# Towards a Standard Reference Database for Computer-aided Mammography

Júlia E.E. Oliveira[a,b], Mark O. Gueld[a], Arnaldo de A. Araújo[b], Bastian Ott[c], Thomas M. Deserno[a,1]

[a] Department of Medical Informatics,
Aachen University of Technology (RWTH), 52057 Aachen, Germany
[b] Computer Science Dept., Federal University of Minas Gerais, 31270-010, Belo Horizonte, Brazil
[c] Department of Diagnostic Radiology, RWTH Aachen, Germany

## ABSTRACT

Because of the lack of mammography databases with a large amount of codified images and identified characteristics like pathology, type of breast tissue, and abnormality, there is a problem for the development of robust systems for computer-aided diagnosis. Integrated to the Image Retrieval in Medical Applications (IRMA) project, we present an available mammography database developed from the union of: The Mammographic Image Analysis Society Digital Mammogram Database (MIAS), The Digital Database for Screening Mammography (DDSM), the Lawrence Livermore National Laboratory (LLNL), and routine images from the Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen. Using the IRMA code, standardized coding of tissue type, tumor staging, and lesion description was developed according to the American College of Radiology (ACR) tissue codes and the ACR breast imaging reporting and data system (BI-RADS). The import was done automatically using scripts for image download, file format conversion, file name, web page and information file browsing. Disregarding the resolution, this resulted in a total of 10,509 reference images, and 6,767 images are associated with an IRMA contour information feature file. In accordance to the respective license agreements, the database will be made freely available for research purposes, and may be used for image based evaluation campaigns such as the Cross Language Evaluation Forum (CLEF). We have also shown that it can be extended easily with further cases imported from a picture archiving and communication system (PACS).

**Keywords:** Database Construction, Mammography, Ground Truth, Computer-Aided Diagnosis

## 1. INTRODUCTION

In occidental countries, breast cancer represents one of the main causes of death among women [1]. Its symptoms are nodules or tumors in the breast (Brazil National Cancer Institute, http://www.inca.gov.br). The early detection is the most effective way to reduce mortality, and mammography is the best method of screening for breast cancer because it can show lesions in their initial phases. The diagnosis of these lesions is made by a radiologist and due to the fast and continuous advances in computer technology as well as the conversion to the digital format of mammographies there is an increasing interest in computer-aided diagnosis (CAD) systems [2].

An effective CAD system, i.e., a system that clearly identifies position, size, and staging of lesions like microcalcifications and masses in x-ray mammographies, must be evaluated using a large number of reference images with approved diagnostics (ground truth), e.g., [3].

However, current studies are based on rather small sets of data. For instance, Zwiggelaar et al. used synthetic data and 15 mammographies of the Mammographic Image Analysis Society digital mammogram database (MIAS, http://peipa.essex.ac.uk/ipa/pix/mias/) to detect linear structures and classify them into the anatomical types: vessels, ducts, and spicules [4]. Christoyianni et al. detected the exact location of circumscribed masses in 22 images of MIAS

[1] Corresponding author: Prof. Dr. Thomas M. Deserno, né Lehmann, Department of Medical Informatics, Aachen University of Technology (RWTH), Pauwelsstr. 30, D - 52057 Aachen, Germany, email: deserno@ieee.org; web: http://irma-project.org/deserno, phone: +49 241 80 88793, fax: +49 241 80 33 88793.

Medical Imaging 2008: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer
Proc. of SPIE Vol. 6915, 69151Y, (2008) · 1605-7422/08/$18 · doi: 10.1117/12.770325

| Database | Resolution | | | | | | File type | |
|---|---|---|---|---|---|---|---|---|
| | x-min | x-max | y-min | y-max | g-min | g-max | format | standard |
| DDSM | 1,411 | 5,641 | 3,256 | 7,111 | 12 bit | 12 bit | LJPEG | no |
| MIAS | 334 | 1,000 | 802 | 1,024 | 8 bit | 8 bit | PNG | yes |
| LLNL | 700 | 4,494 | 2,828 | 6,874 | 12 bit | 12 bit | ICS | no |
| RWTH | 1,582 | 4,129 | 3,382 | 5,928 | 12 bit | 12 bit | DICOM | yes |

**Table 1:** Resolution and image type of databases

| Database | Anatomy | | Direction | | Tissue type | Tumor staging | Lesion description | |
|---|---|---|---|---|---|---|---|---|
| | left | right | CC | ML | | | position | multiple |
| DDSM | 4,918 | 4,915 | 4,917 | 4,916 | ACR | yes | chain code | yes |
| MIAS | 161 | 161 | 0 | 322 | yes | yes | circle | yes |
| LLNL | 93 | 91 | 92 | 92 | yes | yes | chain code | yes |
| RWTH | 81 | 85 | 85 | 79 | ACR | BI-RADS | none | --- |

**Table 2:** Anatomy, direction, and biosystem information

database [5]. Strickland & Hahn employed wavelet decomposition in 40 mammographies of Nijmegen database to detect micro-calcifications [6]. However, this database is no longer available. Arodz et al. achieved pattern recognition methods such as adaptive boosting and support vector machines to detect micro-calcifications and masses in 168 abnormal mammographies taken from the Digital Database for Screening Mammography (DDSM, http://marathon.csee.usf.edu/mammography/database.html) [7]. Also based on DDSM, Eltonsy et al. tried to localize potentially malignant masses in a set of 540 images [8].

Although even large databases for mammography – such as DDSM with about 10,000 images – are public available, the problem for researchers needing reference data are manifold. Resolution, tissue classification and pathology description and/or localization of the region of interest (ROI) are not standardized, and may be partly erroneous. Resulting from different formats, databases cannot be merged, and for particular algorithms and CAD approaches, an insufficient number of cases may exist in each of the database, individually. Therefore, generation of classified mammography collections is still in the focus of recent research [9].

Regarding the large number of images, image handling and image data management becomes difficult, too. Researches must face the problem of selecting appropriate cases and presenting results of the computation visually in a transparent and user-friendly way. However, novel methodologies of visual data management are currently developed in the field of content-based image retrieval (CBIR) and image data mining [10]; for reviews on CBIR and medical CBIR cf. the publications of Smeulders et al. [11] and Müller et al. [12], respectively.

Based on the Image Retrieval in Medical Applications (IRMA, http://irma-project.org) framework [13], we aim at defining a unified database structure and coding scheme for mammographic radiographs associated with diagnostic information that can be filled consistently with images from the various databases available as well as mammography data from Picture Archiving and Communication Systems (PACS) as they are used in routine of hospitals and health centers. The unified reference database will be used for CAD system evaluation and other evaluation campaigns such as the automatic image annotation task in the Cross Language Evaluation Forum (CLEF, http://www.clef-campaign.org) [14, 15].

## 2. MATERIALS AND METHODS

In this section, we present details of the databases that we combine and explain how the images were imported into the IRMA system and unambiguously classified using the IRMA coding scheme.

### 2.1. Freely Accessible Mammography Databases

An overview of the data available for this study is given in Tables 1 and 2.

**DDSM.** The DDSM database [16] officially contains 2,479 studies (695 normal, 870 benign, and 914 cancerous cases). Each study includes two images of each breast, acquired in craniocaudal (CC) and mediolateral (ML) views that have been scanned from the film-based sources by four different scanners with a resolution between 50 and 42 microns. This results in total of 9,916 radiographs. These images are coded with an algorithm according to the lossless Joint Pictures Expert Group (JPEG) standard and had to be converted into a standard file format with special software that is provided in C source code at the DDSM web page.

For all cases, there are additional plain text files with information on the type of digitizer and a staging of the tissue density according to the American College of Radiology (ACR). Where appropriate, information about the lesions types and a chain code with the localization and delineation of these lesions are also provided.

**MIAS.** The MIAS database [17] is available only for researches purposes and contains 322 mammography images, all of them acquired in mediolateral view. Initially scanned from film with a resolution of 50 microns, all images were reduced to 200 microns and clipped / padded so that they fit into a 1,024 x 1,024 bounding box. The image files are available in the portable network graphics (PNG) format and annotated with the following details: a database reference number indicating left and right breast, character of background tissue, pathology, class of lesion present and coordinates as well as size of these lesions.

**LLNL.** The LLNL database [18] contains 197 mammography images (4 images per patient), all of them digitized at 35 microns per pixel. The images are stored in the image cytometry standard (ICS) format and had to be converted into a standard file format with a provided source code for a program that converts images in the ICS format to the portable grey map (PGM) format. For 190 images there is available a plain text file containing patient stats, biopsy results and ground truth comments.

**RWTH.** In order to evaluate the extensibility of mammogram reference resources, 170 cases were extracted arbitrarily from the PACS at the Department of Diagnostic Radiology, University Hospital, Aachen University of Technology (RWTH), Aachen, Germany. These images were acquired digitally using a General Electric Senographe operating with low beam energy about 26 to 32 kV and with a phosphor storage system from Fuji/Philips capable of recording 7 lp/mm. The cassette was read using a Philips PCR Eleva CosimaX. If available, a free text diagnosis in German describing the breast examination, pathology, type of tissue and lesion was included along with the digital imaging and communications in medicine (DICOM) files.

**OTHERS.** There are only few other databases that are public available and have been used for research. In some researches, the Nijmegen Database was used (e.g. by Heinlein et al. [19]), but since March 2000 it isn´t available anymore. Others databases, e.g., the massive database provided by National Digital Medical Archive (NMDA) that holds over a million mammography images [20], are not freely available, and might be included in a second step.

### 2.2. The IRMA System

The IRMA project aims at developing and implementing high-level methods for content-based image retrieval (CBIR) with prototypal application to medico-diagnostic tasks on a radiological image archive [13]. Beyond the mammographies, there are currently more than 20,000 diagnostic images in the IRMA database, which are used for image retrieval [10,14].

In IRMA, all images are coded according to a mono-hierarchical, multi-axial coding scheme [21]. The four axes, each having three to four hierarchical positions, describe the

- *technique*: image modality,
- *direction*: body orientation,
- *anatomy*: body region examined, and the
- *biosystem*: biological system examined,

and result in a unique string of 13 digits: TTTT-DDD-AAA-BBB.

| IRMA code | | Tissue density |
|---|---|---|
| xBB | ACR | description |
| 0 | --- | unspecified |
| 1-c | --- | already in use |
| d | ACR-1 | fat transparent system |
| e | ACR-2 | fibroid glands system |
| f | ACR-3 | heterogeneously dense system |
| g | ACR-4 | extremely dense system |
| h | ACR-3/4 | dense system |

| IRMA code | | Tumor staging |
|---|---|---|
| BxB | BI-RADS | description |
| 0 | BI-RADS 0 | unspecified |
| 1 | BI-RADS 1 | normal |
| 2 | BI-RADS 2 | benign |
| 3 | BI-RADS 3 | probably benign |
| 4 | BI-RADS 4 | suspiciously abnormal |
| 5 | BI-RADS 5 | malignant |

| IRMA | Type of lesion |
|---|---|
| BBx | description |
| 0 | unspecified |
| 1 | calcification, unspecified |
| 2 | micro-calcification |
| 3 | macro-calcification |
| 4 | circumscribed mass |
| 5 | spiculated mass |
| 6 | other mass |
| 7 | architectural distortion |
| 8 | asymmetry |

**Table 3:** IRMA codes for the biosystem

**Technique.** The IRMA technical code for mammography is TTTT = 11xx, where 11 means x-ray, plain radiography, and the two remaining code positions are used to capture the nature of images (1 = directly digital, 2 = secondarily digitized) and their resolution (e.g., 42, 43.5, 50, or 200 microns).

**Direction.** According to the coding scheme, the directions for breast imaging, i.e., CC and ML, are denoted by DDD = 310 (axial – craniocaudal – unspecified) and DDD = 410 (other orientation – oblique – unspecified), respectively.

**Anatomy.** The anatomy axis of the IRMA code is used to differ left from right breast using the codes AAA = 610 (breast or mamma – right breast – unspecified) and AAA = 620 (breast or mamma – left breast – unspecified), respectively.

**Biosystem.** The biosystem axis was extended in order to capture tissue density, tumor staging, and lesion description (Table 3). The first position describes the tissue type according to the ACR classes. For instance, if the breast is almost entirely fat or rather scattered with fibro glandular densities, ACR-1 or ACR-2 is appropriate, respectively. The ACR-3/4 class (dense system) was defined to import the MIAS images. The second IRMA code position captures the tumor staging according to the ACR breast imaging reporting and data system (BI-RADS) [22, 23], where, for example, a breast that needs additional imaging evaluation is coded (BI-RADS 0). Finally, the third code position refers to type of lesion. According to the BI-RADS system, eight descriptions currently are defined (Table 3) which, according to the structure of the IRMA code, can be extended easily if demanded.

**Lesion Localization and Morphologic Description.** The IRMA framework provides a database that hosts images as well as their derived image features. For instance, texture signatures are associated with each image in order to allow fast CBIR access. Internally, this link is established based on the IRMA identifier, which is uniquely assigned to each database element. This relation is used to store one or more lesions descriptions, in form of

1. *circle*, described by its center coordiantes and radius;
2. *contour points*, a list of (x,y)-coordinates;
3. *chain code*, a starting coordinate (x,y) followed by a sequence of numbers in [1..8] describing the direction to the adjacent contour point; or
4. *masking image*, a binary image with the same x-,y-dimensions as the mammography, where 0 and 1 denote "background" and "lesion", respectively.

### 2.3. Integration of Databases

Integrating the images from DDSM, MIAS, LLNL and RWTH into the IRMA system, the IRMA code was set automatically, using the descriptions from the databases. The database-specific conversion problems are addressed in this chapter.

| Code alteration | Description | Number |
|---|---|---|
| ACR 1 → ACR 2 | fat transparent → fibroid glands | 98 |
| ACR 1 → ACR 3 | fat transparent → heterogeneoulsy dense | 12 |
| ACR 1 → ACR 4 | fat transparent → extremely dense | 1 |
| ACR 2 → ACR 1 | fibroid glands → fat transparent | 102 |
| ACR 2 → ACR 3 | fibroid glands → heterogeneously dense | 36 |
| ACR 2 → ACR 4 | fibroid glands → extremely dense | 2 |
| ACR 3 → ACR 1 | heterogeneously dense → fat transparent | 10 |
| ACR 3 → ACR 2 | heterogeneously dense → fibroid glands | 136 |
| ACR 3 → ACR 4 | heterogeneously dense → extremely dense | 4 |
| ACR 4 → ACR 1 | extremely dense → fat transparent | 1 |
| ACR 4 → ACR 2 | extremely dense → fibroid glands | 12 |
| ACR 4 → ACR 3 | extremely dense → heterogeneoulsy dense | 61 |
| Total | | 475 |

**Table 4:** Correction of DDSM tissue classification

**DDSM.** The IRMA code was automatically generated based on bitmap file name as it contains anatomic and directional information, and the ".ics" file, which contains density information, and – if present – the "overlay" file, which stores diagnosis-related data, including the chain-code contour data for marking regions of interest (ROIs) with pathologies.

We encountered some problems with the data: first of all, 10 cases and 5 cases are missing from volumes benign_01 and benign_02, respectively. Furthermore, case 1658 from volume cancer_11 was inaccessible for download. Therefore, 695 normal, 855 benign, and 913 malignant cases were available, with a total of 2,463 cases and 9,852 images. Processing the medical information yielded unproven pathologies for 15 images, which were excluded. In addition, case 1825 of volume cancer_11 has density 0. Images with pathology "benign_without_feedback" were classified as "benign".

The original format used to encode the images is LJPEG, and the authors also provide source code to extract the raw image data. Although the website states that each image uses 12bit quantization for grayscale intensities, 3,130 images violate the intensity boundaries of [0..4095]. Unfortunately, there are different patterns among these violations. Some images contain single outliers, while others fully utilize the extended value range. Therefore, an automatic conversion algorithm was applied, which first checks for outliers by inspecting a median-filtered version of the image. If only outliers caused the violation, these pixels are simply cut off. If a violation still exists, a check is performed if the image using the range fully. If the resulting image is still too dark (meaning that the found boundaries are still caused by surviving outliers), a grayscale stretching is performed based on a histogram which drops 1/64th of the pixels at the upper end. After the re-estimation of the upper grayscale boundary, the images were cropped and scaled to fit a 1024×1024 bounding box, and transferred into an 8bit grayscale intensity range.

Handling the data within the IRMA system, it became obvious that tissue classification partly was inconsistent and crosschecked by a trained radiologist. In total, ACR tissue code was altered by the physician for 475 images (Table 4).

**MIAS.** The codification was done manually according to the database description on the Internet. Initially the ACR tissue type was converted as follows:

- F (MIAS fatty) → d (ACR1);
- G (MIAS fatty-glandular) → e(ACR2);
- D (MIAS dense-glandular) → h (ACR3/4).

However, it turned out that this conversion does not result in correct ACR codes. From the 322 images, 187 had to be corrected by an experienced radiologist (Table 5). Furthermore, three images were corrected in BI-RADS classification from normal to benign with micro-calcifications.

**LLNL.** According to the information provided by text files, like direction of x-ray, breast anatomy, tissue type, pathology and lesion, the IRMA codes were associated. From the 190 images available, two images showing amputated

| Code alteration | Description | Number |
|---|---|---|
| ACR 1 → ACR 2 | fatty → fibroid glands | 24 |
| ACR 1 → ACR 3 | fatty → heterogeneoulsy dense | 1 |
| ACR 2 → ACR 3 | fatty-glandular → heterogeneously dense | 45 |
| ACR 2 → ACR 4 | fatty-glandular → extremely dense | 4 |
| ACR 3/4 → ACR 2 | dense glandular → fibroid glands | 5 |
| ACR 3/4 → ACR 3 | dense glandular → heterogeneously dense | 38 |
| ACR 3/4 → ACR 4 | dense glandular → extremely dense | 70 |
| Total | | 187 |

**Table 5:** Correction of MIAS tissue classification

breasts and another four images with breasts containing silicone were excluded. Also, 2 images that had no diagnostic were inspected and codified manually by the trained radiologist.

**RWTH.** All the IRMA codes were manually extracted from the DICOM files. From the 170 available images, 54 were provided with full coding information, 81 had incomplete codes (tissue density or type of lesion was missing), and for 35 a diagnosis was not included. A physician inspected all these incomplete images and set the IRMA code.

## 3. RESULTS

In total 10,509 images (without differentiating the resolution) are now available for evaluation of mammography CAD systems. As it can be observed from Tables 6-8, the merged database holds examples for all categories and their combinations, which is not the case when restricting to only one data source.

Figures 1 and 2 exemplify the benefits of the new mammography reference database. All web-interfaces can be directly used to explore the database. For example, Fig. 1 shows a screen shot of the IRMA code statistics, a web-interface that lists all available code classes. Merging of groups can be easily computed using the wildcards in the parameter field [24]. Via the "view" buttons, this interface is linked to the IRMA code browser (Fig. 2), a web-interface for interactive exploration of the images. Nonetheless, the new resource is easily extensible, and further databases will be included as they become freely available for scientific research.

In total 10,509 images out of 323 code classes (disregarding the image resolution) are now available for evaluation of mammography CAD systems. As it can be observed from Tables 6-7, the merged database holds examples for all categories and their combinations, which is not the case when restricting to only one data source. As can be further deduced from Table 1, the majority of illustrations are still published in grayscale (87.86%). If multi-panel illustrations that contain at least one colored component were counted as if all components are colored, the number of grayscale panels is still above 80%. Similarly, the majority of illustrations are annotated with text, arrows, or other symbols which may cover image information and affect the textural feature extraction.

## 4. DISCUSSION

Although evaluation of medical image analysis should rely on large sets of ground truth data, research in computer-assisted mammography reading is often based on less than 30 images [4, 5]. Based on the IRMA framework, we defined a coding scheme according to the ACR standards and showed how existing image resources are integrated (DDSM, MIAS, LLNL and RWTH). More precisely, we provided a scheme to integrate available mammography databases using standardized description of imaging modality and resolution, orientation and view, left and right position of breast, tissue type, tumor staging and lesion description as well as lesion positions regardless whether these positions are coded by some boundary points, a bounding circle, or a complete chain code. Since the proposed scheme is based on international ACR and BI-RADS codes, it is extensible and can be used for future evaluation of CAD system. However, some problems arose from the fact that existing resources does not provide complete coding information, or hide this information in file names and additional description files.

| Database | Nominal | IRMA code BBB = xBB | | | | | Converted | |
|---|---|---|---|---|---|---|---|---|
| | | d | e | f | g | h | total | percent |
| DDSM | 9,916 | 1,252 | 3,691 | 2,896 | 1,994 | 0 | 9,833 | 99.2 |
| MIAS (before corrections) | 322 | 80 (105) | 84 (104) | 84 (0) | 74 (0) | 0 (113) | 322 | 100.0 |
| LLNL | 197 | 12 | 84 | 68 | 20 | 0 | 184 | 93.4 |
| RWTH | 170 | 48 | 78 | 42 | 2 | 0 | 170 | 100.0 |
| IRMA | 10,615 | 1,395 | 4,043 | 3,048 | 2,023 | 113 | 10,509 | 99.0 |

**Table 6:** Statistics of issue classes after integration

| Database | IRMA code BBB = BxB | | | | | | Converted |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | total |
| DDSM | 0 | 6,181 | 1,848 | 0 | 0 | 1,804 | 9,833 |
| MIAS (before corrections) | 0 | 206 (209) | 64 (61) | 0 | 0 | 52 | 322 |
| LLNL | 6 | 47 | 111 | 6 | 0 | 14 | 184 |
| RWTH | 2 | 69 | 87 | 6 | 6 | 0 | 170 |
| IRMA | 8 | 6,503 | 2,110 | 12 | 6 | 1,870 | 10,509 |

**Table 7:** Statistics of tumor staging after integration

| Database | IRMA code BBB = BBx | | | | | | | | | Converted |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | total |
| DDSM | 6,181 | 1,488 | 0 | 0 | 478 | 547 | 1,139 | 0 | 0 | 9,833 |
| MIAS (before corrections) | 206 (209) | 23 | 0 | 3 (0) | 23 | 19 | 14 | 19 | 15 | 322 |
| LLNL | 65 | 112 | 4 | 0 | 0 | 0 | 0 | 0 | 3 | 184 |
| RWTH | 71 | 0 | 40 | 43 | 4 | 2 | 0 | 10 | 0 | 170 |
| IRMA | 6,523 | 1,623 | 44 | 46 | 505 | 568 | 1,153 | 29 | 18 | 10,509 |

**Table 8:** Statistics of lesion types after integration

Building a mammography database with a large number of images is proposed by some researches [9], which intend also to provide information about the patients, like age, menstrual history and breast cancer occurrences in the family plus a ground truth data in XML format. With our database, containing until now 10,509 reference images, all the information needed by radiologists to their aid or for the implementation of retrieval and CAD systems is already given by the codified tissue density, lesion staging and type of lesions, without the need of extra toolboxes.

## 5. CONCLUSION

Based on international standards such as ACR and BI-RADS, we provided a scheme to integrate available mammography databases using standardized description of imaging modality and resolution, orientation and view, left and right position of breast, tissue type, tumor staging and lesion description as well as lesion positions regardless whether these positions are coded by some boundary points, a bounding circle, or a complete chain code. Integrating different resources that are freely available in the Internet, our database currently holds 10,509 images from 232 different code classes. Based on this unified database, researches in CAD systems can improve with the implementation of algorithms for tissue characterization.

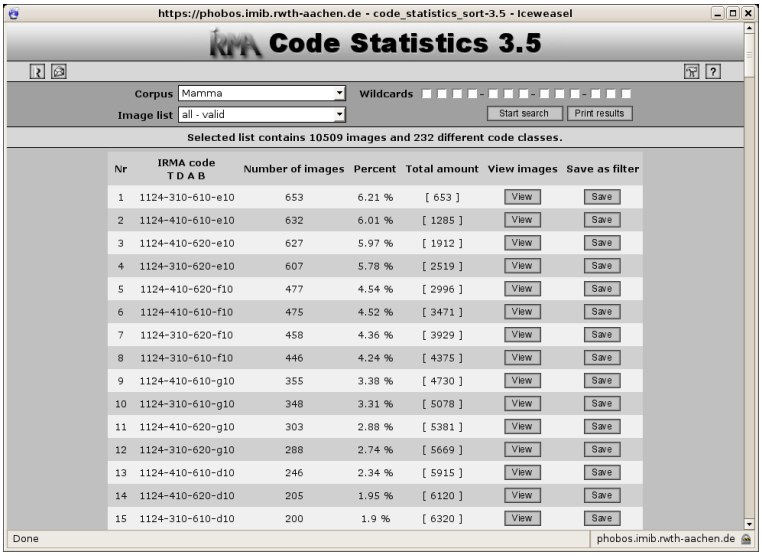


**Figure 1:** IRMA Web-interface for code statistics. Currently, the corpus "Mamma" contains 10,509 images from 232 different code classes.

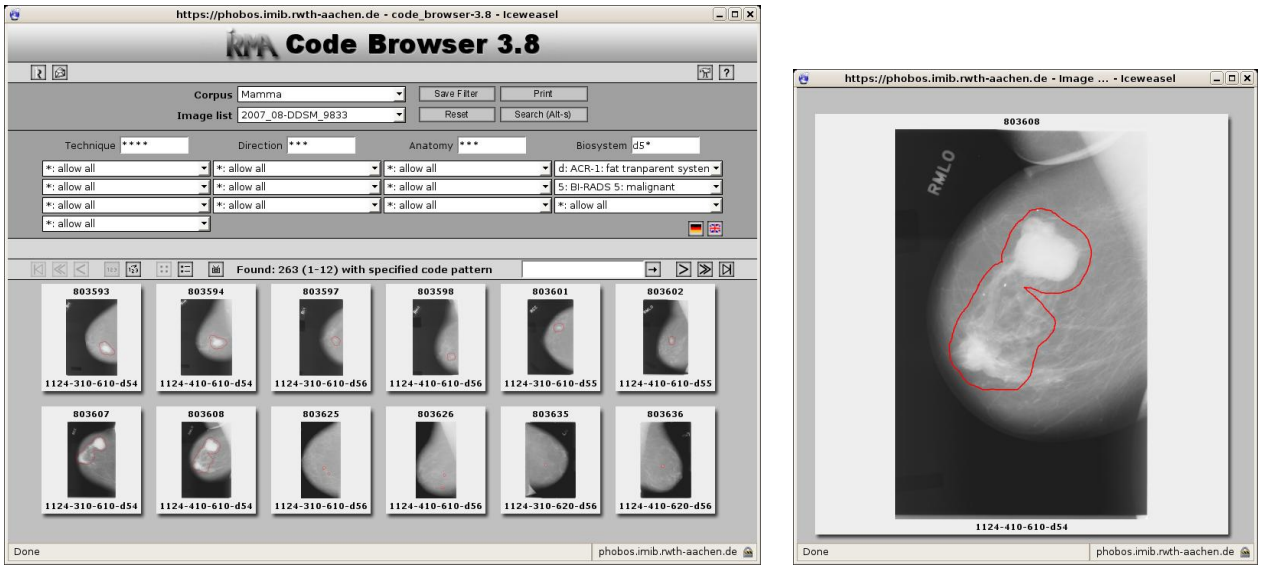


**Figure 2:** IRMA Web-interface for database browsing. Clicking on any icon in the overview (left) opens a separate window with more details (right). Diagnostic information is automatically overlaid.

## ACKNOWLEDGEMENT

This research was supported (in part) by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), process number BEX 0107/07-7 and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The IRMA project is funded by the German Research Foundation (DFG), grant no. Le 1108/4 and Le 1108/9.

## REFERENCES


1. Xue F, Michels KB. Intrauterine factors and risk of breast cancer: a systematic review and meta-analysis of current evidence. Lancet Oncol 2007; 8(12): 1088-100.
2. Burhenne LJW, Wood SA, D'Orsi CJ, Feig SA, Kopans DB, Castellino RA. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiol 2000; 215: 554-62.
3. Elter M, Schulz-Wendtland R, Wittenberg T. The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Med Phys 2007; 34(11): 4164-72.
4. Zwiggelaar R, Astley SM, Boggis CRM, Taylor CJ. Linear structures in mammographic images: detection and classification, IEEE Trans Med Imaging 2004; 23(9): 1077-86.
5. Christoyianni I, Dermantas E, Kokkinakis G. Automatic detection of abnormal tissue in mammography. Proceedings ICIP 2001; 877-80.
6. Strickland RN, Hahn HI. Wavelet transforms for detecting microcalcifications in mammograms. IEEE Trans Med Imaging 1996; 15(2): 218-29.
7. Arodz T, Kurdziel M, Sevre EOD, Yuen DA. Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms. Comput Programs Biomed 2005; 79: 135-49.
8. Eltonsy NH, Tourassi GD, Elmaghraby AS. A concentric morphology model for the detection of masses in mammography. IEEE Trans Med Imaging 2007; 26(6): 880-89.
9. Elter M, Horsch A, Schulz-Wendtland R, Sittek H, Athelogou M, Schmidt G, Wittenberg T. A modern benchmark case database for the computer-aided diagnosis of breast cancer. Int J Comput Assist Radiol Surg 2007; 2(S1): 514.
10. Lehmann TM, Güld MO, Deselaers T, Keysers D, Schubert H, Spitzer K, Ney H, Wein BB. Automatic categorization of medical images for content-based retrieval and data mining. Comput Med Imaging Graph 2005; 29(2): 143-155.
11. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 2000; 22(12): 1349-80.
12. 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.
13. Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohnen M, Schubert H, Wein BB. Content-based image retrieval in medical applications. Methods Inform Med 2004; 43(4): 354-61.
14. Deselaers T, Müller H, Clough P, Ney H, Lehmann TM. The CLEF 2005 automatic medical image annotation task. Intl J Comp Vis 2007; 74(1): 51-8.
15. Müller H, Deselaers T, Deserno TM, Clough P, Kim E, Hersch W. Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks. Lect Note Comp Sci 2007; 4730: 595-608.
16. Heath M, Bowyer KW, Kopans D, et al. Current status of the digital database for screening mammography. In: Digital Mammography, Kluwer Academic Publishers 1998; 457-60.
17. Suckling J, et al. The Mammographic image analysis society digital mammogram database", Exerpta Medica International Congress Series 1994; 1069: 375-8.
18. Center for Health Care Technologies Livermore. Lawrence Livermore National Library / UCSF Digital Mammogram Database. Livermore, CA, USA.
19. Heinlein P, Drexl J, Schneider W. Integrated wavelets for enhancement of microcalcifications in digital mammography. IEEE Trans Med Imaging 2003; 22(3): 402-13.
20. Gardner D. Breast Cancer database provides faster access to patient record. Grid technology is at the heart of this massive database that holds over a million mammography images. Information Week 2005; Nov 18: article 174400322
21. Lehmann TM, Schubert H, Keysers D, Kohnen M, Wein BB. The IRMA code for unique classification of medical images. Proc SPIE 2003; 5033:440-51.
22. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS©). Atlas, 2006.
23. Fischer U, Helbich T. ACR BI-RADS: Illustrierte Anleitung zur einheitlichen Erstellung von Mammographie, Mammasonographie, MR Mammographie. Stuttgart: Thieme, 2nd ed, 2006.
24. 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, in press.