USING FREE TECHNOLOGY TO STORE AND RETRIEVE MAMMOGRAPHIC IMAGES BY INTERNET

Fátima L. S. Nunes[1], Homero Schiabel[2], Claudio Eduardo Góes[2], José Alaor de Oliveira Jr[1], Rodrigo H. Benatti[2],

[1]Centro Universitário Eurípides de Marília, Marília (SP), Brasil
[2]Laboratório de Análise e Digitalização de Imagens, Departamento Engenharia Elétrica – EESC/USP, São Carlos, SP, Brasil
fatima@fundanet.br

ABSTRACT

This paper presents a system to manage a mammographic image database through the Internet. The first goal is making available a large number of digitized mammographic images with characteristics that can serve the needs of computer-aided diagnosis schemes. The second goal is allowing the use of such images as a tool to aid training programs in this area. Four stages were necessary to construct the computational tool: requisites definition, images storage, programs development and Internet publication. The results demonstrate that is possible to build computational system with this purpose using free technology.

1. INTRODUCTION

Computer-aided diagnosis (CAD) schemes have been developed by many research groups around the world, aiming aid the early detection of breast cancer. It is well known that when this disease is discovered in its initial developing phase the chances of cure are bigger [1]. Full schemes supply the identification, localization and classification of structures such as microcalcifications and nodules. Thus, they can contribute to the early detection even in assimptomatic patients. In previous papers [2] a computational scheme aiming to provide useful information to radiologist diagnosis relative to microcalcifications cluster detection was presented. It has used techniques to segment, locate and classify those signals.

One of the main difficulties found during the developing of these techniques is the procedures evaluation. Measuring the efficacy of a technique is not a simple task, since the results can vary according to the images set used during the tests [3]. In order to attest the feasibility of a technique, a wide set of images is necessary. It should include, preferentially, images with diverse acquisition characteristics and also attend the requirements of the technique purposes, that is, they must contain the structures expected by the detection process. Organizing an image database as required involve an intense research in hospitals and clinics in order to obtain the mammograms and the corresponding medical reports. It also requires a hard work for digitizing these images and an efficient and systematic cataloguing to allow a fast and accurate retrieval according to their characteristics. Also the images selection and retrieval are not easy tasks, demanding many time and human resources. Some research centers have offering sites where images used in diagnosis are available for download, but the most of them offers a small quantity of images or images without enough information for CAD schemes. From these considerations, this paper presents a system aiming: (1) to provide a mammographic images set which has been structured during some years [4]; (2) to allow the insertion of new images and information in this database and (3) to allow easy access to these images by Internet. To reach these objectives, an accurate management was defined by building a suitable database to store information from medical reports and enabling a fast image and information retrieval by using a friendly research system.

2. MATERIALS AND METHODS

The tool presented here was developed in four steps. We should emphasize that all tasks were developed using free technologies: Linux operational system, PHP programming language and MySql database. Thus,

expanding the use of this system is possible to any place having an Internet connection without additional costs.

In the first step, the images characteristics as well as the digitization equipment, patients and mammographic exams were studied in order to define a database relational model to store the necessary data in a consistent manner, without redundancies. With the goal of establish the system architecture, a survey of potential user types was made. The attributes that must be recorded in the tables were defined from studies about queries that could be performed by all kind of user. Figure 1 shows the resultant data model.

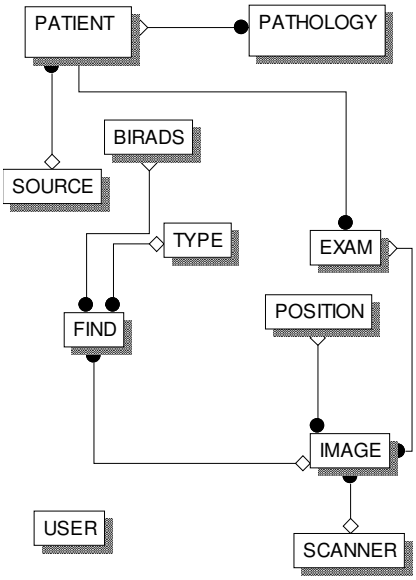

Figure 1 – Database Model.

The main task of the second stage was the programs development in order to allow the information storage and retrieval. From the system requisites analysis, some available technologies options were studied, always with the requirement of allowing the project development without additional costs relative to software acquisition and maintenance. As mentioned before, the Linux operational system was chosen, besides PHP language and MySQL database. The developed software allows 3 levels of users: common (can retrieve and copy images); special (can retrieve, copy and include new images) and administrator (can authorize new users and also include, change and eliminate any information of the database). A user must be registered to obtain access authorization. During the registration, the user must supply some information which will be used later in order to verify the effective use of the images database. First, programs to allow the maintenance of all tables were developed, complying with their relationships. Thus, administrator can include, alter or exclude data without worrying with consistency problems. Then, a program to perform a generic query was implemented aiming to retrieve information from the most diverse storage attributes. The interface of this query is shown in Figure 2.

The database composition was the third phase. Two laser scanners were used to digitize the mammographic films with 12 bits of contrast resolution: a Lumiscan 50 (*Lumisys, Inc.)*, with 0.15mm pixel size and a Lumiscan 75, with 0.075 mm pixel size. The images were recorded using TIFF format, without loss of information. Regarding the possibility of their visualization in the Internet and also to decrease the amount of storage space, the images were converted to JPEG format too.

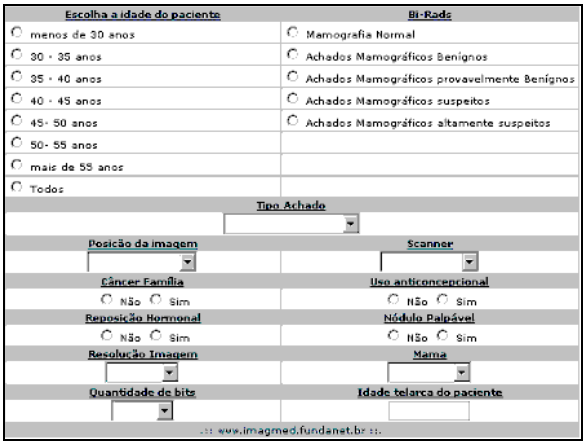

Figure 2 – Generic query interface.

The last step was to place the images set in the World Wide Web. The tasks of this stage were preparing a WWW server, filling the tables with data from digitized images of previous stage and developing home-pages for accessing the system, considering the three levels of users mentioned before. The filling of the tables was performed by a program to convert the previous database [4] into the current model.

3. RESULTS AND DISCUSSION

Until now about 3500 images were included in the database. All of them have medical reports available. These images were obtained from five hospitals, with different acquisition characteristics. The structures present in the images are too diverse: there are mammograms of normal cases (without indicated anomalies in the medical reports) and many images with sparse and clustered microcalcifications, calcifications, suspect masses and asymmetric densities. Also the age interval of patients has a big variety: there are lots of dense breast mammograms, usually related to young women (less than 40-45 years

old) and mammograms of women between 40 up to 80 years old.

The approach of maintenance allows to the administrator to include, eliminate or modify data from any place where there is an Internet connection point. Due to the built interface, performing these operations is veruy easy. The administrator do not need to know about database systems or programming languages to operate. The structuring in levels of users provides more security to the system. Therefore the access to each group of users is controlled by the operational system. Common users can only consult the database and download the images from it; intermediate users can also include images in the database, after proper authorization.

To assure the consistency of included data to this kind of user, the programs are executed in a previous determined order: firstly the patient data must be included, then the exams of the patient and the respective images and, finally, the found anomalies related to each image. If any data is erroneously included, the administrator can alter or eliminate it.

The retrieval by various parameters helps users to find the kind of images is suitable for their needs. Once the parameters are supplied, the user can choose which image format he wants to copy (TIFF or JPEG). He also can choose to download one image of all the set of images that attend one requirement. Figure 3 shows this functionality.

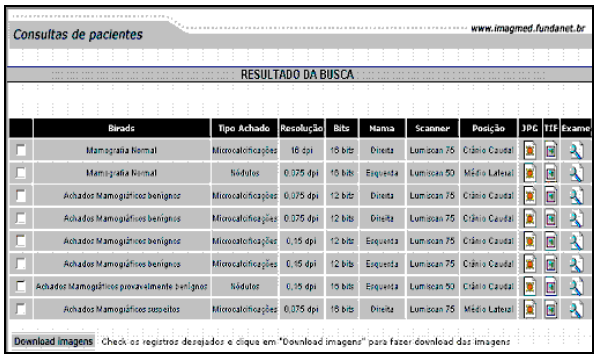

Figure 3 – Download of selected images.

4. CONCLUSIONS

We could verify that developing sites with this purpose is possible using only free technologies. In our tests we realized that the performance of the global system is very good. Since the database does not have a heavy and complex information data set, the main requirement of the equipment is space of storage and a fair memory volume to turn fast the queries answer, even if a big quantity of users is using the system at the same time. Thus, making images database available by Internet demand more hardware than software requirements. Due to these characteristics, the technology without additional costs, as the used in this project, reaches the desired goals.

The second conclusion is regarding to the usefulness of providing medical images in the Internet. As mentioned previously, one of the main challenges of CAD schemes is the validation of its efficacy. Hence, different results can be obtained when different images sets are used in the tests. Organizing images set for testing is not an easy task because it should have lots of images and include the most diverse structures. In addition, the images should be digitized with some basic required acquisition characteristics. Beyond the time to digitize and store the images, some specifications for the digitizer equipment should be attempted, which usually increases the costs. We believe that computational system of this type can be the initial point to establish standardization in the CAD schemes evaluation. The performance of different schemes can thus be compared since they can be tested with the same images set.

It is also important to emphasize the easiness of using database images by Internet with the features presented here: interested users from any part of the world can access and download images in a short time. The architecture of the system facilitates the retrieval and also the inclusions of new cases, allowing improve the database content. This feature can make the system also useful as an educational tool, regarding the large variety of structures presented in the database. The images can constitute complementary didactic material in courses related with this knowledge area.

A last topic to point out is that the proposed relational model also aims the expansion of the current database. It can be made from inclusion of other types of images, coming from other modalities (ultrasound and magnetic resonance, for instance) or images of other parts of human body that can be useful to the researchers community in this area. By making use of this structure, new tables can be added to the model, allowing to elaborate suitable queries to such images. Thereby, we believe that this project can contribute to the development of researches in CAD, by minimizing a challenge: the construction of material sets to evaluate these schemes.

5. REFERENCES

[1] S.A. Feig and R.E. Rhendrick, "Risk, benefit and controversy in mammographic screening", in *A Categorical Course in Physics. Technical Aspects of Breast Imaging,* A.G. Haus and M. J. Yafee, Eds, Radiological Society of North America, 1993, p. 21-33, 79th scientific assembly and annual meeting of RSNA.

[2] H. Schiabel, F.L.S. Nunes, P.M. Azevedo Marques, A.F. Frère, "A computerized scheme for detection of clusters of microcalcifications by mammograms image processing", *Medical & Biolological Engineering & Computing*, v.35 (Supplement Part 2), p. 705, 1997.

[3] R.M.Nishikawa et al., "Performance of automated CAD schemes for the detection and classification of clustered microcalcifications", *Digital mammography.* A.G.Gale et al., eds. Elsevier Science Publishers B.V., p. 13-20,1994.

[4] H. Schiabel, F.L.S. Nunes, M.C. Escarpinati, C.E.Góes, "Investigations on the effect of different characteristics of images sets on the performance of a processing scheme for microcalcifications detection in digital mammograms", *Journal of Digital Imaging,* v.14 (2), p.224-225, 2001.