

Computer-aided diagnostics of screening mammography using content-based image retrieval

Thomas M. Deserno^{a1}, Michael Soiron^a, Júlia E. E. de Oliveira^b, Arnaldo de A. Araújo^c

^aDepartment of Medical Informatics, RWTH Aachen University, 52057 Aachen, Germany

^bCentro de Desenvolvimento da Tecnologia Nuclear, 31270-901 Belo Horizonte, MG, Brazil

^cDepartment of Computer Science, UFMG, 31270-010 Belo Horizonte, MG, Brazil

ABSTRACT

Breast cancer is one of the main causes of death among women in occidental countries. In the last years, screening mammography has been established worldwide for early detection of breast cancer, and computer-aided diagnostics (CAD) is being developed to assist physicians reading mammograms. A promising method for CAD is content-based image retrieval (CBIR). Recently, we have developed a classification scheme of suspicious tissue pattern based on the support vector machine (SVM). In this paper, we continue moving towards automatic CAD of screening mammography. The experiments are based on in total 10,509 radiographs that have been collected from different sources. From this, 3,375 images are provided with one and 430 radiographs with more than one chain code annotation of cancerous regions. In different experiments, this data is divided into 12 and 20 classes, distinguishing between four categories of tissue density, three categories of pathology and in the 20 class problem two categories of different types of lesions. Balancing the number of images in each class yields 233 and 45 images remaining in each of the 12 and 20 classes, respectively. Using a two-dimensional principal component analysis, features are extracted from small patches of 128 × 128 pixels and classified by means of a SVM. Overall, the accuracy of the raw classification was 61.6 % and 52.1 % for the 12 and the 20 class problem, respectively. The confusion matrices are assessed for detailed analysis. Furthermore, an implementation of a SVM-based CBIR system for CADx in screening mammography is presented. In conclusion, with a smarter patch extraction, the CBIR approach might reach precision rates that are helpful for the physicians. This, however, needs more comprehensive evaluation on clinical data.

Keywords: Content-based image retrieval (CBIR), Computer-aided diagnosis (CAD), Support vector machine (SVM), 2D Principle component analysis (PCA), Screening mammography, Breast density, Breast lesion

1. INTRODUCTION

In our world, cancer is the leading cause of death, and the World Health Organization (WHO) projects deaths from cancer worldwide to continue rising, with an estimated 11 million deaths in 2030 [1]. For women, breast cancer is the leading cause of death (age: 15 – 54). Every 13 min a woman dies on breast cancer, and 12.6 % of all women will develop breast cancer during their life [1]. In the past decade, screening mammography has been established worldwide for early detection of breast cancer. Computer-aided detection (CADE), where suspicious regions are marked to guide the radiologists, has been introduced and already proven to support screening mammography [2,3]. Computer-aided diagnostics (CADx) systems aims at providing a classification of suspicious regions, e.g., labeling the lesion according to Breast Imaging Reporting and Data System (BI-RADS) assessment categories [4]. Relying on a large repository of medical images annotated with ground truth data, content-based image retrieval (CBIR) also is seen as promising technology in assisting diagnosis [5].

In a previous work, we have reviewed the state-of-the-art in CBIR-based CAD for mammography [6]. Table 1 summarizes the state-of-the-art in breast cancer CADE and CADx. The techniques usually are applied on a region of interest (ROI) and include artificial neural network (ANN), linear discriminant analysis (LDA), principle component

¹ Corresponding author: Thomas M. Deserno (nè Lehmann), Department of Medical Informatics, RWTH Aachen University, Pauwelsstr. 30, D - 52057 Aachen, Germany, email: deserno@ieee.org; phone: +49 241 80 88793, fax: +49 241 80 33 88793.

Author	Ref.	Year	Problem	Classes	Features	Classifier	Images	Source	Results
Bovis & Sing	[12]	2002	Tissue	2	ROI	ANN	377	DDSM	40 % – 71 %
Oliver et al.	[13]	2009	Tissue	2	ROI	LDA / PCA	54 322	private MIAS	90 %
Tagliafico et al.	[14]	2009	Tissue	4	Histogram	Thresholding	160	private	80 % – 90 %
Subashini et al.	[15]	2010	Tissue	3	ROI	SVM radial kernel	43	private	95 %
Eltonsy et al.	[16]	2007	Lesion	2	ROI	Concentric layer model	540	DDSM	96 %
Elter & Hasslmeyer	[17]	2008	Lesion	2	ROI, meta	Generic algorithm Euclidian metric	360	DDSM	86 % (ROC)
Tao et al.	[18]	2010	Lesion	1	ROI	Multi-phase, pixel level	54	private	69 %
Verma et al.	[19]	2010	Lesion	2	ROI	Soft-clustered direct learning	200	DDSM	97 %
Mazurowski et al.	[9]	2011	Lesion	2	ROI	Random mutation hill climbing	1,852	DDSM	49 % – 83 %
Wei et al.	[8]	2011	Lesion	2	ROI	SVM radial kernel	2,563	DDSM	72% – 74 %
Tao et al.	[20]	2011	Lesion	2	ROI	Curvature scale space, CBIR Local linear embedding metric	415 476	private private	75 % (ROC) 80 % (ROC)
Oliver et al.	[21]	2010	Both	4	ROI	LDA / PCA	184	private	92 % – 94 %
Lesniak et al.	[11]	2011	Both	2	ROI	SVM radial kernel	10,397	private	66 % – 67 %
Sadaf et al.	[10]	2011	Both	10	ROI	Non-statistical analysis	127	private	91 %
Deserno et al.	[6]	2011	Both	12	ROI	SVM Gaussian kernel	2,796	IRMA	80 %

Table 1: State of the art in computer-sided diagnosis of screening mammography. The acronyms refer to:

ANN = artificial neural network; CBIR = content-based image retrieval; DDSM = Digital Database for Screening Mammography; IRMA = Image Retrieval in Medical Applications; MIAS = Mammographic Image Analysis Society; ROC = receiver operator characteristics; ROI = region of interest; LDA = linear discriminant analysis; PCA = principle component analysis; SVM = support vector machine

analysis (PCA), and the support vector machine (SVM). Results are reported in terms of accuracy, precision, sensitivity and area under the receiver operator characteristics (ROC) curve. They are based on data from the Digital Database for Screening Mammography (DDSM), the Mammographic Image Analysis Society (MIAS), and the Image Retrieval in Medical Applications (IRMA) project [7].

In addition to the papers reviewed in [6], Wei et al. [8] have recently performed experiments on 2,563 DDSM images with different lesions using a SVM. A relevance feedback learning method is presented. A three-dimensional feature vector is used to describe the lesions containing the shape of the mass and the margin and the density of the grey values. The results show that Gaussian kernel-based SVM delivers the best result compared to polynomial and spline kernels. The precision varies around 70 % depending on the given problem.

Furthermore, Mazurowski et al. [9] showed a template-based recognition algorithm for breast masses. Their data is built from 1,852 images, partly taken from the DDSM collection and from the Duke University Medical Center, Durham, NC, USA. The algorithm calculates likelihood maps matching templates with masses to identify the malignant region. The results in sensitivity vary from 83.2 % \pm 4.4 % for malignant masses and 71.9 % \pm 4.7 % for average cases to 0.503 % \pm 0.204 % for cases with architectural distortions.

Sadaf et al. [10] evaluated the performance of different systems on 127 mammograms from internal sources. The results vary for different classes. The best results were observed at 100% with calcifications and transparent tissue. The worst results come from mammograms with architectural distortions and are at 75 %.

At last years SPIE meeting, Lesniak et al. [11] compared the performance of an SVM-based classification to several nearest neighbor (NN) algorithms. Their images collect to 10,397 mammographic examinations from a private source and were split into two classes. NN searches result from 23 % to 64.8 %, while the SVM-based classification was found superior providing an accuracy of 66.0 % – 67.6 %.

Combining CBIR and CAD, Tao et al. established a mass retrieval platform [20] with the performance of two experiments. The first one used a database consisting of 415 mammograms containing masses (244 malignant and 171

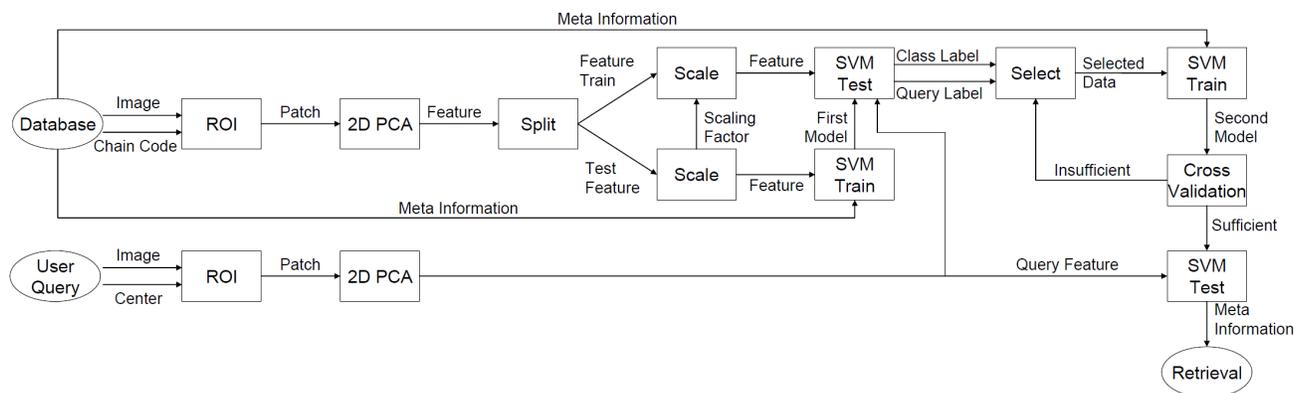


Figure 1: Flowchart of the proposed CBIR algorithm

benign) from University of Michigan, and the second one used a reference library non identified containing 476 masses (219 malignant and 257 benign). Prior to feature extraction, a segmentation of the masses was performed using the multi-level learning-based segmentation method [18]. Regular, lobulated, and irregular shapes were computed with the following features: third order moments, curvature scale space descriptors (CCSD), radial length statistics, and region-based shape. Regarding mass margin, they considered circumscribed, microlobulated, indistinct, and spiculated, extracting texture features from them. Similarity was calculated using a locally linear embedding (LLE) method with the performance indicated by a ROC analysis. For the first database, an area under the ROC curve of 75 % was obtained, and for the second dataset, an average area of 80 % was achieved.

Despite the recent reports of Mazurowski et al., Wei et al., or Lesniak et al., it becomes obvious that previous experiments have been performed on different, rather small, and also selective or private datasets. Selection procedures as well as separation of test and training data are ambiguous, in particular with respect to the small sample sizes. Furthermore, most authors use private data, not being available for others, and manifesting the need of large image repositories as it has been claimed by others, too [22]. Another problem arises from the small number of categories, disregarding whether tissue, assessment, or mixed classes are considered. In most papers, a simple two-class problem is investigated, which reflects clinical practice insufficiently.

Therefore, we analyze up to 20 diagnostic relevant classes, which are related to texture pattern extracted from a large database of more than 10.000 reference images with proven ground truth. In this paper, we present a novel testing environment for screening mammography that is based on a large set of ground truth data of up to 20 diagnostic classes and applying a combination of multi-class SVM and CBIR methods to support CADx.

2. METHODS

Based on the work described in [6], a novel scheme for SVM/CBIR-based CADx is proposed (Fig. 1). In particular, the following steps are performed:

- *ROI location:* Conserving their aspect ratio, all mammographic images have been reduced in pixel resolution to fit a bounding of 1024 x 1024 pixel, and 128 x 128 pixel patches were extracted centering the lesion. For images of BI-RADS category 1, the patches were positioned arbitrarily within the breast tissue, excluding the pectoral muscle, nipple and background areas.
- *Feature extraction:* The two-dimensional principle component analysis (2DPCA) method was performed on all the patches. The principal components related to the first (i.e., the largest) eigenvector of the covariance matrix, composed of 128 values, were used in the experiments.
- *Split & scale:* The data is divided into 60 % for training and 40 % for testing. To improve the performance of the SVM classifiers, the features are scaled into the interval $[-1, \dots, 1]$, and the scaling factor is used to normalize the testing features, too.

- *First model*: 5-fold cross validation is performed on the 60 % training data in order to obtain the best parameters of the SVM, which then are used to build the first model. All hyper planes are stored in a matrix.
- *Selection*: The user chooses a query image that also goes through the previous steps using its classification result and its distance to the hyper-plane, the data is then divided into a relevant and a non-relevant class.
- *Second model*: These classes are used to train the SVM producing a second model. This first SVM model is used for classification of the query image on the test data. A two class problem with the classes relevant and not relevant to the query is created.
- *Evaluation*: If the initial classification of the query is incorrect, this will yield poor overall results. The selection process is therefore repeated with another algorithm if the second model performs to bad in the cross validation.
- *Retrieval*: Using the second model, the whole dataset is tested and the best resulting relevant images are presented to the user.

The system was implemented using MatLab through the image processing toolboxes and the LIBSVM library (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>). Feature extraction was executed on an IntelCore2Quad 2.66 GHz processor with 8 GB of RAM operated with Microsoft Windows 64 bit system. Image retrieval was performed on an Intel-Core2Duo 2 GHz processor with 3 GB of RAM under Microsoft Windows 32 bit.

3. RESULTS

We have performed several experiments to analyze the performances of the 2DPCA feature extraction method, the classification based on combined SVM classifiers, and the CBIR-based CADx approach using two SVM models.

3.1. Experiments with 2DPCA

In theory a 2DPCA feature should be a good description of the image it is extracted from. Previous experiments [23, 24] showed that it delivers quite satisfying results using only the first 4 principal components, which decreased the runtime of the experiments, significantly. Since an incomplete set of eigenvectors loses information about the original image, it seems necessary to fully understand the dependencies.

To test if the 2DPCA works for the SVM, we performed several tests with visually different image classes. Photographs of stamps and radiographs of human hands were scaled down to the size of the patches delivered from mammography. The results are close to 100 % for each test. In comparison, each class is tested against all other classes. The mean value was 97.87 % with a variance of 4.97 % for the stamps and 97.61 % \pm 5.97 % for the hands. The one against all experiments delivered a mean of 71 % \pm 7.4 % (Tab. 2). Since all results are close to 100 % we can assume that the features used deliver sufficient information for the SVM to separate different image classes. Also, it seems that the medical relevant classes are more difficult to classify than obviously different classes.

To test if the 2DPCA appropriately works for the SVM, we compared the performance of 2DPCA-based features with simple two-dimensional feature vector composed of mean and standard deviation of the pixel's gray scales within the patches. Based on twelve classes (Tab. 2), the accuracy increased about 8.3 percentage points when using the advanced 2DPCA features. This improvement is lower than expected and might indicate that the overall performance is improved using more sophisticated features.

3.2. Evaluation of multiple stages classification

From the previous validation experiment, we concluded that SVM and feature extraction are working for simple tasks. However, we need to confirm that the classification is good enough to distinguish BI-RADS lesions and tissue types. Hence, we have tested how the SVM splits classes that – in a first approximation – should not be separable. The main reason to do this is to exclude systematic problems which perhaps would cause a high accuracy independent from the real problem. Accordingly, datasets of one class were divided into two sets, labeled differently and then classified. Since

Mammography	Stamps	Hands	One-against-others	One-against-self
I-1	100.0 %	99.47 %	81.26 %	51.28 %
I-2	98.40 %	84.68 %	75.63 %	64.10 %
I-5	95.75 %	100.0 %	78.31 %	61.54 %
II-1	94.68 %	94.68 %	72.64 %	47.09 %
II-2	97.34 %	97.34 %	69.81 %	55.56 %
II-5	99.47 %	96.80 %	67.54 %	47.86 %
III-1	96.80 %	96.80 %	60.39 %	55.56 %
III-2	97.34 %	100.0 %	59.87 %	43.59 %
III-5	100.0 %	97.34 %	61.54 %	54.70 %
IV-1	97.34 %	100.0 %	67.56 %	64.10 %
IV-2	97.34 %	99.47 %	69.75 %	64.96 %
IV-5	100.0 %	94.68 %	87.71 %	50.43 %
Mean	97.87 % \pm 4.97 %	97.61 % \pm 5.97 %	71.00% \pm 7.40 %	54.42 % \pm 7.33 %

Table 2: Accuracy of SVM classification using 2DPCA features. The patches from screening mammography are coded according to the BI-RADS tissue density classes: I = almost entirely fatty; II = scattered fibro glandular; III = heterogeneously dense; IV = extremely dense; the assessment categories: 1 = no finding (negative); 2 = benign; 5 = highly suggested malignant

we didn't expect any systematic, the result of these experiments should be close to 50 %. This is confirmed by our results, yielding a mean of 54.42 % \pm 7.33 % (Tab. 2).

3.3. Experiments with 20 classes

Calcification and masses yield different pattern in mammography, and from the medical point of view, deciding on malignancy must be accompanied with appropriate classification of the lesion type. Within the IRMA reference data, these types of lesions are precisely determined. Therefore, we extended the experiments to 20 classes, where – with respect to the tissue type – the lesions are explicitly modeled. To have equal numbers of images in each class, the size of the data reduced to 45 instances per class, which leaves us a testing sample of 18 images (40 %). The overall accuracy yielded 52.1 %, which is significant lower than in the previous experiments. This is expected because it is harder for the classifier to split between masses and calcifications than just identify pathology.

Table 3 shows the confusion matrix for the experiments with 20 classes. As we have seen in the previous experiments, low tissue density is easier to classify. Tissue type IV does not appear to be separable in this experiment. This finding confirms that of previous investigations [6]. Also the classifier does not only switch the different mass types with each other but also fails to recognize the right tissue and pathology. The 20 class model therefore seems to be unsuitable for further applications as classifier.

3.4. CBIR-based CADx approach using two SVM models

Since the feature extraction was already implemented in Matlab, these feature were used to run the algorithm on a selected dataset. The main algorithm is implemented in c++ and the user front end is a web interface based on PHP and Smarty Templates, as provided within the IRMA framework [25].

As it is depicted in Fig. 2a, the user first selects one of the images within the dataset, which consists of 233 images per class resulting in a total number of 2,796 images (12 class experiment). The images are sorted by class and the class is indicated to the user. In the next step, the selection of the user is used as query image. Using the algorithm described in Section 2 (Fig. 1), a retrieval image is returned (Fig. 2b). The image classes from the ground truth are printed on each response image along with a score. This score is related to the distance from the hyper-plane of the second classifier. Only the best 40 results – sorted by their score – are shown (relevance facts). However, the presented algorithm does not support learning on the user's feedback. A query composed of several images is more reliable than querying only one

	I-1	I-2c	I-2m	I-5c	I-5m	II-1	II-2c	II-2m	II-5c	II-5m	III-1	III-2c	III-2m	III-5c	III-5m	IV-1	IV-2c	IV-2m	IV-5c	IV-5m
I-1	15						3													
I-2c	2	14			1	1														
I-2m			14	2																
I-5c			6	3							5	2								1
I-5m					12		1			2		1	2							2
II-1	7	3				3	3									2				
II-2c	6					1	11													
II-2m	4	1						12		1										
II-5c									10		4		1						1	2
II-5m							1			16										1
III-1			1								8	5							1	3
III-2c					2						5	11								
III-2m					4						4	2	1					5		2
III-5c	2	1												6	3	6				
III-5m	1													2	13	1				1
IV-1	1						1								1	15				
IV-2c			2	2							5	4	1					4		
IV-2m			3						3				1							11
IV-5c	5	3			2		2	1							4					1
IV-5m	4		3						3		2	1						3		2

Table 3: Confusion matrix of 20 class experiment (the corresponding confusion matrix for the 12 class experiment can be obtained from [6]). The pathology classification codes are: c = calcification; m = mass.

image. Therefore, the user can interactively indicate relevant images (query refinement) by moving the feedback slider on the selected images appropriately (relevance feedback). Using IRMA-based interfaces, positive feedback is pointed to the right and negative feedback to the left, where the amplitude indicates the relevance. Retrieval with multiple images is shown in Fig. 2c. It is to be mentioned that a non-consistent set of query images delivers bad or no result, since the SVM cannot form reliable relevant/non-relevant classes.

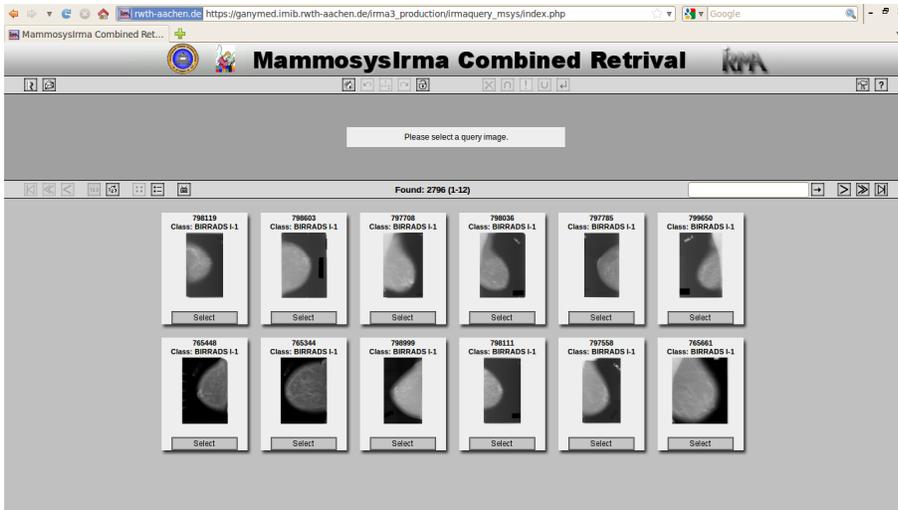
3.5. Evaluation of CBIR-based CADx

In contrast to automatic classification, the physician stays in the loop when interacting with a CBIR system. Therefore, CBIR-based diagnosis assistance is often required to display at least one image of the correct class within all retrieved images that are shown on the screen. For each of the 12 classes, we depicted arbitrarily one image and performed the retrieval experiment. The rate of good retrievals was obtained from repeating 100 experiments for each class, and counting “good” if at least one image was within the return set of the CBIR engine.

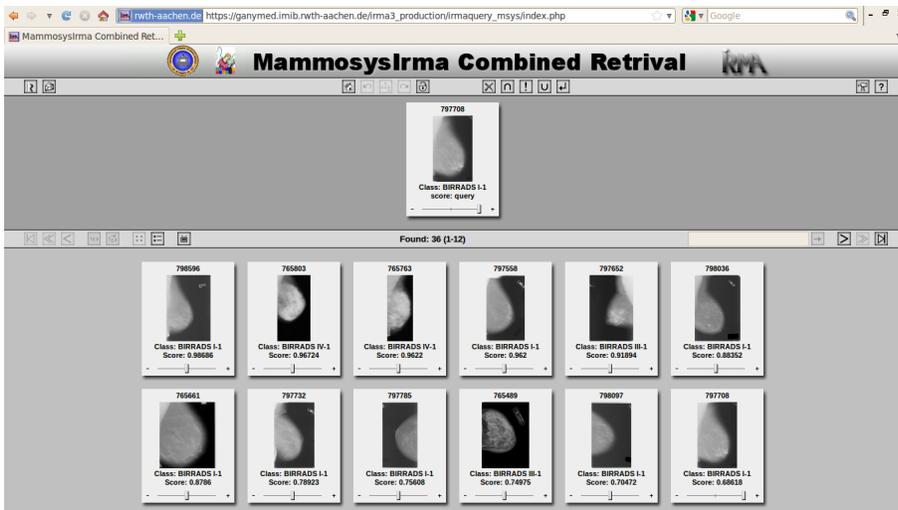
Figure 3 summarized the result. The best results are obtained from class I-1, and class II-2 performs worse. In any case, the correct class is displayed to the physician if at least 13 images are displayed. In a usual setting depicting ten result images, the correct class is given with a likelihood above 95%, which may yield the requirements in diagnostics.

4. CONCLUSION

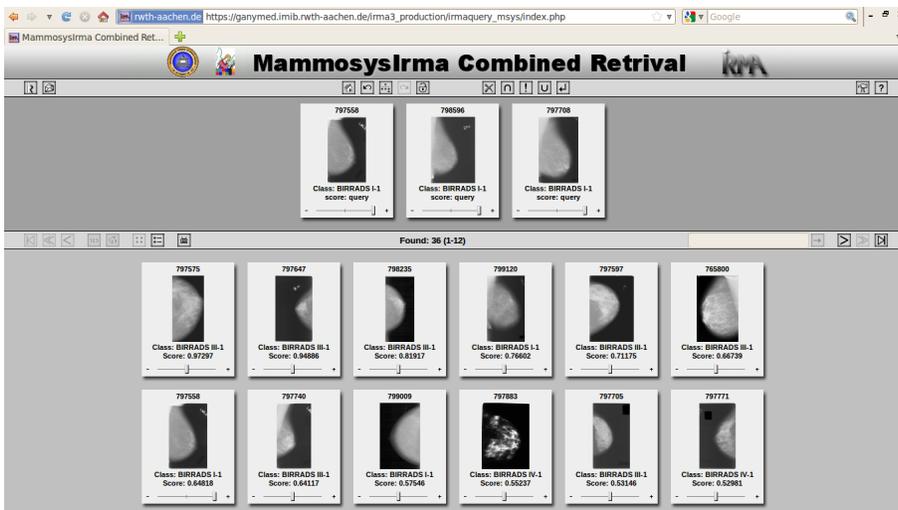
In conclusion, we have presented comprehensive experiments based on a large database, and using a high number of classes, including a non-finding class of arbitrarily extracted patches with healthy tissues. While automated classification in terms of automatic diagnostics yielded poor performance, the CBIR-based approach, where the physician interacts with the system inspecting the patches along with the according validated diagnosis, might be considered as helpful for clinical routine.



(a) Selection of a query



(b) Result of one image query



(c) Result of multiple image query

Figure 2: GUI of the IRMA-based implementation of CBIR for CADx in mammography

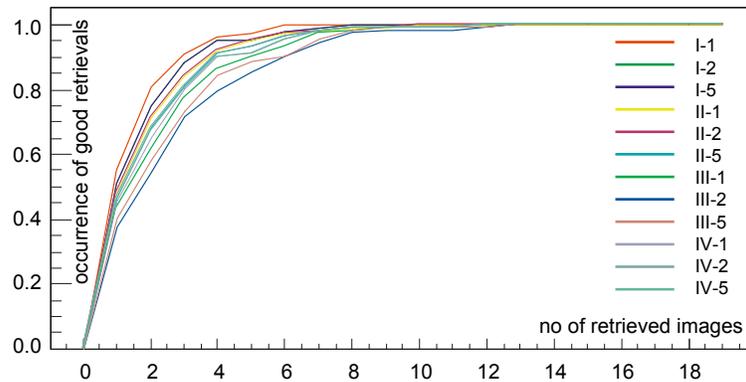


Figure 3: Result of repeated CBIR-based CADx (n = 100 repetitions).
The curves from classes II-1 tie II-2, II-5 tie IV-2, and I-2 tie I-5.

In future, moving towards a further improved CADx in mammography, the patch extraction can be improved, the image database with ground truth information should be extended, and the different perspectives needs to be coupled in order to improve the diagnostic value of mammographic CBIR. In another step, different modalities and meta-information might be merged for a combined image and text retrieval [26].

5. ACKNOWLEDGEMENT

This work is supported by CAPES and CNPq, the Brazilian research funding agencies, and the Federal Ministry of Education and Research Germany (DLR-BRA 009/045). The IRMA project has been funded by the German Research Foundation (DFG), Le 1108/4 and Le 1108/9.

REFERENCES

1. World Health Organization (ed). Cancer. WHO Fact Sheet 2011; no. 297, last assessed on 2011-08-04 at <http://www.who.int/mediacentre/factsheets/fs297/en/>
2. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology* 2009; 253: 9–16
3. Tang J, Rangayyan RM, Xu J, Naqa IE, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed* 2009; 13: 236–51
4. Elter M, Horsch A. CADx of mammographic masses and clustered micro-calcifications: a review. *Med Phys* 2009; 36: 2052–68
5. Muller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Int J Med Inform* 2004; 73: 1–23
6. Deserno TM, Soiron M, de Oliveira JEE, de A. Araujo A. Towards computer-aided diagnostics of screening mammography using content-based image retrieval. *Proceedings SIBGRAPHI 2011*; pp. 211-9
7. Oliveira JEE, Güld MO, de A. Araújo A, Ott B, Deserno TD. Towards a standard reference database for computer-aided mammography. *Procs SPIE* 2008; 6915: 1Y1-1Y9.
8. Wei C, Li Y, Huang P. Mammogram retrieval through machine learning within BI-RADS standards. *J Biomed Inform* 2011; 44: 607-14.
9. Mazurowski MA, Lo JY, Harrawood BP, Tourassi GD. Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis. *J Biomed Inform* 2011; 44(5): 815-23.
10. Sadaf A, Crystal P, Scaranelo A, Helbich T. Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. *Eur J Radiol* 2011; 77(3): 457-61.

11. Lesniak J, Hupse R, Kallenberg M, Samulski M, Blanc R, Karssemeijer N, Székely G. Computer aided detection of breast masses in mammography using support vector machine classification. Proceedings of SPIE 2011, in press
12. Bovis K, Singh S. Classification of mammographic breast density using a combined classifier paradigm. Proc MIUA (Medical Image Understanding and Analysis), 2002.
13. Oliver A, Lladó X, Pérez E, Pont J, Denton E, Freixenet J, Marti J. A statistical approach for breast density segmentation. J Digit Imaging 2009; 23: 55–65.
14. Tagliafico A, Tagliafico S, Tosto F, Chiesa C, Martinoli LE, Calabrese M. Mammographic density estimation: comparison among BIRADS categories, a semi-automated software and a fully automated one. Breast 2009; 18: 35–40
15. Subashini TS, Ramalingam V, Palanivel S. Automated assessment of breast tissue density in digital mammograms. Comput Vis Image Underst 2010; 114(1): 33–43
16. Eltonsy NH, Tourassi GD, Elmaghraby AS. A con-centric morphology model for the detection of masses in mammography,” IEEE Trans on Med Imag 2007; 26: 880–889
17. Elter M, Hasslmeyer E. A knowledge-based approach to CADx of mammographic masses. Proc SPIE 2008; 6915: 0L
18. Tao Y, Lo SCB, Freedman MT, Makariou E, Xuan J. Multilevel learning-based segmentation of ill-defined and speculated masses in mammograms. Med Phys 2010; 37(11): 5993
19. Verma B, McLeod P, Klevansky A. Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. Expert Syst Appl 2010; 37(4): 3344–51
20. Tao Y, Lo SCB, Hadjiski L, Chan HP, Freedman TM. Birads guided mammographic mass retrieval. Proc SPIE 2011; 7963: 2
21. Oliver A, Lladó X, Freixenet J, Marti R, Pérez E, Pont J, Zwiggelaar R. Influence of using manual or automatic breast density information in a mass detection cad system. Acad Radiol 2010; 17(7): 877–83
22. Horsch A, Hapfelmeier A, Elter M. Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies. Int J Comput Assist Radiol Surg, 2011; DOI 10.1007/s11548-011-0553-9
23. de Oliveira JEE, Machado AMC Chavez C, Lopes APB, Deserno TM, de A. Araujo A. MammoSys: a content-based image retrieval system using breast density patterns. Comput Meth Prog Biomed 2010; 99: 289–97
24. de Oliveira JEE, de A. Araujo A, Deserno TM. Content-based image retrieval applied to BI-RADS tissue classification in screening mammography. World J Radiol 2011; 3(1): 24–31
25. Deserno TM, Güld MO, Plodowski B, Spitzer K, Wein BB, Schubert H, Ney H, Seidl T. Extended query refinement for medical image retrieval. J Digit Imaging 2008; 21(3): 280-9
26. Muller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications. Clinical benefits and future directions. Int J Med Inform 2004;73(1):1-23.