive method highly associated with stress, is the main goal in the clinical application. Although our results showed a high specificity of 91%, the subject is still needs more investigation at the molecular level to find out the possible inherent biological overlap (Chen *et al.* 2002, Yilmaz *et al.* 2002, Huber *et al.* 2001).

In conclusion, we have established a logistic discriminantor approach which was able to predict the probability of malignancy of breast cancer using features extracted from ultrasonic measurement on ultrasound imaging. Four of the chief attractions of logistic discrimination are: (i) The model is simple and few distributional assumption are made. (ii) It is applicable with either contiuous or discrete predictor variables, or both. (iii) It is very easy to use with fewer computaiona demands. (iv). Once the parameters have been estimated, the allocation of a fresh individuals requires only the caculation of a linear function.

REFERENCES

1. Anderson J.A. (1972). Separate sample logistic discrimination. *Biometrika,* **59:** 19-3

2. Abdolmaleki P, Yarmohammadi M, Gity M. *(*2004). Comparison of logistic regression and neural network models in predicting the outcome of biopsy in breast cancer from MRI findings. *Iran. J. Radiat. Res., 1:* 217-228*.*

3. Chen D, Chang RF, Huang YL. (2000). Breast cancer diagnosis using self-organizing map for sonography. *Ultrasound Med. Biol.* **26:** 405-11.

4. Chen D.R., Chang R.F., Kuo W.J., Chen M.C., Huang Y.L. (2002). Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound Med. Biol.,* **28:** 1301-10.

5. Chou Y.H., Tiu C.M., Hung G.S., Wu S.C., Chang T.Y., Chiang H.K. (2001). Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis. *Ultrasound Med. Biol.,* **27:** 1493-8.

6. Duck F.A. (1980). Physical properties of tissue. Academic Press, 73-167.

7. Rosenfield A.T., Taylor K.J.W., Jaffe C.C. (1980). Clinical applications of ultrasound tissue characterization. *Radiologic Clinics of North America,* **18:** 31-58.

8. Edmonds P.D., Mortensen C. L., Hill J. R. (1991). Ultrasound tissue characterization of breast biopsy specimens. *Ultrason imaging,* **13:** 162-185.

9. Equihua M. (1988). Analysis of finite mixture of distributions: a statistical tool for biological classification problems. *Comput Appl Biosci.,* **4:** 435-40.

10. Hata T., Takahashi H., Watanabe K., Taka-hashi M., Taguchi K., Itoh T., Todo S. (2004). Magnetic resonance imaging for preoperative evaluation of breast cancer: a comparative study with mammography and ultrasonography. *J. Am. Coll. Surg. ,* **198:** 190-7.

11. Huber S., Vesely M., Zuna I., Delorme S.Czembirek H. (2001). Fibroadenomas: computer-assisted quantitative evaluation of contrast-enhanced power Doppler features and correlation with histopathology. *Ultrasound Med. Biol.,* **27:** 3-11.

12. Hosmer D.W., Lemeshow S. (1989). Applied logistic regression. Wiley, New York.

13. Ikeda O., Yamashita Y., Morishita S., Kido T., Kitajima M., Okamura K., Fukuda S., Takahashi M. (1999). Characterization of breast masses by dynamic enhanced MR imaging. A logistic regression analysis. *Acta Radiol.,* **40: 585-92.**

14. Jiang Y., Metz C.E., Nishikawa R.M. (1996). A receiver operating characteristic partial area index for highly sensitive diagnostic tests. Radiology, **201: 745-750.**

15. Lefebvre F., Meunier M., Thibault F., Laugier P., Berger G. (2000). Computerized ultrasound B-scan characterization of breast nodules. *Ultrasound Med. Biol.*, **26:** 1421-8.

16. Lumachi F., Ermani M.Brandes A.A., Basso S., Basso U., Boccagni P. (2001).Predictive value of different prognostic factors in breast cancer recurrences: multivariate analysis using a logistic regression model. *Anticancer Res., 21 (6A): 4105-8.*

17. Mokhtari-Dizaji M. (2001). Tissue-mimicking materials for teaching sonographers and evaluation of their specifications after three years. *Ultrasound Med. Biol., 27: 1713-6.*

18. Nagelkerke N.J.D. (1991). A note on general definition of the coefficient of determination. *Biometrika*, **78: 691–2.**

19. Sivaramakrishna R., Powell K.A., Lieber M.L., Chilcote W.A., Shekhar R. (2002). Texture analysis of lesions in breast ultrasound images. *Comput Med. Imaging Graph.*, **26: 303-7.**

20. Samuels J.R., Haffty B.G., Lee C.H., Fischer D.B. (1992). Breast conservation therapy in patients with mammographically undetected breast cancers. *Radiology*, **185: 425-427.**

21. Weiwad W., Heinig A., Goetz L., Hartmann H., Lampe D., Buchmann J., Millner R., Spielmann R.P., Heywang-Koebrunner S.H. (2000). Direct measurement of sound velocity in various specimens of breast tissue. *Invest. Radiol.*, **35: 721-6.**

22. Weinreb J.C., Newstead G. (1995). MR imaging of the breast. *Radiology*, **196: 593-610.**

23. Yang W.T., Tse G.M. (2004). Sonographic, mammographic, and histopathologic correlation of symptomatic ductal carcinoma in situ. *AJR Am. J. Roentgenol.*, **182: 101-10.**

24. Yilmaz E., Sal S., Lebe B. (2002). Differentiation of phyllodes tumors versus fibroadenomas. *Acta Radiol.*, **43: 34-9.**