Hellenic Journal of Surgery 2011; 83: 3

# Breast Density: Computerized Analysis on Digitized Mammograms

## Original Article

A. Papaevangelou, S. Chatzistergos, K. S. Nikita, G. Zografos
Received 02/02/2011 Accepted 31/03/2011


## Abstract

*Background-Aim:* Mammographically dense breast tissue is related to a higher risk of breast cancer. We aim to evaluate a computerized system, assess whether it can provide an accurate and objective estimation of the breast density and if it can accurately classify the mammograms according to the ACR/ BIRADS system.
*Methods:* We retrospectively reviewed the mediolateral oblique (MLO) and cranial-caudal (CC) views of 83 normal mammograms and classified them, both manually and with the use of computerized texture analysis (CTA), according to their density. We grouped the mammograms either into two (ACR 1-2, ACR 3-4) or four categories (ACR 1 to 4). An inter-rater reliability analysis was performed using the kappa statistic to determine consistency among the radiologist and the CTA.
*Results:* The best matching was observed for the MLO view when the classification involved 2 groups (94%). The equivalent matching for the CC view was 92.8%.
When we used all 4 ACR categories the matching was lower: i.e. 84.3% for the MLO view and 79.5% for the CC view. For older patients (>50 years old) the best matching was for the MLO views while for the younger patients equal matching was observed for both views. Overall, substantial to almost perfect agreement was observed between the two methods of assessment.
*Conclusion:* CTA is a reliable and accurate form of computerized assisted diagnosis. If a single view is to be used, it should be the MLO view since the addition of CC view does not seem improve the sensitivity of the method.


A. Papaevangelou (Corresponding author)
- Radiology Department, Thriassio Hospital Athens, GreeceS. Chatzistergos, K. S. Nikita
- Department of Electrical and Computer Engineering, National Technical University of Athens, Greece
G. Zografos
- First Department of Propaedeutic Surgery, Hippokratio Hospital, Medical School University of Athens, Greece
e-mail: gpapaevangelou@gmail.com

*Key words:*
Digitized mammography, Mammographic density and breast cancer, Mammographic parenchymal patterns and breast cancer, Estimation of mammographic tissue, Computer-aided diagnosis, Statistical features, Image analysis, and Classification of breast density.

## Introduction

Breast cancer remains the leading cause of cancer death in women in their 40s [1,2]. Studies have shown that women with dense tissue in more than 60-75% of their breast have a four to six-fold risk of breast cancer than those with densities in less than 5% [3,4,5]. Breast density actually refers to the amount of "white areas" of the breast on mammogram while the balance of "white" and "black" areas reflects the breast composition of glandular tissue, connective tissue, and fat [6]. The American College of Radiology (ACR) developed the Breast Imaging Reporting and Data System (BI-RADS), which classifies the mammograms according to the breast density into four categories and has largely become universally accepted (Fig. 1). According to this system, the percentage of "white areas" is estimated and the mammogram is classified as ACR 1 if the density is less than 25%, as ACR 2 if it is between 25-50%, as ACR 3 if it is between 51-75% and finally as ACR 4 if it is more than 75% [6, 7]. Apart from the subjective manual method of assessing and classifying mammograms into ACR categories, attempts have been made to use software in order to identify and outline the white areas, compare them with the total breast area and obtain objective estimation of glandular density [8,9,10]. These automatic methods for classification of breast tissue according to its density could be used and justified as an automatic risk assessment in screening population. [11,12,13].

For computers, it is simpler and easier to solve the normal tissue recognition problem rather than the tumour detection problem because of the variability of the appearance of the tumours. However, it is very important to assess and validate computer performance during normal tissue recognition [1, 14,15].

"Text analysis" is a method used to discover topics in a document using the "bag-of words" document representation [16]. Anna Bosch et al adopted this method to be used in the analysis of medical images [2, 17,18,19]. We based our approach on the aforementioned principal treating the mammographic images as "documents" full of "visual words". The basic elements used in our method were "textons" based on statistical and textural characteristics [20,21]. We aimed to evaluate this computerized system and assess whether it can provide an accurate and objective estimation of the breast density and accurately classify the mammograms according to the ACR- BIRADS system.

## Materials and Methods

For the purpose of the study, we used the mediolateral oblique (MLO) and cranial-caudal (CC) views of 83 digitized mammograms randomly selected from the files of the Radiology Department of Thriasio Hospital. Only mammograms initially reported as "normal mammograms" by the radiologists of the radiology department were included, while mammograms with pathologic findings or bad quality images were excluded.

## Methodology

All mammograms were blindly reviewed by two independent radiologists (Rad) and classified according to the BIRADS–ACR system. This assessment was used as the standard to which the computerized analysis would be compared.

For the computer-assisted diagnosis (CAD) we used the adapted method of "text analysis" for medical images (computerized text analysis-CTA). The basic principal of "bag of words" document representation was followed.

In the analysis of breast imaging, the" images" would be the mammograms, the "topics" would be the different densities of the tissue and finally the "words" would be a number of local area image characteristics. In this way, the mammographic image was treated as a document full of "visual words" where the target was to find the "topic" of the document. These "visual words" were created from automatically extracted local textural descriptors.

### *Breast density analysis was performed in five stages, as follows:*

- All available mammographic images were preprocessed in order to distinguish the actual breast region from its surroundings. An experienced radiologist performed this step manually.
- All mammograms were digital images of 1024 x1024 pixels with a pixel size of 200μm. Around each pixel, a NXN neighbourhood square was selected to form a "texton" which was the basic texture particle (N=7 in our series). The pixels contained in this square were reordered and aligned to form a vector in an N2 dimensional feature space. Each vector was characterized by fourteen statistical and textural characteristics calculated on the 7X7 square (mean value, standard deviation, kurtosis, skewness, entropy, median absolute deviation 1 and 0, Igr, inertia, clustershade, clusterprominence, Haralickcor, correlation and IDF). The procedure was applied to the entire image to create a set of textons (i.e . the fundamental microstructures composing the mammographic image/ the descriptors).
- K-means, a well-known alogorithm, was subsequently used to cluster the textons, group them together and finally create a texton vocabulary. The vectors separated the image dataset into a number of clusters by defining an equal number of cluster centres. These cluster centres constituted the actual texton vocabulary of the image dataset. K-means aims to cluster data by rearranging cluster centres so as to minimize the summated distance of all vectors from their closest cluster centre.
- After defining the textons, we calculated (a) the number of appearances of each texton in every single image and (b) the total number of textons present on each image. By dividing a/b we obtained the texton frequency vector (frequency matrix) of the image, which was used for the classification of the breast tissues.
- To perform classification of the breast tissue we used five different classifiers provided by MAT LAB (Matrix Laboratory): Classification trees, Bayes Classifier, SVM (Support vector machines), Probabilistic Neural Networks, and Pattern Recognition Network

In order to evaluate the effectiveness of the proposed algorithm we used the "leave-one-out" method. According to this method, one by one, each image was removed from the image dataset and analyzed separately by a classifier trained using the remaining images (training set). After the training was completed, each test image was classified according to the estimated parenchymal breast density into a different category. The number of correct estimations of the computerized classification divided by the total number of images indicates the success rate of the algorithm. The concordance rate between the 2 studies represents the sensitivity of the method under study ( i.e. CAD/CTA). The statistical analysis was carried out using SPSS for Windows version 17 software package (Statisti-cal Package for Social sciences; Inc, Chicago, IL).

We applied the inter-rater reliability analysis using the Kappa statistic in order to determine the consistency among the two assessment methods ( i.e CAD/CTA and Rad).

## Results

The mean age of the patients in this series was 55 years (29-77), of whom 28 patients were under 50 years of age with a mean age of 43.9 years (29-50) and 55 were over 50 with a mean age of 60.5 (51-77). We compared the radiologist reports (Rad) with the computerized reports (CAD). Initially, we classified the mammograms into 2 wide categories: Group A comprised the ACR 1 and 2 mammograms and group B the ACR 3 and 4 mammograms. We did a subgroup analysis applying the same comparisons for both patients under and over 50 years of age.

### *MLO view (group A and group B) (N=83)*

Overall, there was a 94% concordance between the Rad and the CAD classification. In 1.2% of the cases the Cad underestimated, while in 4.8% it overestimated the breast density.

The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.855, ($p<0.001$). The concordance rate was higher for group B (97.6%) than for group A (90.2%).

For patients over 50 years of age (N=55), the overall concordance rate was 94.5% and the CAD overestimated the breast density in 5.5% of the cases. In this age group, there was 100% concordance for group B as opposed to 90% for group A. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.891, ($p<0.001$).

For patients under 50 years of age (N=28), the overall concordance rate was 92.9% but the CAD both overestimated and underestimated the breast density in 3.6% of the cases. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.850, ($p<0.001$). In this age group there was 94.1% concordance for group B and 90.9% for group A.

### *CC view (group A and group B)*

Overall, there was a 92.8% concordance between the Rad and the CAD classification. In 2.4% of the cases the Cad underestimated while in 4.8% it overestimated the breast density. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.855,($p<0.001$). Once more, the concordance rate was higher for group B (95.2%) than for group A (90.2%).

For patients over 50 years of age, the overall concordance rate was 92.7%. The CAD overestimated the breast density in 5.5% of the cases and underestimated it in 1.8%. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.854,($p<0.001$). In this age group, there was 96% concordance for group B and 90% for group A.

For patients under 50 years of age, similar results to the MLO view were observed, i.e. the overall concordance rate was 92.9% but the CAD both overestimated and underestimated the breast density in 3.6% of the cases. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.850,($p<0.001$). In this age group, there was 94.1% concordance for group B and 90.9% for group A.

In the next step of our study, we compared the results of all BIRADS categories (i.e. 1 to 4) for the same views and subgroups

### *MLO view (ACR 1 to 4)*

Overall, there was an 84.3% concordance between the Rad and the CAD classification. The Cad underestimated in 3.6% and overestimated in 12% the breast density. The inter-rater reliability for the rates was found to be substantial since the Kappa was 0.710, ($p<0.001$). The concordance rate was higher for BI-RADS 3 (88.5%) followed by BI-RADS 2 (87.5%), BI-RADS 4 (81.3%) and BI-RADS 1 (66.7%).

For patients under 50 years of age, there was an overall 85.7% concordance between the RAD and the CAD classification. The CAD both underestimated and overestimated the breast density in 7.1% of the cases. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.802, ($p<0.001$). In this age group, the concordance rate was higher for ACR 1 (100%) followed by ACR 3 (88.9%), ACR 2 (87.5%) and ACR 4 (75%).

For patients over 50 years of age, the overall concordance rate was 83.6 %. The CAD overestimated in 14.5% and underestimated in 1.8% the breast density. The inter-rater reliability for the rates was found to be substantial since the Kappa was 0.757, ($p<0.001$). In this age group, the concordance rate was similar for ACR 2 (87.5%), ACR 3 (88.2%), and ACR 4 (87.5%) while it was low for ACR 1 (50%).

### *CC view (ACR 1 to 4)*

Overall, there was a low concordance between the RAD and the CAD classification (79.5%). The CAD underestimated in 8.4% and overestimated in 12% the breast density. The inter-rater reliability for the rates was found to be substantial since the Kappa was 0.710, ($p<0.001$). The concordance rate was higher for ACR 2 (81.3%) and ACR 3 (80.8%), followed by ACR 1 (77.8%) and ACR 4 (75%).

For patients under 50 years of age, there was an overall 85.7% concordance between the Rad and the CAD classification. The inter-rater reliability for the rates was found to be outstanding since the Kappa was 0.802,($p<0.001$). The Cad underestimated in 3.6% and overestimated in 10.7% the breast density. In this age group, the concordance rate was higher for ACR 1 (100%) followed by ACR 3 (87.5%), ACR 4 (87.5%) and ACR 3 (77.8%).

Among patients over 50 years of age, the overall concordance rate was once again low (76,4%). The inter-rater reliability for the rates was found to be substantial since the Kappa was 0.855, ($p<0.001$). The CAD overestimated in 12.7% and underestimated in 10.9% the breast density. In this age group, the concordance rate was higher for ACR 2 (79.2%) and ACR 3 (82.4%) while it was lower for ACR 1 (66.7%) and ACR 4 (62.5%).

## Discussion

Our method was applied to 83 digitized mammograms. Initially we used the MLO views only, but later also applied the method to the CC views, to assess whether there was a benefit in the sensitivity of the method. Many researchers grouped ACR 1 with 2 and ACR 3 with 4 and used two wide categories of parenchymal breast density in order to reduce the dimensionality of the problem [3,14,18]. We also used this approach to test our algorithm. In this way, there was one category with low density and another with high breast density.

According to the BI-RADS-ACR system, women who have dense breasts are at an increased risk of developing breast cancer compared with women whose breasts are of average density. The consideration of the breast physiology and the cancer risk factors with the image data analysis produces a more effective use of a CAD system [3,4,5].

The use of our CAD system is very important in the estimation of breast density and the early detection of breast cancer. The results of previous studies demonstrated the feasibility of estimating mammographic breast density using computer-aided techniques. Zhou et al achieved reproducibility of breast density estimation and managed to improve the diagnostic accuracy of breast density classification in comparison with the subjective visual assessment [18].

For various reasons, it is very important for the radiologist to be able to provide the surgeon with the facts regarding the composition of the breast tissue, irrespective of whether or not breast cancer is present. This can facilitate the decision making process in terms of diagnosis and screening as well as management and subsequent follow-up [22].

In this study, we noted a high concordance rate and substantial to almost perfect agreement between the observations of the radiologists and the CAD reports in all comparisons.

When dealing with only two broad categories of breast density, there was almost perfect agreement between the radiologist and the CAD assessment and the concordance rate was more than 92.8%, irrespective of the age group. There was a trend for higher concordance when dealing with the MLO view, although this was not observed when analyzing the mammograms of the younger (<50 years) patients.

When all four ACR categories were used, it seemed that concordance was overall better for the MLO views. When the dataset was broken down according to the age group, an agreement was observed between the radiologists and the CAD that was almost perfect and around 85.7% for both the CC and the MLO view for the younger patients. On the contrary, the agreement was somewhat lower, but still substantial, for the elder patients (>50 years). There was a substantial discrepancy between the concordance rate of the MLO and the CC views in the elder patients with the higher rates being observed for the MLO views. This may well have affected the overall results.

Were we to classify the mammograms in all 4 ACR categories, using the MLO view would be the ideal since it seems to be associated with a higher concordance rate than that of CC views. This limitation does not apply to younger patients where any view could be used.

Hence, our method is overall reliable and sensitive in classifying the mammograms according to the glandular density. Any view can provide equally good results if we want to classify the mammograms into two broad categories (i.e. dense vs. fatty).

A correct classification by a similar CAD system could decrease the workload of the experienced radiologists who would only need to review the dense mammograms leaving the fatty ones for less experienced clinicians. Moreover, untrained radiologists who tend to overestimate the breast density would stand to benefit from using CAD procedures [21]. Finally, this method could ultimately reduce the radiologist's workload and improve screening efficiency. It could also be used as a pre-screening system, which would detect normal mammograms, classify them according to the density and indicate the high risk (dense) mammograms which would then be assessed by a radiologist.

In this paper, we used a new computerized procedure for texture analysis based on statistical measures and textons which provided an automatically accurate and objective estimation of the parenchymal breast density. The breast tissue was successfully classified into four categories according to the ACR-BIRADS system. It remains for future studies to assess whether or not this method could become a safe substitute for the screening radiologists.

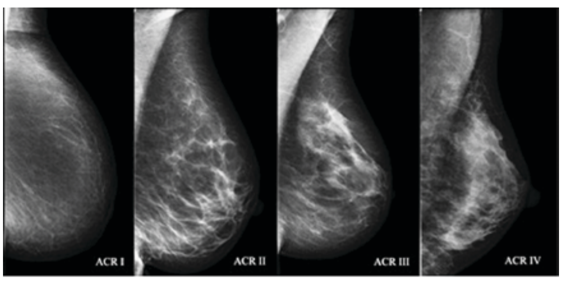


**Fig. 1** Digitized Mammograms of the 4 BIRADS-ACR categories

## *Conflict of interest*

The authors declare that they have no conflict of interest.

## References

**1.** Eurostat: Health statistics atlas on mortality in the European Union. Official Journal of the European Union, 2009

**2.** July 27, 2010 • Volume 7 / Number 15 National Cancer Institute.

**3.** Boyd NF, Lockwood GA, Byng JW, et al. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers 1998;7:1133-44

**4.** J.N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," Cancer, vol. 37, pp. 2486–2492, May 1976.

**5.** C. Tulin, E. Fishell, H. Wedad , P. Sun, E. Rawlinson, S. Narod, D. McCready. Mammographic density and the risk of breast can-cer recurrence following breast-conserving surgery. Cancer, Pub-lished Online: November 9, 2009

**6.** Sheng Liu, Babbs C., Delp E. Normal mammogram analysis and recognition. Perdue Cancer Center 1998. West Lafayette, Indiana.IEEE 1998:727-731.

**7.** American College of Radiology (ACR) (2003) Illustrated breast imaging reporting and data system (BI-RADS), 4th edition. American College of Radiology, Reston, VA

**8.** Yaffe Martin. Review. Mammographic density. Measurement of mammographic density. Breast Cancer research 2008, 10:209.

**9.** Petroudi Styliani, Kadir Timor, Brady Michael. Automatic clas-sification of mammographic parenchymal patterns: a statistical approach. Conf.Rec. 2003 IEEE Int. Conf. Eng. Med. Biol. Soc., pp416-423.

**10.** N.F. Boyd, J.W. Byng, R.A. Jong, E.K. Fishell, L.E. Little, A.B. Miller, G.A. Lockwood, D.L. Tritchler, and M.J. Yaffe., "Quantitative classification of mammographic densities and breast cancer risk: Results from the canadian national breast screening study," J. Natl. Cancer I., vol 87, pp. 670–675, May 1995.

**11.** T. Freer, M. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12860 patients in a community breast center," Radiology vol. 220, pp. 781–786, Sept. 2001.

**12.** R. Birdwell, D. Ikeda, K. O'Shaughnessy, E. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," Radiology, vol. 219 pp. 192–202, Apr. 2002.

**13.** A. Oliver, J. Freixenet, R. Zwiggelaar, "Automatic classification of breast density," presented at the 2005 Int. Conf. Image Pro-cessing, Genoa, IT.

**14.** A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. Denton, R. Zwiggelaar, "A Novel Breast Tissue Density Classification Methodology," IEEE T. Inf. Technol. B., vol. 12, pp. 55-65 , Jan. 2008.

**15.** L. Basset and R. Gold, Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging, Grune & Stratton, New York, 1987.

**16.** Mihran Tuceryan, Anil K. Jain Texture Analysis. The Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau,. P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co., 1998.

**17.** Bosch Anna, Munoz Xavier Oliver Arnau, Marti Joan. Modeling and classifying breast tissue density in mammograms. IEEE Computer society conference on computer vision and pattern recognition 2006.

**18.** Zhou, C., Chan, H.: Computerized image analysis: Estimation of breast density on mammograms. In: Med. Phys. Volume 28. (2001) 1056–1069.

**19.** S. Cha, S. Shihari. "On measuring the distance between his-to-grams," Pattern Recogn., vol. 35, pp. 1355-1370 , Jun. 2001.

**20.** S.Chatsistergos, J.Stoitsis, A.Papaevangelou., G.Zografos., K.Nikita. Parenchymal Breast density estimation with the use of statistical characteristics and textons. Journ suppl. ITAB 3-5 Nov, 2010, Corfu.

**21.** Martin Katherine, Helvie Mark, Zhou Chuan, et al. Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. Radiology 2006; 240(3): 656-665.

**22.** Tulin C, Eve Fishell, Wedad H, Ping Sun, E. Rawlinson, S.A. Narod, David R. McCready. Mammographic Density and the Risk of Breast Cancer Recurrence After Breast-Conserving Surgery-Cancer. Cancer Vol. 115, Issue 24, pp. 5780–5787, Dec 2009.

# Υπολογιστικά Υποβοηθούμενη Ανάλυση της Ψηφιακής Μαστογραφίας για Εκτίμηση της Πυκνότητας του Μαστού

## Πρωτότυπο Άρθρο

Α. Παπαευαγγέλου, Σ. Χατζηστέργος, Κ. Σ. Νικήτα, Γ. Ζωγράφος

### Περίληψη

*Εισαγωγή:* Σκοπός της εργασίας είναι η ανάλυση της υφής του μαζικού αδένα στη ψηφιοποιημένη εικόνα της μαστογραφίας, η μέτρηση της πυκνότητας του παρεγχύματος με την βοήθεια τεχνικών ψηφιακής επεξεργασίας καθώς και η ταξινόμηση των μαστογραφιών σύμφωνα με το ACR-BIRADS σύστημα. Έχει αναφερθεί ότι η πυκνότητα του μαζικού αδένα είναι ένας ουσιαστικός δείκτης που σχετίζεται με τον καρκίνο του μαστού.

*Μέθοδος:* Μελετήθηκαν 83 μαστογραφίες (CC και MLO λήψεις) και ταξινομήθηκαν, σύμφωνα με την πυκνότητά τους, αρχικά απο έμπειρους ακτινολόγους και στη συνέχεια απο υπολογιστικό πρόγραμμα ανάλυσης υφής του μαζικού αδένα. Ταξινομήθηκαν είτε σε δύο ομάδες (ACR 1-2, ACR 3-4) είτε σε τέσσερις κατηγορίες (ACR 1 μέχρι 4) και συγκρίναμε στατιστικά τα αποτελέσματα.

*Αποτελέσματα:* Το μεγαλύτερο ποσοστό συμφωνίας μεταξύ ακτινολόγων και υπολογιστικής μέτρησης ήταν 94% και παρατηρήθηκε στις MLO στηνταξινόμηση των δύο ομάδων. Το αντίστοιχο ποσοτό για τις CC λήψεις ήταν 92,8% ενώ στις τέσσερις κατηγορίες ήτων 84,3% και 79,5% για τις MLO και CC αντίστοιχα. Σε γυναίκες >50 ετών υπήρχε μεγαλύτερη συμφωνία στις MLO λήψεις ενώ σε μικρότερης ηλικίας γυναίκες δεν υπήρχε ουσιαστική διαφορά μεταξύ των δύο λήψεων.

*Συμπέρασμα:* Η υπολογιστικά υποβοηθούμενη ανάλυση της υφής του μαζικού αδένα αποτελεί μια ακριβή και αξιόπιστη μέθοδο κατά την οποία είναι δυνατή η μέτρηση και η ταξινόμηση της πυκνότητάς του.

### *Λέξεις κλειδιά*

Ψηφιοποιημένη μαστογραφία, Πυκνότητα μαζικού αδένα και καρκίνος μαστού, Ταξινόμηση πυκνότητας μαζικού αδένα, Ανάλυση υφής μαζικού αδένα, Υπολογιστικά υποβοηθούμενη ανάλυση εικόνας

-Ακτινολογικό Τμήμα Θριασίου Νοσοκομείου Αθήνα
-Τμήμα Ηλεκτρολόγων-Μηχανικών Πληροφορικής, Εθνικό Τεχνικό Πανεπιστήμιο Αθηνών
-Χειρουργική Προπαιδευτική Κλινική, Ιπποκράτειο Νοσοκομείο, Ιατρική Σχολή Πανεπιστημίου Αθηνών