A multi-image approach to CADx of breast cancer with integration into PACS

Matthias Elter[a], Thomas Wittenberg[a], Rüdiger Schulz-Wendtland[b] and Thomas M. Deserno[c]

[a]Image Processing and Medical Engineering Department,
Fraunhofer-Institute for Integrated Circuits IIS,
Am Wolfsmantel 33, 91058 Erlangen, Germany
[b]Institute of Radiology, Gynaecological Radiology,
University Erlangen-Nuremberg,
Universitätsstraße 21-23, 91054 Erlangen, Germany
[c]Department of Medical Informatics,
RWTH Aachen University,
Pauwelsstr. 30, 52057 Aachen, Germany

ABSTRACT

While screening mammography is accepted as the most adequate technique for the early detection of breast cancer, its low positive predictive value leads to many breast biopsies performed on benign lesions. Therefore, we have previously developed a knowledge-based system for computer-aided diagnosis (CADx) of mammographic lesions. It supports the radiologist in the discrimination of benign and malignant lesions. So far, our approach operates on the lesion level and employs the paradigm of content-based image retrieval (CBIR). Similar lesions with known diagnosis are retrieved automatically from a library of references. However, radiologists base their diagnostic decisions on additional resources, such as related mammographic projections, other modalities (e.g. ultrasound, MRI), and clinical data. Nonetheless, most CADx systems disregard the relation between the craniocaudal (CC) and mediolateral-oblique (MLO) views of conventional mammography. Therefore, we extend our approach to the full case level: (i) Multi-frame features are developed that jointly describe a lesion in different views of mammography. Taking into account the geometric relation between different images, these features can also be extracted from multi-modal data; (ii) the CADx system architecture is extended appropriately; (iii) the CADx system is integrated into the radiology information system (RIS) and the picture archiving and communication system (PACS). Here, the framework for image retrieval in medical applications (IRMA) is used to support access to the patient's health care record. Of particular interest is the application of the proposed CADx system to digital breast tomosynthesis (DBT), which has the potential to succeed digital mammography as the standard technique for breast cancer screening. The proposed system is a natural extension of CADx approaches that integrate only two modalities. However, we are still collecting a large enough database of breast lesions with images from multiple modalities to evaluate the benefits of the proposed approach on.

Keywords: Breast cancer, CADx, PACS, RIS, mammography

1. INTRODUCTION

To support a radiologist with the diagnosis of mammographic lesions, we have recently developed a knowledge-based system for the computer-assisted diagnosis (CADx) of mammographic masses.[1–3] Within this CADx-system, content-based image retrieval (CBIR) methods as well as case-based reasoning (CBR) techniques[4] are applied to retrieve similar lesions with known diagnoses from a library, an adequately organized picture archiving

Further author information: (Send correspondence to M. Elter.)
M. Elter: E-mail: matthias.elter@iis.fraunhofer.de, Telephone: +49 9131 7767327
T. Wittenberg: E-mail: thomas.wittenberg@iis.fraunhofer.de, Telephone: +49 9131 7767330
R. Schulz-Wendtland: E-mail: ruediger.schulz-wendtland@idr.imed.uni-erlangen.de, Telephone: +49 9131 8533460
T. Deserno: E-mail: tdeserno@mi.rwth-aachen.de, Telephone: +49 241 8088793

system or from an annotated image data base of reference lesions.[5] Based on the histological validated ground truth diagnoses of the retrieved lesions, a diagnosis proposal for the unknown query lesion is generated and displayed to the radiologist. Furthermore, to make the reasoning process more transparent to the operator, the retrieved lesions, their corresponding diagnoses and additional annotations[6] are presented to the radiologist in an adequate way together with the generated diagnosis proposal.

While our CADx system currently works only on the lesion level for the retrieval of related images and information, the diagnosis of a suspicious mammographic lesion is usually based on further information. Such additional information consists of demographic and anamnestic data available from the patients health care record, the appearance and location of the lesions in corresponding mammographic views (e.g. the craniocaudal (CC) and mediolateral-oblique (MLO) views), and the appearance of the lesions in additional imaging modalities, such as ultrasound or MRI. In other words, to discriminate benign and malignant mammographic lesions, the experienced radiologist does not operate on the lesion level (as most CADx-systems do) but on the full case level, and thus integrates all available sources of information into the process of decision making. Several groups have already proposed CADx approaches that integrate at least some of the modalities and other information sources mentioned above. Veldkamp et al.[7] integrate features extracted from two mammographic views for the discrimination of benign and malignant mammographic microcalcifications. Gupta et al.[8] integrate texture features from two mammographic views for the classification of mammographic masses. The work of Drukker et al.[9] is one of the few examples of CADx approaches that integrate lesion features from multiple modalities. More precisely, they integrate features extracted from ultrasound images and multiple mammographic views to classify benign and malignant breast lesions. Although their case database is relatively small (100 lesions), they conclude that computerized classification of cancer significantly improves when lesion features from ultrasound and mammography are combined.

Digital mammography systems are currently the gold-standard for the acquisition of mammograms. In the past years, they have been established worldwide as the successors of analogue film-based mammography systems. The characterization of mammographic masses based on analogue or digital mammograms is based on a pair of two-dimensional X-ray images of each breast obtained from two different angles (e.g. CC and MLO). Since mammographic masses are often more dense than their surrounding tissue, they usually show up as irregular white areas commonly referred to as shadows. Within two-dimensional X-ray images, such shadows are often hidden behind a dense breast tissue, which leads to a superposition of mass and tissue in the resulting mammogram. Thus, manual or computer assisted reading of such mammograms are often error-prone, as such a superpositions may either result in a false-negative or a false-positive diagnosis. However, digital breast tomosynthesis (DBT) systems, which have been recently introduced as a new imaging modality for mammography have a high potential to succeed digital mammography systems as the standard imaging technique for breast cancer screening in the next decade.[10–13] With DBT, low-dose mammograms are acquired at a number of different projections over a limited range, and the 3D breast volume can thus be reconstructed.[14] Hence, since DBT allows the reconstruction of a stack of two-dimensional slices, the depicted mammographic lesions are free of tissue superpositions and can thus can be better interpreted by both, man and machine.[15] Even though only prototype DBT systems are available so far, considerable work on the computer aided detection (CADe) of breast lesions in DBT images has already been done.[16–20] However, work on the computer aided diagnosis (CADx) based on DBT data is sparse.[21]

Considering these observations and developments, the next generation of CADx systems for breast cancer diagnosis are expected to mimic the process of decision making and diagnosis of radiologists more closely and thus will rely on the full case data instead of isolated images from the lesion level. Furthermore, we expect, that DBT will soon be FDA-approved and thus be available for a broader clinical use. Accordingly this paper describes the required extensions to extend our CADx-system from the lesion to the case level.

2. MATERIALS AND METHODS

From our point of view, three main extensions are required to migrate CADx for mammographic lesions from the lesion level to the full case level:

1. The characteristic features describing a lesion, which are currently extracted from relevant image regions from two-dimensional mammograms, must be extended to be extracted from several sources. These data

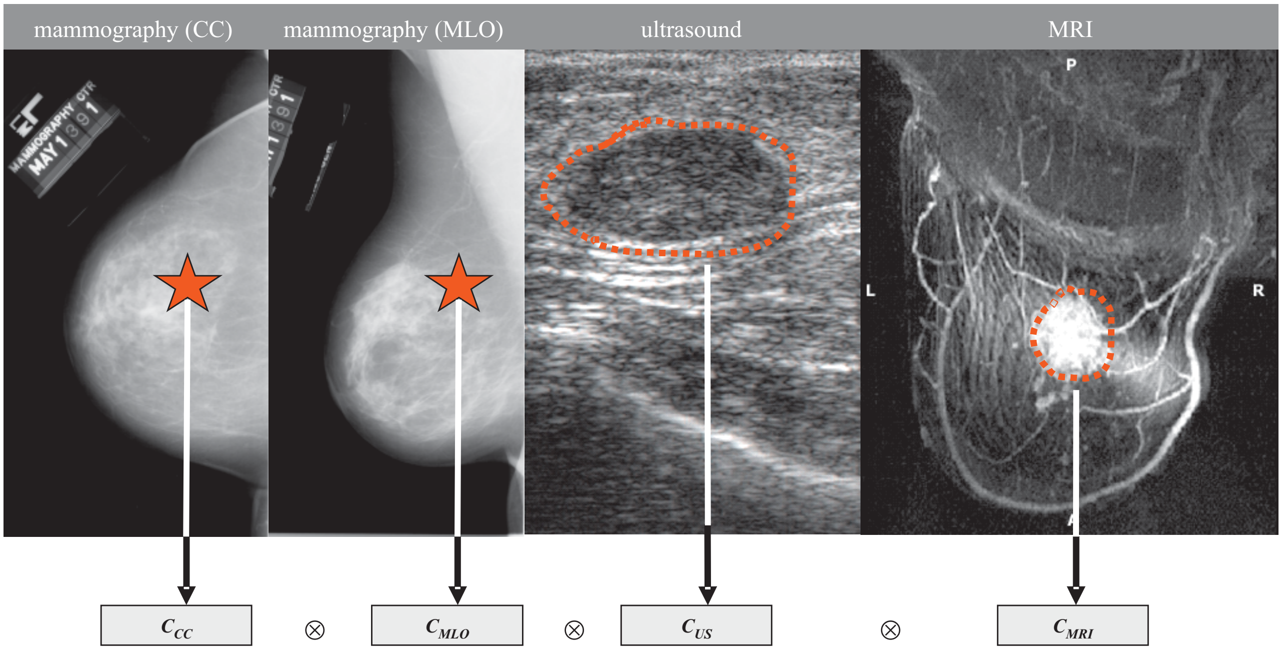


Figure 1. Generation of an extended feature vector characterizing a mass lesion from multiple modalities, including the craniocaudal (CC) and mediolatreal-oblique (MLO) projections of mammograms, ultrasound and MRI.

sources feeding a highly dimensional feature vector include the description of the same lesion from multiple 2D mammographic projections (CC and MLO), multiple modalities (mammography, ultrasound, MRI) as well as multiple frames of the 3D image stacks from DBT and MRI. Furthermore, available demographic and anamnestic patient data must be considered as an essential part of the extended feature vector.

2. The architecture of the CADx system must be extended in such a way as to allow the fusion of characteristic attributes from multi-medial sources related to the mammographic case. These sources include annotated and validated reference images, alphanumerical text as well as class indices taken from ontologies or standardized vocabularies such as the Breast Imaging - Reporting And Data System (BI-RADS^{TM}) standard of the American College of Radiology (ACR).[22] The work of Drukker et al.[9] suggests that classification performance strongly depends on the specific method for combining features from multiple images of a lesion.

3. Finally, in order to access and retrieve the additional information sources and thus support the radiologist during the process of clinical decision making, the case-based CADx approach must be integrated into the infrastructure of the radiology information system (RIS), as well as the picture archiving and communication system (PACS). It should also support the integration with health care enterprise (IHE) profiles.

2.1 Multi-image features

The first migration stage of the proposed extended CADx system will allow to extract and combine characteristic lesion attributes from corresponding regions-of-interest (ROIs) depicting the lesion in different projections in two-dimensional mammograms, such as the CC and MLO views. This approach mimics the idea that a radiologist will always use both projections of a breast for the diagnosis of a mammographic mass to partially resolve superpositions of masses and tissue. This stage of course requires linking regions of interest in different mammographic views that correspond to the same lesion. We have implemented an approach based on the method proposed by Engeland et al.[23] to accomplish this task. Furthermore, optional features from the same mass lesion obtained from additional breast imaging modalities will be able to be added to the feature vector on demand (cp. Figure 1). Ultrasound and MRI are two common imaging modalities, which are routinely used by radiologists to provide further information about suspicious breast lesions previously detected in mammograms.[24] As the superposition of masses and surrounding breast tissue is especially problematic for an image-based computer-assisted diagnosis of mammographic masses, it is important that characteristic features obtained from a stack of superposition-free two-dimensional slices from DBT images can be integrated and added to the extended feature vector. Even though we will provide this capability to our system, we expect that a broad distribution, availability, and acceptance of DBT will still take a couple of years. Furthermore, all available demographic (sex, age, race, etc.) and anamnestic data (smoker, children, inheritance, etc.) about the patient must be described as categorical data and be considered as an essential part of the extended feature vector. Previous experiments have shown that clinical decision making based on such patient data can significantly improve the diagnosis suggestions of a CADx system.[25]

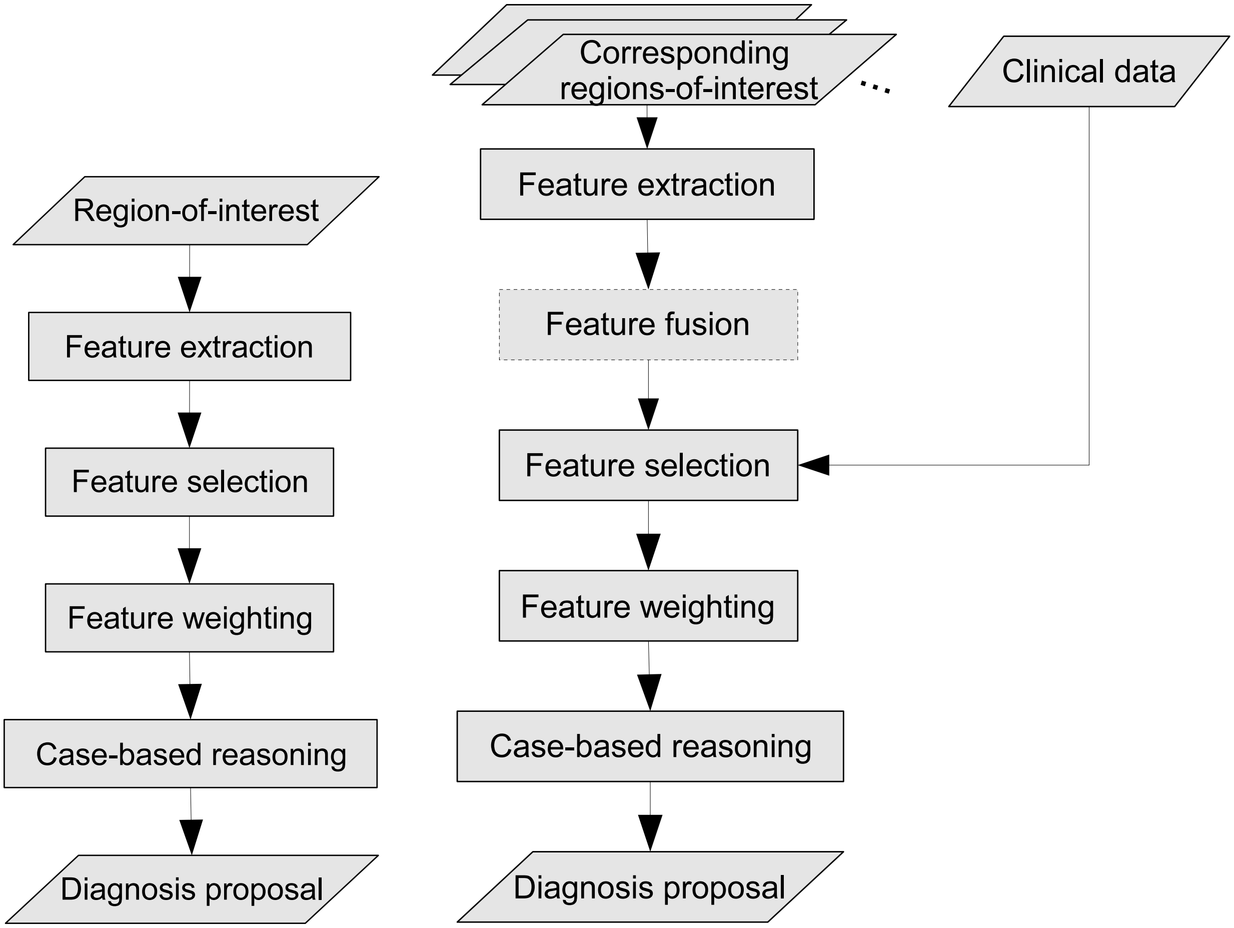
2.2 CADx system architecture

In order to handle image features calculated from multiple views and multiple modalities as described in Section 2.1, the current architecture of the CADx system must be revised. The original *image-to-image* relation for the comparison and retrieval of similar images in the reference database on the *lesion level* must be extended towards a *set of images* (SOI)-*to-SOI* relation, where the described multi-view and multi-modality features can be embedded. Especially the extended comparison function between such SOI's must be newly defined, allowing the needed possibility to compare only partially filled set of images. This special case must be considered, since a SOI may not always be complete. Most often, only mammographic images are available for a SOI, as ultrasound and MRI modalities are only applied to a patient if there is found indication for a mass lesion. Accordingly, the integration of features obtained from DBT slices must also be supported, since ROIs depicting a mass in different DBT slices must be treated similar to ROIs depicting the same mass in different mammographic projections. Additionally, image registration and geometric standardization for DBT images is required as pre-processing steps.[26]

Furthermore, the data structures capturing medical meta information (e.g., previously annotated ground truth based on histological findings) will be extended. While the knowledge base of our current CADx system consists of a structured collection of interactively annotated ROIs together with histology-based ground-truth diagnosis, the knowledge-base needed for the proposed extended CADx system must be capable to handle all information sources associated with a given mammography case. The image and information sources making up a typical case of a breast mass lesion consist in the different related images (mammography, ultrasound, DBT, MRI) stored on (distributed) RIS and PACS servers as well as standardized classification ontologies the description of tissue type and lesions (e.g. BI-RADS^{TM}) and alphanumerical free text and categorical data (anamnestic and demographic patient data as well as additional patient information) stored on the hospital information systems (HIS) server.

2.3 PACS integration

Besides the extension of the feature vector describing a mammographic lesion and the thus related system architecture as described above, furthermore, the CADx system must be integrated with PACS, radiological (RIS) and hospital information systems (HIS). In previous work, we have already proposed a possibly framework for the integration of content-based image retrieval in medical applications (IRMA) with PACS and RIS[27] which is currently extended to support the integration with health care enterprise (IHE) profiles.[28] Generic communication modules that communicate with the RIS, HIS and PACS servers over standard protocols such as Digital Imaging and Communications in Medicine (DICOM) and Health Layer 7 (HL7) have been used for this integration. The application programming interface (API) of the viewing software used by the radiologist when reading mammograms has been used to interconnect the web-based IRMA interfaces providing extended query refinement and relevance feedback.[29]



(a) (b)

Figure 2. Comparison of the system architecture of the previously proposed CADx system that works on the lesion level (a) and the system proposed in this work that is extended to work on the full case level (b).

3. RESULTS

Figure 2 illustrates the system architectures of the previously proposed CADx system and the proposed extended version. In the extended version the reasoning process has been enhanced by an additional module that allows the fusion of characteristic lesion attributes extracted from corresponding ROIs from different image views, image frames and image modalities. This module is optional as an alternative way of handling lesion attributes from corresponding ROIs is to simply concatenate the attribute vectors. Lesion attributes extracted from textual clinical data such as patient information or histological data can also be added to this attribute vector.

We have previously evaluated the performance of our current CADx approach on the *lesion level* using ROC analysis. The resulting area under the ROC curve was $A_z = 0.86 \pm 0.03$ on the publicly available DDSM mammography database.[30] This database contains over 2000 cases acquired in the early 1990s. Each case includes two images of each breast, along with the patient's age, breast density rating, and BI-RADS standard lexicon compliant descriptions of abnormalities. The database contains normal, benign, and malignant cases. As the DDSM only consist of pure (analogue) mammography cases, the construction and validation of the proposed extension of the CADx system to the *case-level* requires an extended collection of mammographic cases that includes clinical data as well as images acquired using additional modalities such as DBT, MRI and ultrasound.

Currently we are compiling an appropriate multi-modal database and while our experiments with the case-level CADx approach proposed in this work indicate a significant increase in ROC performance, these results so far are based on a very limited dataset. Merging the existing IRMA mammography database[31] and the collection of mammographic cases from the University Hospital of Erlangen, approximately $11,000$ images have been collected so far.

4. CONCLUSIONS

In this work, an extended computer assisted system (CADx) for improved breast cancer diagnosis based on multiple image views, image modalities and image stacks has been proposed, which operates on the medical case level rather than the individual lesion level. For easy access to the required information and image sources such as mammography, ultrasound, DBT, MRI, we propose an integration of the CADx system with PACS, RIS and HIS using the existing IRMA framework. We show that our concept of multi-image features is also applicable to lesion attributes extracted from multiple DBT slices. The system evaluation will be based on a large ground truth of annotated and histologically validated images integrated from different reference databases. While the proposed system is a natural extension of CADx approaches, like the work of Drukker et al.,[9] which integrate only two modalities, we are still collecting a large enough database of breast lesions with images from multiple modalities to evaluate the benefits of the proposed approach on.

ACKNOWLEDGMENTS

The authors are grateful to Charles E. Metz, PhD, for providing the ROCKIT program for ROC analysis. This work was supported in part by the *Bayerische Forschungsstiftung* within the scope of the project *Mammo-iCAD*.

REFERENCES

[1] Elter, M. and Haßlmeyer, E., "A knowledge-based approach to the CADx of mammographic masses," in [*Medical Imaging 2008: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6915**, 69150L–1–69150L–8, SPIE (2008).

[2] Elter, M. and Held, C., "Semiautomatic segmentation for the computer aided diagnosis of clustered microcalcifications," in [*Medical Imaging 2008: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6915**, 691524–1–691524–8, SPIE (2008).

[3] Elter, M. and Bergen, T., "Incorporating a segmentation routine for mammographic masses into a knowledge-based CADx approach," in [*Medical Imaging 2009: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **7260**, 7260–75, SPIE (2009).

[4] Lehmann, T., Güld, M., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., Kohnen, M., Schubert, H., and Wein, B., "Content-based image retrieval in medical applications," *Methods Inf Med* **43**(4), 354–61 (2004).

[5] Wittenberg, T., "The need of annotation for reference image data sets," in [*Proc's 19th Int. Congress and Exhibition Computer Assisted Radiology and Surgery (CARS) 2005*], Lemke, H., Inamura, K., Doi, K., Vannier, M., and Farman, A., eds., 453 – 458 (2005).

[6] Wittenberg, T., Elter, M., and Schulz-Wendtland, R., "Complete digital iconic and textual annotation for mammography," in [*Bildverarbeitung für die Medizin 2007*], Horsch, A., Deserno, T., Handels, H., Meinzer, H.-P., and Tolxdorff, T., eds., *Reihe Informatik Aktuell*, 91 – 95, Workshop, 25 - 27.03.2007 in München, Springer Verlag, Heidelberg (2007).

[7] Veldkamp, W. J., Karssemeijer, N., Otten, J. D., and Hendriks, J. H., "Automated classification of clustered microcalcifications into malignant and benign types.," *Med Phys* **27**, 2600–2608 (Nov 2000).

[8] Gupta, S. and Markey, M. K., "Correspondence in texture features between two mammographic views.," *Med Phys* **32**, 1598–1606 (Jun 2005).

[9] Drukker, K., Horsch, K., and Giger, M. L., "Multimodality computerized diagnosis of breast lesions using mammography and sonography.," *Acad Radiol* **12**, 970–979 (Aug 2005).

[10] Schulz-Wendtland, R., Wenkel, E., Lell, M., Böhner, C., Bautz, W., and Mertelmeier, T., "Experimental phantom lesion detectability study using a digital breast tomosynthesis prototype system," *Rofo* **178**(12), 1219–1223 (2006).

[11] Diekmann, F. and Bick, U., "Tomosynthesis and contrast-enhanced digital mammography: recent advances in digital mammography.," *Eur Radiol* **17**, 3086–3092 (Dec 2007).

[12] Dobbins, J. T. and Godfrey, D. J., "Digital x-ray tomosynthesis: current state of the art and clinical potential.," *Phys Med Biol* **48**, R65–106 (Oct 2003).

[13] Gur, D., "Tomosynthesis: potential clinical role in breast imaging.," *AJR Am J Roentgenol* **189**, 614–615 (Sep 2007).

[14] Zhang, Y., Chan, H., Goodsitt, M., Schmitz, A., Eberhard, J., and Claus, B., "Investigation of different PV distributions in digital breast tomosynthesis (DBT) mammography," in [*Proc's 9th Int. Workshop on Digital Mammography*], *LNICS* **5116**, 593 – 600, July 20-23, 2008, Tucson, AZ, USA, Springer Verlag, Heidelberg (2008).

[15] Andersson, I., Ikeda, D. M., Zackrisson, S., Ruschin, M., Svahn, T., Timberg, P., and Tingberg, A., "Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and birads classification in a population of cancers with subtle mammographic findings.," *Eur Radiol* **18**, 2817–2825 (Dec 2008).

[16] Chan, H.-P., Wei, J., Sahiner, B., Rafferty, E. A., Wu, T., Roubidoux, M. A., Moore, R. H., Kopans, D. B., Hadjiiski, L. M., and Helvie, M. A., "Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience.," *Radiology* **237**, 1075–1080 (Dec 2005).

[17] Jerebko, A., Yuan, Q., Merlet, N., Ratner, E., Singh, S., Lo, J. Y., and Krishnan, A., "Feasibility study of breast tomosynthesis cad system," in [*Medical Imaging 2008: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6514**, 651414, SPIE (2007).

[18] Chan, H.-P., Wei, J., Zhang, Y., Helvie, M. A., Moore, R. H., Sahiner, B., Hadjiiski, L., and Kopans, D. B., "Computer-aided detection of masses in digital tomosynthesis mammography: comparison of three approaches.," *Med Phys* **35**, 4087–4095 (Sep 2008).

[19] Park, S. C., Zheng, B., Wang, X.-H., and Gur, D., "Applying a 2d based cad scheme for detecting micro-calcification clusters using digital breast tomosynthesis images: an assessment," in [*Medical Imaging 2008: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6915**(1), 691507, SPIE (2008).

[20] Reiser, I., Nishikawa, R. M., Edwards, A. V., Kopans, D. B., Schmidt, R. A., Papaioannou, J., and Moore, R. H., "Automated detection of microcalcification clusters for digital breast tomosynthesis using projection data only: a preliminary study.," *Med Phys* **35**, 1486–1493 (Apr 2008).

[21] Palma, G. J., Peters, G., Muller, S., and Bloch, I., "Masses classification using fuzzy active contours and fuzzy decision trees," in [*Medical Imaging 2008: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6915**(1), 691509, SPIE (2008).

[22] American College of Radiology, "Breast imaging reporting and data system (BI-RADS©) atlas," (2006).

[23] van Engeland, S., Timp, S., and Karssemeijer, N., "Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views.," *Med Phys* **33**, 3203–3212 (Sep 2006).

[24] Dasmahapatra, S., Dupplaw, D., Hu, B., Lewis, H., Lewis, P., and Shadbolt, N., "Facilitating multi-disciplinary knowledge-based support for breast cancer screening," *Int. J. Healthc. Tech. & Management* **7**(5), 403–20 (2006).

[25] Elter, M., Schulz-Wendtland, R., and Wittenberg, T., "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process," *Medical Physics* **34**, 4164 – 4172 (Nov. 2007).

[26] Bakic, P. R., Vujovic, N., and Maidment, A. D. A., "Establishing correspondence in mammograms and tomosynthesis projections," in [*Medical Imaging 2007: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6514**(1), 651425, SPIE (2007).

[27] Fischer, B., Deserno, T., Ott, B., and Günther, R., "Integration of a research cbir system with ris and pacs for radiological routine," in [*Medical Imaging 2008: Advanced PACS-based Imaging Informatics and Image-Guided Therapy*], Liu, K. M. S. B. J., ed., **6919**, 691914–691914–10, SPIE (2008).

[28] Welter, P., Deserno, T., Fischer, B., Wein, B., Ott, B., and Guenther, R., "Integration of a CBIR system in radiological routine in accordance with ihe," in [*Medical Imaging 2009: Advanced PACS-based Imaging Informatics and Image-Guided Therapy*], Liu, K. M. S. B. J., ed., **7264**, 7264–3, SPIE (2009).

[29] Deserno, T., Güld, M., Plodowski, B., Spitzer, K., Wein, B., Schubert, H., Ney, H., and Seidl, T., "Extended query refinement for medical image retrieval," *J Digit Imaging 2007* , online–first, DOI 10.1007/s10278–007–9037–4 (2007).

[30] Heath, M., Bowyer, K., Kopans, D., Moore, R., , and Kegelmeyer, P., "The digital database for screening mammography (DDSM)," in [*Proc's 5th Int. Workshop on Digital Mammogrphy*], Madison, WI, Medical Physics Publishing, Toronto, Canada (2000).

[31] Oliveira, J., Güld, M., Araújo, A., Ott, B., and Deserno, T., "Towards a standard reference database for computer-aided mammography," in [*Medical Imaging 2008: Computer-Aided Diagnosis*], Giger, M. L. and Karssemeijer, N., eds., **6915**, 69151Y–1–69151Y–9, SPIE (2008).