Automatic classification of breast density according BIRADS categories using a clustering approach

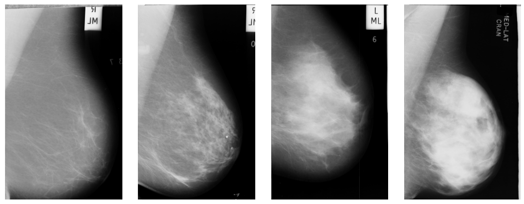
Jordi Freixenet, Arnau Oliver, Joan Martí
University of Girona
{jordif, aoliver, joanm}@eia.udg.es

Josep Pont
Josep Trueta Hospital
rd.jpont@htrueta.scs.es

Reyer Zwiggelaar
University of Wales
rrz@aber.ac.uk

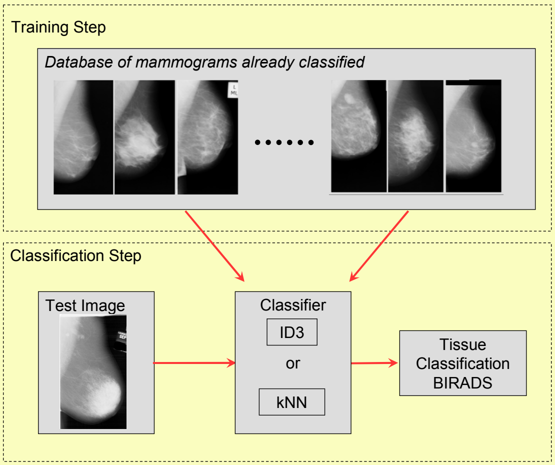
Abstract

In this work we propose a new algorithm to classify breasts according their internal tissue. Moreover, and in contrast with most of the previous works, the classification follows the BIRADS standard. Our proposal consists on grouping those pixels with similar tissue by using a clustering algorithm. Subsequently, mammograms are classified by comparing textural and morphological features extracted from each cluster



Methodology

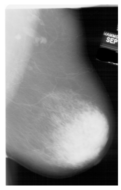
Classifier Approach



Constructing the data

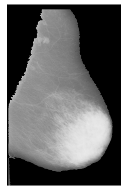
0) Database of already classified mammograms

The method needs as starting point a subset of previously classified mammograms



1) Breast Segmentation

Two-step algorithm: an adaptive threshold extracts the background and the annotations, while the pectoral muscle is segmented by using a region growing algorithm

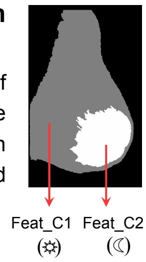


2) Cluster those pixels with similar appearance

We used Fuzzy C-Means with two seeds depicting fatty and dense tissue. Both seeds were initialized by using histogram information

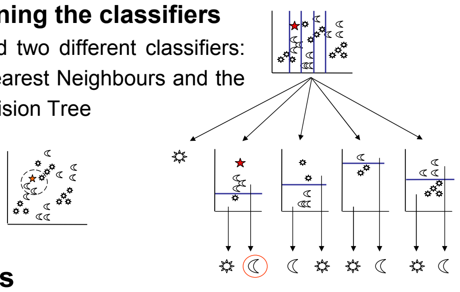
3) Extract features for both clusters

We extracted two different sets of features: morphologic set (centre of masses, relative area, mean intensity) and textural set (derived from co-occurrence matrices)



4) Training the classifiers

We used two different classifiers: the k-Nearest Neighbours and the ID3 Decision Tree



5) Fuzzy combination of both classifiers

In some cases, both classifiers give us complementary results. Hence, we constructed a new classifier which takes advantage of both approaches by converting this classifier approach into a probabilistic framework

$$P_{FzC}(mammo \in B_i) = \alpha P_{kNN}(mammo \in B_i) + (1-\alpha) P_{ID3}(mammo \in B_i)$$

$$P_{kNN}(mammo \in B_i) = \sum_{\substack{j \in B_i \\ j \in kNN}} \frac{1}{dist(j, mammo)}$$

$$P_{ID3}(mammo \in B_i) = \sum_{j \in B_i} P(mammo \in leaf_j)$$

$$P(mammo \in leaf_j) = \sum_{k\ Nodes} \frac{w_k}{dist(feat_{ID3}, feat_{mammo})}$$

Results

Testing the model

Leave-one-woman-out

We used a leave-one-out procedure: each mammogram of the set was classified by using all the other mammograms except the opposite one (the other mammogram of the same woman) in the training set

The method was tested using 300 mammograms of the DDSM database, as well as 320 mammograms of the MIAS database, which were classified in BIRADS categories by an expert

MIAS

	kNN				ID3			
	B-I	B-II	B-III	B-IV	B-I	B-II	B-III	B-IV
B-I	**57**	37	22	12	**74**	32	13	9
B-II	18	**26**	25	8	14	**39**	15	10
B-III	4	28	**32**	6	8	17	**28**	17
B-IV	4	6	17	**17**	6	9	17	**12**

	kNN		ID3	
	Low	Dense	Low	Dense
Low	**138**	68	**159**	47
Dense	42	**72**	40	**74**

DDSM

	kNN				ID3				Fc (kNN,ID3)			
	B-I	B-II	B-III	B-IV	B-I	B-II	B-III	B-IV	B-I	B-II	B-III	B-IV
B-I	**17**	24	5	4	**24**	20	3	3	**23**	21	3	3
B-II	27	**35**	30	8	34	**49**	10	7	32	**48**	12	8
B-III	3	25	**54**	18	18	17	**40**	25	3	22	**52**	23
B-IV	3	9	23	**15**	9	5	19	**17**	5	7	20	**18**

	kNN		ID3		Fc (kNN,ID3)	
	Low	Dense	Low	Dense	Low	Dense
Low	**103**	47	**127**	23	**124**	26
Dense	40	**110**	49	**101**	37	**113**

Conclusions

We have proposed a novel approach to classify mammograms according their internal tissue, and into BIRADS categories. We tested it using MIAS and DDSM databases, obtaining an overall percentage of correct classification of 66% and 73% for kNN and ID3 classifiers when testing the MIAS database, and almost 80% for the Fuzzy combination when testing the DDSM database