World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:10, 2009

Integration of Image and Patient Data, Software and International Coding Systems for Use in a Mammography Research Project

V. Balanica, W.I.D. Rae, M. Caramihai, S. Acho and C.P. Herbst

Open Science Index, Computer and Information Engineering Vol:3, No:10, 2009 publications.waset.org/8003/pdf

***Abstract*—**Mammographic images and data analysis to facilitate modelling or computer aided diagnostic (CAD) software development should best be done using a common database that can handle various mammographic image file formats and relate these to other patient information.

This would optimize the use of the data as both primary reporting and enhanced information extraction of research data could be performed from the single dataset. One desired improvement is the integration of DICOM file header information into the database, as an efficient and reliable source of supplementary patient information intrinsically available in the images.

The purpose of this paper was to design a suitable database to link and integrate different types of image files and gather common information that can be further used for research purposes. An interface was developed for accessing, adding, updating, modifying and extracting data from the common database, enhancing the future possible application of the data in CAD processing.

Technically, future developments envisaged include the creation of an advanced search function to selects image files based on descriptor combinations. Results can be further used for specific CAD processing and other research. Design of a user friendly configuration utility for importing of the required fields from the DICOM files must be done.

***Keywords*—**Database Integration, Mammogram Classification, Tumour Classification, Computer Aided Diagnosis.

I. Introduction

IN order to incorporate mammographic images and data from different medical environments, a procedure usually needed when doing modelling or computer aided diagnostic (CAD) software development, it is necessary to have a common database that can handle various mammographic image file formats. This integration will facilitate the use of the common database as a primary reporting system as well as the enhanced source of information for the extraction of research data. One desired improvement in this field is the integration of DICOM files header information [1] into the database, as an efficient and reliable source of supplementary patient information which is intrinsically available in the images.

Such a research database should be modular and able to integrate the requirements of a record system and should be able to link to images in different formats. This should not necessarily be a reporting system as the image display environment may not be suitable. It should further integrate internationally acceptable norms and codes, namely the BI-RADS (Breast Imaging Reporting and Data System) [2] classification, Tabár's pattern classification [3] and ICD10 (International Classification of Diseases and Related Health Problem 10th Revision) codes [4] for easy and consistent patient classification. The database should also allow for running CAD software to aid analysis of the patient images and information, and the development of tumour models [5].

The aim of this paper is to design a suitable interface and database structure which will cater for the requirements of the future research needs in the field of CAD and tumour modelling.

V.Balanica and M.Caramihai are with the Univerity of Politechnics Bucharest, CO 060042, Bucharest, Romania (Tel: +4021 402 9167; e-mail:vicord20011@gmail.com, m.caramihai@ieee.org)

W.I.D Rae, S. Acho and C.P.Herbst are with the Univeristy of Freestate, CO 9300, Bloemfontain, South Africa (Tel: +27 51 4052825; e-mail: raewid.MD@ufs.ac.za, Gnbich.MD@ufs.ac.za, gnbisa.md@ufs.ac.za).

II. Methods

A. Development Language

The interface was developed in C-Sharp programming language. The choice of selecting C-Sharp over Java was made based on the speed of code execution, and the range of well-known features for easy design of the user interface, data access, database connectivity and network communications. The ease of integration of C-Sharp with other Microsoft Products like Microsoft Office and Microsoft SQL Server was thought to be especially important.

The database for storing, integrating and accessing all available oncologic information about patients was designed and implemented using Microsoft SQL Server 2005. The high performance, scalability and the reliability of the SQL Databases combined with the ease of integration with the .NET Framework was the basis for the choice over the Oracle Server.

Open Science Index, Computer and Information Engineering Vol:3, No:10, 2009 publications.waset.org/8003/pdf

B. Database Structure

The database structure consists of five main tables, namely: patient demographics, study request, study report, image information and lesion information. Provision is made for more than one request per patient, and a request may have more than one associated image, but there will be only one report for each specific request. The database structure is illustrated in figure 1.

The interface was designed to add or change information in any of these tables separately, with feedback to other tables if any of the associated fields are changed.

C. Disk File Structure

The file structure on the disk was designed to accommodate future full migration to DICOM, but the current database has to accommodate DICOM files and also other usual non-DICOM formats (e.g. TIFF, JPEG etc.). As the mammography image files can be very large, processing time cannot be disregarded. For faster processing, thumbnails of two different sizes were additionally created from the originals imported and stored in separate folders. Images were not incorporated into the database, but links were provided to the image file. DICOM file headers provide the information for linking the image to the correct patient and request, while this information needs to be provided by the user for other image file formats. In our interface we considered only two types of image files DICOM and JPEG, but extension to other formats would be easily achieved.

D. Interface

Reports require Tabár [3], and BI-RADS classification [2] as well as ICD10 coding [4]. The Tabár and BI-RADS classifications are subjective assessments of the patient images generally decided on by the radiologist or oncologist. The former is based on the overall mammographic appearance and implies the selection of one of the five (I-V) parenchymal breast patterns, and the BI-RADS unitary system comprises the evaluation of the breast tumour pathology in a concise, unambiguous and standardized way and places the image into one of the six (0-5) possible assessment categories.

The ICD10 coding is selected by the doctor and involves the choice of a standardized ICD10 code classifying the specific patient's diagnosis of breast cancer.

E. Integration With Other Software

Image display was designed to be done directly from the database by selecting the image file from the interface and running public domain display software in a shell from within the database software. Freeware such as XnView [6] was tested for the display of JPEG files, and Sante DICOM viewer (http://www.santesoft.com/) was tested for DICOM image display during development.

Provision is made for automatic analysis of image features (the CAD component) by independent software that will also run within a shell of the interface with image feature extraction transferred indirectly to the database through a temporary text file. This will allow the integration in a standardized manner of the CAD results and allow a better modularization of the interface application.

III. Results

A. Database Structure

The SQL relational database was designed to facilitate the storage and easy access of the required information. The patient demographics were stored in a Patients Table, the study request data in a Requests Table, the study report data in a Reports Table, the image information data in a Films Table, and the lesions information for each film in a Lesions Table (see figure 1). The fields of each of the tables were chosen carefully taking into account both the nature and characteristics of available data and the required fields for research purposes.

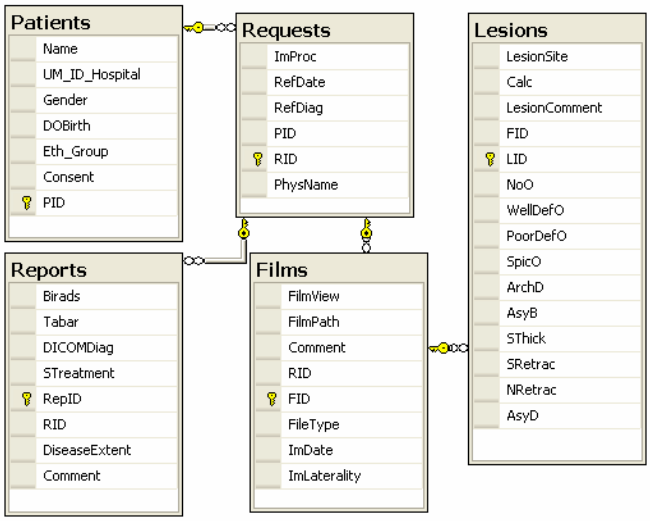


Fig. 1 The database structure consisting of the five main tables showing the links between the tables.

The integration of ICD10 codes is also done by using database resources. A database table is used for storing all ICD10 codes and their descriptions. In the report section, the ICD10 Diagnostic field will display only the codes that contain the words "mammary" or "breast", for easy access to the required data.

B. Disk File Structure

As described above, the available mammographic images that we used are either JPEG or DICOM files. The integration of both image types in a common database necessitates the availability of a common set of patient data information. After the required information was provided by the user (in the case of JPEG files) or imported from the image file (in the case of

Open Science Index, Computer and Information Engineering Vol:3, No:10, 2009 publications.waset.org/8003/pdf

DICOM files), a final procedure was successfully incorporated to check the consistency of the information in the database and to automatically update the required fields.

Simultaneously, the image files, either DICOM or JPEG, undergo a series of transformations to standardize them, such as: integration of the image files in a folder structure that follows the guidelines of the standardized DICOM folder, changing the filenames according to our naming scheme and conversion of the files to two different sized thumbnail images for easy display in the interface. In storing the thumbnail images, the same folder structure as for the original image files was used with only a change in the parent folder name. This makes access to the stored data transparent, based on the image type and on the original image file path.

The folder structure *for DICOM files* will be then:

- "DICOM\PatientXX\StudyXX\imagefilesXX" for all original DICOM files,
- "DICOMTHUMBNAIL\PatientXX\StudyXX\files" for the converted thumbnail images.

and *for JPEG files*:

- "JPEG\PatientXX\StudyXX\imagefilesXX" for all original JPEG files,
- "JPEGTHUMBNAIL\PatientXX\StudyXX\files" for the converted thumbnail images

C. Interface

The layout of the interface comprises four panels that show the available patients, requests, reports and image files. The functionality of the interface is straightforward meaning that it follows the normal workflow structure of a reporting radiologist. The interface first displays all patients available in the database. When a patient is selected in the patient's panel, the corresponding available study requests are shown in the request panel. Once a request is selected, its corresponding available reports and image files are displayed as lists in the appropriate panels.

Searching is possible using several database fields, individually or in combination, allowing ease of image selection for reporting or processing.

Each of the panels has a standardized functionality for adding, deleting and viewing or updating the database information. These are all supervised activities, as all new or updated information will be validated against the database information in order to preserve the uniqueness and consistency requirements of the database information; deletion of information will be accompanied by a warning, raising the level of consciousness about the deleted item, and no information will be deleted if the item selected for deletion has other related data stored in the database.

Adding or importing an image file requires a stepwise workflow specific to each JPEG or DICOM image. As opposed to the DICOM images that contain patient and study information that is automatically imported, JPEG files require the user to enter patient and study data to define the links for the JPEG file. This step can be avoided if prior selection of the patient and study is done using the interface. For a DICOM file the interface will automatically check the database for the patient information found in the DICOM header. The information from the DICOM header will always update the existing information within the database if the date of the DICOM image file is more recent than the patient creation date. In this way, the database will always be updated to the newest available information.

The interface also provides support for manual image file display and manipulation. All images available for a certain study can be reviewed in a separate window, where, for each of the images, evaluation of any observed lesions can be done. Each lesion has to be described in terms of the lesion site, presence of calcifications and the observed appearance and these findings should be entered in the database table fields. After review the reporter can proceed to the report section where the Tabár, BI-RADS classification and the ICD10 code can be selected. All this information is stored in the database in the appropriate tables.

D. Integration With Other Software

The conversion tool used for creating the thumbnail images of the original image files is NConvert software, a tool included in the XnView package. The two thumbnail images are 160x120 and 320x240 pixels in size and are displayed in the interface during the review process.

A detailed review of the images can be done by the integration of public domain display software that is being run from the shell using the original image files.

IV. Discussion

The purpose of this paper was to design a suitable database that can link and integrate different types of image files and gather common information that can be further used for research purposes. Moreover, an interface was developed for accessing, adding, updating, modifying and extracting data from the common database, enhancing the future possible application of the data in CAD processing.

From the technical perspective, future developments are envisaged including the creation of an advanced search function that selects image files based on advanced combinations of descriptors. The results can be further used for specific CAD processing and other research. Design of an easy-to-use configuration utility for the required fields imported from the DICOM files must be done.

Acknowledgment

This material is based on work supported by the SA National Research Foundation (NRF) under Grant Number UID 67953.

This material is based on work supported by the Romanian National Research Foundation under Project Number: 101/08.

Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF does not accept any liability in regard thereto.

Ethical approval for this study has been given by the Harmonised Ethics Committee of the University of the Free State and is carried out as study: Mammography Computer Aided Diagnosis Development and Implementation ETOVS NR 72/07A and Sub-study: Mammary Malignancy Modelling using Artificial Intelligence ETOVS NR 72/07B.

REFERENCES

[1] DICOM Home page, Accessed at http://medical.nema.org/ on 24/08/2009.
[2] BI-RADS, *What is BIRADS*? Accessed at http://www.birads.at/info.html on 24/08/2009.
[3] L. Tabár, T. Tot, P.B. Dean, *Breast Cancer The Art and Science of Early Detection with Mammography* Thieme, Stuttgart, 2005, Ch. 1.
[4] World Health Organisation, ICD 10, Version 2007, Accessed at http://apps.who.int/classifications/apps/icd/icd10online/ on 24/08/2009.
[5] M. Elter, A. Horsch, "CADx of mammographic masses and clustered microcalcifications: A review," Med. Phys. 2009, 35 (6) 2052-2068.
[6] XnView, Accessed at http://www.xnview.com/ on 24/08/2009.

Open Science Index, Computer and Information Engineering Vol:3, No:10, 2009 publications.waset.org/8003/pdf