THE DIGITAL DATABASE FOR SCREENING MAMMOGRAPHY

M. Heath[1], K. Bowyer[1], D. Kopans[2], R. Moore[2] and P. Kegelmeyer Jr.[3]
E-mail: heath@csee.usf.edu, kwb@csee.usf.edu,

[1] Department of Computer Sci. & Eng., University of South Florida, Tampa, FL 33620, USA
[2] Department of Radiology, Massachusetts General Hospital, Boston MA 02114
[3] Center for Computational Eng., Sandia National Labs, Livermore, CA, 94551

1.0 INTRODUCTION

The Digital Database for Screening Mammography (DDSM) is a database of digitized film-screen mammograms with associated ground truth and other information. The purpose of this resource is to provide a large set of mammograms in a digital format that may be used by researchers to evaluate and compare the performance of computer-aided detection (CAD) algorithms. The database was completed in the fall of 1999. It contains 2620, four view, mammography screening exams. Since that time, the database has been enhanced through the addition of new software tools that simplify the extraction of image data to other file formats, allow simplified access to the data in the ground truth files and simplify the steps necessary to evaluate a CAD algorithm.

This paper provides an overview of the completed database, information on tools that are available to access this data, suggestions on common performance measures and information on how to acquire data and software tools from the database.

2.0 MOTIVATION

The evaluation of a CAD algorithm often begins with a retrospective evaluation of cancer cases. A preliminary evaluation of this type is more time and cost effective that a prospective evaluation in a clinical setting. While the data for a retrospective CAD performance evaluation may be obtained by digitizing mammograms from the case files at a mammography center, this task may be time consuming and thus expensive to achieve. This expense may be avoided by using data from the DDSM.

When investigators of CAD methods utilize the resources in a common database, rather than using their own data, much more may be learned by a performance evaluation. Provided the same data, the same performance measure and the same train and test methodologies are followed, results from different algorithms can be compared to find the relative strengths of each algorithm. This in turn may lead to the development of new or combined approaches to the problem that yield superior performance (te Brake and Karssemeijer 1998).

3.0 CONTENTS OF THE DDSM

The DDSM contains mammograms obtained from Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital and Washington University of St. Louis School of Medicine. The four standard views (medio-latral oblique and cranio caudal) from each

Digital Database for Screening Mammography

Volume: cancer_11 Case: A–1252–1

A_1252_1.LEFT_MLO A_1252_1.RIGHT_MLO A_1252_1.LEFT_CC A_1252_1.RIGHT_CC
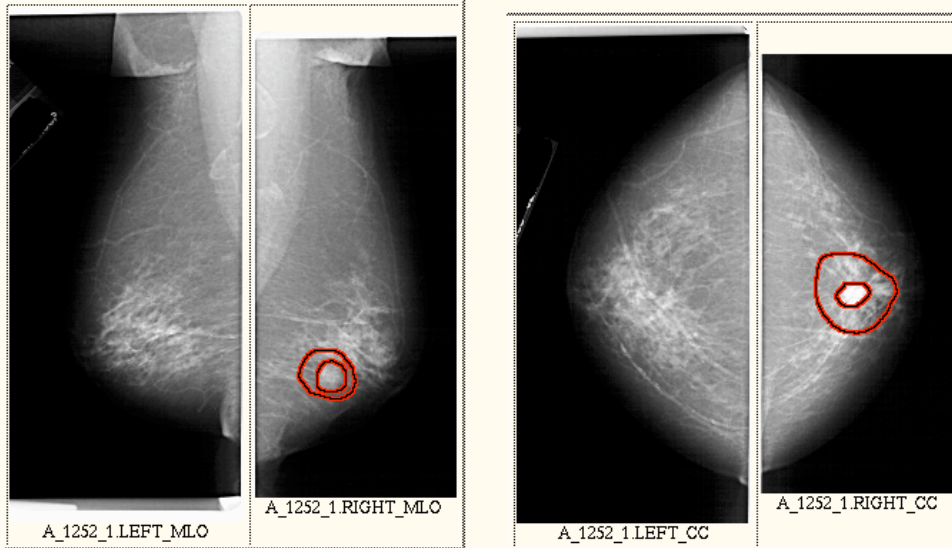
```
ics_version 1.0
filename A-1252-1
DATE_OF_STUDY 16 10 1991
PATIENT_AGE 74
FILM
FILM_TYPE REGULAR
DENSITY 3
DATE_DIGITIZED 23 7 1998
DIGITIZER HOWTEK 43.5
SEQUENCE
LEFT_CC LINES 5956 PIXELS_PER_LINE 2776 BITS_PER_PIXEL 12 RESOLUTION 43.5 NON_OVERLAY
LEFT_MLO LINES 6871 PIXELS_PER_LINE 3106 BITS_PER_PIXEL 12 RESOLUTION 43.5 NON_OVERLAY
RIGHT_CC LINES 5251 PIXELS_PER_LINE 2326 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
RIGHT_MLO LINES 6226 PIXELS_PER_LINE 2701 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
```

```
FILE: A_1252_1.RIGHT_MLO.OVERLAY

TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC DISTRIBUTION CLUSTERED
LESION_TYPE MASS SHAPE IRREGULAR MARGINS SPICULATED
ASSESSMENT 5
SUBTLETY 4
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 2
BOUNDARY
        CORE
```

```
FILE: A_1252_1.RIGHT_CC.OVERLAY

TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC DISTRIBUTION CLUSTERED
LESION_TYPE MASS SHAPE IRREGULAR MARGINS SPICULATED
ASSESSMENT 5
SUBTLETY 4
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 2
BOUNDARY
        CORE
```

Figure 1: The web page showing the thumbnail version of the images for case1252 in the DDSM. The boundary of the region in each image is marked by the larger marking and the core of the spiculated mass is marked by the smaller region inside the boundary. The abnormality is a malignant spiculated mass with calcifications.

Table 1: The sampling rate, number of gray levels and the equation to estimate optical density (OD) from gray values (GV) for each scanner used to digitize mammograms for the DDSM.

Digitizer	Sampling Rate (microns)	Gray Levels (bits)	Optical Density Calibration Equation
DBA M2100 ImageClear	42	16	$OD = 4.26700423014133 + (-0.90303289757264) * \log10(GV)$
Howtek 960	43.5	12	$OD = 3.78928997845071 + (-0.00094568009377) * GV$
Lumisys 200 Laser	50	12	$OD = 4.05977749300340 + (-0.00099080941710) * GV$
Howtek MultiRad850	43.5	12	$OD = 3.96604095240593 + (-0.00099055807612) * GV$

case were digitized on one of four different. Table 1 lists some characteristics for each of these scanners and provides calibration equations to convert pixel values to optical densities. The cases were all from mammography exams conducted between October of 1988 and February of 1999.

3.1 CASE CATEGORIZATION

The cases were assigned to volumes according to the severity of the finding. Normal volumes contain mammograms from screening exams that were read as normal and had a normal screening exam four years later (plus or minus 6 months). Benign without callback volumes contain exams that had an abnormality that was noteworthy but did not require the patient to be recalled for any additional workup. Benign volumes contain cases in which something suspicious was found and the patient was recalled for some additional work-up that resulted in a benign finding. Cancer volumes contain cases in which a histologically proven cancer was found. Each volume may contain cases that include less severe findings in addition to the more severe findings that resulted in the assignment of a case to a particular volume. Table 2 shows the breakdown of the 2620 cases in the database by mammography facility and volume type.

Table 2: Contents of the DDSM database by institution, digitizer and type of the case as expressed by the most severe findings in the case. The institutions are where the mammograms were recorded: Massachusetts General Hospital (MGH), Wake Forest University School of Medicine (WFU), Sacred Heart Hospital (SH) and Washington University of St. Louis Medical Center (WU).

| Institution | Digitizer | Number of Cases By Most Severe Finding | | | | Total |
		Normal	Benign Without Callback	Benign	Malignant	
MGH	DBA M2100 ImageClear	430	0	0	97	527
	Howtek 960	78	0	446	323	847
WFU	Lumisys 200	82	93	126	159	460
SH	Laser	0	48	202	234	484
WU	Howtek MultiRad850	105	0	96	101	302
Total		695	141	870	914	2620

Every case in the DDSM contains the patient age, the screening exam date, the date on which the mammograms were digitized and the ACR breast density that was specified by an expert radiologist. Cases in all volumes other than the normal volume contain pixel level ground truth markings of abnormalities. Each marking contains a subtlety value and a description that was specified by an expert mammography radiologist using the BI-RADS™ (ACR 1998) lexicon.

3.2 CASE PROCESSING

After digitization, mammograms were automatically cropped to remove much of the background (non-breast tissue) area. They were then manually processed to darken (digitally zero) pixels in regions that contained patient identifiers and were stored in files using a truly loss-less compression algorithm. The ground truth data was entered into computer readable format using custom software that allowed one to draw free form digital curves of the radiologist identified ground truth regions.

4.0 SOFTWARE TOOLS

Several software tools are linked to the main web page for the database. These tools simplify the most common tasks such as case selection, data extraction and performance evaluation.

A search engine is available that uses a web interface to obtain a list of DDSM cases that meet a search criteria defined by the volume type, ACR breast density, scanner, lesion pathology, assessment and subtlety as well as by the BI-RADS™ keywords. This tool may be used to identify a set of cases for evaluating an algorithm or may simply be used as a quick index to overviews of each case. The web page containing the form to initiate a search is illustrated in figure 2.

A software package that contains programs and 'C' source code for decompressing images, converting them to 8 or 16 bits, rescaling them, re-mapping them to optical density, resizing them to other resolutions, and for writing them in Portable Gray Map (PGM) or Tagged Image File Format (TIFF) files is available through the database. This same package also provides a program for rendering an image of user selectable ground truth regions at a user specified resolution. Together, these software tools can be used to extract data from the database at a user specified resolution in a simple format. This allows users to focus more of their attention on the CAD algorithms and less on studying the details of the DDSM.

Other programs in the software package include an automated mass detection algorithm and software to automatically generate a Fractional Receiver Operating Characteristic (FROC) analysis of its performance on a user specified set of cases. Elements of this process can be easily

Figure 2: The DDSM search page. This form allows the user to select the attributes of cases and returns a list of all the cases that match the search criteria. The matching cases are listed in a table that provides links to the thumbnail web pages for each case and links to the cases that are currently available for downloading using ftp.

reconfigured to evaluate the performance of any detection algorithm that generates prompts of suspicious locations in an image. The purpose of the included mass detection algorithm is to provide a simple baseline algorithm with verifiable results and to illustrate the use of some of these tools.

5.0 ACCESS TO THE DDSM RESOURCE

The web page for the DDSM resource is marathon.csee.usf.edu/Mammography/Database.html. From this web page, it is possible to browse thumbnail versions of the images in all 2,600 cases and to download associated software. The thumbnail versions of the images in each case appear similar to those seen in figure 1. Note that the radiologist's outline of the abnormality is overlaid on the thumbnail image. The ground truth data for the abnormality is available with the full-resolution images for each case.

At any given time, several volumes of the full-resolution images are available for ftp transfer. All volumes are available on 8mm ExaByte tape; details for ordering tapes are available on the web page.

6.0 PERFORMANCE EVALUATION OF CAD ALGORITHMS

A major anticipated use of the DDSM resource is to facilitate comparison of CAD algorithms through evaluation using a common database. Toward this end, special volumes of data have been extracted from DDSM for use in evaluating CAD algorithms that focus on detection of masses or of clustered micro-calcifications. These volumes are named BCRP_MASS_0, BCRP_MASS_1, BCRP_CALC_0 and BCRP_CALC_1. Each volume contains 40 to 50 cases of data. More data about the specific contents of each volume is available on the web page.

The large common data set is the most obvious pre-requisite for sound performance evaluation, and it is probably the most difficult and expensive element to acquire. But now that a large common data set is available, other elements need to be addressed as well. One is the criteria for declaring detections as true positives and false positives and false negatives. There are many ways to do this even for masses, and the issue is more complex for clusters of calcifications (Kallergi et al 1999).

Another subtle issue is what elements of ground truth to count. That is, should an algorithm aim to find only proven cancers, all abnormalities suspicious enough to merit additional workup, or some other criterion? Furthermore should a mass or micro-calcification detection algorithm be evaluated using abnormalities of all types, or just ones that the algorithm was designed to detect?

The DDSM contains the results of a performance evaluation for a new mass detection algorithm using the BCRP volumes. The purpose of including it was to demonstrate a method for evaluating a CAD algorithm and to provide a baseline performance against which other algorithms can be compared. However, in order to directly compare the performance of algorithms, one must use the same data and performance assessment methods.

The method used to evaluate the algorithm was as follows. Each detection region was represented by the location of its maximum suspiciousness value. A detection region was scored as false positive if the pixel location of the abnormality fell outside all ground truth regions. A detection that fell inside a ground truth region counted as a true positive detection. In the event where multiple detections fell inside a ground truth region, one counted as a true positive

detection, and the others counted as false positive detections. This was done because only one

prompt is desired for each ground truth region. Any ground truth region that did not correspond

to a detection location was counted as a false positive detection. Finally, all benign and

malignant masses that were present in the ground truth were used for the performance

evaluation.

Any mass detection algorithm that can output detection sites to a file as pixel locations ordered

by suspiciousness can be evaluated using software available on the DDSM web site. Provided

the algorithm is trained with the BCRP_MASS_0 dataset and the BCRP_MASS_1 dataset is not

used until the algorithm and its parameters are fixed, a performance evaluation done with the

afore mentioned software tools in the DDSM can be directly compared to the results presented

on the DDSM website for this data.

7.0 ACKNOWLEDGEMENTS

REFERENCES

ACR 1998. The ACR Breast Imaging Reporting and Data System (BI-RADS™), Third Edition, http://www.acr.org/departments/stand_accred/birads-a.html.

Kallargi, M., G. Carney, and J. Gaviria. 1999. Evaluating the performance of detection algorithms in digital mammography. Med. Phys. 26(2):267-275.

te Brake, G.M., and N. Karssemeijer. 1998. Comparison of three mass detection methods. In *Digital Mammography*. Dordrecht, The Netherlands: Kluwer Academic Publishers pp:119-126.