

MammoApplet: an interactive Java applet tool for manual annotation in medical imaging

Christian Mata, Arnau Oliver, Albert Torrent, and Joan Martí

Dept. of Computer Architecture and Technology

University of Girona

Campus Montilivi, Ed. P-IV, 17071, Girona (Spain)

Email: {cmata,aoliver,atorrent,joanm}@atc.udg.edu

Abstract—Web-based applications in computational medicine have become increasingly important during the last years. The rapid growth of the World Wide Web supposes a new paradigm in the telemedicine and eHealth areas in order to assist and enhance the prevention, diagnosis and treatment of patients. Furthermore, training of radiologists and management of medical databases are also becoming increasingly important issues in the field. In this paper, we present *MammoApplet*, an interactive Java applet interface designed as a web-based tool. It aims to facilitate the diagnosis of new mammographic cases by providing a set of image processing tools that allow a better visualization of the images, and a set of drawing tools, used to annotate the suspicious regions. Each annotation allows including the attributes considered by the experts when issuing the final diagnosis. The overall set of overlays is stored in a database as XML files associated with the original images. The final goal is to obtain a database of already diagnosed cases for training and enhancing the performance of novice radiologists.

Index Terms—Web-based application, e-learning tool, medical database, medical imaging.

I. INTRODUCTION

The emergence of large digital volumes of information in medical imaging and the rapid growth of computer-based clinical exams have demanded to increase the amount of available on-line medical imaging systems. Digital imaging technologies have become beneficial to modern medical practices and health care systems, providing powerful tools for diagnosis, treatment, and surgery.

The Picture Archiving and Communication Systems (PACS) technology along with the Digital Imaging and Communications in Medicine (DICOM) protocol were developed and quickly became the most common means for acquisition, storage, communication, and display of digital images during the last decade [1], [2]. The access to a PACS system is based on a client-server architecture, and it is a common procedure for the radiological and eHealth communities to access PACS systems not only in a local environment but also remotely. In this latter case, learning purposes, assessment of singular clinical cases, and medical decision support [3] are

This work was partially supported by the Spanish Science and Innovation grant nb. TIN2011-23704. C. Mata holds a grant provided by the Reference Network in Advanced Production Technologies (XaRTAP) of Generalitat de Catalunya.

the main usage scenarios. Medical support decision systems are becoming essential in order to assist the experts in issuing the diagnoses [4], [5]. Current research efforts are directed to incorporate learning capabilities for training and guidance of newly employed or resident radiologists. One of the challenges is designing web-search tools in order to obtain digital images according to the different attributes used when issuing a diagnosis. Although several systems are available [6], [7], [8], most of them are of general purpose, even if they are tailored to a thematic image collection [9].

In this paper a new web-based medical tool intended to facilitate the inclusion of medical findings on existing mammographic cases is presented. *MammoApplet* is implemented as an interactive JAVA applet that allows to create XML files containing the manual annotations provided by the experts (or even the automatic findings provided by automatic CAD algorithms [10]) and store them as a part of the medical case being analyzed. The aim of a database of already diagnosed cases is twofold: it serves as a training platform to train resident and young radiologists as well as a medical support decision system which retrieves similar cases to the one being analyzed.

II. FRAMEWORK

MammoApplet is part of the MammoDB tool [11], a web-based tool with the purpose of training and assisting radiologists in the diagnosis of mammographic images. It is characterized by the ability to retrieve relevant images from a digital mammographic database based on their visual, semantic, and lexical contents. These criteria are based on a query form with the purpose to detect mammographic findings according to their number, location, shape, size, margins, density, and distribution. In addition to a comprehensive list of clinical cases contained in the database, the query forms allow to perform a search according to specific medical criteria: reported diagnosis, BIRADS classification, and presence of abnormalities.

In this paper we present the *MammoApplet* tool as the first step to create a comprehensive list of annotated clinical cases and store them in the database. As can be seen in Figure 1, *MammoApplet* is a web-based platform that allows drawing accurate overlays of mammographic findings, like masses or

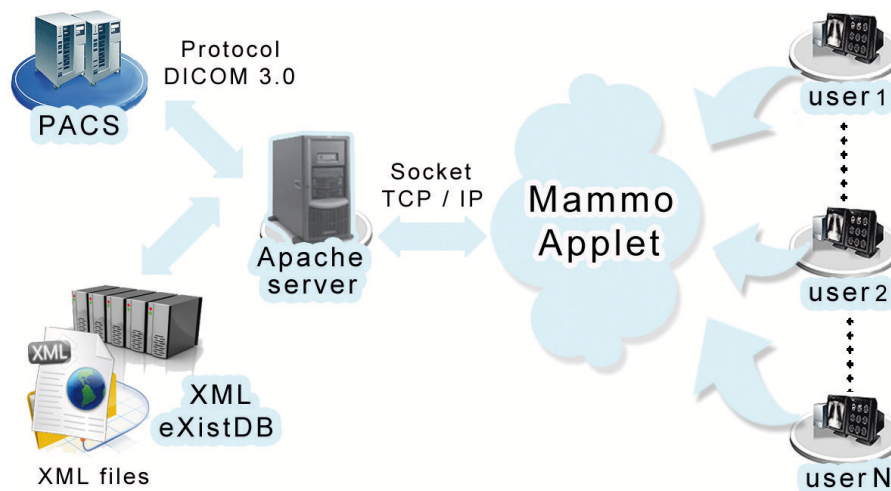


Fig. 1. Interaction between PACS server, MammaApplet and XML database.

micro-calcifications lesions, and provide facilities to annotate other features related with lesions, like the subtlety of the lesions, or related to the breast characteristics, like its internal density. Once the case is fully annotated the platform stores them in a database of annotations without modifying the original DICOM images.

A. Objectives

The main technological features when developing *MammaApplet* are:

- Guarantee the connection and the security access to the PACS server using the DICOM standard format.
- Use of a search engine of mammographic images based on their lexical information (patients, studies and/or images).
- Implement a JAVA applet tool in order to manage the most common actions in image processing and visualization (zoom, contrast, BIRADS lexicon, brightness,...).
- Allow the annotation of findings in the mammograms using polygonal marks.
- Characterize in an automatic way the morphological features (location, shape, size) and the imaging ones (intensity distribution, texture) of the annotated region.
- Issue a final diagnosis and store all the annotations of a mammogram in a single XML file associated to the mammographic study.
- Allow the annotation by different experts. It is important that more than one expert analyze the same image in order to obtain robust annotations.
- Integration of the JAVA applet with a database of medical images. Although this platform has been initially developed for mammography, we are currently extending the application to other imaging modalities.

The annotations are stored as XML files, one per mammogram (not per case). In this way part of the XML file is repeated for the images of the same patient, but it allows a

faster access to the overlay information. Therefore, the XML file has the following structure:

```

<?xml>
<root>
  <patient>
    Information related to the patient
  <study>
    Information related to the study
    Information related to the breast
  <series>
    Information related to the series
    <list of overlays>
      <overlay 1>
        Information of the overlay
      </overlay 1>
      ...
      <overlay N>
        Information of the overlay
      </overlay N>
    </list of overlays>
  </series>
</study>
</patient>
</root>

```

B. DICOM query levels

DICOM is a standard format in medical image communication developed with the goal to store and to easily retrieve the internal information associated not only with the image but also with the patient. Hence, a DICOM query may be used to retrieve information about patients, studies, series and images. The DICOM attributes are divided into different level groups (PatientID, StudyInstanceUID, SeriesInstanceUID or InstanceUID). According to Pinykh [12] each level is defined as:

Login Search Visor			
Patient ID	Patient Name	Institution Name	Year of Birth
hjt001	anon_hjt_001	Hospital Josep Trueta	1937
hjt002	anon_hjt_002	Hospital Josep Trueta	1952
pts001	anon_pts_001	Parc Tauli de Sabadell	1945
pts002	anon_pts_002	Parc Tauli de Sabadell	1948
pts003	anon_pts_003	Parc Tauli de Sabadell	1939
pts004	anon_pts_004	Parc Tauli de Sabadell	1950
pts005	anon_pts_005	Parc Tauli de Sabadell	1946
pts006	anon_pts_006	Parc Tauli de Sabadell	1938

Study ID	Manufacturer	Acquisition Date	Number of Series
0	LORAD / Mammo-Clinical	04/10/2004	1
1	LORAD / Hologic Selenia	09/10/2006	2
2	LORAD / Hologic Selenia	04/09/2008	2

Series ID	Acquisition Date	DICOM images	XML files
1	04/09/2008	4	1
2	04/09/2008	2	2

Patients
 Studies
 Series

Fig. 2. Example of a standard query provided by the search engine.

- Patient: Contains personal data of the patient (patient name, date of birth, patient comments ...). The key attribute in this level is “PatientID” which normally is associated with his medical historic number and indispensable for each DICOM.
- Study: Contains the information of the study (study date, acquisition number, study description ...). The key attribute is “StudyInstanceUID” which should be unique for each study.
- Series: Contains the information of a series (modality, series number ...) and is provided by a sequence of images for each study. The key attribute is “SeriesInstanceUID”.
- Instance image: Contains the information of the instance image. The key attribute is “InstanceUID” which is unique for each image.

As noticed by Dreyer et al. [13], the DICOM format establishes a hierarchy for the four information levels: one patient has one or more studies, and each study is associated with one or more series, each of them associated with one or multiple instance images. Therefore, *MammoApplet* uses this hierarchy in order to retrieve all the images of a patient, as can be seen also in the XML file structure. In the case of mammography, a series usually consists in four images, obtained according to the two views (CC and MLO) of both breasts. Sometimes it is necessary to obtain an additional series with a different view (ML), hence consisting of two additional images.

III. MAMMOAPPLET: AN INTERACTIVE JAVA APPLET TOOL

MammoApplet has been developed in Java, which is an object-oriented, multi-platform, and free distribution programming language. The interface has been created as an applet,

therefore it is visible as a web page and can be executed from a web browser. Hence, it is not necessary to install any additional program on the computers and is independent of the operating system. This allows a full portability of the application.

Next sections explain the architecture of the system and the access security policy, the search engine system, and the main characteristics to annotate a mammographic study and store it in the XML format.

A. Connection and security access

The interface is designed as an applet, hence the system needs a web server to maintain the webpage. This server allows also the communication of the users with both the PACS and the XML databases. A Java applet, for security reasons, can only establish communication with another application which is stored on the same machine as the web server (where the applet is executed), and cannot store data on remote machines. Thus, the user interface communicates with the server through sockets using the standard TCP/IP protocol and then the server communicates with the PACS using the DICOM 3.0 protocol, and with the XML database through XML queries, as it is graphically shown in Figure 1. Notice that the XML database may be stored in the same PACS machine or in a different one [14].

Note also in the figure that different users can connect to the server simultaneously. This is another advantage of using the World Wide Web paradigm, since multiple users can work simultaneously with the same database. Note, however, that to avoid collisions a protocol must be defined, and requests are served in the order they arrive to the server.

In this moment the security of the system is done via username and password, requested by the JAVA applet. Once

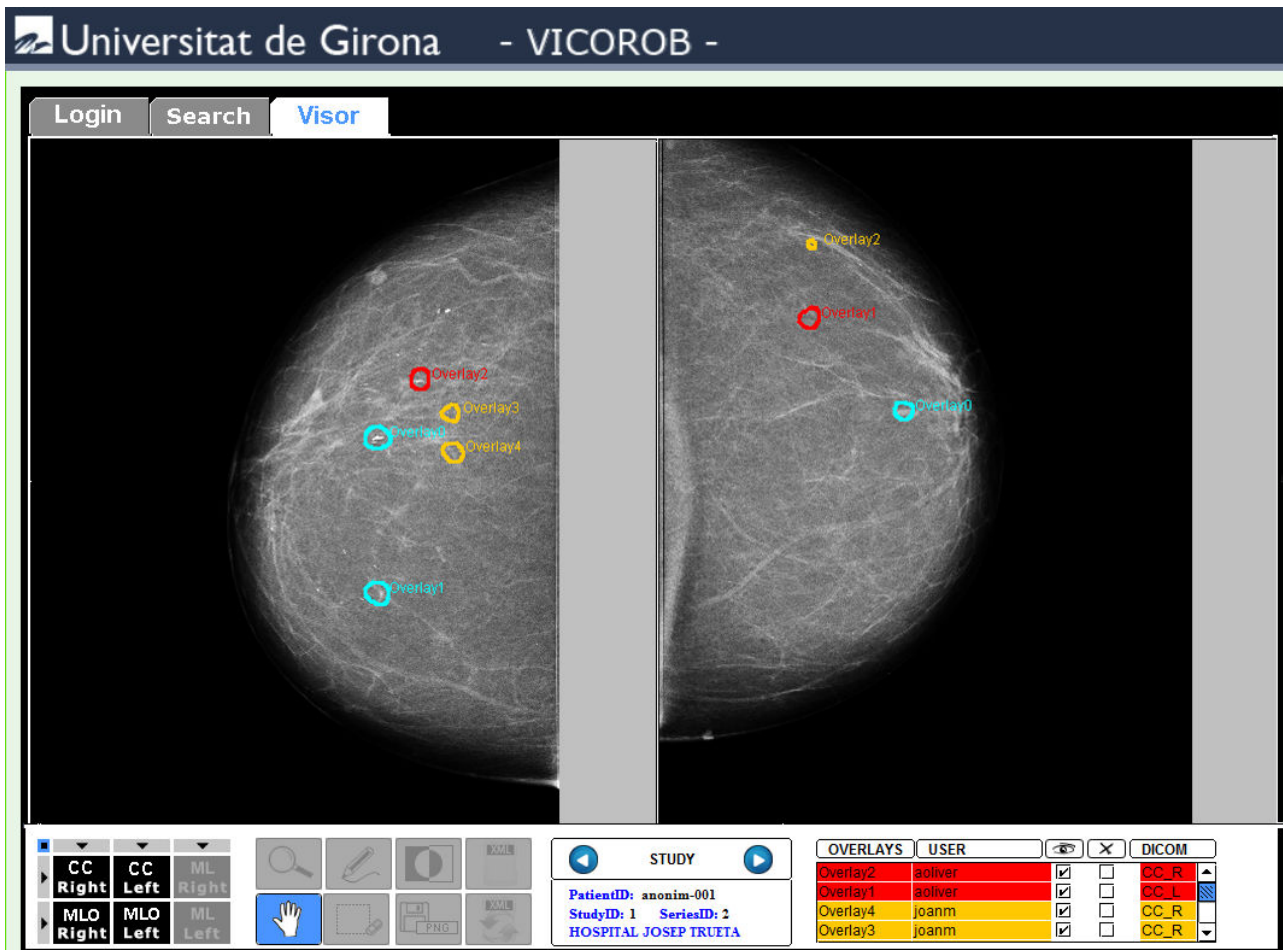


Fig. 3. Example of a visualization of two mammographic images with their overlays. This example shows the CC view of right and left mammograms.

the user is validated, a home window is displayed to welcome to the *MammoApplet* tool. We are currently working on improvements on this security issues by encrypting the data.

B. Search engine

Searches within *MammoApplet* are performed through requests to the PACS using the DICOM protocol. The obtained data is then arranged according to the patient, studies, and series attributes, which later enables to refine the search. Searches can be exhaustive, where the PACS returns a list with all the patients that have undergone a study, or selective, where only a reduced subset of cases which fulfills the search condition is returned (note that the selection criteria can affect the patient, the institution, or the year of birth).

Once a patient is selected, *MammoApplet* returns all the studies contained in the PACS server associated to that patient. The information concerning those studies is related to the following DICOM attributes: studyID, manufacturer, acquisition date, and the associated number of series.

Once a study is selected, *MammoApplet* returns all the series associated to that study. Now, the information concerning that study is arranged according to the following series attributes:

SeriesID, acquisition date, number of images, and number of XML files.

Figure 2 shows an example of the search process. Notice that the interface is divided into three different areas: from up to down, the first one displays the list of patients, the second one displays the studies associated to the selected patient, and finally, the third one corresponds to the series that belong to the selected study. When a series is selected, the visualization panel is displayed on a new tab showing the DICOM images associated to the case, as well as the XML files containing the manually annotated overlays.

C. Visualization

The *MammoApplet* interface has been designed in a simple and intuitive way for radiologists, with an easy access to all the tools available. The interface panel is divided into two parts, as it can be seen in Figure 3. The upper (and larger) part is dedicated to the display of the DICOM images, while the bottom part contains a set of buttons used to perform different actions. From left to right: visualization tools, image processing tools, information related to the study, and information related to the manual annotations (overlays).

The visualization tools offer different possibilities to arrange the images. By default, the first option is the visualization of the four images usually associated with a standard study. The panel is provided with different buttons to visualize two images according to the selected view or breast, and also users have the possibility to visualize just a single image. This latter way of visualization offers to the radiologists the possibility to create overlays with the full resolution of the mammograms. Figure 3 shows an example of the visualization of the CC view of left and right mammograms. This view is useful for the radiologists because it allows detecting breast asymmetries between the right and left breasts.

The image processing tools offer a set of tools like zoom control, scroll contrast, inverse image, import the images, or scroll brightness. These tools are important for the radiologists since they can see the behavior of the breast in different conditions.

The third part of the bottom panel is the box that shows the main data of the patient, study, series and the institution where the study was acquired. A novelty of this applet is that it allows scrolling the previous and the following studies of the same patient. This is very useful since the comparison of previous studies, i.e. the analysis of the breast evolution, is a procedure used for the experts to diagnose the patient.

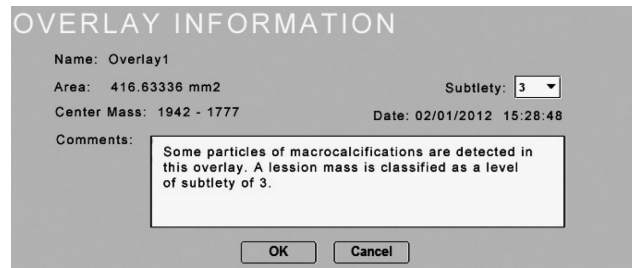
The last part of the panel is related to the annotation facilities, and is explained in the next section.

D. Manual annotation

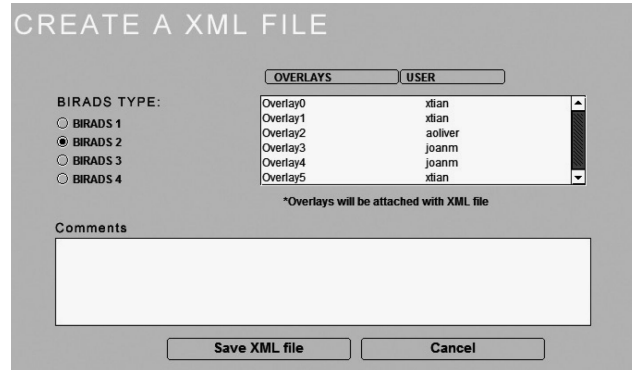
The visualization panel allows different actions related to the creation and manipulation of overlays. As can be seen in Figure 3, overlays are displayed in the images as well as in the overlays table, located in the right part of the bottom panel. Since different experts can analyze the mammogram, we decided to show the overlays in a different color per expert. This way, it is easier for radiologists to know which overlays have been made by themselves and, besides, it allows a fast comparison of the performance of the different experts.

We have implemented the following options to view and manipulate the overlays:

- Draw an overlay. Radiologists can add new areas of interest in the mammograms by simply clicking left and right buttons of the mouse.
- Annotate an overlay. Once an overlay is created, a pop-up window like the one shown in Figure 4(a) appears. This window displays information related to the annotation, like the area and the center mass location, which are automatically computed, and also allows the expert to introduce extra information, like the subtlety of the lesion (degree of difficulty to find the lesion) or any additional comments regarding the lesion.
- Delete an overlay. Radiologists can delete any overlay by selecting the area where it is located.
- Create/update XML files. In order to store information, a pop-up window like the one shown in Figure 4(b), is displayed. The radiologists select the breast density



(a)



(b)

Fig. 4. Examples of pop-up windows corresponding to (a) overlay information and (b) XML file.

(according to the BIRADS lexicon) and may introduce comments regarding the final diagnose.

All the information related to the annotations is also shown in the overlays table (right part of the bottom panel of Figure 3). This table shows the lists of overlays created for each case, specifying the username that created the overlay and the mammogram where the overlay is. It also allows to visualize or to delete each overlay, although this latter option is only available to the expert who created it. It is easy to identify the overlays of each expert because the table shows a different color for each user. Selecting an overlay of the table allows to display its related information (the annotations filled in the pop-up window by the expert).

IV. AN INDICATIVE EXAMPLE OF USE

In this section, an indicative example of how to use *MammoApplet* is described. The first step is to connect with the interface using a web browser, and login to the interface using the identification process.

The next step is to use the search form in order to retrieve the patients, studies and series, as is shown in Figure 2. Subsequently, the user should select which study and series he wants to analyze. This study will be displayed in the visualization panel, with the associated DICOM images. Moreover, all the existing overlays (if any) will be painted surrounded the suspicious part of the image and all the attributes are updated according to the overlays.

Subsequently, the user can perform different actions, such as visualizing the DICOM images with different views, or

using the bottom panel to perform different image processing actions. The most important action, however, is to use the provided annotation tools to mark the overlays on the images. The users can draw a polygonal area to the suspicious regions of the image, and associate these regions to the manual annotations through the pop-up window that is displayed when closing the polygon.

Finally, when the user finishes, the created overlays are saved in a XML file and automatically stored in the XML database for posterior analyses. Therefore, *MammoApplet* provides annotated cases for a medical decision support tool for radiologists and health communities.

V. CONCLUSIONS AND FUTURE WORK

We have presented an interactive web-based application, implemented as an applet interface, with the purpose to provide a manual annotation tool for DICOM images. The final goal is to help radiologists to diagnose more effectively the breast cancer in digital mammography by providing similar already diagnosed cases. We would like also to remark that although the presented applet has been designed for mammography, it is easily portable to other kind of medical images. For instance, we have extended to deal with MRI data, and nowadays it is connected with a database containing both MRI and spectroscopy prostate data sets, allowing the visualization and diagnostic of the prostate through different views and sequences (i.e. T2, perfusion, diffusion, spectroscopy) [15].

We are working on adding more image processing algorithms in order to incorporate new state of the art techniques to help radiologists to visualize and manage the images. Concurrently, we are working on data encryption to improve the security policy. Finally, we are planning the integration of effective e-learning tools to train new radiologists and to improve the interaction between the expert and the novel physicians.

An exhaustive experimental study is currently being carried out in order to evaluate the scalability, complexity and response speed of the proposed tool.

REFERENCES

- [1] H. Huang, "Short history of PACS – part I: USA," *Eur. J. Radiol.*, vol. 78, pp. 163–176, 2011.
- [2] H. Lemke, "Short history of PACS – part II: Europe," *Eur. J. Radiol.*, vol. 78, no. 2, pp. 177–183, 2011.
- [3] K. Fridell, L. Edgren, L. Lindsköld, P. Aspelin, and N. Lundberg, "The impact of PACS on radiologists work practice," *J. Digital Imaging*, vol. 20, pp. 411–421, 2007.
- [4] R. Carvalho, R. Isola, and A. Tripathy, "MediQuery: An automated decision support system," *Int. Symp. Computer-Based Medical Systems*, pp. 1–6, 2011.
- [5] S. Karim and I. Bajwa, "Clinical decision support system based virtual telemedicine," in *Intelligent Human-Machine Systems and Cybernetics*, 2011, pp. 16–21.
- [6] D. S. Kaster, P. H. Bugatti, M. Ponciano-Silva, A. J. M. Traina, P. M. Azevedo-Marques, A. C. Santos, and C. Traina Jr, "MedFMI-SiR: A powerful DBMS solution for large-scale medical image retrieval," in *Information Technology in Bio- and Medical Informatics*, 2011, pp. 16–30.
- [7] J. Miñarro-Giménez, T. Miranda-Mena, R. Martínez-Béjar, and J. Fernández-Breis, "Exploitation of translational bioinformatics for decision-making on cancer treatments," in *Information Technology in Bio- and Medical Informatics*, 2011, pp. 1–15.
- [8] P. Welter, T. Deserno, B. Fischer, R. Gunther, and C. Spreckelsen, "Towards case-based medical learning in radiological decision making using content-based image retrieval," *J. BMC Medical Informatics and Decision Making*, vol. 11, no. 1, p. 68, 2011.
- [9] V. Castelli and L. D. Bergman, *Image Databases: Search and Retrieval of Digital Imagery*. John Wiley and Sons, 2002.
- [10] A. Oliver, X. Lladó, J. Freixenet, R. Martí, E. Pérez, J. Pont, and R. Zwiggelaar, "Influence of using manual or automatic breast density information in a mass detection CAD system," *Acad. Radiol.*, vol. 17, no. 7, pp. 877–883, 2010.
- [11] J. Martí, C. Mata, X. Lladó, and A. Oliver, "MamoDB: A web-based tool for training radiologist in the diagnosis of digital mammography," in *Int. Conf. Education and New Learning Technol.*, 2011, pp. 2359–2367.
- [12] O. S. Pinykh, *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide*. Springer, 2011.
- [13] K. Dreyer, A. Mehta, and J. Thrall, *PACS: a guide to the digital revolution*. Springer, 2002.
- [14] eXist official web page, "<http://www.exist-db.org/>," accessed on 02/02/2012.
- [15] C. Mata, P. Walker, A. Lalonde, F. Brunotte, A. Oliver, and J. Martí, "Prostateanalyzer: Gui in medical domain with management of dicom images of prostate cancer," in *Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology*, 2012.