

A comparison of breast tissue classification techniques

Arnau Oliver¹, Jordi Freixenet¹, Robert Martí¹, and Reyer Zwiggelaar²

¹ Institute of Informatics and Applications, University of Girona
Campus Montilivi, Ed. P-IV, 17071, Girona, Spain
{aoliver,jordif,marly}@eia.udg.es

² Department of Computer Science, University of Wales
Aberystwyth SY23 3DB, UK
rrz@uw.ac.uk

Abstract. It is widely accepted in the medical community that breast tissue density is an important risk factor for the development of breast cancer. Thus, the development of reliable automatic methods for classification of breast tissue is justified and necessary. Although different works in this area have been proposed in recent years, only a few are based on the BIRADS classification standard. In this paper we review different strategies for extracting features in tissue classification systems, and demonstrate, not only the feasibility of estimating breast density using automatic computer vision techniques, but also the benefits of segmentation of the breast based on internal tissue information. The evaluation of the methods is based on the full MIAS database classified according to BIRADS categories, and agreement between automatic and manual classification of 82% was obtained.

1 Introduction

Breast cancer is considered a major health problem in western countries, and constitutes the most common cancer among women in the European Union. It is calculated that between one in eight and one in twelve women will develop breast cancer during their lifetime [1]. Mammography is still the preferred and most efficient method for detecting breast cancer at early stages, a crucial issue for a high survival rate.

Computer Aided Diagnosis (CAD) systems are being developed to assist radiologists in the evaluation of mammographic images [2, 3]. However, recent studies have shown that the sensitivity of these systems is significantly decreased as the density of the breast increases while the specificity of the systems remains relatively constant [4]. From a medical point of view, these studies are disappointing, because it is well-known that there is a strong positive correlation between breast parenchymal density in mammograms and breast cancer risk [5]. Therefore, automatic classification of breast tissue will be beneficial, not only to estimate the density of the breast, but also to establish an optimal strategy to follow if, for example, the user is looking for masses.

Automated parenchymal pattern classification is not a recent topic. Effectively, a number of different proposals can be found in literature [6, 7]. However, only a small number of those papers have classified the density according to BI-RADS categories [8–11], *which is the U.S. standard in radiology, and the one that many European countries are introducing*. All these methods are based on extracting features from the breast, which can be related to texture or just gray-level information. A main difference between these works is the regions from which the features are extracted. For instance, Bovis and Singh [8] extract texture features as if the breast has homogenous texture, Karssemeijer [7] extracts features from regions selected according to the distance to the skin-line, and Oliver et al. [10] segment the breast according to internal tissue information. Figure 1 shows graphically these different strategies. In this work we implemented all those strategies and compared them using the MIAS database [12], with the aim to demonstrate the effectiveness of a segmentation strategy. The remainder of this work is structured as follows. The next section presents the implemented strategies, while Section 3 shows the obtained results. The paper ends with a discussion and conclusions section.

2 Breast density classification

The aim of this paper is to evaluate the application of a segmentation step prior to the extraction of features in the mammogram. Thus, the strategies are divided in (a) no segmentation of the breast, (b) segmentation of the breast according to the distance of each pixel to the skin-line and (c-e) segmentation of the breast according to internal tissue information, where (c) uses Fuzzy C-Means, (d) fractal analysis, and (e) a statistical approach. The following subsections provide more details on these strategies, while the remaining subsections describe the extracted features, as well as the classifier used in this work.

However, the first step is the segmentation of the profile of the breast. The aim of this initial segmentation is to isolate the breast area from other objects present in a mammographic image such as background, annotations and pectoral muscle in MLO images. *We used a previous developed algorithm for such proposal [13].*

2.1 Strategies on breast tissue density segmentation

No segmentation The first approach is the extraction of features of the global breast, without any kind of segmentation. Note that this approach implicitly considers the breast area as homogeneous, with the same texture for the whole breast.

Segmentation according to the distance to the skin-line This approach, graphically shown in Figure 1(c), was first suggested by Karssemeijer [7]. The main idea for such approach is the assumption that a strong correlation exists between tissue thickness and distance to the skin line. To compute the distance

to the skin line of the breast a distance transform is used. First, a binary object is formed by merging the breast tissue and the pectoral (or the equivalent segmented region). This object is eroded repeatedly using a circular structuring element. The number of erosions that it takes for a pixel to remove it from the object represents its taken as the distance to the skin-line.

Segmentation through Fuzzy C-Means In contrast with the previous approach, the approach proposed by [10] aims to divide the breast into two different clusters according to the internal breast density, as is shown in Figure 1(d). The Fuzzy C-Means algorithm [14] was used in order to group the pixels of the breast in two separate categories: fatty and dense tissue. As the authors suggested, instead of using a random placement of the initial seeds, the initialization is done by using the gray level values that represent 15% and 85% of the accumulative histogram of the breast pixels.

Segmentation using Fractal analysis Another work segmenting the breast using breast tissue information is the work of Raba et al. [15], in which the breasts were divided using a fractal scheme. The fractal coding recursively splits the image in quadrants depending on the information that is contained in each region. The function to determine if a region should be split is based on a local histogram measure. A broad histogram means that the region contains intensities that covers a large spectrum from dark to lighter values, and can be split in four regions. These regions are recursively evaluated until the decision function determines that the region has not to be split, thus obtaining regions with uniform properties of tissue. Figure 1(e) shows some results using this algorithm.

Segmentation via Statistical analysis This novel approach is based on a statistical analysis of the breast. Thus, windows of 25×25 pixels of one mammogram are extracted and used as the ground-truth in order to segment the other breasts mammograms. Some of the windows represent dense breast tissue while others represent non-dense tissue. Hence, using the fisherfaces approach [16], these windows are used to construct a model from each part of the mammogram, and subsequently, each subwindow of the mammogram is classified as one of those regions. Thus, we finally obtain a segmentation of the breast in two regions, which represents fatty and dense tissue, as is shown in Figure 1(f).

2.2 Extracted Features

Once the breast is divided in regions, a set of morphological and texture features is extracted from each one.

As morphological features, the relative area and the four first histogram moments for the fatty and dense regions were calculated.

On the other hand, a set of features derived from co-occurrence matrices [17] were used as texture features. Co-occurrence matrices are essentially

two-dimensional histograms of the occurrence of pairs of gray-levels for a given displacement vector. Formally, the co-occurrence of gray levels can be specified as a matrix of relative frequencies P_{ij} , in which two pixels separated by a distance d and angle θ have gray levels i and j . In this work we use four different directions: 0° , 45° , 90° , and 135° , and three distances equal to 1, 5, and 9 pixels. For each co-occurrence matrix the following statistics were used: contrast, energy, entropy, correlation, sum average, sum entropy, difference average, difference entropy, and homogeneity features. Thus, we deal with a set of 113 features for each class.

2.3 Classification

In order to train the system, we used a Bayesian classifier which combines two well-known classifiers: the k-Nearest Neighbours (kNN) algorithm and the C4.5 decision tree. Briefly, the kNN classifier [18] consists of classifying a non-classified vector into the k most similar vectors presents in the training set, while a decision tree recursively subdivides regions in feature space into different subspaces, using different thresholds in each dimension to separates the classes “as much as possible”. In this work we used the C4.5 algorithm developed by Quinlan [19], improved by a boosting procedure.

Thus, when a new case is studied, it is classified according to the classic Bayes equation:

$$P(x \in B_i | A(x)) = \frac{P(A(x) | x \in B_i)P(B_i)}{\sum_{j=1..4} P(A(x) | x \in B_j)P(B_j)} \quad (1)$$

Translating this formula into words, we consider the probability of a mammogram (x), with a set of features $A(x)$, to belong to the class B_i as the posterior problem. The prior is the probability of the mammogram to belong to a class before any observation of the mammogram is done. Here we used as the prior probability the number of cases that exists in the database for each class, divided by the total number of cases. The likelihood estimation is calculated by using a non-parametric estimation, which is explained in the next paragraph. Finally, the evidence is just a normalization factor, needed to ensure that the sum of posteriors probabilities for each class is one. The value of j indicates the number of BIRADS classes.

The likelihood estimation is based on the combination of the explained classifiers. This is achieved following a soft-assign approach where binary (or discrete) classification results are transformed into continues values which depict class membership. For the kNN classifier, the membership value of a class is proportional to the number of neighbors belonging to this class. The membership value for each class will be the sum of the inverse Euclidean distances between neighboring patterns belonging to that class and the unclassified pattern. A final normalization to one between all the membership values is required. On the other hand, in the traditional C4.5 decision tree, a new pattern is classified by using the vote of the different classifiers weighted by their accuracy. Thus, in

order to achieve a membership for each class, instead of considering the voting criteria we take into account the result of each classifier. Adding all the results for the same class and normalizing all the results, the membership for each class is obtained.

3 Results

In order to test the proposed methods we used the MIAS database [12], which is a public database of digitized mammograms. An expert has classified all the 322 mammograms according to the BIRADS lexicon, obtaining 129 BIRADS I, 79 BIRADS II, 70 BIRADS III, and 44 BIRADS IV samples. To evaluate the different methods, a leave-one-woman-out procedure was used, in which each sample was analyzed by a classifier which was trained using all other samples except for those from the same woman. Note that this is necessary in order to not bias the results as, typically, both mammograms for a single women have similar internal morphology.

The confusion matrix for the first strategy (no segmentation) is shown in Table 1(a). This should be read as follows: rows indicate the object to be recognized (the true class) and columns indicate the label the automatic classifier associates with this object, thus obtaining the correct classified mammograms in the diagonal of the matrix. Therefore, the performance of this approach is 62%. Moreover, we can see that the mammograms better classified are those belonging to BIRADS I, while mammograms belonging to BIRADS IV are the worst classified. *The Cohen' kappa (κ) coefficient [20] is also included in the results. A κ coefficient equal to one means a statistically perfect model whereas a value equal to zero means every model value was different from the actual value. A commonly used interpretation of the various κ values [21] is: < 0 poor, $[0, 0.20]$ slight, $[0.21, 0.40]$ fair, $[0.41, 0.60]$ moderate, $[0.61, 0.80]$ substantial, and $[0.81, 1.00]$, almost perfect agreement.*

Table 1(b) shows the results obtained by the second approach, which is the segmentation of the breast in regions according to the distance to the skin-line. Using such approach, the results are drastically increased compared with the no-segmentation approach, obtaining 79% correct classification **and** $\kappa = 0.71$. Although the performance of all the classes are improved, the highest improvement is found in BIRADS IV.

Finally, Table 2 shows the results obtained by using a segmentation of the breast according to the internal breast tissue, explained above. Here (a) shows the results obtained by the Fuzzy C-Means approach, (b) based on the Fractal approach, and (c) using the Statistical approach. Note that the overall performance for each algorithm is similar: 82%, 81%, and 80%, **and** $\kappa = 0.75$, $\kappa = 0.73$, **and** $\kappa = 0.73$, respectively, and all are better than the results obtained by the other two approaches.

The results obtained by the Fuzzy C-Means and the Statistical approach are quite similar, with more than 90% of correct classification for mammograms

Table 1. Confusion matrix for the classification of the mammograms of MIAS database (a) without segmentation of the breast and (b) segmenting according to the distance to the skin-line.

		Aut. Classif.			
		B-I	B-II	B-III	B-IV
Truth	B-I	98	9	15	7
	B-II	11	44	20	4
	B-III	11	8	46	5
	B-IV	5	9	18	12

$\kappa = 0.47$

		Aut. Classif			
		B-I	B-II	B-III	B-IV
Truth	B-I	117	6	6	0
	B-II	10	55	14	0
	B-III	8	6	55	1
	B-IV	0	3	13	28

$\kappa = 0.71$

Table 2. Confusion matrices for MIAS mammogram classification by using the internal breast density as a segmentation strategy: (a) Fuzzy C-Means, (b) Fractal, and (c) Statistical approaches.

		Aut. Classif			
		B-I	B-II	B-III	B-IV
Truth	B-I	118	7	4	0
	B-II	8	59	9	3
	B-III	6	11	53	0
	B-IV	0	3	6	35

$\kappa = 0.75$

(a)

		Aut. Classif			
		B-I	B-II	B-III	B-IV
Truth	B-I	112	11	4	2
	B-II	9	61	7	2
	B-III	2	4	52	12
	B-IV	0	2	7	35

$\kappa = 0.73$

(b)

		Aut. Classif			
		B-I	B-II	B-III	B-IV
Truth	B-I	116	8	5	0
	B-II	6	59	10	4
	B-III	2	8	52	8
	B-IV	0	3	8	33

$\kappa = 0.73$

(c)

belonging to BIRADS I. This percentage is reduced for the rest of classes to values around 75%. In contrast, in the Fractal approach, the percentage of correct classification for BIRADS I, is lower compared to the other approaches, but the other classes are slightly better classified.

4 Discussion and Conclusions

This paper has presented five different automatic methods to classify mammograms according to the BIRADS standard. The paper has focused on the strategy used *for the breast tissue density segmentation*. Thus, we distinguished among no breast segmentation, segmentation of the breast according to the distance to the skin-line, and segmentation of the breast according to their internal breast tissue. Moreover, we compared three different ways to segment the breast according to this latter strategy.

Figure 1 shows the segmentation of the breast according to those strategies. The three last columns (which corresponds to the segmentation algorithms that use breast tissue information) show similar results, therefore the classification results for these strategies are also similar. Analysing the results in more detail, one can conclude that the fractal approach results in a pixelated segmentation, while the statistical approach obtains larger and clearly separated regions. On the other hand, the Fuzzy C-Means performance is an intermediate solution

and classification results are slightly improved compared to the other two. *The different behaviour of the algorithms can be explained by their own nature. Thus, the fractal approach recursively splits the image using a quadtree structure. When there is not sufficient variance into a quadrant, the algorithm stops splitting, thus obtaining quadrangular regions. On the other hand, with the statistical analysis of the breast, small tissue variations are insufficient to modify the assigned density class. In contrast, these small variations vary the result of the Fuzzy C-Means algorithm.*

The obtained results show that the segmentation step increases the performance of the classification, improving the results by more than 15%. Moreover, we have shown that using the segmentation according to the breast tissue outperforms the segmentation according to the distance to the skin-line. We have also noted, that the strategy used to segment the internal breast tissue do not result in major variation in the results, with the Fuzzy C-Means based results slightly better than the other ones.

Bovis and Singh [8] and Petroudi et al. [9] have classified breast tissue according to BIRADS categories. While Bovis and Singh reached 71% of correctly classified mammograms, Petroudi et al. obtained 90%, 64%, 70% and 78% of correct classification for mammograms belonging to BIRADS I, BIRADS II, BIRADS III and BIRADS IV respectively, achieving a total of 76% correct classification for all the cases. In contrast, using a segmentation according to the internal breast tissue, these results are clearly improved, obtaining 82% of correct classification. It should be noted that our results are based on a leave-one-woman-out methodology.

Acknowledgments

This work was partially supported by the Spanish Ministry of Education and Science under the grant TIN2005-08792-C03-01.

References

1. Eurostat: Health statistics atlas on mortality in the european union. Official Journal of the European Union (2002)
2. Birdwell, R., Ikeda, D., O'Shaughnessy, K., Sickles, E.: Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology* **219** (2002) 192–202
3. Freer, T., Ulissey, M.: Screening mammography with computer-aided detection: Prospective study of 12860 patients in a community breast center. *Radiology* **220** (2001) 781–786
4. Ho, W., Lam, P.: Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities. *Clinical Radiology* **58** (2003) 133–136
5. Wolfe, J.: Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* **37** (1976) 2486–2492

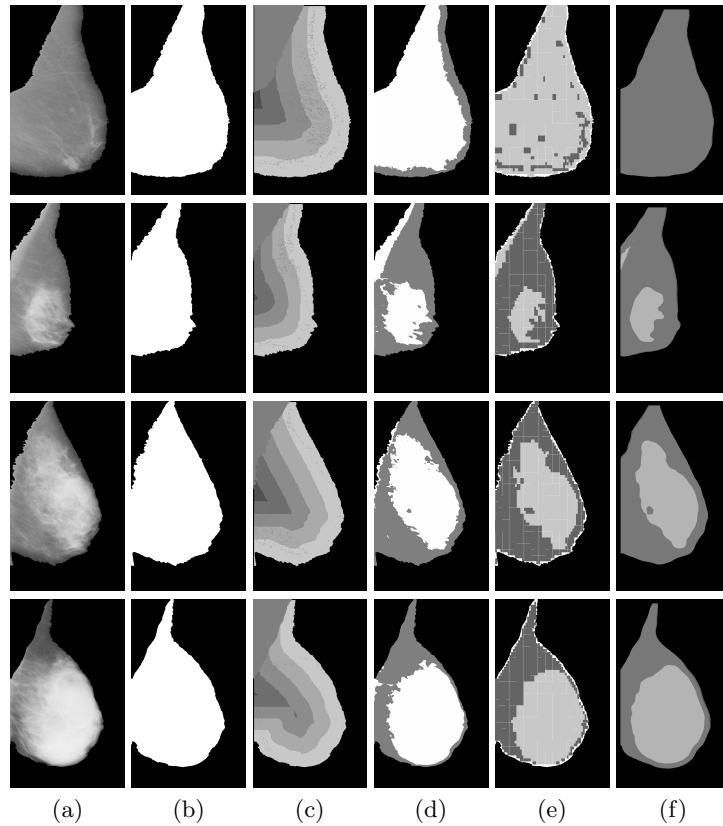


Fig. 1. The reviewed strategies for dividing a mammogram into regions in mammograms belonging to BIRADS I the upper row to BIRADS IV the lower. (b) Segmentation using a single breast area, (c) based on the distance between the pixels and the skin-line, (d) based Fuzzy C-Means clustering of pixels with similar appearance, (e) based on the fractalization of the image, and (f) using the statistical approach.

6. Miller, P., Astley, S.: Classification of breast tissue by texture analysis. *Image and Vision Computing* **10** (1992) 277–282
7. Karssemeijer, N.: Automated classification of mammographic parenchymal pattern. *Phys. Med. Biol.* **28** (1998) 365–378
8. Bovis, K., Singh, S.: Classification of mammographic breast density using a combined classifier paradigm. In: *Proc. Med. Image Underst. and Anal.* (2002) 177–180
9. Petroudi, S., Kadir, T., Brady, M.: Automatic classification of mammographic parenchymal patterns: A statistical approach. In: *IEEE Conf. Eng. Med. Biol. Soc.* Volume 1. (2003) 798–801
10. Oliver, A., Freixenet, J., Zwiggelaar, R.: Automatic classification of breast density. In: *IEEE Int. Conf. on Image Proc.* Volume II. (2005) 1258–1261
11. Bosch, A., Muñoz, X., Oliver, A., Martí, J.: Modeling and classifying breast tissue density in mammograms. In: *IEEE Conf. on Comp. Vision and Patt. Rec., New*

- York, U.S. (2006)
12. Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., Savage, J.: The Mammographic Image Analysis Society digital mammogram database. In: *Int. Work. on Dig. Mammography*. (1994) 211–221
 13. Raba, D., Oliver, A., Martí, J., Peracaula, M., Espunya, J.: Breast segmentation with pectoral muscle suppression on digital mammograms. In: *Lect. Not. in Comp. Sc. Volume 3523*. (2005) 471–478
 14. Bezdek, J.: *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
 15. Raba, D., Martí, J., Martí, R., Peracaula, M.: Breast mammography asymmetry estimation based on fractal and texture analysis. In: *Comp. Aided Rad. and Surg.*, Berlin, Germany (2005)
 16. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.* **19** (1997) 711–720
 17. Haralick, R., Shanmugan, K., Dunstein, I.: Textural features for image classification. *IEEE Trans. Syst., Man, Cybern.* **3** (1973) 610–621
 18. Duda, R., Hart, P., Stork, D.: *Pattern Classification*. 2 edn. John Wiley & Sons, New York (2001)
 19. Quinlan, J.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, New York (1993)
 20. Cohen, J.: A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20** (1960) 27–46
 21. Landis, J., Koch, G.: The measurement of observer agreement for categorical data. *Biometrics* **33** (1977) 159–174