Universitat
de Girona

WEB-BASED APPLICATION FOR MEDICAL IMAGING MANAGEMENT

Christian MATA MIQUEL

Dipòsit legal: Gi. 1967-2015
http://hdl.handle.net/10803/323093

PhD Thesis

Web-based Application for Medical Imaging Management

Christian Mata Miquel

2015

PhD Thesis

Web-based Application for Medical Imaging Management

Christian Mata Miquel

2015

Doctoral Programme in Technology

Supervised by:

Joan Martí and Arnau Oliver

Alain Lalande and Paul M. Walker

Work submitted to the University of Girona and University of Burgundy in partial
fulfilment of the requirements for the degree of Doctor of Philosophy

A tots els que m'han ajudat a arribar fins aquí,
als que ens han deixat però sempre hi seran.

Dr. Joan Martí, from Universitat de Girona, and
Dr. Alain Lalande, from Université de Bourgogne,

DECLARE

That the work entitled *Web-based Application for Medical Imaging Management*, presented by *Christian Mata* to obtain the degree in Doctor of Philosophy, has been developed under our supervision and complies with the requirements needed to obtain the International Mention.

Therefore, in order to certify the aforesaid statement, we sign this document.

Girona, April 2015.

Agraïments

Quantes vegades esperava aquest moment per poder agrair de forma meritòria a tots els que m'han ajudat dia a dia a arribar fins aquí. La vida és una obra d'art on cada pinzellada compta. Gràcies a cadascun de vosaltres que heu fet possible que aquest treball que aquí presento s'hagi fet realitat.

En primer lloc vull agrair amb molt d'afecte i de forma explícita als meus directors de tesi Joan Martí i Arnau Oliver pel suport, la paciència i l'ajuda per portar a bon port tanta feina realitzada durant aquests anys. Especialment, donar les gràcies a en Joan per donar-me l'oportunitat de realitzar el doctorat i confiar amb mi quan més ho vaig necessitar. I a l' Arnau, entre moltes altres, per aconseguir-me les beques de mobilitat per les meves estades franceses (ho feies per la mostassa de Dijon eh?). El camí no ha sigut gens fàcil i malgrat els múltiples entrebancs que han sorgit per tirar endavant tot això, mai m'heu fallat.

De la mateixa manera donar les gràcies a en Xevi Cufí (arigató tó tó...) per buscar sempre solucions i aconseguir una beca de la XARTAP per finançar el doctorat. També a en Robert Martí per donar-me un cop de mà amb projectes de col.laboració, a en Jordi Freixenet i Xavi Lladó pels seus consells i suport, a l'Albert Torrent (el que va iniciar el projecte), a en Quintana i l'Aina (per compartir les ratafies de cada divendres), a en Ricard Prados (el Magyver del laboratori), a la Joseta (per ser la meva guardiana de les sabates de la sort), i a cadascun de vosaltres de la família VICOROB (em deixaria algú i em sabria greu!)

No puc oblidar-me tampoc dels meus directors de tesi de França, Alain Lalande i Paul Walker que tant bé m'han tractat i ajudat. Per la seva disposició i voluntat des del primer moment de poder realitzar doctorat de co-tutela amb el seu grup de recerca. Així com en Fabrice Meriaudeau, en David Fofi i tot el grup de recerca de LE2i de Le Creusot. També a tot el departament de radiologia i cardiologia de

l'hospital de Dijon amb qui he treballat per facilitar i recolectar les imatges amb les quals realitzem les nostres tasques d'investigació. Sense ells la validació del projecte no hagués estat possible. Finalment, a totes (que han sigut moltes!) les infermeres "C'est nickel" que m'han cuidat tant carinyosament aquests anys.

Puc considerar-me afortunat que després de deu anys (i espero que molts més), el Servei d'Esports és ja per a mi com la meva segona família. Treballar cada dia amb vosaltres és un plaer i m'heu aportat tot el necessari per sentir-me sempre estimat i recolzat tant en els bons, com en els mals moments. Mil gràcies.

A tots els meus amics d'universitat i màster, al carcamal d'en Sik (per ajudar-me i comptar sempre amb mi), Sebas, Sam i Josep Aulinas (sex on the beach team!), Jan Margeta (January), Lucca, Gildardo Lozano i Karla (guardo tu bob esponja mexicano!), Bruno, Marcos y Sergio (gracias por vuestro cariño y estima durante todo estos años), Hayyan (underdeep), Naji (gha-gha), Nadhir (abuyafer) and Imen (Thanks my dear friends, you are the best of the rest, mumtaz gha-gha habibi). Jhimly and Soumya (thanks for the great moments), and finally Sakura and Joyce (Arigatou gozaimasu, watashi wa shiawase desune to gambatte ne!).

I menció a part a en Fredi (a pasa boca?) que tantes coses he d'agrair-li i sap que avui no estaria escrivint aquests agraïments si no fos per ell.

En Ferran i la Tere (Els amics de veritat no existeixen distàncies), l'Oriol (bomber-man), l'Òscar, Marta i Roc (é yoyic) i els petitons de la casa, a l'Oliver i Sonia (perquè us estimo com la familia), PapaToni y MamaLou (mis papas de Ibiza), Laurinski (Stá er tré?), l'Alba (per ser tant especial), a Laura Andrés i en Jordi i tota la colla de TERRORISTES (nosinmianto!), Josep i Dolors, l'Ariadna, Laura, Uli (Mr. Octopus), Bea (be positive!), Irina (MySpecialOne), Jordi i Miriam Gelabert (família espàrrec!).

Dedicar els agraïments a les persones més importants de la meva vida. Donar les gràcies als meus pares per fer-me sentir a prop quan la distància s'ha fet llarga. L'amor, el suport, la paciència, la confiança, la sinceritat, l'amistat com tantes d'altres, fan que per a mi sigueu més que uns pares. Incomprensiblement són tantes les vegades que no diem el que estimem i que alhora són insuficients. US ESTIMO.

Al meu germà preferit. Ja sabem que l'Adrià és l'únic germà que tinc, però tot i que li dono canya, el faig emprenyar, i trenquem vidres sempre que ens barallem..., no podria tenir-ne un de millor. Saps que estic molt orgullós de tu, que malgrat no estigui d'acord amb moltes de les teves decisions confio amb tu i estic segur que aconseguiràs trobar el teu camí. Gràcies per estar quan t'he necessitat.

Donar les gràcies als meus avis. Malauradament, ens va deixar una de les persones que més he admirat, estimat i m'hagués agradat tant que un dia com el d'avui estigués aquí al meu costat. Però allà on estiguis espero que et sentis orgullós de mi.

He seguit els teus consells, vas confessar-me coses que sé que mai havies dit a ningú, i que passat el temps m'he adonat la raó que tenies. Que la vida posa tothom al lloc on es mereix. Avi Joan, vull que aquest sigui el teu millor reconeixement. Àvia Rosa, Àvia Mercè i Avi Miquel gràcies per estimar-me tant, així com la resta de la família.

Quan mires enrere i recordes les decpecions, els mals moments i les situacions límit que has hagut de superar durant tots aquests anys, t'adones que qui t'estima mai et falla. Al final t'adones que el millor encara havia d'arribar. Així que "la vida és un boomerang" i sense adonar-te el millor arriba quan menys ho esperes. Gràcies a TU per arèixer a la meva vida, ajudar-me, estimar-me i fer-me sentir especial.

Dedicar unes paraules a l'amic que m'ha vist néixer, créixer i fins a dia d'avui el considero el germà gran que no he tingut mai. Dani (Kameni), és un honor que formis part de la meva vida. Sergio (toloko) i tota la seva família (niña ven pa cá) poques paraules em queden per expressar perquè ets un dels meus millors amics. I al Dr. Beto, que desde el primer dia y pasado todos estos años me has demostrado que pase lo que pase, y estemos donde estemos la amistad es y será infinita. La vida la fa cadascú, i com a tal, formeu part de la meva família perquè sou com els meus germans.

Sé que he fet molt d'esforç, i com molts sabeu he hagut de lluitar amb mi mateix en moments molt difícils. Per a mi ha sigut més que una tesi doctoral, ha sigut la tesi de la meva vida pel que ha costat arribar fins aquí. Però això no s'acaba aquí, tot final te un nou principi i vull que en els nous reptes que vindran en sigueu formant part. Gràcies a tots, i als que no he citat i també hi heu estat.

Acknowledgments

List of publications

Here it can be found a list of scientific contributions already published during the course of this PhD along with a list of undergoing publications as outcome of this dissertation.

Published contributions

Publications in Journals:

- C. Mata, P. M. Walker, A. Oliver, F. Brunotte, J. Martí, and A. Lalande. ProstateAnalyzer: a web-based medical application for the management of MR imaging techniques and MR spectroscopy for prostate cancer diagnosis. Informatics For Health And Social Care, to be published. February 2015. DOI: `http://dx.doi.org/10.3109/17538157.2015.1008488`.

Under review in journals

- C. Mata, A. Lalande, P. M. Walker, A. Oliver and J. Martí. Web-based applications in medical imaging. Submitted to Journal in Biological Informatics, January 2015.

- C. Mata, A. Lalande, P. M. Walker, and A. Oliver and J. Martí. Interest of a collaborative work in the case of evaluation of the prostate cancer from MRI. Submitted to IRBM, May 2015.

Published contributions

- C. Mata, J. Martí, and X. Lladó. Web-application system using a Digital Mammographic Database: MamoDB. In MICCAT: The First Workshop in Medical Imaging. Girona, Spain. October 2010.

- C. Mata, X. Lladó, R. Martí, J. Freixenet, A. Oliver, and J. Martí. Aplicación web para la gestión de una base de datos pública en mamografía digital: MamoDB. In Congreso Anual de la Sociedad Española de Ingeniería Biomédica, pp 175-178. Madrid, Spain. November 2010.

- C. Mata, X. Lladó, and J. Martí. MamoDB as a new e-learning tool in Digital Mammography. In MICCAT: The Second Workshop in Medical Imaging. Barcelona, Spain. November 2011.

- J. Martí, C. Mata, X. Lladó, and A. Oliver. MamoDB: a web-based tool for training radiologists in the diagnosis of digital mammography. In International Conference on Education and New Learning Technologies, pp 2359-2367. Barcelona, Spain. July 2011.

- C. Mata, P. Walker, A. Lalande, F. Brunotte, A. Oliver, and J. Martí. Prostate-Analyzer: 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. Lisboa, Portugal. October 2012.

- C. Mata, A. Oliver, A. Torrent, and Martí, J. MammoApplet: an interactive Java applet tool for manual annotation in medical imaging. In International Conference on Bioinformatics & Bioengineering, pp 34-39. Larnaca, Cyprus. November 2012.

- C. Mata, A. Oliver, and Martí, J. A web-based tool for training radiologists in medical imaging. In Workshop de Reconocimiento de Formas y Anáisis de Imágenes, pp 5-8. Madrid, Spain. September 2013.

List of Tables

List of Figures

List of Acronyms

ACR American College of Radiology

ACS The American Cancer Society

ADC Apparent Diffusion Coefficient

AOV Area Overlap

AFT Anterior Fibromuscular Tissue

BI-RADS Breast Imaging-Reporting and Data System

CAD Computer Aided Diagnosis

CADe Computer-Aided Detection

CBIR Content-Based Image Retrieval

CC Cranio-Caudal

CE Zend Server Community Edition

CLV Cleavage view

CNRS Centre National de la Recherche Scientifique

CT-scan Computed Tomography

CZ Central Zone

DCE Perfusion-weighted imaging

Dcm4chee Dcm4chee project

DICOM Digital Imaging and Communications in Medicine

DIMSE DICOM Service Element

DM Digital Mammography

DSC Dice Similarity Coefficient

DWI Diffusion-Weighted Imaging

EHR Electronic Healthcare Record

EMR Electronic Medical Record

EPR Electronic Patient Record

ERC Endorectal Coil

ESUR European Society of Urogenital Radiology

eXist-db eXist-db XML Server

FFDM Full-Field Digital Mammography

FN False Negative

FP False Positive

HD Hausdorff Distance

HIS Hospital Information System

HTTPS Secure Web Site

ISO International Organization for Standardization

HIPAA Health Insurance Portability and Accountability Act

IIOP Internet Inter-ORB Protocol

IOD Information Object Definitions

IUID Instance Unique Identifier

Java Java

JDK Java Development Kit

JBoss JBoss Application Server

JRE Java Runtime Environment

JSC Jaccard Similarity Coefficient

JSP Java Server Pages

JVM Java Virtual Machine

LE2I Laboratoire Electronique, Informatique et Image

LM Latero-Medial

MML Medical Markup Language

MamoDB Web-based application based on a Digital Mammography

MamoApplet Mamo Java applet tool

MD Minimum Distance

MGF Magnification

ML Medio-Lateral

MLO Medio-Lateral Oblique

MRI Magnetic Resonance Imaging

MRS Magnetic Resonance Spectroscopy

MySQL MySQL Community Server

NEMA National Electrical Manufacturers Association

NMR Nuclear Magnetic Resonance

PACS Picture Archiving and Communication System

PET Position Emission Tomography

PHP Hypertext Pre-Processor

phpMyAdmin phpMyAdmin "Bringing MySQL to the web"

PI-RADS Prostate Imaging-Reporting and Data System

PixelMed PixelMed Java DICOM Toolkit

PPA Pelvic Phased Array

PCa Prostate Cancer

PSA Prostate Specific Antigen

PZ Peripheral Zone

RIS Radiology Information System

ROI Region Of Interest

SC Spot Compression

SFM Screen-Film Mammography

SI Similarity Index

SOP Service-Object Pair

SQL SQL Language

SD Standard Deviation

SSL Secure Socket Layer

T2WI T2-Weighted Imaging

TE Echo Time

TN True Negative

TP True Positive

TR Repetition Time

TV Tumor Volume

TZ Transition Zone

T2WI 3D T2-Weighted Imaging

UID Unique Identifiers

UML Unified Modeling Language

US Echography with Ultra Sound

XML eXtensible Markup Language

XPath XML Path Language

XQuery XML Query Language

VICOROB Computer Vision and Robotics group

VPN Virtual Private Network

WADO Web Access to DICOM Persistent Objects

ZEND Zend Server

Contents

Abstract

The rapid growth of digital technology in medical fields during the last few years has increased the need for applications able to manage patient medical records, imaging tests, and chart information [127]. Technologies related with digital database architectures, storage and transmission protocols, management of large volumes of data, and security are being improved every day, allowing the possibility of reading, analyzing, and even diagnosing a case far from the medical center where the information was acquired.

Furthermore, the inclusion of web-based applications applied to the medical imaging field has growing demand over the last few years in the medical community [101, 199]. The study of how applications can store and exchange this huge amount of information is becoming a new paradigm to solve present day limitations to managing medical data. Indexing tools based on the standard protocols allows a fast retrieval of the information and securities issues. Web-based applications in computational medicine have become increasingly important over the last few years [128, 108]. It supposed a new paradigm in the telemedicine and e-Health areas in order to assist and enhance the prevention, diagnosis and treatment of ailments [201]. Furthermore, the training of radiologists and the management of medical databases are also becoming increasingly important issues in the field.

In this domain, firstly, this thesis analyzes the main advantages and drawbacks of the reviewed work in web-based applications applied to the medical imaging field. Then, a web-based medical application including a system architecture based on digital imaging is presented. One of the main purposes of this PhD thesis is to offer an easy accessibility to web-based applications for users using a new system architecture. They can connect from any part of the world, providing this to be a time-saving and efficient mechanism to find information. In fact, this information can be saved directly in computers without the tiring task of searching in libraries or alternative databases. However, the access of the previous process is also available with the possibility of downloading medical annotations using query forms. Then collaborative work is possible by considering several opinions from different experts. Our system, being dedicated to medical imaging, manages the corresponding specific

Digital Imaging and Communications in Medicine (DICOM) format. Finally, our web-based application also provides a high level security system.

Prostate and breast cancer are the most common cause of cancers in men and women, respectively. Medical imaging plays an important role in breast and prostate cancer detection and evaluation. Then to prove that our web-based medical application could be applied in different medical disciplines, the main part of this thesis is the implementation of two frameworks as a Java applet interface designed as a web-based tool in the domains of mammography from X-rays in radiology, and of prostate imaging from an Magnetic Resonance Imaging (MRI). This aims to facilitate the diagnosis of new mammographic and prostate 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 (overlays). 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 eXtensible Markup Language (XML) files associated with the original images. Finally, an exhaustive evaluation of the results is also discussed in this thesis. For the application on mammography, the experimental study is performed in order to evaluate the scalability, complexity and response speed at the proposed tool. For the application on MRI of prostate cancer, the evaluation focused on the decrease of the variability of the expert assessments when collaborative work is performed. To conclude, a new architecture with the main goal of managing patient databases with potentially multi-modal imaging is presented such as for an MRI of the prostate cancer and evaluation from potentially several experts.

Resum

El ràpid creixement de la tecnologia digital en el camp mèdic ha generat la necessitat de desenvolupar aplicacions que siguin capaces de gestionar grans volums de dades [127]. Conseqüentment, les bases de dades digitals, els protocols d'emmagatzement, la transmissió i seguretat es van millorant cada dia possibilitant la lectura, anàlisi i gestió de diagnòstics clínics tant en el mateix centre mèdic o externament.

La inclusió d'aplicacions web aplicades a la pràctica mèdica s'ha incrementat notablement mitjançant els avenços tecnològics dedicats a l'anàlisi i processat d'imatges digitals [128, 108]. En aquest sentit, el camp de la imatge mèdica te un rol molt important en aquesta última dècada gràcies a eines basades en protocols que compleixen un format estàndard i que permeten una ràpida recuperació de la informació. Malgrat això, un dels majors inconvenients és l'intercanvi i emmagatzematge de gran quantitat d'informació provinent de centres mèdics [201]. Ha suposat sens dubte un nou paradigma en el camp de la telemedicina en trobar aplicacions que puguin ajudar i millorar la prevenció, diagnòstic i tractament de pacients. Així com la formació de radiòlegs i l'aprenentatge per la gestió de base de dades mèdiques cada vegada és més necessària en el camp mèdic.

Hi ha molts avantatges en l'ús d'aplicacions web en l'àmbit mèdic [199]. La principal avantatge es deu a la fàcil accessibilitat per als usuaris que poden connectar-se des de qualsevol lloc del món, l'estalvi de temps i l'eficient mecanisme per la gestió de la informació. A més, proporcionen eines que faciliten la descàrrega d'anotacions mediques així com alts nivells de sistema de seguretat [101]. En el camp mèdic és imprescindible tenir sistemes de seguretat i protecció de la informació dels pacients. Per tant, les tecnologies d'imatges digitals s'han convertit beneficioses per les practiques mèdiques modernes i sistemes d'atenció de salut proporcionant ajuda per al diagnòstic, el tractament i la cirurgia.

Durant els últims anys, la taxa d'incidència en la detecció de càncer s'ha reduit lleugerament en els homes i s'ha mantingut estable en les dones, mentre que les taxes de mortalitat per càncer es va reduir en un 1,8% anual en els homes i un 1,4% per any en les dones [171, 164]. Aquesta reducció es deu en gran part a l'anàlisi i processament d'imatges mèdiques i tot aquest conjunt d'aplicacions d'ajuda al

diagnòstic del càncer. D'una forma molt notòria contribueix a la seva detecció precoç mitjançant cribratge, diagnòstic, biòpsia guiada per imatges i arquitectures per la gestió de dades mèdiques tals com base de dades d'imatges digitals [100]. En realitat, els càncers de pròstata i mama són la causa més comú de mort per càncer comparat amb els altres. Per aquest motiu, el principal objectiu és millorar les eines i aplicacions actuals per proporcionar una millor instrumentació per obtenir una millor qualitat d'imatges, així com, metodologies i procediments amb la finalitat de millorar la interpretació de les lectures de la imatge.

En aquesta tesi es realitza una revisió bibliogràfica de les principals publicacions recents en els últims anys en aplicacions mèdiques basades en web. Aquest estudi analitza els avantatges i inconvenients dels treballs d'investigació en el camp de la imatge mèdica, així com les arquitectures de base de dades per a la gestió d'imatges digitals. La part principal d'aquesta tesi és la implementació d'una eina basada en la web amb la finalitat de demostrar la integritat i aplicació en diferents disciplines mediques. En aquest sentit, l'aplicació proposada en aquest projecte de tesis ha sigut implementada com a eina d'ajuda al diagnòstic de càncer de mama i pròstata. L'objectiu és facilitar el diagnòstic proporcionant un conjunt d'eines de processat d'imatge que permetin una millor visualització de les imatges, i un conjunt d'eines d'anotació de regions sospitoses o malignes (superposicions). Cada anotació permet incloure tots els atributs i especificacions considerades pels experts a l'emetre el diagnòstic final. S'han dissenyat diferents arquitectures per a la gestió de base de dades (per exemple PACS per emmagatzemar imatges monogràfiques). Per altra banda, el conjunt global d'anotacions s'emmagatzemen en una base de dades d'arxius XML associats a les imatges originals. Conseqüentment, aquesta nova arquitectura es presenta amb l'objectiu d'obtenir una base de dades de casos diagnosticats i validats per radiòlegs experts per a la formació de radiòlegs novells. Finalment, conclusions i noves línies d'investigació associades al projecte com a treball futur són presentades en aquesta tesi.

Resumen

El rápido crecimiento de la tecnología digital en el campo médico ha generado la necesidad de desarrollar aplicaciones que sean capaces de gestionar grandes volúmenes de datos [127]. Por consiguiente, las bases de datos digitales, los protocolos de almacenamiento, la transmisión y seguridad se están mejorando cada día permitiendo la posibilidad de leer, analizar y gestionar diagnósticos tanto en el mismo centro médico o remotamente.

La inclusión de aplicaciones web aplicadas a la práctica médica se han incrementado notablemente mediante los avances tecnológicos dedicados al análisis y procesamiento de imágenes digitales [128, 108]. En este sentido, el campo de la imagen médica tiene un rol muy importante en esta última década gracias a herramientas basadas en protocolos estándar que permiten una rápida recuperación de la información. Sin embargo, uno de los mayores problemas es el intercambio y almacenamiento de la gran cantidad de información que se genera diariamente en un hospital o centro médico [201]. Consecuentemente, ha supuesto un nuevo paradigma para el campo de la telemedicina encontrar aplicaciones que puedan ayudar y mejorar la prevención, diagnóstico y tratamiento de pacientes. Así como la formación de radiólogos y el aprendizaje para la gestión de base de datos médicas donde cada vez se necesitan nuevas herramientas para el campo médico.

Hay muchas ventajas del uso de aplicaciones web en el ámbito médico [199]. La principal ventaja se debe a la fácil accesibilidad para los usuarios ya que pueden conectarse desde cualquier lugar del mundo, el ahorro de tiempo y el eficiente mecanismo para la gestión de la información. Además, proporcionan herramientas que facilitan la descarga de anotaciones médicas así como altos niveles de sistema de seguridad [101]. En el campo médico es imprescindible tener sistemas de seguridad y protección de la información de los pacientes. Por tanto las tecnologías de imágenes digitales se han convertido beneficioso las prácticas médicas modernas y los sistemas de atención de salud, proporcionando herramientas de gran alcance para el diagnóstico, el tratamiento y la cirugía.

Durante los últimos años, la tasa de incidencia en la detección de cáncer se ha reducido ligeramente en los hombres y se mantuvieron estables en las mujeres,

mientras que las tasas de mortalidad por cáncer se redujeron en un 1,8% anual en los hombres y un 1,4% por año en las mujeres [171, 164]. Esta reducción se debe en gran parte al análisis y procesamiento de imágenes médicas y todo este conjunto de aplicaciones de ayuda al diagnóstico del cáncer. De una forma notoria contribuye a su detección precoz mediante cribado, diagnóstico, biopsia guiada por imágenes, aplicaciones de imágenes médicas para el análisis y arquitecturas para la gestión de datos médicos tales como bases de datos de imágenes digitales [100]. En realidad, el cáncer de próstata y de mama son la causa más común de muerte por cáncer. Por este motivo, el principal objetivo es mejorar las herramientas y aplicaciones actuales para proporcionar una mejor instrumentación para obtener una mejor calidad de imagen, así como, metodologías y procedimientos con el fin de mejorar la interpretación de las lecturas de imagen.

En esta tesis se realiza una revisión bibliográfica de las principales publicaciones recientes en los últimos años en aplicaciones médicas basadas en web. Este estudio analiza las ventajas e inconvenientes de los trabajos de investigación en el campo de la imagen médica. Así como las arquitecturas de bases de datos para la gestión de imágenes digitales. La parte principal de esta tesis es la implementación de una herramienta basada en la web con la finalidad de demostrar la integridad y aplicación en diferentes disciplinas médicas. En este sentido, la aplicación propuesta en este proyecto de tesis ha sido implementada como herramienta de ayuda al diagnóstico de Cáncer de Mama y Próstata. El objetivo es facilitar el diagnóstico proporcionando un conjunto de herramientas de procesamiento de imágenes que permiten una mejor visualización de las imágenes, y tambíen de herramientas de anotación de regiones sospechosas o malignas.

Cada anotación permite incluir todos los atributos y especificaciones consideradas por los expertos al emitir el diagnóstico final. Se ha diseñado diferentes arquitecturas para la gestión de base de datos para imágenes médicas (por ejemplo PACS para almacenar imágenes mamográficas). Por otra parte, el conjunto global de superposiciones se almacenan en una base de datos como archivos XML asociados a las imágenes originales. En consecuencia, esta nueva arquitectura se presenta con el objetivo de obtener una base de datos de casos diagnosticados y validados por radiólogos y expertos para la formación y el desempeño de radiólogos novatos. Finalmente, las conclusiones y nuevas líneas de investigación para trabajo futuro son presentadas en esta tesis.

Résumé

L'utilisation de la technologie numérique dans le domaine médical a conduit à la nécessité de développer de nouvelles applications pour la gestion d'un grand nombre de données [127]. Dans ce cadre, les bases de données numériques, les protocoles de stockage, et la transmission et la sécurité dans les canaux de communication sŠaméliorent chaque jour pour aider les experts dans la lecture, l'analyse et la gestion des examens médicaux, sur site ou à distance. L'intégration des applications web appliquées à la pratique médicale s'est accrue notablement grâce aux avancées technologiques dédiées à l'analyse et au traitement des images numérisées [128, 108]. Cependant, l'un des problèmes majeurs dans le contexte actuel est l'échange et le stockage des grandes quantités d'informations qui sont régulièrement générées dans les hôpitaux ou dans les centres médicaux [201]. Ces problématiques demandent donc un changement de paradigme dans le domaine de la télémédecine ainsi que de nouvelles applications pour faciliter la prévention, les diagnostics et le traitement des patients. De plus, ces nouvelles approches permettront la formation des médecins ainsi qu'un changement important dans les méthodes de gestion des bases de données médicales.

Il y a plusieurs avantages dans lŠutilisation des applications web dans ce contexte [199]. La plus évidente est l'accès simplifié aux données par des utilisateurs, qui peuvent y accéder depuis n'importe quel ordinateur dans le monde, ainsi qu'un gain de temps considérable, sans compter le gain en efficacité dans la gestion de l'information. De plus, ces outils fournissent des moyens pour le téléchargement des annotations médicales sur les examens, tout en assurant les plus hauts niveaux de sécurité [101]. Dans tous les domaines, mais surtout dans les applications médicales, la sécurité est d'une importante capitale, notamment pour la protection des données des patients. Plusieurs méthodes pour garder la confidentialité des données existent et peuvent être mises en œuvre.

Dans le cadre de cette thèse, premièrement, nous avons réalisé une étude bibliographique exhaustive des principales publications récentes dans le contexte des applications web dans le domaine médicale. Cette étude analyse les avantages et les inconvénients des travaux préexistants dans le champ de l'imagerie médicale.

Ensuite, la contribution principale de notre travail a consisté dans l'implémentation d'un outil web avec comme finalité l'utilisation de ces outils dans différentes disciplines médicales. L'objectif principal est de faciliter le diagnostic en fournissant un ensemble d'outils pour faciliter le traitement des images, pour améliorer leur visualisation ainsi que des moyens pour la mise en œuvre d'annotation des régions suspects. Les utilisateurs pourront se connecter et avoir accès à l'information depuis n'importe quel endroit dans le monde. Chaque annotation permet l'inclusion des spécificités considérées comme essentielles par les experts pour l'édition d'un diagnostic final. L'ensemble des annotations sont stockées dans une base de données comme des fichiers eXtensible Markup Language (XML) associés aux images originales. De plus, l'accès aux traitements précédents est disponible avec la possibilité de charger des annotations utilisant des outils de recherche. Ainsi, des travaux collaboratifs sont possibles en considérant plusieurs avis de plusieurs experts, potentiellement de sites différents. Notre système est dédié à l'imagerie médicale, et donc gère le format d'image correspondant (format Digital Imaging and Communications in Medicine (DICOM)). Enfin, notre système intègre un haut niveau de sécurité et de protection de l'information.

Les cancers de la prostate et du sein sont parmi les plus communs chez l'adulte. L'imagerie médicale joue un rôle important dans la détection et l'évaluation de ces cancers. De ce fait, nous nous sommes focalisés dans le contexte de la détection et diagnostique des cancers du sein et de la prostate, que nous présentons comme cas d'étude. Une importante partie de notre travail est donc l'implémentation de deux applications développées à partir de notre outil, pour la gestion de mammographie et d'IRM dans le cadre du cancer de la prostate. Le but est de faciliter le diagnostic de nouveaux cas en fournissant une série d'outils permettant dŠannoter les zones pathologiques ou suspectes. L'ensemble des annotations est enregistré dans une base de données utilisant le langage XML, chaque fichier étant lié à l'image initiale. Enfin, une évaluation des logiciels développés est effectuée. Pour l'application sur les mammographies, l'étude expérimentale est effectuée pour évaluer l'évolutivité, la complexité et la vitesse de réponse de l'outil proposé. Pour l'application sur l'IRM du cancer de la prostate, l'étude est focalisée sur la diminution de la variabilité du diagnostic par les experts quand un travail collaboratif est effectué.

En conclusion, une nouvelle architecture est présenté avec comme but principal de gérer des bases de données patients, avec potentiellement une imagerie multi-modale (comme pour l'IRM du cancer de la prostate) et d'intégrer une évaluation potentiellement à partir de plusieurs avis.

Chapter 1

Introduction

Breast cancer and prostate cancer are two of the major public health issues among all cancers in men and women in most developed countries. An introduction of the medical imaging concepts focusing on these areas is presented in this chapter. The context, motivations, main objectives and aims of this thesis are also explained. Finally, a summary and organization of the rest of this document is detailed.

The key to growth is the introduction of higher dimensions of consciousness into our awareness.

Lao Tzu

1.1 Medical imaging management

Over the past decade, the emergence of large digital volumes of information in medical imaging and the rapid growth of computer-based clinical examinations have increased the number of available on-line web-based medical applications. Moreover, digital imaging technologies have become essential in modern medical practices and health care systems, providing powerful tools for diagnosis, treatment, and surgery. Health care consumers have begun to benefit from new web-based applications as tools to guide decision making on treatments and tests [166], or even to enhance the capabilities and physical activity in various chronic disease settings [100].

The increase in medical data supposes a new paradigm to develop new system architectures and digital databases to store patient information along with medical images. In this sense, the Digital Imaging and Communications in Medicine (DICOM) protocol has become a standard for the storing and exchange of images and related

information [18], although some problems prevail with respect to standardization [79, 129]. The main characteristic of the DICOM format is that it is focused on patient information, which is stored jointly with the actual image(s) [94].

The importance of developing database models and architecture systems in order to store medical images has increased over the last decade. The connection and secure access with the Picture Archiving and Communication System (PACS) server using the DICOM standard are common in many hospitals and health care systems. From an informatics perspective, it is evident that the fields of PACS and imaging informatics are synergetic; developments in both PACS and in image processing had an undeniable impact on imaging informatics. Several reviews focused their efforts on the issues related to the management of image information, including image study summarization, standardization, and navigation in image databases [173].

In fact, images are produced in ever-increasing quantities, and they allow diagnosis and follow-up of patients with their examinations. The importance of developing database model, and architecture centric systems in order to store medical images has been widespread during the last decade. Evidently, PACS supports clinical decision-making that would ease the management of clinical data [127].

On the other hand, the interest in Content-Based Image Retrieval (CBIR) has grown due to the limitations inherent in metadata-based systems, as well as the large range of possible uses for efficient image retrieval. The large amount of both visual and data information along with the new on-line tools underlined the need to create thematic accesses that offered more than simple text-based queries or requests based on matching exact database fields [127]. CBIR technology has been proposed to aid clinical care, bio-medical research and education [131, 194]. However, despite the initial widespread enthusiasm for CBIR in the engineering research community, its application has still to solve practical problems, including optimization stages in the indexing and retrieval work-flow [110].

Besides, systems based on web applications [128, 108, 201] are particularly interesting due to recent advances in computer technology and communications systems and protocols, which have allowed the development of online software and data recovery protocols in a reliable and inexpensive way [93]. The main advantages of web-applications are obvious: ease of accessibility, users can connect from anywhere in the world, time-saving and efficient mechanism to find medical data [199]. Moreover, they provide a high level security system, easy of downloading medical annotations using query forms and, they can provide a high level of interactivity among health professionals.

One of the long standing problems is finding ways to share medical data across a variety of media [49]. In this sense, eXtensible Markup Language (XML) has emerged as a leading enabler. Some applications have been developed on the inte-

gration of frameworks and data management using queries to examine information from multiple data types and large data sets [101]. Although the XML is provided with predefined tags, one of its advantages is its extensible use. Finally, the inclusion of the XML databases facilitates the management of XML files by storing them in an efficient way.

Finally, security is an important requirement in the medical informatics field. In order to manage medical data, a security system is essential for local and remote access, and protection of patient data (by encryption of the data). Issues related to security should also be addressed.

1.1.1 Data acquisition

The medical imagine is a discipline is qualified as one of the most important medical developments of the past thousand years since medical imaging provides physicians with in vivo images describing anatomy, physiology and functionality of organs [8].

One of the important points of medical imaging is the image standardization techniques. According to Sinha et al. [173], we refer to image standardization techniques as those that enable quantitative comparisons of image data sets acquired across patient populations. This includes the creation of standard data sets and methods for mapping image data sets similar to standard terminologies for textual data as DICOM format (see subsection 1.1.2).

Figure 1.1 summarizes the three main blocks of a typical medical imaging system. The first block, data acquisition techniques, is summarized in this section. The pre-processing module is based on the treatment of medical images such as segmentation, classification, and pre-processing tools after the data acquisition. This part is relatively beyond the scope of our work. Concretely, only a part of the processing and analysis focused on the extraction of relevant parameters (e.g. web-based medical applications) are described in Chapters 4 and 5. Finally, the last block will be covered in the main chapters of this thesis.

The past two decades have witnessed an explosion of digital imaging modalities: Scintigraphy, X-ray radiology (conventional radiology), Magnetic Resonance Imaging (MRI), Computed Tomography (CT-scan), Position Emission Tomography (PET), Echography with Ultra Sound (US), to name the principle ones [198]. Next, sub-sections describe just a few of these, in particular, X-ray radiology and MRI since they are the ones involved in this thesis.
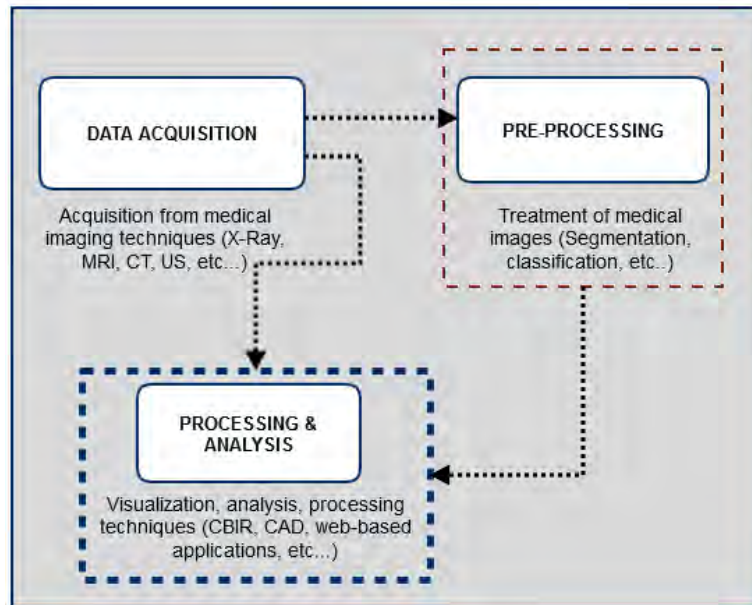
Figure 1.1: The standard model used in medical imaging.

X-Ray radiology

Imaging with x-rays involves exposing a part of the body to a small dose of ionizing radiation to produce pictures of the inside of the body. X-rays are the oldest and most frequently used form of medical imaging[1]. X-rays are a type of electromagnetic radiation with in extremely short wavelength and then a high frequency (wavelengths ranging from about 10^{-8} to 10^{-12} metre and corresponding to frequencies from about 10^{16} to 10^{20} hertz (Hz)[2]). X-rays are commonly produced by the interaction of particles (e^-) with high kinetic energy and devices (with high atomic mass numbers); examples include a beam of electrons striking a metal plate in an X-ray tube and a circulating beam of electrons in a synchrotron particle accelerator or storage ring. In addition, highly excited atoms can emit X-rays with discrete wavelengths characteristic of the energy of electronic levels (at the level of electrons).

An X-ray tube sends an X-ray photon beam toward the body. The images are recorded on a computer or film. Attenuation of the X-ray beam depends on the density of the crossed structures (and their size) and also on the energy of the X-rays [58]. X-rays can provide information about obstructions, tumors, and other diseases [46]. An example of an X-ray acquisition machine is depicted in Figure 1.2.

---

[1]RadiologyInfo. URL: http://www.radiologyinfo.org

[2]Wilhem Conrad Röntgen and the discovery of X-ray beams, 2014. URL: https://explorable.com/wilhem-conrad-rontgen
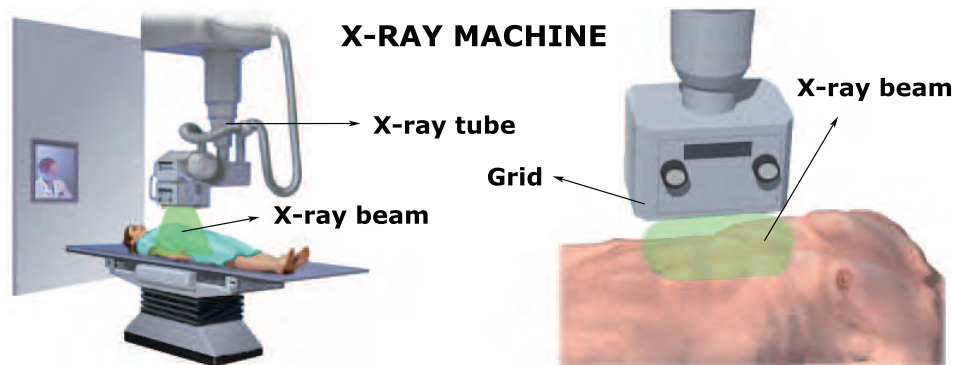
Figure 1.2: Example of an X-ray machine acquisition (image adapted from [46]).

Mammography is a particular X-ray radiology used as a screening tool to detect early breast cancer in women experiencing no symptoms. It can also be used to detect and diagnose breast disease in women experiencing symptoms such as a lump, pain or nipple discharge.

A mammography unit is a rectangular box that houses the tube in which x-rays are produced. An example of an X-ray acquisition applied to mammography is depicted in Figure 1.3. The unit is used exclusively for x-ray exams of the breast, with special accessories that allow only the breast to be exposed to the x-rays. Attached to the unit is a device that holds and compresses the breast and positions can be obtained images at different angles to increase the contrast before soft tissues. Also, its low energy allows using a diminished radiation.

Magnetic Resonance Imaging (MRI)

MRI is a medical imaging technique used to investigate the anatomy and physiology of the body in both health and disease. MRI uses a powerful magnetic field and radio frequency pulses to produce detailed pictures of organs, soft tissues, bone and virtually all other internal body structures. In difference with X-rays, MRI does not use ionizing radiation[3]. Frequently, the differentiation of abnormal (diseased) tissue from normal tissues is better with MRI than with other imaging modalities such as X-Ray, CT-scan and ultrasound [46], in particular for soft tissues.

On the other hand, it is important in MRI to consider "weighted" images, in particular, the terms T1-Weighted Imaging (T1WI) and T2-Weighted Imaging (3D T2-Weighted Imaging (T2WI), that could be in 3D) are linked with two specific relaxation parameters [122].

---

[3]MRI: A Guided Tour Web site.
URL: http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri
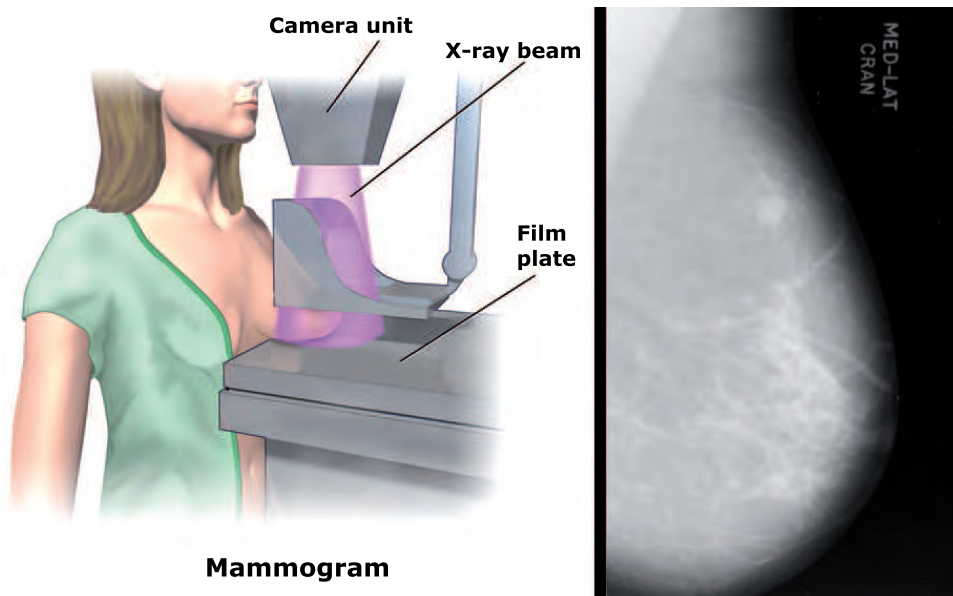
Figure 1.3: Example of an X-ray machine acquisition in mammography (left image adapted from [175]).

In general terms, T1 and T2 are among the most overused concepts in MRI [45]. In the broadest sense, these terms are used to communicate to other physicians the type of MR pulse sequence employed to generate a series of images. In a narrower sense, an implication exists that a single intrinsic tissue parameter (T1, T2, spin-density, diffusion, susceptibility, chemical shift, flow, perfusion, etc.) dominates the image contrast observed.

Some features related to MRI concepts are focused on the pulse sequence. An MRI pulse sequence is a programmed set of changing magnetic gradients. Different pulse sequences allow the user to image the same tissue in various ways, and combinations of sequences reveal important diagnostic information about the tissue in question. The two most important parameters are the Echo Time (TE) and the Repetition Time (TR). The TE refers to time between the application of radiofrequency excitation pulse and the peak of the signal induced in the coil. It is measured in milliseconds. The amount of T2 relaxation is controlled by TE. On the other hand, the TR is the time from the application of an excitation pulse to the application of the next pulse. It determines how much longitudinal magnetization recovers between each pulse. It is also measured in milliseconds.

In general, T1-weighted sequences provide the best contrast for paramagnetic contrast agents. T1 weighting tends to have short Echo Time (TE) and short Repetition Time (TR) times. The TR and the TE are used to control image contrast and the "weighting" of the MR image. T1 weighting tends to have short TE and TR
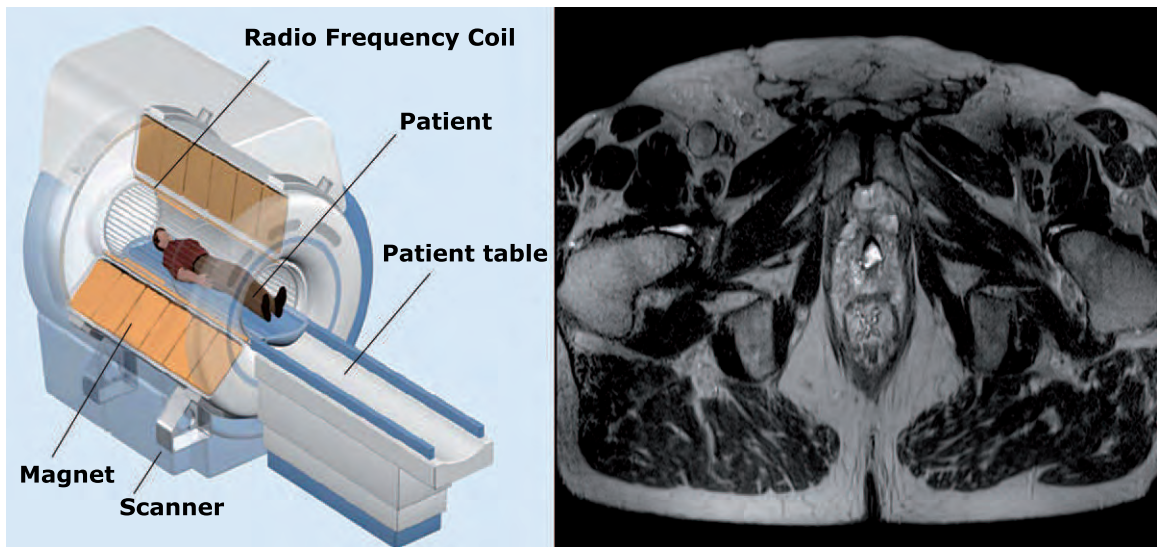
Figure 1.4: Example of an MRI scanner cutaway for a prostate acquisition in an axial orientation (left image adapted from Magnet[3]).

times. In general, T1-weighted sequences provide the best contrast for paramagnetic contrast agents. On the other hand, T2WI tends to make images with a sequence with long TR and TE to show contrast in tissues with varying T2 relaxation times. For example, water gives a strong signal.

Also, another way to understand the effect of TE on T2-weighting is to consider the signals generated by two tissues with different T2 values [143]. When TE is short, the echo occurs when there has been too little time for T2-decay to have taken place and hence the tissues are not differentiated. If TE is long, the relative differences in signal decay between the two tissues become more noticeable, and hence more "T2-weighting". On the other hand, T1 is directly linked to short TR. The echo occurs when there has been too little time for T1-recovery to have taken place and hence the tissues could be differentiated according to this relaxation time. If TR is long, the relative differences in signal recovery between the two tissues become no noticeable because all tissues have recover their signal.

MRI is the most commonly used technique to image the prostate gland. The primary indication for MRI of the prostate is the evaluation of prostate cancer. The test is commonly used to evaluate the extent of prostate cancer in order to determine if the cancer is confined to the prostate, or if it has spread outside of the prostate gland. Figure 1.4 shows an example of a T2WI MR image of the prostate gland. In fact, MRI has been established as the best imaging modality for detection and localization on account of its high resolution and excellent spontaneous contrast of soft tissues and the possibilities of multiplanar and multiparameter scanning [29]. As

such, four MRI techniques (3D T2-Weighted Imaging (T2WI), Diffusion-Weighted Imaging (DWI), Perfusion-weighted imaging (DCE) and 3D Magnetic Resonance Spectroscopy (MRS)) are used in this thesis.

1.1.2   DICOM standard

A Digital Imaging and Communications in Medicine (DICOM) standard specifies a non-proprietary data interchange protocol, a digital image format, and a file structure for biomedical images and image-related information [18]. It has become the standard format in the medical image domain. DICOM enables the transfer of medical images in a multi-vendor environment and facilitates the development and expansion of picture archiving and communication systems [11] [86].

DICOM was developed in the 80's to make medical images and associated data standardized for easier interchange. Besides that, DICOM defines network oriented services for transfer or printing of the images, media formats for data exchange, work-flow management, consistency and quality of presentation and requirements of conformance of devices and program [129].

The main advantages of the DICOM format are the following:

- For network communications, a set of protocols to be followed by devices claiming conformance to the standard.

- The syntax and semantics of commands and associated information which can be exchanged using these protocols.

- For media communication, a set of media storage services to be followed by devices claiming conformance to the standard, as well as a file format and a medical directory structure to facilitate access to the images and related information stored on interchangeable media.

- Transmission and persistence of complete objects such as images, waveforms and documents.

- Access to administrative and technical information thanks to the header (TAG).

- Query and retrieval of objects with information content.

- Offers quality and consistency of image appearance with the possibility of being displayed or printed.
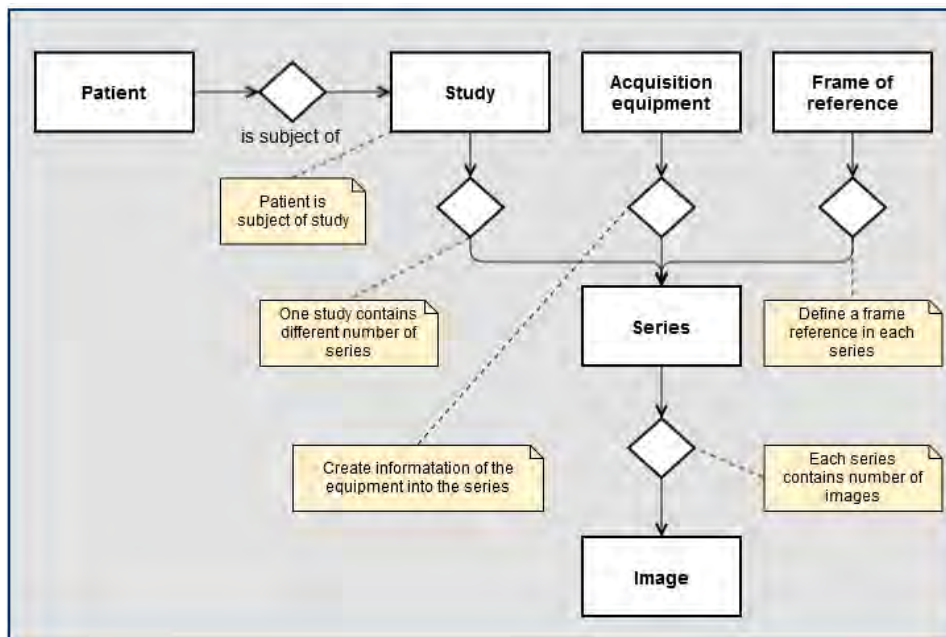
Figure 1.5: The DICOM composite instance IOD information model (adaptation from [11]).

Definition and data structure

Information Object Definitions (IOD)s are introduced in the standard to define attributes that describe a certain characteristic of the image. These attributes vary when comparing devices from different manufacturers and different modalities (X-Ray, MRI, CT-scan, US, etc.). There are also private attributes that can be used by equipment vendors to save proprietary data that can not be used by other manufacturer's workstations [120].

Figure 1.5 shows the DICOM composite instance of IOD information according to the model presented by Güld et al. [62]. In fact, IOD represents parts of several entities included in the DICOM model. Normally, each patient has a number of medical studies. Each study is composed of different series containing a set of medical images. It provides a common framework to ensure consistency between the various IOD defined by the DICOM standard [11].

According to the National Electrical Manufacturers Association (NEMA) [86], a DICOM structure is represented in Figure 1.6. This structure of a DICOM image includes the following elements of specification as attributes:

- DICOM media storage Service-Object Pair (SOP) classes and associated IODs.
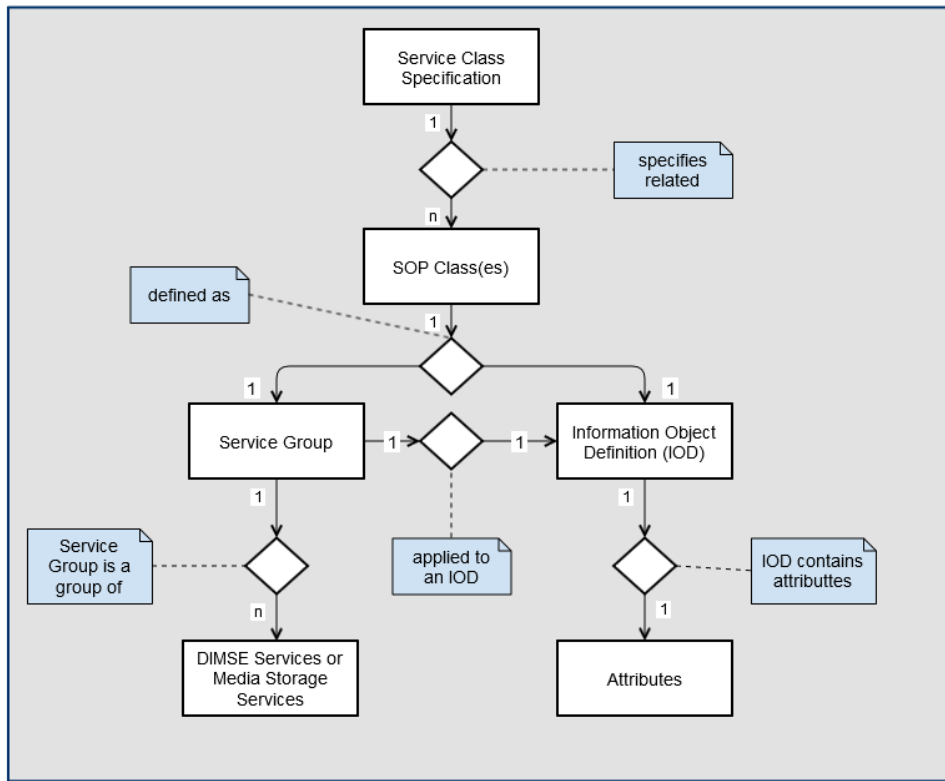
Figure 1.6: The structure of a DICOM information model (adaptation from [11]).

- The DICOM file format.

- The secure DICOM file format.

- The DICOM media storage directory SOP class.

- DICOM media storage application profiles.

- DICOM security profiles for media storage.

All these attributes define SOP classes as the union of an IOD and a DICOM Service Element (DIMSE). Examples of DIMSE are the most common functions for DICOM protocol (such as store, get, find, move, etc.). Moreover, the SOP class definition contains the rules and semantics which may restrict the use of the services in the DIMSE Service Group or the Attributes of the IOD.

Figure 1.7: Example of a DICOM header (adaptation from Dcm4cheeGuide[4])

DICOM header format

A DICOM file contains a file header portion, a file meta information portion, and a single SOP instance. The header is made up of a 128 byte preamble, followed by the characters DICOM, all uppercase[4]. The preamble must contain all zeroes if it is not used (sometimes applications will use it for proprietary data). Following the header is the file meta information. This portion follows a tagged file format, and contains information about the file, the series and the study it belongs to, and the patient that it belongs to.

Figure 1.7 shows an example of a DICOM header and its associated "TAG". In this example some of the common representative tags classified according to the data model depicted in the previous section is shown. DICOM makes extensive use of Unique Identifiers (UID)s as depicted in the first column of the example which contains a TAG ID. Almost every entity in the DICOM Data Model has a UID. In DICOM, every SOP Class has its UID which represents the identification of a unique ID number in a 2x4 hexadecimal digits. Finally, the ID number is linked to

---

[4]Dcm4chee Open Source Image Manager.
URL: http://www.dcm4che.org/docs/reference/dcm4chee/html/index.html

the TAG name of the DICOM header as depicted in Figure 1.7.

A typical example of a DICOM image object has patient, study, series, and image modules is a UIDs generated for each Series Instance UID and also for the SOP Instance UID which that is the Image UID. The Study Instance UID is provided by the acquisition system through a DICOM service called Modality Worklist but devices can default when creating the Study Instance UID. The Series Instance UID and SOP Instance UID are always generated by the Imaging device. The definition of a DICOM series is a set of DICOM Instances (images) generated together by the same equipment in the same operation.

DICOM query levels

A DICOM is a standard format in medical image communication developed with the goal of storing and easily retrieving 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. DICOM's attributes are divided into different level groups; PatientID, StudyInstanceUID, SeriesInstanceUID and InstanceUID. According to Pianykh [141] each level is defined as:

- Patient: Contains the patient's personal data, such as patient's name, date of birth or patient comments among others. The key attribute in this level is "PatientID" which is normally associated with his or her medical history number and indispensable for each DICOM .

- Study: Contains information about the study such as the study date, acquisition number or study description among others. The key attribute is "StudyInstanceUID" which should be unique for each study.

- Series: Contains information about a series, ie. modality, series number..., and is provided by a sequence of images for each study. The key attribute is "SeriesInstanceUID".

- Instance image: Contains information about the instance image. The key attribute is "InstanceUID" which is unique for each image.

Web access to DICOM and security

The NEMA and the International Organization for Standardization (ISO) decided to introduce the Web Access to DICOM Persistent Objects (WADO) service for web distribution and display of medical images because of the limitations of the point-to-point DICOM transfer syntax [98].
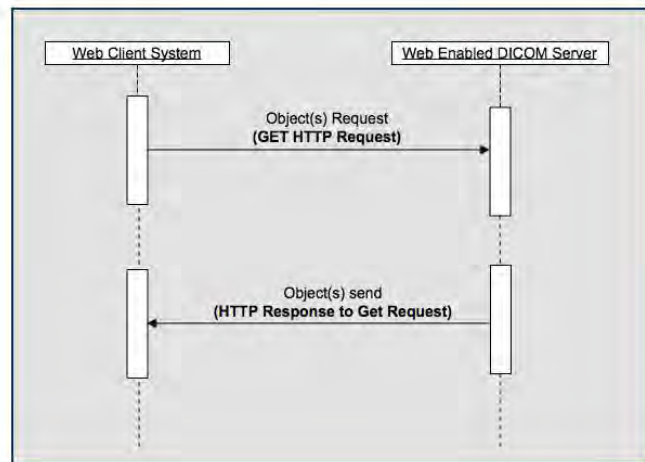
Figure 1.8: Example of an interaction information model.

As shown in Figure 1.8, the interaction is done between a web client system and a web embedded DICOM server using HTTP requests. The HTTP request should use a GET method to obtain parameters provided by a WADO query. Then, in the HTTP response, the "Accept" field of the GET method request will specify the media type(s) acceptable to the web client system. The content of the message-body varies according to the media type [11].

On the other hand, the DICOM standard does not address issues of security policies, though clear adherence to appropriate security policies is necessary for any level of security. The standard only provides mechanisms that can be used to implement security policies with regard to the interchange of DICOM objects between application entities [86]. However, DICOM is provided with a system management profile designed to support automation of the configuration management. Note that the system management profiles are generally addressed by the implementation, not the application entities. The same profiles need to be supported by the different applications in the network. This entails that the security and protection of the DICOM information be implemented by an external system in the server or web-client application. Normally, the DICOM store service is used to send images or other persistent objects (structured reports, etc.) to an image database or workstation. Details of such operations will be explained in the next subsection.

1.1.3   Medical databases

The increment of medical data supposes a new paradigm in the development of new system architectures and digital databases to store patient information along with his

or her medical images [28]. However, digital images are produced in ever-increasing quantities, and they improve the diagnosis and follow-up of patients with, usually, a non-invasive physical examination. The importance of developing database models and architecture systems in order to store medical images has seen a significant increase over the last decade. The integration of the Picture Archiving and Communication System (PACS) with the DICOM protocol has enabled to offer solutions to manage and store digital images in huge medical databases.

Picture Archiving and Communication System (PACS)

PACS is an evolving healthcare technology for short and long term storage, retrieval, management, distribution and presentation of medical images. The universal format for PACS image storage and transfer is DICOM. PACS, first originated in the early 80's, became a major digital radiology development in the 90's. The components of a full-scale PACS include medical images, a Radiology Information System (RIS), data acquisition computers, file storage systems, radiologisy workstations, and sometimes Hospital Information System (HIS), interconnected by computer networks [198].

Hence, the availability of PACS has revolutionized the practice of radiology over the past two decades [77] and will eventually increase productivity in radiology and medicine [113]. Moreover, PACS has established itself in Europe as a key tool for the management of diagnosis, particularly in radiology driven healthcare environments. Increasingly, however, an enterprise-wide PACS integrated with the Electronic Medical Record (EMR) is conceptualized and implemented in a growing number of selected hospitals all over Europe [106].

PACS systems consist of software and hardware parts, which directly get pictures from acquisition machines. The images are transferred to a workstation for viewing purposes and issuing radiological reports. The PACS viewer is a software tool installed at the workstation that the radiologist uses to receive and display radiological images or even reports. The images are then archived in the PACS server for later downloading to other workstations.

The major components in PACS consist of an image and data acquisition gateway, a PACS server and archive, and several display workstations integrated together by digital networks. PACS can be further connected with other healthcare information systems by database gateways and communication networks. After an image or an image set has been captured in digital form from an acquisition device, it is transmitted to a storage device for long-term archiving. A digital image file requires a large storage capacity for archiving [82].

PACS features and usability

The basic management of a PACS structure is performed using techniques such as quering, retrieving of data, and routing to the database. These queries are defined to search the database for a number of examinations using various criteria, such as the patient's name, medical record number or accession number [182]. The database is the component of the PACS system used to store all the information [74].

In fact, there are two basic PACS architectures: one is centralized with a central cache and controller, and the other is distributed with a distributed cache and central controller. However, the basic components of a PACS system[5] can be described as follows:

- PACS central server: the hardware core of the system.

- PACS workstation: provides an interactive display of digital images to a PACS viewer.

- Database system: responsible for managing the storage of all the information and images in the PACS system.

- DICOM server: responsible for all the DICOM based communication with imaging modalities, other PACS servers and DICOM workstations.

- Storage system: the physical storage required to store the DICOM images in the PACS system.

- Interfaces to RIS/HIS: aggregates all the data of the patient from different departments and provides an end-to-end patient care workflow.

- Web server for remote access: with this the images and other PACS data can be accessed using internet browsers.

Figure 1.9 shows an example of a typical PACS network structure with a central server that stores a database containing images connected to one or more clients through a local area network. These images can be provided by different DICOM modalities from different acquisition machines connected with the PACS server. Also, it should provide a single point of access for images and the data associated with them. One of the major advantages is that it allows a healthcare organization (such as a hospital) to capture, store, view and share all types of images internally and externally. Moreover, many clinical and diagnostic workstation can be connected with the PACS server [176].

---

[5]ActualMed PACS. URL: http://www.actualmed.com
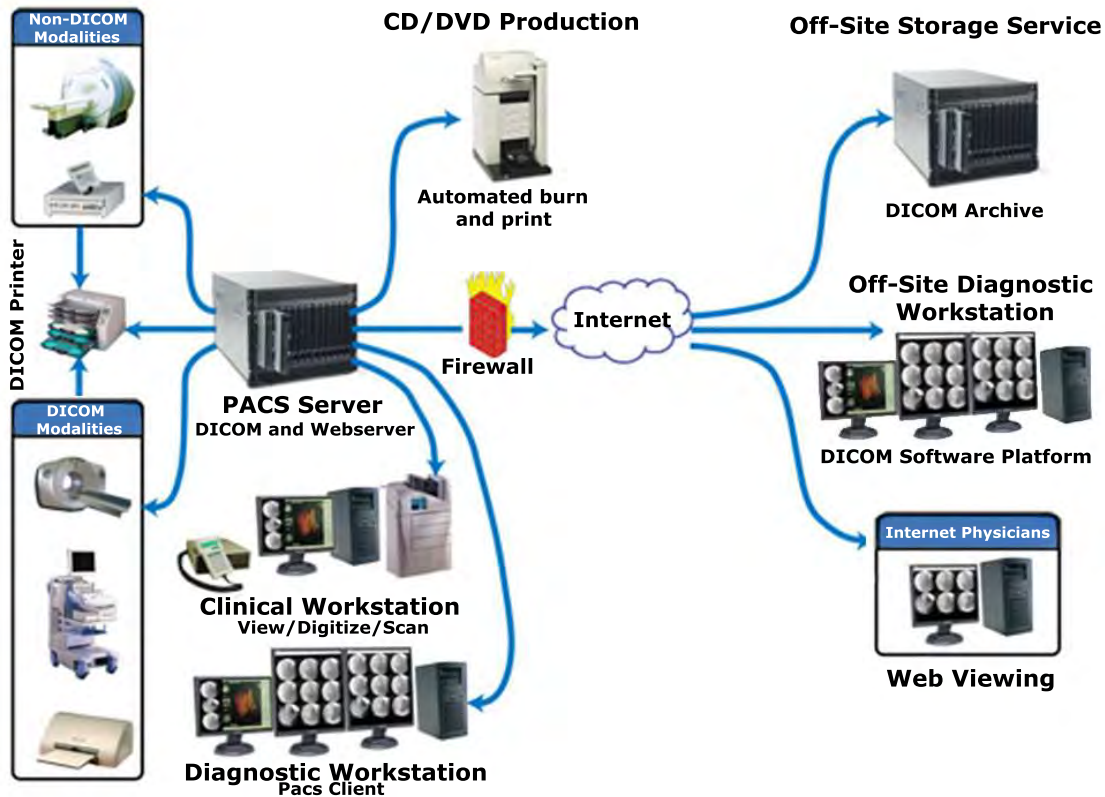
Figure 1.9: Example of a PACS network connection [87]

PACS are now starting to include web-based interfaces so clients can connect remotely over a Virtual Private Network (VPN) or Secure Web Site (HTTPS) [87, 41]. Remote access is regarded as more convenient for several different practitioners to view the same information simultaneously. Figure 1.10 shows another example of an extension of a typical PACS network architecture.

In summary, the inclusion of the new technologies provides PACS systems multiple advantages over film based reporting as listed in the following:

- Reduces the cost of radiology by eliminating the need to have CD or hardcopy films and the associated storage space required for the same. The cost of digital storage has decreased drastically over the past few years.

- PACS provides a faster and more reliable way of retrieving prior images.

- PACS enables remote access of the images, helping the radiologist to manage his/her time effectively.
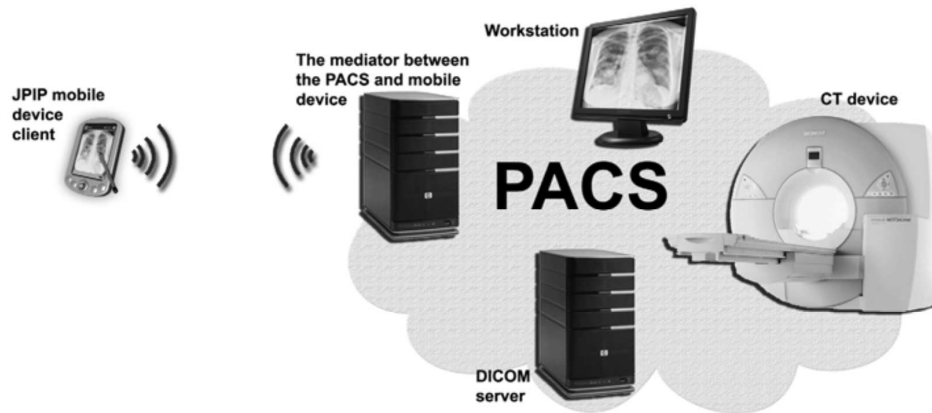
Figure 1.10: Example of a PACS network structure applied to modern technologies [41].

- PACS provides easy integration of images into the hospital information systems. This enables an effective way of managing comprehensive patient information providing a single point of access to all the patient's information.

- As the images and reports are transmitted digitally, when combined with an effective workflow, this can deliver a fast report turnaround. This also enables a more rapid diagnosis and potentially enhances patient care.

- It enables faster peer review of the reports and sub-specialty second opinions resulting in a significant improvement in accuracy of diagnosis and patient care.

- Access simultaneously to images and reports.

1.2 Context of research

The Computer Vision and Robotics group (VICOROB) of the University of Girona (Girona, Spain) has been working on medical image analysis since 1996. The Image Analysis Lab is committed to developing and optimizing methods (including both hardware and software developments) for the analysis of data, with particular focus on the study of medical images. To this end, computer algorithms were developed which include techniques such as image segmentation, pattern recognition, object detection and characterization, statistical signal processing, and registration.

The research projects are focused on breast, prostate and skin cancers and brain imaging. Some of the most relevant projects of the Image Analysis Laboratoy are

related to breast cancer topics. Different detection research projects based on screening programs in high-risk populations is a major asset in the struggle against breast cancer. Currently, screening programs in most developed countries focus on analyzing digital mammography images. Moreover, a variety of modalities such as X-ray radiology, magnetic resonance imaging, ultrasound, and computed tomography have also been developed. Currently, the lab is engaged in Computer-Aided Detection (CADe) and Computer Aided Diagnosis (CAD) of breast cancer in mammography from MRI and US.

On the other hand, the Laboratoire Electronique, Informatique et Image (LE2I) of the University of Burgundy (Dijon, France) is associated with the INSIS (main) and INS2I (secondary) departments of the Centre National de la Recherche Scientifique (CNRS) (FRE CNRS 2309 in January 2001, UMR CNRS 5158 in January 2003 and UMR CNRS 6306 in January 2012 (IMAC group).

The medical imaging research group of the LE2I is mainly based in the Faculty of Medicine of Dijon. They have access to imaging equipment of the University Hospital of Dijon (CHU), especially in nuclear medicine and Nuclear Magnetic Resonance (NMR). The purpose is to develop medical tools to help in medical decisions (classification of patients according to the diagnosis and prognosis, as well as therapeutic monitoring). Relevant research projects are focused on prostate cancer topics among other projects on, for example, cardio-vascular imaging. MRI and MRS allow better visualization of the zonal anatomy of the prostate and better delineation of tumor location, area and volume. In fact, it is important to determine whether a relationship exists between the Tumor Volume (TV) or research domain using MRS at $3T$ and the Prostate Specific Antigen (PSA) parameters for patients with known prostate cancer.

1.3 Motivations

Breast cancer and prostate cancer are two of the major issues among all cancers in men and women in most developed countries[6]. According to Breast Cancer Statistics, about $40,000$ of the new cases in women in the U.S[7] were expected to die in 2014 from breast cancer, though death rates have been decreasing since 1989 (with larger decreases in women under 50). These decreases are thought to be the result of treatment advances, earlier detection through screening, and increased awareness. In this sense, many automatic and semiautomatic techniques for cancer detection have been studied in the last few years including topics such as segmentation, description, and classification of medical issues.

---

[6]The American Cancer Society (ACS). URL: http://www.cancer.org
[7]U.S Breast Cancer Statistics. URL: http://www.breastcancer.org

Any diagnostic tool to help improve the sensitivity of breast cancer detection is highly valued. The usefulness of mammography in the symptomatic patient is undisputed; mammography is primarily used to demonstrate the presence of breast cancer and, specifically, to indicate the size and location of the tumor. There is also considerable evidence indicating the ability of mammography to detect cancer.

On the other hand, prostate cancer is the most common cancer among men, with the exception of skin cancer[8]. In 2014, an estimated $233,000$ men in the United States will be diagnosed with prostate cancer. Most prostate cancers (93%) are found when the disease is confined to the prostate and nearby organs. Prostate cancer is rarely found in men younger than 40. According to The American Cancer Society (ACS) estimates for prostate cancer in the United States for 2015, around $220,800$ new cases of prostate cancer will be diagnosed[9].

The probability is about 1 man in 7 will be diagnosed with prostate cancer during his lifetime. Moreover, around $29,000$ people died from prostate cancer give a period of time here. Prostate cancer can be a serious disease, but most men diagnosed with prostate cancer do not die of it. In fact, a total of $1,665,540$ new cancer cases and $585,720$ cancer deaths were projected to occur in the United States in 2014 [171].

Different MRI techniques and X-ray radiology are used for the image acquisition. Moreover, MRI has also been established as the best imaging modality for detection and localization on account of its high resolution and excellent spontaneous contrast of soft tissues along with the possibilities of multiplanar and multiparameter scanning [29]. As such, three MRI techniques (3D T2-Weighted Imaging (T2WI), Diffusion-Weighted Imaging (DWI) and Perfusion-weighted imaging (DCE)) and 3D Magnetic Resonance Spectroscopy (MRS) are the most commonly used.

One of the major motivations in doing this PhD thesis was to develop a generic medical application for the medical community, with the support of medical staff and hospitals for the diagnoses of cancer. A generic application confirms that the development of generic tools that underpin effectiveness and adaptability in different image modalities. The use and support of standard tools facilitate the development of generic applications. In order to prove that the proposed generic medical tool can be adapted to different image modalities, this thesis has focused on breast and prostate cancer. Furthermore, the support of the University of Girona (for breast cancer) and University of Burgundy (for prostate cancer) is the key to my encouragement to develop this project with a cotutelle program.

---

[8]Journal of Clinical Oncology. URL: http://jco.ascopubs.org
[9]CANCER.NET. URL: http://www.cancer.net

1.4   Thesis objectives

The main objective of this PhD thesis is to develop and design an architecture model in order to manage medical diagnosis data using a web-based medical imaging application. Hence, a centralized database of medical images digitally acquired is presented.

> The aim is to provide a useful tool for the medical and scientific communities in order to manage database images including their associated diagnoses. Moreover, the goal is to develop a medical tool focused on cancer issues with a collaborative work between experts.

The focus is on issues related to medical databases in medical imaging. It is important to remark on the importance of managing medical data associated with imaging studies. For this reason, the use of the XML environment is also studied with special attention to radiology reports, integration of diverse data sources including imaging sources, and, finally, secure access to the associated data.

A study of the most recently published research work in the areas of web-based application, digital and medical databases, the system management of medical data, and security is also presented. From this study, we will select and implement two applications to demonstrate the usability using different image modalities. In this sense, we aim to demonstrate the advantages of using our proposed system architecture implementing a web-based application for a mammographic and prostate images, as well as to point out significant conclusions.

The aims of this PhD thesis can be summarized as:

- Conduct a literature review to obtain a basis and fundamentals for web-based medical applications and classify research work in the literature.

- Propose and test a new system architecture to develop the web-based application related with a digital medical database.

- Build a web-based application to manage medical images and diagnoses for different image modalities.

- Apply the implemented web-based application to X-ray radiology to aid the diagnosis of mammographic findings.

- Apply the implemented web-based application prostate cancer studies from MRI.

- Evaluate the developed web-based applications.

- Suggest future improvements connected to this research work.

Framework implementation

This section describes the specific objectives of the implementations of the proposed web-based application. Two frameworks are developed to prove the viability of the proposed architecture model. The frameworks are commited to the management of data related to breast and prostate cancers.

In this sense, the main features are:

- To develop and design an architecture model in order to maintain a system management and diagnostic oriented to a web-application.

- Framework provides an architecture to allow annotating mammographic findings. In mammography some of the most common features to annotate are location, shape, size, margins, density, and distribution of others.

- Use a Breast Imaging-Reporting and Data System (BI-RADS) classification to classify mammographic findings.

- Visualization of the most common views of mammographic images.

- Visualize and analyze different MR modalities for prostate images: T2WI, DWI, DCE and MRS.

- Framework provides an architecture to allow annotating prostate findings. In prostate cancer some of the most common features to annotate are location, size, margins, and distribution of others. Moreover, the identification of the prostate gland zone such as the central, peripheral or transition zone are also evaluated.

- Use a Prostate Imaging-Reporting and Data System (PI-RADS) classification to classify prostate studies.

- Create a network-based database system allowing the management of both MRI and spectroscopy data sets.

- Provide a collaborative platform with the following goal: aid for the experts to make a medical diagnosis, collaborative work between different hospitals and health care systems, and a training, learning and support platform for novice radiologists.

- Perform an exhaustive evaluation of data analysis and design effort benchmarks for the implemented web-based applications.

- Enhance interactivity with medical experts and radiologists.

1.5   Thesis summary and organization

A brief summary of each chapter is presented as follows:

Chapter 2 presents an extensive survey of web-based medical applications in medical imaging published in the literature. It analyzes the state-of-the-art of the most relevant research projects related to web-based applications with medical databases, exchange and management of medical data, and security.

Chapter 3 is focused on the developed system architecture framework . Moreover, a view model and a detailed description are also presented.

Chapter 4 presents the first framework application dedicated to the analysis of mammographic cases. This chapter covers all the parts of the proposed framework as well as reporting the experiments carried out and a discussion of the outcome.

Chapter 5 presents the second framework based on the analysis of an MR examination of the prostate cancer. This chapter covers all the components of the proposed framework as well as reporting the experiments carried out and a discussion of the outcomes.

This thesis concludes with Chapter 6 in which conclusions and the ways in which further work will be directed are covered.

Moreover, Appendix A is given to explain the evaluation parameters used to evaluate the proposals.

Chapter 2

Web-based medical applications applied to digital databases and digital imaging

The rapid growth of digital technology in medical fields over recent years has increased the need for applications able to manage patient's medical records, imaging data, and chart information. Technologies related to architectures of digital databases, storage and transmission protocols, management of large volumes of data and security are being improved every day, allowing the possibility of reading, analyzing, and even diagnosing away from the medical center where the information was acquired. The objective is to review web-based applications applied to the medical imaging field, analyzing how applications store and exchange the huge amount of information. Indexing tools based on the XML protocol, allowing a fast retrieval of the information, and security issues are also studied. The purpose of this chapter is to point out the main advantages and drawbacks of the work reviewed. This chapter is an adaptation of a paper submitted to the Journal JBI[1].

> It is not enough to discover and prove a useful truth previously unknown, but that it is necessary also to be able to propagate it and get it recognized.
>
> Jean B. Lamarck

---

[1]This Chapter is an adaptation of a paper under review "Web-based applications in medical imaging" that is submitted to the Journal of Biomedical Informatics.
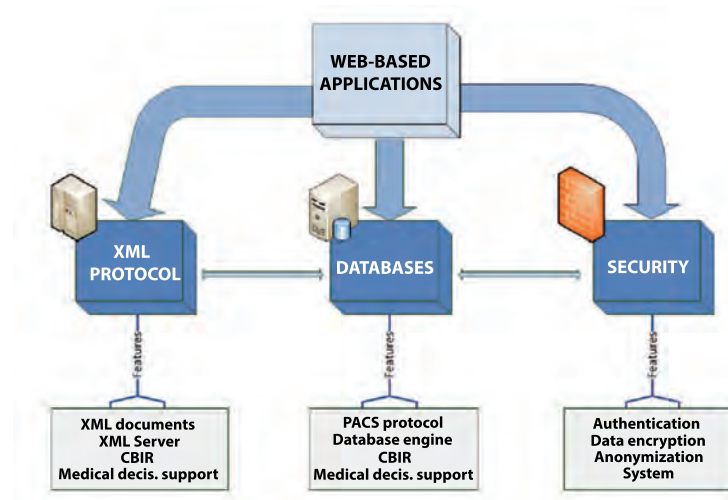
Figure 2.1: Main areas related to web-based medical applications.

2.1   Literature classification

2.1.1   Introduction

We present an overview of the most relevant research projects using web-based applications in medical informatics. This chapter starts with a presentation of the different aspects of the development of web applications, graphically depicted in Figure 2.1. Firstly, medical databases are reviewed, highlighting PACS solutions since this is the main architecture in medical information storage and how web-based medical applications are related to the PACS. Secondly, the review covers the management of information, mainly the management of medical data associated with imaging studies. In this area, the use of the XML environment is studied with special attention to the inclusion of radiology reports and the integration of diverse data sources such as images. And thirdly, the more common security systems and access to the associated data are also described. In order to manage medical data, security systems are essential for local and remote access, protection for the patient data, or encryption of the data. Indeed, encrypted solutions are absolutely necessary in medical applications because of the confidentiality of the personal information.

Then, this chapter is structured as follows. Firstly, we present the search strategy using key words concerning the medical web-based applications. Subsequently, in Section 2.2, we present a classification and discussion on each relevant topic: database management, exchanging information and security. In Section 2.3 we describe the most interesting aspects of the implementation of medical web-based applications.

2.1.2 Search strategy

The literature review was performed with the following databases: PubMed, Scholar, Embase and Scopus. The main search strategy combined four concepts: Web-based medical applications, PACS environment, XML framework and security. For each concept, the following search terms were used: web-based applications, computational medicine, digital databases, XML protocol, and internet security.

We wanted to identify the main research projects for each area in PubMed and Scopus. As these searches resulted in many duplicates, this dual search strategy was not repeated in Scholar or Embase. To retrieve other relevant publications, we also examined the reference lists of the selected publications and reviews, and have included those projects that met the eligibility criteria.

2.1.3 Web-based applications

We aim to review web-based applications developed during the last decade in the medical environment. The final goal of these applications is to help physicians when diagnosing a case. We can classify these applications into two categories according to the way information is provided. On the one hand, medical decision support systems aim to analyze multiple data and return to the expert, a ranking probability or even a second opinion about the patient to be diagnosed. On the other hand, content-based retrieval systems aim to provide the expert with cases similar to the one being analyzed. When this data is related to imaging technologies, it is referred to as a content-based image retrieval system.

There are recent surveys on both classes of applications [166, 127, 110, 2, 84]. In contrast to all these projects, we centered our literature review on the implementation of web-based applications, independently of whether the goal was a medical decision tool or a CBIR. From a web-based tool point of view, the two approaches are similar. As illustrated in Figure 2.2, they need a web-server to provide the interface between the user-computer and the medical database, consisting of data, but also of images. Web-based applications run on the user's computer, usually through a small application called an applet. From the web-server page, the applet is downloaded to the user's machine and is then executed an environment provided by the Java Virtual Machine (JVM). Notice that once the applet is running, it is a process independent of the web browser itself.
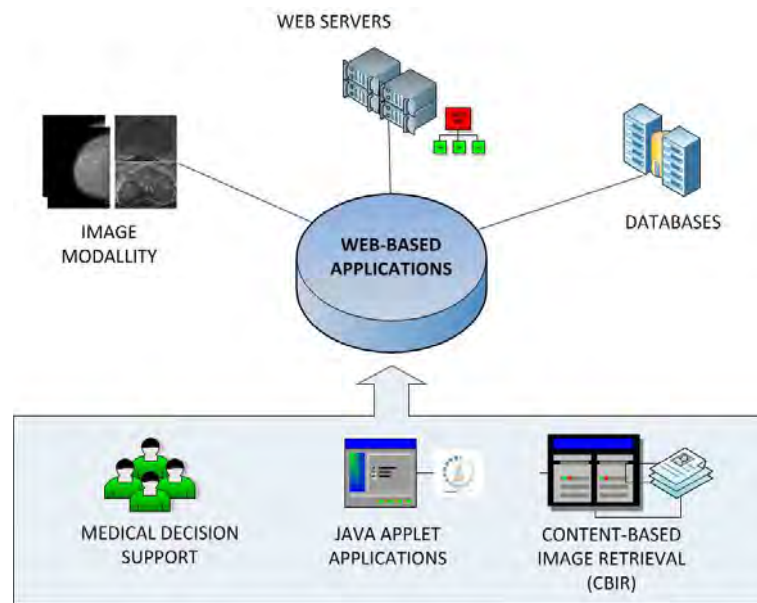
Figure 2.2: Web-based applications allow the creation of a visual interface between the stored medical data and the user applications.

2.1.4   Web-based applications examples

A number of web-based medical applications have been presented over the last decade. Fernàndez-Bayó et al. [48] presented an approach involving a web server that allowed the query and retrieval of images stored in a DICOM format. The images can be viewed a web browser with the use of a small Java (Java) program known as the DICOM Java viewer, which was executed in the browser. The system supported mainly magnetic resonance and computed tomographic images in a very simple graphical interface. The use of a web-browser user interface allowed access, independently of its hardware and of the operating system. The application was based on a client-server intranet system and on multi-platform intranet technologies such as Java. The most significant disadvantage of the system was the response speed to obtain images the web browser.

A similar medical web-based application was presented by Deftereos et al. [35] to facilitate clinical work by allowing the communication of patients'records between experts. The aim of this project was to share patient records of a particular disease at a national level. The advantage of using an automatic tool is not only the exchange of information but also the standardization of the records. The web-application was implemented with Java technology and was connected to a web server and with a patient's database using the SQL Language (SQL). The use of Java makes scalable designs possible, suitable to the needs of health care units differing in size. However,

the application operates on stand-alone computers; therefore the database server is constrained to run on the same machine. These difficulties were encountered because of the lack of more powerful database servers and workstations.

Matsopoulos et al. [118] proposed a web-based medical system developed for the management and processing of obstetrical, gynecological and radiological medical data. The system, implemented in a modular structure, recorded the necessary medical information in terms of patient data, examinations and operations. It also provided the user with advanced image processing tools for manipulation and processing and storage of ultrasound and mammographic images using the DICOM protocol. This system ensured the provision to medical experts of an unified patient management protocol, either in the hospital or outside. On the same thematic, a web portal platform was implemented by Koutelakis and Lymperopoulos [97]. In general, Java Server Pages (JSP) technology was mainly used to satisfy design specifications and data exchanges. The principal advantage proposed in this work is the use of Java applets. In their evaluation, the authors obtained the improvement of DICOM services when they are provided through the web.

Arnold et al. [9] implemented a well-documented and easily accessible communication loop adaptable to different types of image modalities. Images were displayed in a native DICOM format with a Java applet, along with the use of various image manipulation tools. The web-based infrastructure consisted of a server that stored imaging studies and reports, and with web browsers that downloaded and installed optional client software on demand. In contrast with previous projects, their approach allowed the inclusion of Hypertext Pre-Processor (PHP) as the language to implement the web-application. Application logic consists of a set of PHP modules that were accessible using an application programming interface. The system might be adapted to any clinician-specialist communication loop, and, because it integrated radiological standards with web-based technologies, it can communicate and report data in an efficient way.

On the other hand, we proposed [116] an interactive Java applet interface designed as a web-based tool. It aimed to facilitate the diagnosis of mammographic cases by providing a set of image processing tools that allowed a better visualization of the images and a set of drawing tools used to annotate the suspicious regions. Hualei et al. [81] presented a medical image access and presentation system as a web-based system designed for remotely accessing and presenting DICOM images in a teleradiology operation through the internet. One of the main advantages of this work is that they provided multiple DICOM image retrieval methods: text-based retrieval, content-based retrieval and combined retrieval, as well as the implementation of on-demand medical image transmission method in a web environment. Haraguchi et al. [71] presented a web-based image viewer designed for the visualization of PET-CT-scan follow-up studies. They innovated it by defining a series

of dynamically generated image grid layouts that displayed sets of related images together in order to improve the comparison and assimilation of images. A similar web-application tool allowing, for example, to the use of retrospective studies was also presented by Baltasar-Sánchez and González-Sistal [12] for X-ray skeletal images.

A web-based tool for patient-specific consultation in the clinical scenario of breast augmentation was proposed by de Heras Ciechomski et al. [76]. This example shows that the current state of development allows for the creation of responsive and effective Web-based 3D medical tools, even with highly complex and time-consuming computation, by off-loading them to a dedicated high-performance data center. Internet web-based communications tools to guide decision making on treatments and tests were also studied by Schwitzer et al. [166]. They analyzed 4 areas of variability in the use of key web-enabled features in online decision-support tool development. The main features were focused on the presentation of outcome probability data tailored to the individual user, the use of videotaped patient interviews in the final product to convey the experiences of people who have faced similar diagnoses in the past, the ability to interact with others in a social support network and the accessibility of the tool to any health care consumer with an Internet connection. Recently, Moorhead et al. [124] studied the benefits and limitations of social media for health communication among the general public, patients and health professionals, and identified current gaps to provide recommendations for future health communication research.

CBIR systems have been proposed for a variety of pathologies [110]. For instance, Hsu et al. [80] described the implementation of a web-based CBIR for spine pathologies, consisting of four components: a client applet, a gateway, an indexing and retrieval system, and a database of images and associated text data. It allowed the exploration of a large bio-medical database of digitized spine X-ray images using a combination of textual or even visual queries. In the latter case, a query editor enabled users to pose queries by sketching a unique shape or selecting or modifying an existing shape from the database. Ortega et al. [134] presented a CBIR for retinal image analysis consisting the web client interface, of the web server that processes the requests and the service module that performed the image processing tasks.

2.2 Implementation of medical web-based applications

Table 2.1 presents a compact at-a-glance summary of the reviewed projects and their main features. In this literature review we aim to describe the methodological

Year	Paper			Database		Data Exchange			Security					
1999	*Halamka [68]*	✓	✗	SQL-Server	CT	✗	✗	n.d	✓	✓	✓	✗	✓	HL7
2000	*Fernàndez-Bayó [48]*	✓	✓	SQL-Server	CT,MRI	✗	✗	n.d	✓	✓	✗	✗	✗	n.d
2001	*Defteros [35]*	✓	✗	SQL-Server	MRI	✓	✗	EHCR	✗	✗	✗	✗	✗	n.d
	Pappadakis [137]	✓	✓	own DB	n.d	✓	✗	R-EHCR						CORBA
2002	*Sinha [173]*	✓	✓	own DB	CT	✓	✗	HIS.	✓	✓	✓	✗	✗	HIS
	Rosset [159]	✓	✓	n.d	CT,US	✗	✗	n.d	✓	✓	✗	✓	✓	CBIR
	Rassinoux [149]	✓	✗	own DB	n.d	✓	✗	HL7	✓	✓	✗	✗	✓	HL7
	Schwitzer02 [166]	✓	✗	Cochrane, DIPEx	n.d	✓	✗	n.d	✓	✓	✗	✗	✓	n.d
2003	*Zhang [201]*	✓	✓	own DB	CT	✓	✓	IIS	✓	✓	✓	✓	✓	HIS
	Müller [127]	✓	✓	own DB	CT,MRI	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
	van der Haak [65]	✓	✓	IS- H*MED	n.d	✗	✗	n.d	✗	✗	✗	✗	✗	HIS
2004	*Matsopoulos [118]*	✓	✓	own DB	MRI,US	✗	✗	n.d	✓	✓	✗	✗	✗	n.d
	Müller [126]	✓	✓	own DB	CT,MRI	✗	✗	n.d	✓	✗	✗	✗	✗	n.d
	Gritzalis [59]	✓	✓	n.d	n.d	✓	✗	n.d	✓	✓	✓	✗	✓	EHR
	Kluge [95]	✓	✗	n.d	n.d	✗	✗	n.d	✓	✓	✓	✗	✓	EHR
2005	*Hastings [72]*	✓	✓	MAkoDB	MRI	✓	✗	GridPACS	✓	✓	✗	✗	✗	OGSA
	Coeira [148]	✓	✗	MySQL	n.d	✓	✓	Apache	✗	✗	✗	✗	✗	n.d
2006	*Kautelakis [97]*	✓	✓	Oracle9iAS	n.d	✓	✗	n.d.	✓	✓	✓	✗	✗	PACS
	Kurc [101]	✓	✓	MySQL	n.d	✓	✓	MakoDB	✓	✓	✗	✗	✗	CAD
2007	*Güld [61]*	✓	✓	GridPACS	n.d	✓	✗	n.d.	✗	✗	✗	✗	✗	n.d
	Tahmoush [180]	✓	✓	MySQL	MRI	✓	✗	x.d	✓	✓	✓	✗	✓	RMI
	Doi [40]	✓	✓	own DB	CT, MRI	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
	Arnold [9]	✓	✓	MySQL	n.d	✗	✓	XML-RPC	✓	✗	✗	✗	✗	n.d
	Marcos [114]	✓	✓	MEDIMAN	MRI	✗	✓	OGSA-DAI	✗	✗	✗	✗	✗	n.d
	Sucurovic [177]	✓	✗	n.d	n.d	✗	✗	n.d	✓	✓	✓	✗	✓	EHR
2008	*Choe and Yoo [31]*	✓	✗	n.d	n.d	✓	✗	HIS-CAC	✓	✓	✓	✗	✓	HIS
	Oster [135]	✗	✗	CaGrid	CT	✓	✗	GRID	✓	✓	✗	✗	✓	GAARDS
2009	*Hsu [80]*	✓	✓	MySQL	X-rays	✓	✓	H IS-RIS	✓	✓	✓	✗	✗	n.d
	van der Linden [109]	✓	✗	own DB	MRI	✗	✗	n.d	✓	✓	✓	✗	✗	EHR
	Long [110]	✓	✗	own DB	n.d	✓	✗	n.d	✗	✗	✗	✗	✗	n.d
2010	*Wang [193]*	✓	✓	OsiriX	n.d	✗	✗	Retinal	✗	✗	✗	✗	✗	n.d
	Do [39]	✓	✓	MySQL	CT	✗	✗	✗	✓	✓	✓	✓	✓	RIS
	Mahmoudi [112]	✓	✓	n.d	n.d	✗	✗	n.d.	✓	✓	✓	✗	✗	n.d
	Ortega [134]	✓	✗	own DB	CT	✓	✗	XML-RPC	✓	✓	✗	✗	✗	n.d
2011	*Gao [53]*	✓	✓	QBE	X-Rays	✓	✗	n.d	✗	✗	✗	✗	✗	n.d
	Akül [2]	✓	✓	EMR	CT	✓	✗	n.d	✗	✗	✗	✗	✗	n.d
	Korenblum [96]	✓	✓	MySQL	CT	✓	✗	LibXML.	✗	✗	✗	✗	✗	n.d
	Rajala [147]	✓	✓	CRDAS	MRI	✗	✗	n.d	✓	✓	✓	✓	✓	CRDAS
	Bond [23]	✓	✓	ECG	MRI	✗	✓	HL7	✗	✗	✗	✗	✗	n.d
	Neubauer [130]	✓	✓	SQL-Server	n.d	✗	✗	n.d	✓	✓	✓	✓	✗	HSM
	Haas [66]	✓	✗	own DB	n.d	✗	✗	n.d	✓	✓	✗	✓	✓	EHR
	Baltasar [12]	✓	✓	own DB	X-Rays	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
	Haraguchi [71]	✓	✓	own DB	PET-CT	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
2012	*de Hwang [84]*	✓	✓	own DB	CT, X-Rays	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
	de Welter [194]	✓	✓	n.d	X-Rays	✗	✓	HIS-RIS	✗	✗	✗	✗	✗	n.d
	Mata [116]	✓	✓	MySQL	X-Ray	✓	✗	eXistdb	✓	✓	✗	✓	✓	n.d
	Gerstmair [54]	✓	✓	RIS	CT	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
	Carrión [26]	✓	✗	myPHR	n.d	✗	✗	n.d	✓	✓	✗	✓	✓	PHR
	de Heras [76]	✓	✗	own DB	n.d	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
2013	*Jorritsma [85]*	✓	✓	n.d	CT,MRI	✗	✗	n.d	✗	✗	✗	✗	✗	n.d
	Chipman [30]	✓	✗	MySQL	n.d	✓	✗	HL7	✗	✗	✗	✗	✗	n.d
	Agarwal [1]	✓	✗	own DB	n.d	✓	✗	n.d	✓	✓	✗	✗	✗	PHR
	Kuijpers [100]	✓	✗	own DB	MRI	✗	✗	n.d	✓	✓	✗	✗	✗	n.d
2014	*Huang [83]*	✓	✓	n.d	CT,MRI	✗	✗	n.d	✓	✓	✗	✗	✗	n.d
	van de Watering [196]	✓	✓	EMR	n.d	✗	✗	n.d.	✓	✓	✓	✗	✗	n.d
	Schweiger [165]	✓	✗	n.d	n.d	✗	✗	n.d.	✗	✗	✗	✗	✗	n.d
	Biternas [21]	✓	✗	n.d	n.d	✓	✓	SpinXML	✗	✗	✗	✗	✗	n.d
	Kammerer [91]	✓	✓	MySQL	CR	✗	✗	n.d	✓	✓	✗	✓	✓	XAMPP
	Hualei [81]	✓	✓	n.d	CT, X-Rays	✓	✓	eXistdb	✓	✓	✗	✗	✓	EHR
	Sethi [168]	✓	✗	n.d	n.d.	✗	✗	n.d	✓	✓	✓	✓	✓	HSM

Web-based application
PACS
Database
Image modality
XML documents
XML-server
System architecture
Security
Authentication
Data Encryption
Anonimization
Web-security system
System architecture

Table 2.1: Compact at-a-glance summary of the reviewed projects and their main features. 'n.d.' means 'not defined'.
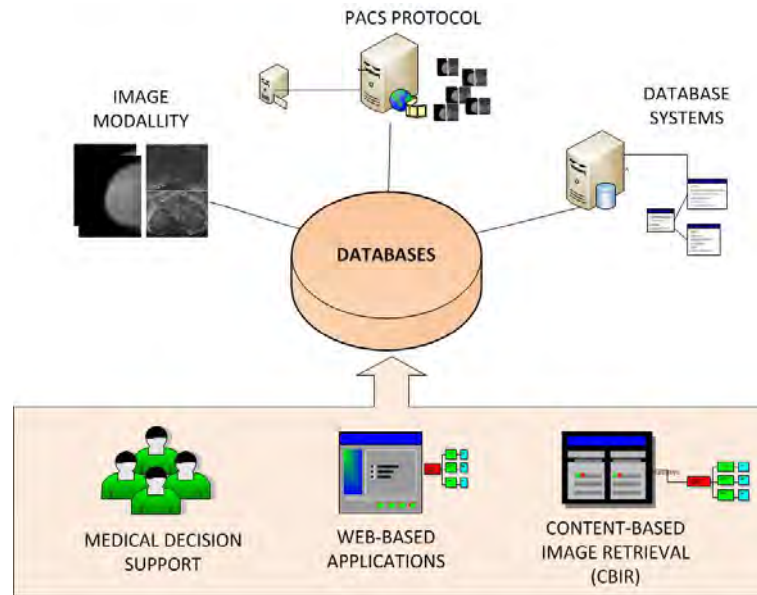
Figure 2.3: Web-based applications rely on the use of databases. These databases may be categorized according to their architecture and the kind of data/images they contain.

aspects of a web-based application and how to tackle the features presented in Figure 2.1. We focus this review on how the database is implemented, using PACS or not, and which images it stores. We also depict how the information is stored and transmitted with or without the use of XML protocol, and finally, we also illustrate the main security aspects.

2.2.1   Databases

Web applications typically run on the user's computer and communicate via standard procedures, i.e. http, to a web server application that usually contains or communicates with another computer that contains the medical data, as is depicted in Figure 2.3. The web server is responsible for searching the medical database and returning the information to the web application. As can be seen in table 2.1, in the medical context this database is currently stored under the PACS architecture, that allows integrally storing the patients' records with their own available digital images acquired at the hospital. Notice that PACS not only refers to a single database but also consists of image and data acquisition, storage, and display subsystems integrated by various digital networks. New digital imaging technologies present a challenge to the management of such large volumes of information, and PACS offers a practical solution. The initial reluctance to PACS has vanished as a result [83].

Different web-applications have described the implementation of a PACS system in their work for the storage, research, transmission and evaluation of medical images [118] even when applied to telemedicine applications [180]. Jorritsma et al. [85] studied the differences in usability terms of PACS with identical functionality, concluding that functional requirements alone are insufficient to determine the overall quality of a PACS. The implementation of the PACS also determines the PACS usability. In the same way, Van der Wetering et al. [196] presented a study based on an integrative model to determine PACS alignment and performance in hospitals, adopting theories and perspectives from the field of information system research and complexity theory.

The advantage of web-based applications is that they allow the easy addition of newer capabilities to the PACS system. Hastings et al. [72] complemented the PACS by providing extra management of the information using XML language, allowing a more friendly environment for large biomedical databases. Similarly, Mahmoudi et al. [112] developed a web-application tool connected to a PACS system to allow direct access to the data. The data were uploaded to the application for processing and the processed data were saved on the client machine. One of the advantages was that the application supported various DICOM formats. Rosset et al. [159] created an image-based electronic system to print film images directly from a PACS.

Another benefit of using a PACS in the clinical routine allows for better tools to support teaching and training tasks. Kautelakis et al. [97] developed a web-based interface to the clinical PACS, simplifying the access to the repository and providing a simple mechanism for accessing DICOM objects through the net. Rajala et al. [147] proposed a new system using an integration of PACS in a Hospital Information System (HIS). It consisted of an alternative implementation to regular PACS systems. Their study described the fusion of a PACS server with an HIS server and their implementation in a large university hospital organization.

In order to facilitate the integration of image-processing and visualization modules, Zhang et al. [201] proposed the integration of a PACS-oriented image data flow and the diagnostic display software architecture. They hypothetized that web technologies and web servers were the best solution for medical image applications in intranet and internet environments. Also, for Fernàndez-Bayó et al. [48], a client-server intranet system was implemented to obtain the images and display them on a web-browser. The authors stated that this approach provided an easier system to manipulate and maintain the data. Wang et al. [193] proposed a browse/server style method using inter-process communication techniques that allowed PACS workstation developers and image processing software developers to cooperate while focusing on different interests. We publish [116] and present in this thesis another alternative to connect PACS database to a Java applet application which will be explained in detail in Chapters 4 and 5.

Other approaches deal with the integration of PACS and CBIR or Computer Aided Diagnosis (CAD) proposals. Güld et al. [61] presented a technical framework to support the development of image retrieval systems in PACS environments. Akgül et al. [2] suggested that data sharing between databases (clinical PACS) should lead to a better performance in CBIRs systems. Hwang et al. [84] reviewed medical image retrieval systems comparing text-based and content-based systems. While text-based methods are widely used and fast, they require precise annotation. In contrast, a content-based approach provides semantic retrieval, although effective and precise techniques still remain elusive. A combination of the two approaches should provide a better and more robust approach. Moreover, the use of standardized terms in radiology reports [162] should increase the performance of such tools. On the other hand, the integration of CAD systems with PACS is still in its infancy, although different authors [40, 194] have already suggested that CBIR systems applied to CAD should integrate their results in a PACS, probably as DICOM structured reporting documents.

Finally, it must be mentioned that there are approaches outside PACS. Most of these approaches use the PACS only as an image database, while their own information is stored in another particular database [39, 96, 54, 91]. The use of an external database avoids interfering with the clinical data [135, 165].

2.2.2 Exchange of information

Once the web application receives the data from the PACS or medical database through the web-server application, it is ready to process according to its own goal. When the application finishes, the resulting outcome can be stored, either locally or transmitted to an external database. Since these results usually contain the analysis of the cases, they can be stored in formats compatible with word processor or spreadsheet tools in order to present the information in a structured way. Although there are different formats for information exchange [39, 54, 165], the XML has emerged as a leading facilitator in exchanging structured information between different platforms and storing information in an efficient way and at a low memory cost. In the column entitled "Data Exchange" of Table 2.1, it is evident that the XML language is also becoming a standard in information exchange in web-based applications. In Figure 2.4, the common language terms used in XML are depicted: files containing XML information are called XML documents and are stored in databases called XML servers, and used by applications or XML systems.

Hastings et al. [72] addressed the work to metadata and data management issues using a generic, distributed, XML-based data management system. According to the authors, the use of XML schemas and XML documents allowed the management and publication data models to efficiently store, query, reference and create virtu-
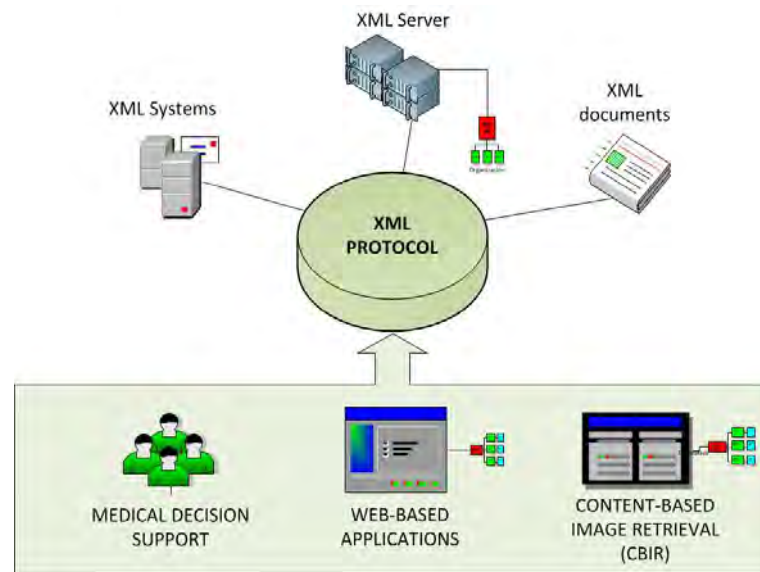
Figure 2.4: The XML protocol is used by many web-based applications to exchange information. XML documents are stored in XML servers and used by XML systems.

alized XML views in distributed and interconnected databases. Kurc et al. [101] showed one of the best examples of the usability of XML schemes oriented to medical research work. The proposed XML-based data management system supported the integration of data from disparate data sources and large data sets. The system facilitated management of XML schemes and on-demand creation and management of XML databases that conform to those schemes. Recently, Chipman et al. [30] presented a software that can convert local structured data into XML-formatted data. Hence, while clinicians and patients benefit greatly from readily available risk predictions, a large pool of multi- institutional family history data can be continually gathered.

According to Choe and Yoo [31], a key point to allow cooperation among heterogeneous repositories at various hospitals is the introduction of a common healthcare-record, and XML allows the efficient storage of this information.XML has progressively become a gold standard for accessing, representing and exchanging information, especially in the health care environment. Defteros et al. [35] integrated the use of the XML documents designed to allow the exchange of Electronic Healthcare Record (EHR) information between different clinical information systems. Complete patient records or record extracts are exported in the form of messages in XML documents. In the same way, Zhang et al. [201] used the XML formatted messages to send browser requests or to receive responses from the Web server through the http protocol. Rassinoux et al. [149] also presented an implementation of XML for the Electronic Patient Record (EPR) and discussed more specifically its growing use in

two areas of the EPR: firstly, as a format for the exchange of structured messages, and secondly, as a comprehensible way of representing patient documents.

There have even been attempts to change traditional binary data to an XML format, arguing that XML files are easily readable, searchable and provide a better form of semantics. For instance, Bond et al. [23] studied the encoding of Electrocardiogram data using an XML format, while Biternas et al. [21] introduced in their work a simple and general XML format for a spin system description in magnetic resonance imaging. The aim was to unify under one unique format the principle spin interaction specification conventions.

The XML standard is also used in CBIR systems as a way to link the retrieval machine to the web-based application, as in Ortega et al. [134] for retinal images or in Hsu et al. [80] for spine images. This latter work provided a simple interface between both systems. The information was transmitted between systems using just three XML primary events. The first one was used to determine whether a desired service was available and to obtain a list of currently available services. The second contained information about the features of the query, while the third one returned the matching images along with the similarity scores. Coiera et al. [148] proposed a similar architecture, where XML was used to represent the queries obtained from users in a consistent way.

In fact, XML schemes can be useful for mainly more complex query languages, like those related to DICOM files. For instance, in Marcos et al. [114], the advantages in representing DICOM header information using XML and storing their data in a XML database is proposed. In the proposal by Arnold et al. [9], XML documents included annotations and descriptive texts related to DICOM images. Authors argued that this capability allowed a more efficient exchange of templates and would represent a step toward standardization. In a similar way, Tahmoush et al. [180] also affirmed that the incorporation of a DICOM 3 capability was necessary in medical image systems and that XML support should also be included. Also, it has become the preferred method for uploading data that is not already in a DICOM format. Finally, Hualei et al. [81] remarked on the advantages of representing DICOM header information in a XML file, highlighting the importance of extracting and storing in the semantic attributes of DICOM images in a XML file(one XML file per DICOM image). Besides, Wang et al. [193] analyzed the use of XML in a PACS workstation managing DICOM images. The advantage is that all the information is embedded in the XML document, thus knowledge of the structure of the PACS workstation is unnecessary. Notice that the necessity of storing additional information related to the images stored in a PACS systems has become crucial, and new applications are appearing in this direction. An example of this was published [116] and will be presented in the next chapter.

2.2.3 Security

Security and privacy protection are essential aspects in web-based medical applications. Figure 2.5 depicts the main aspects in security areas that web-based applications have to deal with, including authentication of the user and encryption and anonymization of the data. Note that aspects related to security and data protection of Internet are beyond of the scope of this review. Under the column "Security" of Table 2.1, there is a summary of the efforts made in the following aspects: authentication, encryption, anonymization and the security architecture proposed.

The research work of Halamka et al. [68] is one of the first projects based on internet security systems related to medical applications. They introduced the concept of authentication in web-based medical applications. A security server authenticates the user by verifying that the correct password is typed. The user's name and password information are stored using encrypted security cookies. Papadakis et al. [137] proposed a secure medical digital library capable of confronting risks through the employment of a number of security services. An authentication system was also implemented to authenticate medical professionals and provide role-based access to the users of the digital library. Moreover, data integrity was achieved through digital signatures and confidentiality was accomplished with the employment of various encryption schemes. In a similar way, van der Haak et al. [65] proposed a model of a signature system for ensuring the integrity of the data. Authenticity of medical documents was developed and implemented in the proposed information system architecture applied to data security, data encryption and protection of the patients' health data. The aim was to identify the specific legal requirements applied to a cross-institutional electronic patient record, and described possible solutions for meeting these requirements.

Gritzalis and Lambrinoudakis [59] presented a security architecture designed to provide authentication and authorization services in web-based distributed systems. The proposed architecture was based on a role-based access scheme and on the implementation of an intelligent security agent per site. The architecture allowed authenticating the users either locally or using a remote connection, and subsequently gave temporary certificates and privileges to the authenticated users in accordance to their role. The system also encrypted the medical data. In a similar environment, Van der Linden et al. [109] applied different security tools to their work as identification, analysis of privacy and security related issues that occurred when health information is exchanged between health care organizations. The system should be able to authorize granting rights, including the granting of access based on access rights, either in a personal capacity or conferred on the basis of a role held by the user. They proposed an approach to allow the patient to hide direct access to
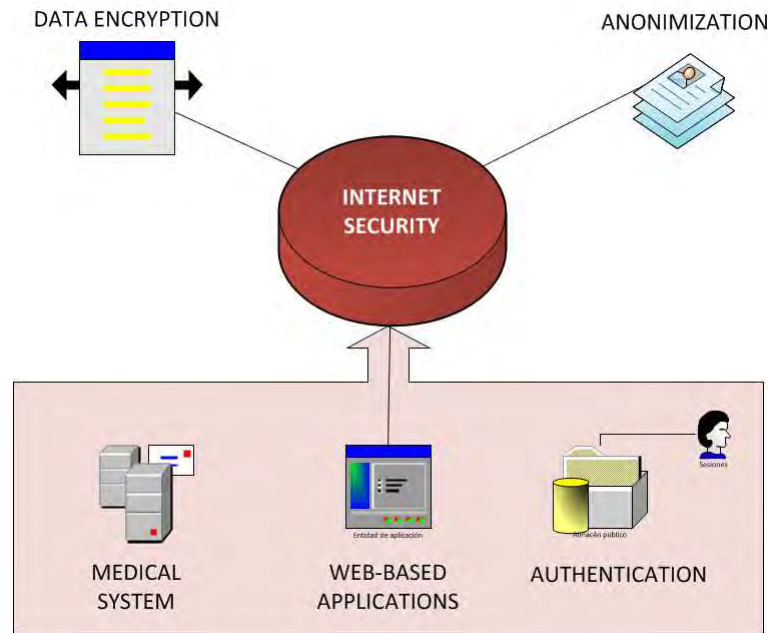
Figure 2.5: Security in web-based applications relates with the user authentication and the encryption and anonymization of the data.

his/her own information by using a mechanism outside the regular access control mechanism. Specifically, physicians need a patient key to access the information, thus requiring patient cooperation for decryption. The security and ethical considerations of EHR has also been discussed in several projects [95, 66]. Koutelakis and Lymperopoulos [97] developed a web portal platform with enhanced security related to DICOM applications. Java technology is mainly used to satisfy design specifications and dynamic data exchanges. Furthermore, Java applets have been developed and introduced in the whole project to serve specific demands. This allows better performance and sending the data using a stream encryption algorithm.

Choe and Yoo [31] proposed the use of a role-based access control system to make a security system to control medical data. Their security system enclosed an efficient repository management by providing methods for access-control, information exchange, user authentication, data integrity and selective encryption. It is important to note that the use of an encryption tree was used for the request data. The encryption tree was used to create the final data based on the healthcare data returned from the server. Finally, the encrypted data is transferred to the client using an attached digital signature. A similar idea was suggested by Sucurovic [177], who presented an authorization policy system based on XML schemas in his work to define the user's role in the system. The implementation of web service security between the clinical and central server and a secure transmission between the

central server and a client was previously defined. Furthermore, an extra security algorithm was chosen for signing and performing the encryption during the transmission of medical data. A similar approach was recently proposed by Hualei et al. [81], where the role-based access control was also applied to protect the medical data and the patients' privacy. Finally, Mahmoudi et al. [112] claimed that the use of a standard XML-based communication framework allows reducing the implementation overhead which made their communication well-defined. In addition, by using this combination, the security settings on the client machine were solved.

Regarding data anonimization, Zhang et al. [201] proposed a security model based on a multi-purpose internet mail extension message. The aim was to transfer medical data between servers and clients. The browser sent the request to the server and decoded the XML formatted messages. Subsequently, it looks into its local database and returns the results including patient and study information. This strategy allowed decoding anonymized data information provided by the DICOM files. Differently, Neubauer et al. [130] proposed a methodology for the anonymization of medical data that stores health data decoupled from the corresponding patient-identifying information, thus allowing privacy-preserving secondary use of the health records in clinical studies without additional anonymization steps. This approach allowed the unlinked storage of medical data and related patient-identifying information, while still making it possible to re-establish the patient/health data relationship for authorized users. Notice that anonymized data can allow protection of confidential information from unwanted connections [168].

2.3   Discussion

This chapter presented a summary of the current status and trends in web-based medical applications linked to PACS databases, the use of XML protocols and security systems. The main observation is the fact that no format currently exists that seeks to satisfy the needs of all potential users. Therefore, as researchers continue to find gaps in current web-based medical informatics applications, they propose their own solutions. A list of web-based medical application research projects have been presented in this review. It has brought together a diverse group of projects drawn from imaging, computer science, informatics and clinical disciplines. In general, the majority of web-based applications have used their own web server with a combination of Java applet tools. Basically, the inclusion of web technologies in web-based medical applications has been a solution to aid in the diagnosis and visualization of medical images. In addition to this, web-based medical applications based on CBIR systems have been found in the literature.

Web-based applications are directly associated with the use of databases, XML

protocol and security systems. All these aspects have been and still are the most common research areas in the field of computer vision over the last 10 years. In spite of their acknowledged importance, several shortcomings in current approaches have prevented their widespread acceptance into medical research, practice and education. In fact, the use of web-based medical applications has became a crucial solution in the medical field. The availability of large and steadily growing quantities of visual and multimedia data and the development of the internet underlined the need to create thematic access methods that offer more than simple text-based queries or requests based on matching exact database fields. In the same way, medical decision support tools are a good complement to web-based medical applications because they use the same databases as PACS systems or XML protocols to store and transfer medical data.

The PACS system is the most common medical database used in medical applications (see Table 2.1 and Section 2.2.1). The inclusion of the DICOM format as the standard medical image protocol was the key to the use of PACS. The support of CBIR, CAD, and HIS / RIS architectures have demonstrated the power of PACS as a common image data model. Progress in the related areas of PACS and image processing will certainly impact strongly on medical imaging. The simultaneous management of several image modalities is also present in all the research projects. Indeed, we can find the appropriate tools to manage these different image modalities, the standard DICOM being the most commonly used. However, the main issues in imaging informatics are indexing and context-specific organization of imaging data and the increment of large DICOM repositories provided by the acquisition systems. Thus, some research projects use parallel architectures as a solution(e.g. mini-pacs, GridPacs, etc.) to improve the management of medical data in an efficient way [97, 72, 61].

It is worth to mention that the cloud computing is emerging as a solution to the challenge of delivering complex services and data interchange over the Internet [90, 179, 140, 158]. Healthcare systems are trying to evolve and benefit from cloud services. This is because information technologies and information-rich services such as medical imaging can be greatly enhanced by use of cloud technologies. While the number of medical imaging studies is increasing approximately 3% - 5% annually, this rate is not a significant driver of growth. However, the size of medical imaging studies, especially CT-scan and MRI, is growing considerably faster, increasing storage requirements from 10% to 25% annually [90]. Cloud storage prices have been dropping faster than enterprise storage prices, and this trend will likely lead to faster cloud adoption for medical image storage.

The benefit of using cloud system demonstrates that it is possible to associate PACS systems with cloud's architecture providing data privacy, integrity, and availability [172]. The characteristics of PACS cloud ensure scalability and reliability of

medical data and can have a significant impact on healthcare institutions. Integration within a healthcare institution is effortless and inexpensive since the solution is fully compatible with DICOM devices. Moreover, another key benefit is that those repositories are vendor-neutral, so aging and PACS upgrading is not an issue. Moreover, this new paradigm maximizes PACS efficiency because the solution can serve multiple sites, facilitating inter-institutional data share and common workflow.

However, there are some disadvantages using cloud services. The use of this technology requires a constant Internet connection. In order for users to connect remotely to documents or applications stored in the cloud, users need to first access the Internet. Moreover, the internet connection needs to be a strong speed connection. For example, if the user is downloading a larger document from a web-based application, it requires a large bandwidth for the connection [3]. Moreover, data stored in the cloud can be unsafe. If the cloud was to go down and the user does not have a backup of their document, users will simply lose the document. It is also hard to trust cloud computing to store confidential document, in that it is possible for the cloud systems be hacked and accessed by unauthorized people [121].

On the other hand, the exchange of information has benefited from the inclusion of the XML protocol, a language that is easily accessible and extensible. Moreover, XML is capable of supporting a rich set of features as validation, indexing, query, retrieval and visualization methods developed using open software and standards (see Table 2.1 and Section 2.2.2). The relatively quick and easy implementation of XML to transfer information is a major asset. In order to improve the performance of transferring the information such as a messaging system, various strategies have been implemented. It is important to note how the XML files can be stored and related to DICOM images. Hualei et al. [81] proposed the use of an open source package to manage XML files. In addition to this, in their research work, a basic attribute and a semantic attribute of a DICOM image are extracted and stored in an XML file (one XML file per DICOM image). Thus, the future seems to be the use of XML and DICOM protocols linked with a web-based medical application.

Finally, the necessity of creating security systems in web-based medical applications in order to protect the privacy of medical data have grown over the last decade (see Table 2.1 and Section 2.2.3). The main goal is to create security architectures designed mainly to provide authentication and authorization services in web-based distributed systems. Most of the recent contributions focus on authentication and data encryption, but few projects focus on the anonymization of the medical data [201, 130, 168]. It is evident that most of the anonymization systems are applied directly to the DICOM protocol after the acquisition process. However, the use of anonymized information in web-based medical applications is not present in recent projects and this aspect could be improved in future work. Moreover, a new architecture system is proposed covering all the lacks appeared in the recent

research works according to the parameters evaluated in Table 2.1. Our purpose is to develop a new web-based application using all the main evaluated features. However, as a future work could update the actual evaluation adding new features such as cloud management, backup repositories or security systems within others in order to improve our proposed work.

2.4 Conclusion

Chapter 2 presents a summary of the most relevant research projects in computational medicine applied to digital databases and digital imaging. It has dealt with the current status and trends in web-based medical applications linked to PACS and databases, the use of XML protocol and security systems in the medical field. The aim of this literature review is to address several of these issues and to convey to the reader that it is a growing field which needs future improvements in information storage, exchange and security.

The lack of a common framework for web-based medical applications related to a medical databases using security protocols to safe-guard the privacy of patient data are studied in this review. However, the inclusion of XML documents used for a clinical diagnosis remains unaddressed, specifically due to the absence of a public image dataset despite its being highly demanded by the scientific community. It demonstrates the necessity of creating a new system architecture to cover the relationship between image processing tools and medical databases with the integration of web-based medical applications.

Chapter 3

System architecture framework

A system architecture framework to cover the relationship between image processing tools and medical databases with the integration of web-based medical applications is proposed. An introduction to the background of our purpose, concepts used to design the proposed architecture, a proposed view model, the advantages and drawbacks of the proposed system architecture framework compared to the others are presented in this chapter.

> Architecture is not an inspirational business. It is a rational procedure to do sensible and hopefully beautiful things.
>
> Harry Seidler

3.1 Introduction

Chapter 2 presented a number of research projects published in the literature over the last decade using web-based applications in medical informatics. However, there are still major problems to solve. The main problem is accessing and sharing information between different centers, and the interaction between hospitals and experts to assist in medical diagnosis. In this sense, the goal of our project is to develop a web-application system using a digital database shared by several hospitals or medical centers.

Development of a database of digital medical images with a high number of representative diagnosed cases provided by radiologists and experts is built in our work. It is essential for the evaluation of aided systems in cancer diagnosis. The aim

of this database of already diagnosed cases is twofold: firstly, one goal is to offer easy accessibility to medical diagnosis using this system architecture. In this sense, the user can access from internal or external connections offering efficient mechanism to find information. Also, it provides collaboration between experts from different hospitals and medical centers offering a high level of interactivity among health professionals. Secondly, the proposed system architecture is related to diagnosing cases for training, formation and enhancing the performance of novice radiologists. It serves as a training platform to train resident and intern radiologists as well as a medical support decision system which retrieves cases similar to the one being analyzed. Development and publication of a database of digital medical images with a high number of representative cases is essential because the purpose is to become a worldwide flag database in the scientific community for the evaluation of aided systems in cancer diagnosis [115]. We present our developed system architecture in this section as a solution to all these requirements.

3.2 System architecture framework

3.2.1 A view model of architecture

A view model of the architecture is designed surrounding a web-based application in a web server which is also used as a database server for data storage. The application links both the Picture Archiving and Communication System (PACS) and the XML servers which store clinical cases as DICOM images and annotation files. All data transmissions between users and the web server are encrypted in order to ensure complete confidentiality of the data. Currently, the Secure Socket Layer (SSL) cryptographic protocol is used. Queries to the web-application are performed through a web form and provide the user with a list of clinical cases according to the query parameters, including the image and annotation files.

Figure 3.1 shows the system architecture model. It is important to emphasize that the application is stored in a web-server which contains a database manager. In addition, once the PACS is installed, a database is also created and therefore directly related to the server. Thus, when a new examination is added to the PACS server, the data are stored inside the server but the information is saved inside the database.

The main idea is to develop a system architecture also related to the XML server. It incorporates access and management via a web interface. We opted for the creation of XML annotation files in the database in order to preserve the integrity of the original DICOM files and to ensure their full support if they come from different acquisition systems. All XML files stored in an XML server database
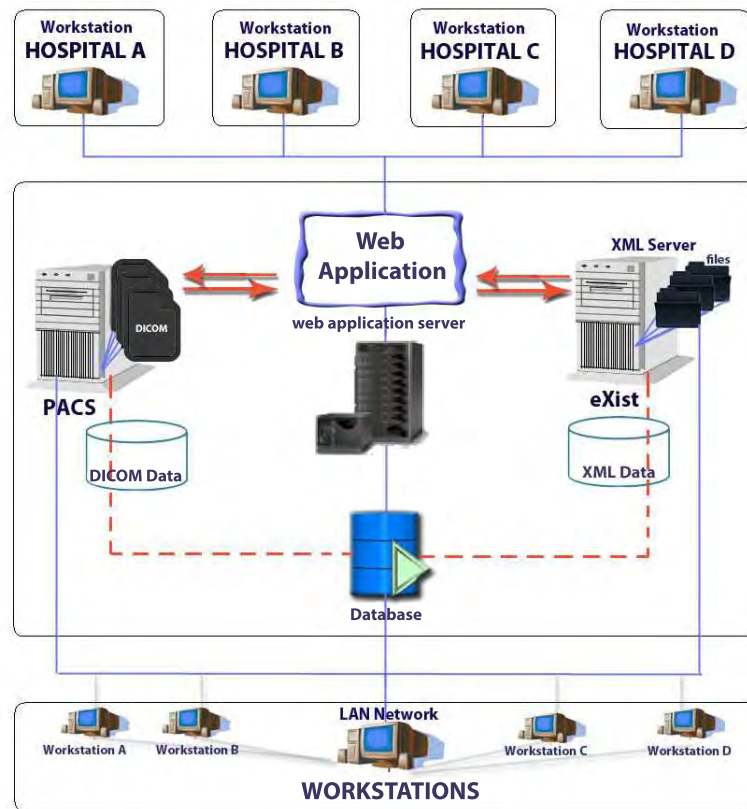
Figure 3.1: The scheme of the developed system architecture model.

also relate to their corresponding medical images. In the same way, these images are related to the files containing the records of diagnosis made by the radiologists. For this reason, as can be seen in the architecture model in Figure 3.1, that there is a direct relationship between the database and the XML server. Finally, we have defined the key field identifier or files with their examinations so the database and the files are correctly matched.

In order to facilitate access to the web-based application, and with the idea of serving both the medical experts present at the hospital itself and any outside professionals, we decided to design a web application. One of the key elements in this application is the data and information security across the network, especially when it comes to confidential data related to medical examinations. In this sense, the three modules shown in Figure 3.1 that constitute our system's architecture (web-server, PACS and XML server) have internal security measures to avoid vulnerabilities during the transfer of information. Therefore, besides using a control User ID private key, the algorithms have systems that protect and encrypt the information during its transmission.
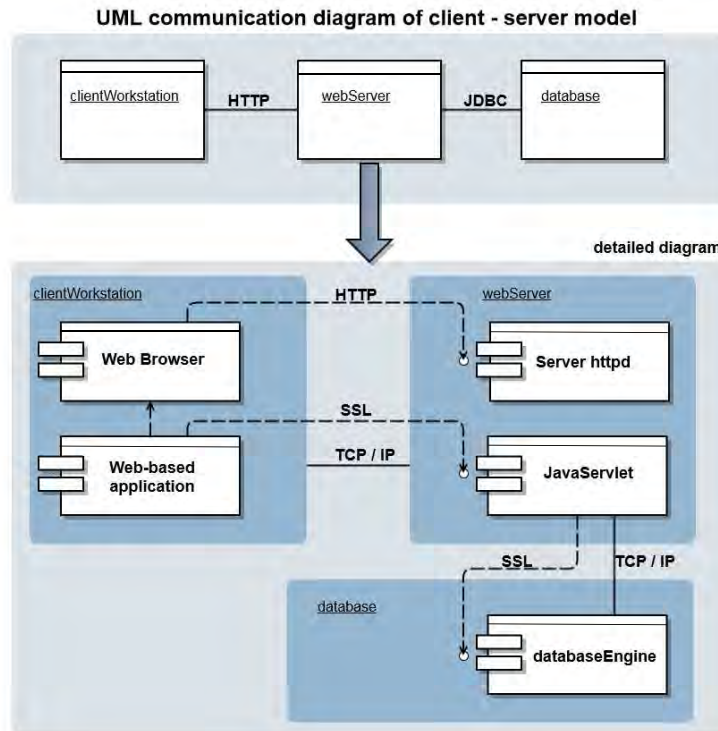
Figure 3.2: UML communication diagram of the proposed system architecture framework.

A client-server model describing the cooperation between the proposed architecture server with web-based applications as a client is used. The server component provides a function or service to one or more workstation (clients), which initiate requests for such services. In fact, clients and servers often communicate over a computer network on separate hardware, but both client and server may reside in the same system. Normally, a server can offer two different services: firstly, as a web server which serves web pages, and secondly, as a file server serving computer files. A shared resource may be any of the server computer's software and electronic components, from programs and data to processors and storage devices. In this sense, our proposed system architecture framework offers both modalities serving the interconnection between a web-server and a file server applied to the medical field.

Figure 3.2 shows the Unified Modeling Language (UML) communication modes employed for the proposed system architecture framework. This UML diagram is based on the client-server model of computing. Clients and servers often communicate over a computer network on separate hardware, but both client and server may reside in the same system. In our case, clients or workstations can reside in an in-

ternal or external way. The methodology of communication between client-server is provided by a host server which runs one or more server programs which share their resources with workstations. Indeed, a client does not share any of its resources, but requests a server's content or service function. Therefore, clients initiate communication sessions with servers which await incoming requests. A client workstation requests that the server connect a web-based application via the SSL cryptographic protocol. Moreover, the server contains a Java applet contained in the web-based application that also connects to the database via the SSL. These connections are provided via TCP/IP.

It might be possible that whether a computer is a client, a server, or both, it is determined by the nature of the application that requires the service functions. Our architecture offers the possibility of a single computer running the web server and file server software at the same time to serve different data to clients making different requests. Then, workstations can also communicate with the server software in the same computer.

3.3 System architecture implementation

The system architecture described in the previous section and illustrated in Figure 3.1 has been implemented in three independent blocks linked by a web-application. The purpose is to store, organize and manage large volumes of data. In this sense, three open source tools were used:

- Block I: Web-based application server.

- Block II: Picture Archiving and Communication System (PACS).

- Block III: XML server.

3.3.1 Block I: Web-based application server.

A web-based application server is implemented using a framework package namely the Zend Server Community Edition (CE)[1]. CE offers an administrator management to install and configure a server (Zend server), and also includes a MySQL Community Server (MySQL) database[2]. They can be installed in different platforms such as Windows, Linux or Mac OS X. In our case, they were installed and

---

[1]Zend Server. URL: http://www.zend.com
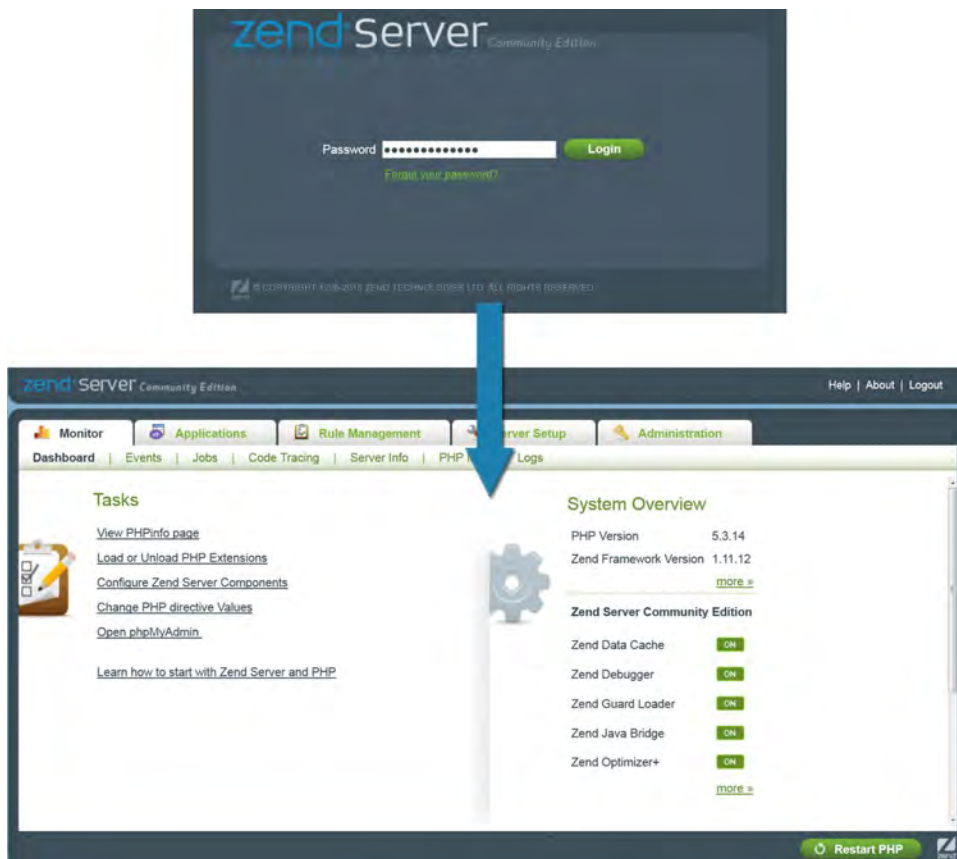[2]MySQL server. URL: http://www.mysql.com

Figure 3.3: ZEND server authentication and internal management.

configured for Windows 7. One important feature is that CE is a free tool pro-
vided for easy installation. Moreover, three important modules are included with
the Zend Server (ZEND) pack installation: ZEND server Apache, language-oriented
Hypertext Pre-Processor (PHP) web applications and MySQL database.

The reason for using CE over other managers is that it provides multiple capabili-
ties to improve the application's response-time and minimize the resource utilization.
Also, it incorporates different settings to increase the performance without changes
in PHP applications. Finally, CE also offers the management of extensions, upgrades
and improvements from its official website's totally open source. Figure 3.3 shows an
example of the authentication panel to access in the administrator management of
the ZEND server. In the management panel, the administrator user can configures
the server according to the needs and features of his web-based applications. Con-
cretely, the system's administrators can configure libraries and add implementation
plugins concerned with the Hypertext Pre-Processor (PHP) configuration. PHP is
a popular general-purpose scripting language especially suited to web development.

Figure 3.4: PHP specifications installed in a ZEND server.

It is an indispensable feature which should be installed in the ZEND server. The PHP module can be checked directly using the server and displaying the installed specifications (see Figure 3.4).

Concerning the installation of the PHP configuration, notice that an important plugin is required in order to execute Java applet tools in web-based applications.

Zend Java Bridge

JavaBridge Server	Enabled
JavaBridge extension version	3.1.21
JavaBridge Server Port	10001
JavaBridge Ints are Longs	No
JavaBridge Encoding	UTF-8
JavaBridge Use Java Objects	No

Figure 3.5: Configuration of a ZEND Java Bridge installed to ZEND server.

This plugin is "Zend Java Bridge v3.1, Copyright (c) 2004-2010, by Zend Technologies". It should be installed and enabled in the ZEND server. Figure 3.5 shows the configuration of this package into the ZEND server. Furthermore, two more requirements were also installed and configured in the server: a Java Development Kit (JDK) and a Java Runtime Environment (JRE). All these packages are available as an open source in the official Oracle webpage[4].

Finally, access to MySQL is provided by phpMyAdmin "Bringing MySQL to the web" (phpMyAdmin)[3]. phpMyAdmin is a free software tool written in PHP, intended to handle the administration of MySQL over the Web. Also, phpMyAdmin supports a wide range of operations on MySQL [36]. Frequently used operations (managing databases, tables, columns, relations, indexes, users, permissions, etc) can be performed via the user interface, while you still have the ability to directly execute any SQL statement. Figure 3.6 shows an example of the PHP specifications installed in a ZEND server.

3.3.2   Block II: PACS

The main features and usability of a Picture Archiving and Communication System (PACS) were presented in Chapter 1. PACS is a combination of hardware and software dedicated to short and long term storage, retrieval, management, distribution, and presentation of images. A PACS server environment, Dcm4chee project (Dcm4chee), installed in our system architecture is presented. The Dcm4chee project is associated with PACS. It is defined as a collection of open source applications and is used for healthcare information. These applications have been developed in Java programming language for performance and portability, supporting deployment on JDK. Moreover, the Dcm4chee project is equipped with a robust implementation of the DICOM standard. The toolkit Dcm4chee-1.x DICOM is used in many production applications worldwide, while the current (2.x) version of the guide has been redesigned for high performance flexibility. In our case, we used the

---

[4]Oracle Java Platform and Standard Edition. URL: http://www.oracle.com
[3]PhpMyAdmin. URL: http://www.phpmyadmin.net
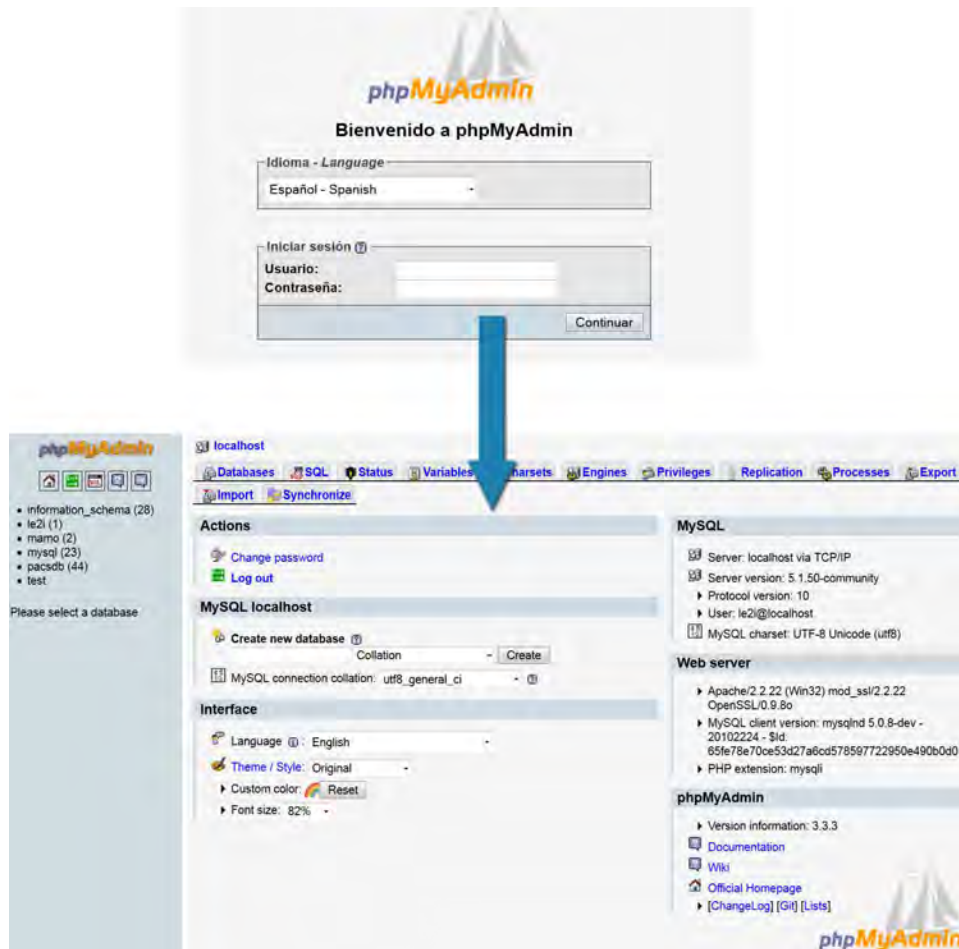
Figure 3.6: phpMyAdmin server authentication and internal management.

2.0.20 Dcm4cheeversion.

The application contains interfaces and services required to provide storage, retrieval and workflow in a healthcare environment. Dcm4chee uses a JBoss Application Server (JBoss) and therefore acts as a server for the PACS environment. JBoss is a Java application server implemented in pure Java open source that can be used with any operating system that supports it. Part *a*) of Figure 3.7 shows an example of the JBoss web server administrator. It contains the database belonging to the collection of DICOM images. Users can make custom queries and manage the database easily thanks to the wide range of features using a manager offered by Dcm4chee.

According to the Dcm4chee guide[5], the folder window can be divided into differ-

---

[5]Dcm4chee Open Source Image Manager: `http://www.dcm4che.org/`
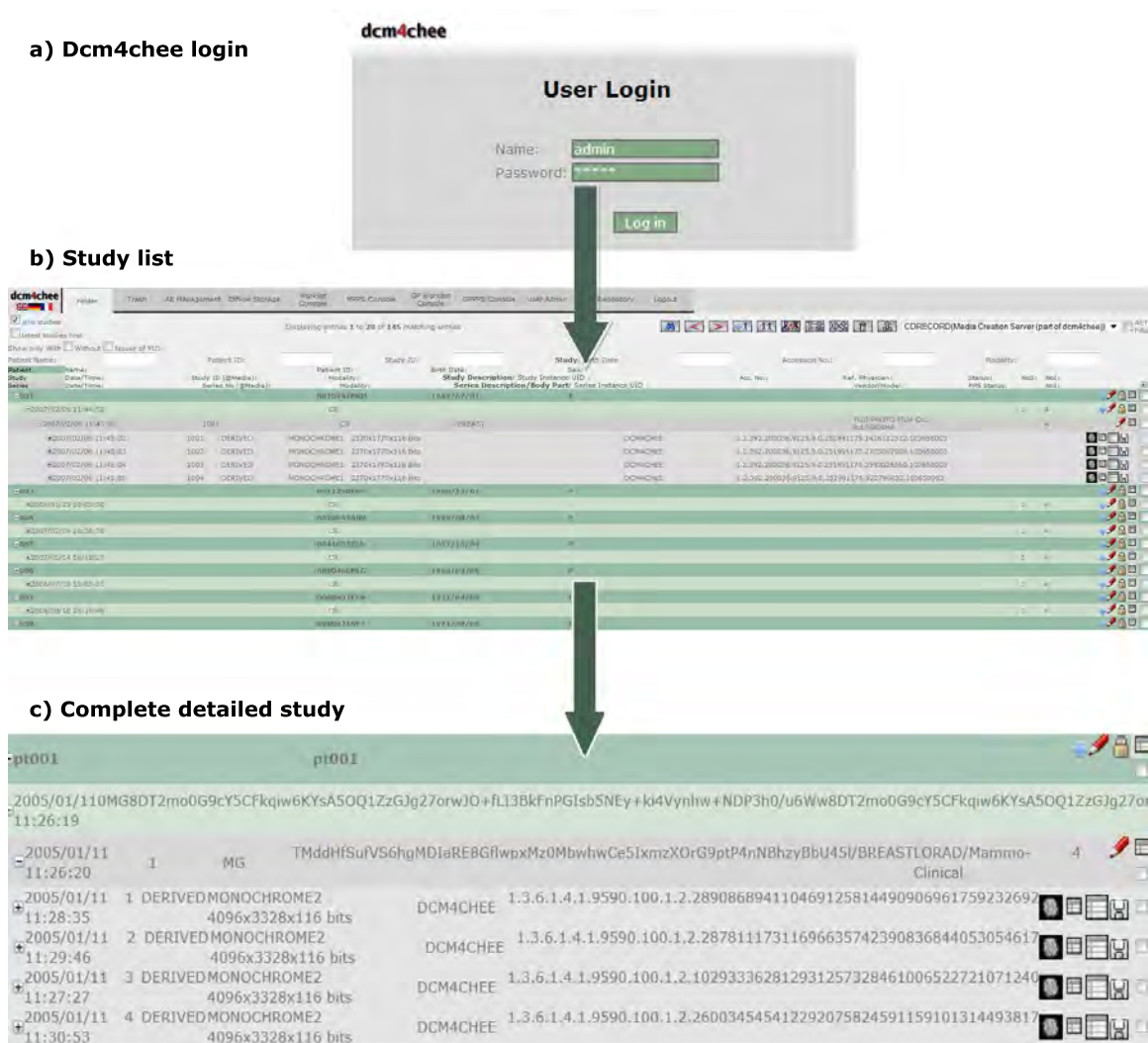
Figure 3.7: PACS server authentication and internal management.

ent logical parts. Among all the logical parts of Dcm4chee, our work is focused on two logical parts: Study list and Search filter. According to the system architecture features, which will be presented in Section 3.4, Study list and Search filter are two of the more important parts because they are essential requirements for the management of DICOM images. A summary of the most relevant details of both parts is shown in following subsections.

Study list

The study list contains information regarding patients, studies, series, and images. An example of a study list contained in the Dcm4chee server is shown in Figure 3.7 (see part *b*)). Each row represents the complete study of a patient's examination containing the following information depicted in part *c*) of the Figure 3.7.

Patient: The first row contains information about the patient (displayed in a green line). Labels for the columns can be seen at the top of each column. The information displayed at the Patient level is Name, Patient ID, Birthdate, and Sex.

Study: The second row contains more specific medical information about the patient (displayed in a light green line). The information displayed at the Study level is Date/Time, Study ID, Modalities, Study Description/IUID Instance Unique Identifier (IUID), Accession Number, Referring Physician, Number of Series, and Number of Instances.

Series: The third row contains series information about the patient (displayed in a grey line). The information displayed at the Series level is Data/Time, Series Number, Modality, Series Description/Body Part/IUID (Series IUID), Vendor/Model, and Number of Images.

Search filter

The search filter specifies matching criteria for listed studies. To filter the records in the study list, search criteria can be specified above the column labels. It is possible to search for Patient Name, Patient ID, Study ID, Study Data, Accession Number, Modality, or a combination thereof. To perform a search just press "Return" on the keyboard or click the New Search button. The result of the search will be displayed in the list.

Figure 3.8 shows an example of the search filter. It can include a specific entry or part of an entry truncated with a question mark (?) or an asterisk (∗). A search with a question mark will only search for one character wherever it is placed, whereas a search with an asterisk will include an unlimited amount of characters. It is possible to combine question marks and an asterisk in the same search.

3.3.3 Block III: XML server

The eXist-db XML Server (eXist-db) is an open source environment based on XML technology, also the called native XML database. Unlike most management systems
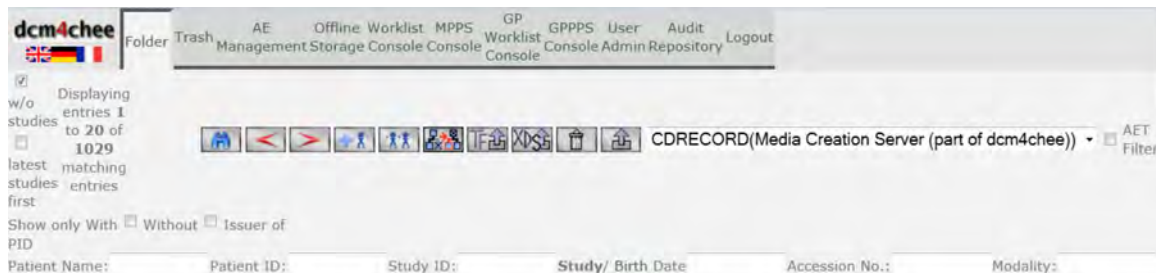
Figure 3.8: Search filter panel of the Dcm4chee server.

of relational databases, eXist-db is provided to XML Query Language (XQuery) statements in order to store and manipulate XML data.

An example of an eXist-db interface is shown in Figure 3.9. As can be seen, three different parts are shown. Firstly, the login access panel of eXist-db is depicted (see part *a*) of the example). With this interface, users can access according to their permissions. In fact, only administrators can access the control panel. After the user is logged, a list of the database collections in the eXist-db server appear. In the left part of Figure 3.9, a menu offers different options of maintenance of our server and database. The administrator can manage the acxml database internally in order to store and delete XML files. Finally, every collection contains a list of XML files (see part *c*) of Figure 3.9.

XQuery operates using a specific logical structure of XML documents rather than its surface syntax[6]. It is designed to be a language in which queries are concise and easily understood. It is also flexible enough to query a broad spectrum of XML information sources, including both databases and documents. Using different formats of queries, such as SQL queries used in databases, a list of XML files with the content introduced in the XQuery expression is obtained. These contexts for a given XQuery expression consist of all the information that can affect the result of the expression (XML Path Language (XPath)). In this sense, XPath is the result of an effort to provide a common syntax and semantics for shared functionality[7]. This syntax is depicted in Figure 3.10 showing a subtree of an XML document[8]. Here, the central node is the context node, and the arrows represent axes referring to the other nodes. In this example, the parent axis refers to the parent node of the context node.

Figure 3.11 shows more details of the axes. Note that we still omit some axes

---

[6]XQuery 1.0: An XML Query Language (Second Edition). URL: http://www.w3.org/TR/xquery

[7]XML Path Language (XPath) Version 1.0. W3C Recommendation. URL: http://www.w3.org/TR/xpath

[8]XML Path Language (JavaNB). URL: http://book.javanb.com

Figure 3.9: eXist-db server authentication and internal management.

defined by the XPath specification. There is an ordering in XPath called document order. The document order is the order in which the start tag of each node appears in the literal XML document. The axes are defined based on the document's order.

Examples of XML structures will be explained in detail in chapters 4 and 5. Each structure is different but always follows a typical XPath model as can be seen in Figure 3.11. This requirement is very important because the access of the XML files content should be structured, and also used for our related web-based applications.
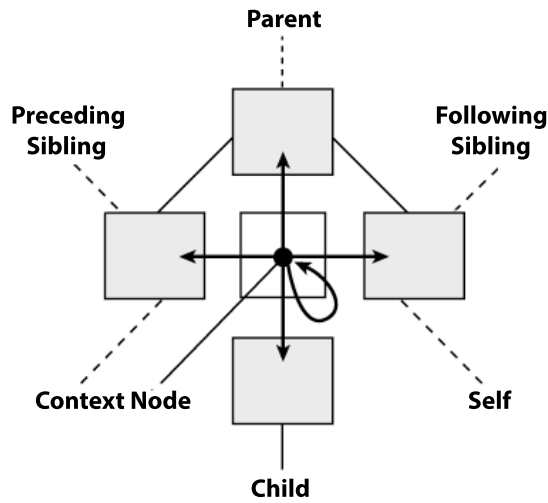
Figure 3.10: Internal structure of a context node in XPath[8]
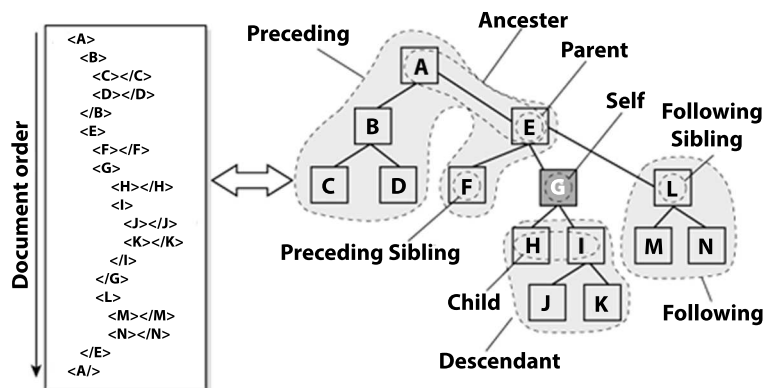
Figure 3.11: Typical XPath[8]

3.4   System architecture features

A system architecture is designed as a web-based application in the ZEND server which is also used as a MySQL database server for data storage. The application links both the PACS and the XML server which store the clinical cases as DICOM and annotation files. Database queries are performed through a web form and provide the user with a list of clinical cases according to the query parameters, including both the image and the annotation files.

In chapter 2 a list of research projects related with web-based medical applications was presented. Table 2.1 showed a detailed classified list of projects according to different features. As a general review, most of the common projects focused on PACS databases. One of the main reasons is due to the fact that PACS databases

are used in many hospitals and health care centers. Also, most of them use different and independent applications to save the medical findings. In this sense, the data exchange used still has problems in centralizing this information in a correct way. It means that shared medical diagnosis associated with DICOM images stored in a PACS database and able to connect with internal applications to get the data still presents limitations. For this reason, a solution to link PACS and data exchange files using web-based applications is presented in this chapter. In this way, experts can use a web-based application to control, manage, visualize medical examinations and build a database in a collaborative way. Furthermore, the proposed system architecture is implemented using standard frameworks which could be adapted to the medical field.

The advantages of our proposed architecture are that it makes collaborative work sharing possible and aids experts in medical diagnosis. Thanks to the introduction of the XML file format, one can exchange information regardless of the platform, thus, introducing a standard format to include information (such as aiding the diagnosis) linked to the images. Hence, it is a new way of using clinical material to help radiologists to learn, and permits the medical expert to use facilitating and supporting skills rather than didactic, directive ones. This paradigm emphasizes the application of knowledge and skills to the solution of problems instead of just recalling the facts. Therefore, given that the training process for radiologists should include a large number of cases with multiple benign and malignant findings, using this system architecture allows testing and training a radiologist's ability to:

- Detect medical findings.

- Properly characterize medical findings according to their number, location, shape, size, margins, density, and distribution.

- Properly characterize medical findings according to their pathological classification depending on the image modality.

Technology in health allows one to record, measure, monitor, manage and provide guidance to the patient, together with the opportunity to offer care services through the Internet in real time. Health systems have enabled one to maximize the collection, delivery and communication of medical information, clinical messaging, and interaction with doctors and medical records [123] [115]. In this context, this system architecture model is intended for medical experts, assistant radiologists, and research users who need the following set of skills:

- Provide management tools to recognize when there is the need for an information request;

- Provide of an accessible way to connect from local or external connections.

- Evaluate and organize information, and have the ability to present medical examinations.

- Provide a consistent security model using a standard environment.

Moreover, the role of the doctor is very important because the user must spend time in teaching users to use the application. In the same way, students must be receptive and willing to learn. If this effort is not mutual the educational action will fail. This system architecture offers the possibility of developing web-applications adapted to medical matters. Accordingly, some advantages are given below:

- Provides radiologists with a wide range of information.

- Facilitates information and content updating.

- Promotes "just in time just for me" training.

- Provides training and collaborative grouping.

- Enhances interactivity with information, with the teacher and among medical experts.

- This project involves the collaboration of different hospitals with the same goal: learning and support in the diagnosis.

It has been presented the advantages and the main features about the proposed architecture system. However, the architecture is also tested in order to determine whether there were common patterns and issues that could be addressed to avoid scalability problems. Scalability is the ability to handle increased workload. In this sense our architecture has been developed as a combination of computing hardware able to providing services to a web-based application in the same or other computers. The proposed architecture is a project making collaboration between two centres as a starting point of this project. At the moment, a good performance on a large volume was evidence of scalability. Also, this system is prepared to add more hospitals or medical centres on the future. However, scalability failures could be occurred when the demand is increased causing some resource to become overloaded or exhausted. Common scalability problems occur when the memory is overloaded, available address space is exceeded or network bandwidth is exceeded. In this sense, the increased number of users may lead to additional database entries causing the system to remain too slow. The terms of efficiency and the response time will be presented in next chapters.

The next chapters will explain in detail two different ways to develop a web-based applications using the proposed system architecture. The aim is to demonstrate the capabilities, advantages and facilities to implement a web-based application in different medical areas and modalities using the proposed architecture framework model. In Chapter 4 the proposed framework is applied to a mammographic environment to aid the diagnosis of breast cancer, while in Chapter 5 it is used along with MRI to aid the diagnosis of prostate cancer. These two modalities have been chosen due to offer the possibility of collaborative work between different universities and hospitals.

Chapter 4

MamoDB: Web-based medical application for the management of breast cancer in digital mammography

In this Chapter, we present an interactive Java applet interface designed as a web-based tool applied to mammographic studies. This Java applet, Mamo Java applet tool (MamoApplet)*, 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 suspicious regions. Each annotation includes specific attributes considered by the experts to help the deteccion of a breast cancer. The overall set of overlays is stored in a database as XML files associated with the original images. The final goal is to implement a new tool based on a web application to aid experts in mammographic diagnosis. Also, it aims to facilitate a collaborative work for training and enhancing the performance of novice radiologists. Part of this chapter is an adaptation of a paper published in the conference IEEE BIBE[1]*

> The better you were able to imagine what you wanted to imagine, the farther you could flee from reality.
>
> Haruki Muramaki

---

[1][BIBE] MammoApplet: an interactive Java applet tool for manual annotation in medical imaging, IEEE International Conference on BioInformatics and BioEngineering, 2012 [116].

4.1 Background

4.1.1 Digital mammography

Mammography is and remains the preferred screening technique for early detection and diagnosis of breast cancer [19]. It is estimated that a 15 to 35% reduction in mortality in breast cancer is due to the wide implementation of screening mammography [25]. In this sense, a web-based application (MamoDB) to facilitate the analysis of breast cancer diagnoses is presented. This application offers the possibility of comparing old mammographic diagnoses with new ones by providing a set of image processing tools. Moreover, a set of drawing tools has been developed in order to annotate suspicious regions.

Until some years ago, mammography was typically performed with screen-film cassettes. Digital Mammography (DM) is the natural evaluation of screening the breast using X-rays and has become the image screening of reference when diagnosing breast cancer [125]. The two types of imaging currently used for mammography are[2]:

- **Screen-Film Mammography (SFM):** DM can either be digitized SFM when the image is obtained as the digitization of an analogical film. The image is commonly referred to as a mammogram.

- **Full-Field Digital Mammography (FFDM):** FFDM is when the image is generated directly by a digital sensor instead of a sensitive film.

Comparing FFDM and SFM one can see that[3]:

- The higher spatial resolution demands of mammography.

- Significant expense increase of equipment.

- DM equipment demonstrates that it is at least as good as screen-film mammography at detecting breast cancers without increasing the number of women recalled for further evaluation.

In that sense, FFDM is gaining importance compared to the conventional film-screen mammography of today due to the fact that digital acquisition, digital storage, and digital display processes may be separated and individually optimized [184].

---

[2]U.S. Food and Drug Administration. URL: `http://www.fda.gov/Radiation-EmittingProducts/`
[3]American College of Radiollogy (ACR). URL: `http://www.acr.org`

In Section 1.1.1 general information on X-ray acquisition was presented. In fact, digital detectors offer higher quantum efficiency and higher resolution than traditional screen-film receptors. This translates into both lower dosage and mammograms with an improved image quality. Digital detectors for mammography can be categorized as indirect or direct conversion detectors according to how the X-ray is captured. In indirect conversion methods (the earliest ones) the energy of the X-ray is captured by a scintillator which converts it to light. Subsequently, this light is captured by an array of thin-film diodes which convert it to an electronic signal which, in turn, is captured using thin-film transistors. The main problem of these systems is that the light created by the scintillator scatters into its neighbourhood, and the same X-ray is captured by different transistors. On the other hand, in direct conversion systems the same photoconductor that captures the X-ray is able to directly generate an electronic signal. Thus, scattering is less severe in these systems.

Mammography views

Standard views are those performed on routine screening mammograms consisting of in two different views of each breast: Medio-Lateral Oblique (MLO) view and Cranio-Caudal (CC) view.

- An MLO is taken from the side at an angle from a diagonally-compressed breast.

- A CC view is taken from above a horizontally-compressed breast.

Figure 4.1 shows both viewpoints: (a) physically shows the viewpoints' direction, while Figures 4.1(b) and 4.1(c) show an example of each mammogram view. It is important to notice that in MLO views there is one region corresponding to a portion of the pectoral muscle which may be present in the left or the right upper corner of the image. Moreover, some annotations and labels can appear in the images. For instance, the R symbol which appears in Figure 4.1 is used to mark the mammogram as *right* as opposed to L (left breast).

The MLO view is the standard mammographic view. It is the most important projection as depicts the most breast tissue. The representation of the pectoral muscle on the MLO view is a key component in assessing the adequacy of the patient's positioning and therefore, the adequacy of the film. The amount of visible pectoral muscle determines the amount of breast tissue included in the image, resulting in a decisive quality factor, very important to reduce the number of false negatives and increase the sensitivity of the mammography[4]. Furthermore, as most breast pathol-

---

[4]Radiopaedia. URL: `http://www.radiopaedia.org`

(a)   (b)   (c)
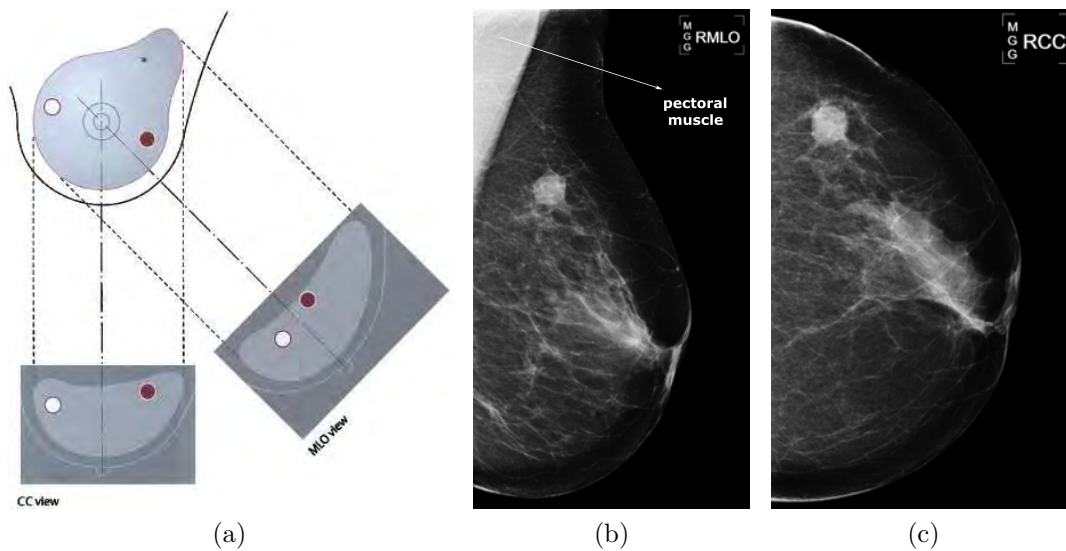
Figure 4.1: Two viewpoints for extracting mammograms MLO and CC. (a) Illustrates the projection of the two most used view points (image from [**Massich14**]), which produces images like the MLO in (b) and the CC in (c).

ogy occurs in the upper outer quadrant, it is important that this area is clearly visible in MLO views.

The CC view is another common projection used in a screening mammography. It must show the medial part as well the external lateral portion of the breast as much as possible[4]. A correctly performed CC projection may show the pectoral muscle at the posterior edge of the breast, indicating that the breast has been positioned as far forward as possible, although this is not always achievable. It will show as much as possible of the glandular tissue (ducts and lobes), the surrounding fatty tissue and the outermost edge of the chest wall muscle. The CC view cannot capture much of the breast tissue in the armpit or the upper chest.

Additional supplementary views could be taken for a diagnostic mammogram [15]:

- **lo!** (**lo!**): Taken from the outside towards the center.

- Medio-Lateral (ML): Taken from the center towards the outside.

- Spot Compression (SC): Compression on only a small area to get more detail.

- Cleavage view (CLV): Both breasts are compressed to see the tissue nearest the center of chest.

- Magnification (MGF): Taking of the borders of structures and calcifications.

4.1.2 BI-RADS classification

The American College of Radiology (ACR) defines guidelines for appropriate use of imaging technologies. In particular, ACR defined Breast Imaging-Reporting and Data System (BI-RADS) [6, 7] providing standard categories in order to classify the breast tissue. BI-RADS was initiated in the United States in the late 1980s to provide a standardized description of imaging features of breast lesions and to relate them to the underlying pathology and risk of malignancy [197]. It is now mandatory to include BI-RADS information in mammography reports. This classification is much used in many research projects (see [132] for a complete description). Following this reference we can classify breast tissue according to its density evaluation:

- BI-RADS I: the breast is almost entirely fatty.

- BI-RADS II: there is some fibroglandular tissue.

- BI-RADS III: the breast is heterogeneously dense.

- BI-RADS IV: the breast is extremely dense.

Figure 4.2 shows an example of each BI-RADS class: (a) corresponds to BI-RADS I, (b) BI-RADS II, (c) BI-RADS III and (d) BI-RADS IV.

Moreover, the most common use of the BI-RADS classification by radiologists in mammography , ultrasound and MRI is to identify anomalies and allow views of what should be done next or to return to testing, close monitoring or biopsy [43].

4.2 Materials and data acquisition

In this section we provide a brief description of the two mammographic databases used in this thesis: the Hospital Dr Josep Trueta (Girona, Spain) and UDIAT Diagnositc Centre of Parc Taulí (Sabadell, Spain) databases.

4.2.1 Hospital Trueta de Girona

The Radiologic Department of Hospital Universitary Dr. Josep Trueta is provided with a Siemens Mammomat Novation digital mammographic unit. This mammograph stores the images in a computerized way in the hospital's PACS system, thus resulting in a large digital mammographic database. The database is composed of 89 cases, where each case is composed of two views of the same breast (18 cases) or the complete case (71 cases), consisting of both Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC) views of both of the patient's breasts.

(a) BI-RADS I **(b) BI-RADS II**

(c) BI-RADS III **(d) BI-RADS IV**
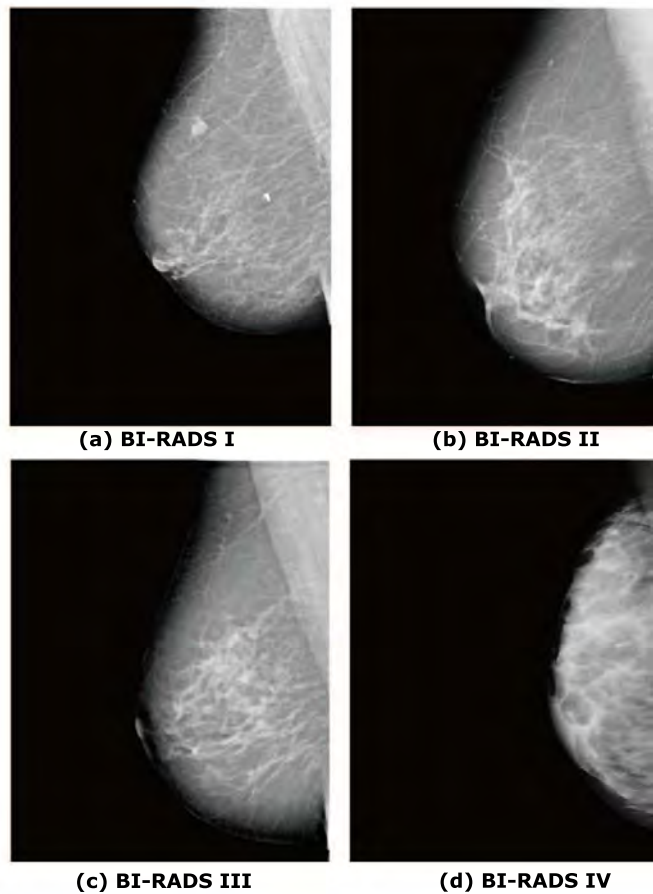
Figure 4.2: An example of 4 mammograms of increasing density. The BI-RADS categories per density allows to group then in four classes.

Image characteristics

Each image in the database has a 70 micron pixel edge and is stored using 12-bits according to the DICOM standard [10]. Two different image sizes are possible: 2560×3328 or 3328×4096 pixels. Although different formats of storing the images are possible using this standard, in this case, they are stored in a raw format: each number of the stored file corresponds to the grey-level value (from 0 to 4096) of their corresponding pixel in the image.

Each case has its own annotations, including the breast's density classified in BI-RADS categories, the number of lesions, and if a lesion is present, the location of its center of mass and the minimum radius to evolve the lesion. Moreover, in cases where a lesion is present, a binary image is stored, where pixels outside the

	B-I	B-II	B-III	B-IV	TOTAL
MLO	46	21	20	2	89
CC	45	20	20	2	87
Total	91	41	40	4	176

Table 4.1: Distribution of the Trueta Hospital database according to the BI-RADS density classification.

lesion are encoded with the value 0 and pixels belonging to the lesion (including the boundary) are encoded with the value 1.

Used mammograms containing masses

A subset of 89 MLO and 87 CC views mammograms containing masses were extracted from this database. Table 4.1 shows their distribution according to the BI-RADS breast density categories.

4.2.2 UDIAT Diagnositc Centre of Parc Taulí

Image characteristics

The dataset used for this work comes from the collaboration between the University of Girona and the UDIAT Diagnostic Centre of Parc Taulí in Sabadell where an image database from UDIAT is being collected in order to make it available to the researchers. The collected database is composed of a total of 332 images. These images were obtained from a Hologic Selenia Mammograph with a spatial ressolution of 70mm x 70mm, a size of 4096 x 3328, and a 12-bit depth.

The overall dataset is composed of DICOM formated images with anonymized metadata information. All the images have at least one lesion and a delineation of the lesion structures in each image provided by an experienced radiologist from UDIAT.

Used mammograms containing masses

This database is composed of a total of 231 normal cases and 101 with masses. Table 4.2 shows their distribution according to the BI-RADS breast density categories.

	B-I	B-II	B-III	B-IV	TOTAL
Normal	49	85	51	46	231
Masses	22	40	19	20	101
Total	71	125	70	66	332

Table 4.2: Distribution of the UDIAT Diagnositc Centre of Parc Taulí database according to the BI-RADS density classification.

4.3　MamoDB framework

The presented Web-based application based on a Digital Mammography (MamoDB) is a tool which has already been developed and published [115]. It is a web-based tool with the purpose of aiding to 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 of detecting mammographic findings according to their number, location, shape, size, margins, density, and distribution. A comprehensive list of clinical cases contained in the database was used. Also, using query forms allows one to perform searches in the database according to specific medical criteria: reported diagnosis, BI-RADS classification, and the presence of abnormalities.

Chapter 3 presented the generic system architecture framework. The proposed view model of the architecture was designed as a web-based application. In this sense, a MamoDB framework is developed as a web-based tool within a web server and related to an XML database server used to store data. For this framework, a digital database of DICOM images containing mammographic studies is also created. Hence, the database used to store them is a PACS system. MamoDB links both the PACS and the XML servers which store clinical cases as DICOM and annotation files. For this reason, the XML database server used to store annotation files containing mammographic findings is an eXist-db XML Server (eXist-db) which will be explained in the next sections of this chapter.

In summary, in this section we present the MamoDB as the first step of creating a comprehensive list of annotated clinical cases and store them in the database. MamoApplet is a web-based platform that allows drawing accurate overlays of mammographic findings, like masses or micro-calcifications lesions, and provide facilities to annotate other features related with lesions, like the subtlety of the lesions, or related to the breast's 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.

4.3.1 Objectives

The main technological features when developing MamoApplet 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, series and/or images).

- Implement a Java applet tool, MamoApplet in order to manage the most common steps in image processing and visualization (zoom, contrast, BI-RADS lexicon, brightness,...).

- Implement a BI-RADS lexicon report for each mammographic study.

- 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 annotation by different experts. It is important that more than one expert analyze the same image in order to obtain a robust evaluation.

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 structure depicted in Figure 4.3.

As noted by Dreyer et al. [42] and presented in Section 1.1.2, 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 which is then associated with one or multiple instance images. Therefore, MamoApplet uses this hierarchy in order to retrieve all the images of a patient, as can also be seen in the XML file structure. In the case of mammography, a series usually consists of 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 Medio-Lateral (ML) or Latero-Medial (LM) view, hence consisting of two additional images.

```
<?xml>
<root>
  <patient>
     Information related to the patient
     <study>
        Information related to the study
        Information related to the breast image
        <birads>Information related to the birads</birads>
        <viewposition>Type of view position</viewposition>
        <series>
          Information related to the series
          <list of overlays>
             <overlay 1>
                <label>User label</label>
                <user>Username</user>
                <listpoints>Coordinates points</listpoints>
                <centermass>Center mass coordinates</centermass>
                <area>Area of the overlay</area>
                <mean>Mean of the overlay</mean>
                <std>Standard deviation</std>
                <subtlety>Subtlety of the overlay</subtlety>
                <data>Data of the overlay</data>
             </overlay 1>
             ...
             <overlay N>
                Information of the overlay
             </overlay N>
          </list of overlays>
        </series>
     </study>
  </patient>
</root>
```

Figure 4.3: Example of the XML structure of a mammographic image.

4.3.2   MamoDB interface

Firstly, the secure access user to the MamoDB application is ensured by a username and password generated by the system administrator. The system incorporates mechanisms of authentication and protection against fraudulent use of identity.

Figure 4.4: Example of the MamoDB login access.

The system also has a three-level security setting for user profiles. The first level (Standard User) allows users to submit queries on the various medical cases stored in MamoDB (images and diagnosis). This level corresponds to a profile of a student or auxiliary medical personnel, who use the application for remote diagnostics, training, etc. The second security level, in addition to the permissions of the first level, allows access to PACS and eXist-db for addition and modification of the mammographic database content and clinical cases (images and diagnosis). Essentially, it corresponds to a profile of an advanced student or a medical specialist partner in the construction of the mammographic database who needs access to the system locally or remotely. The third security level serves for system administrators, adding user account management to the list of permissions.

Once a user is identified correctly in the system, the home page of MamoDB is displayed. Depending on the permission privileges assigned to each user, the control panel shows the different options to perform.

Figure 4.4 shows the control panel, which has been designed to take into account UI ergonomics and simplify access to the options of the system. This home page is designed as a web map; the user can access all the functions in the system in a single click. There are two blocks of access elements: a drop-down menu bar and an expandable sidebar, which the user can make invisible. The latter's role is to
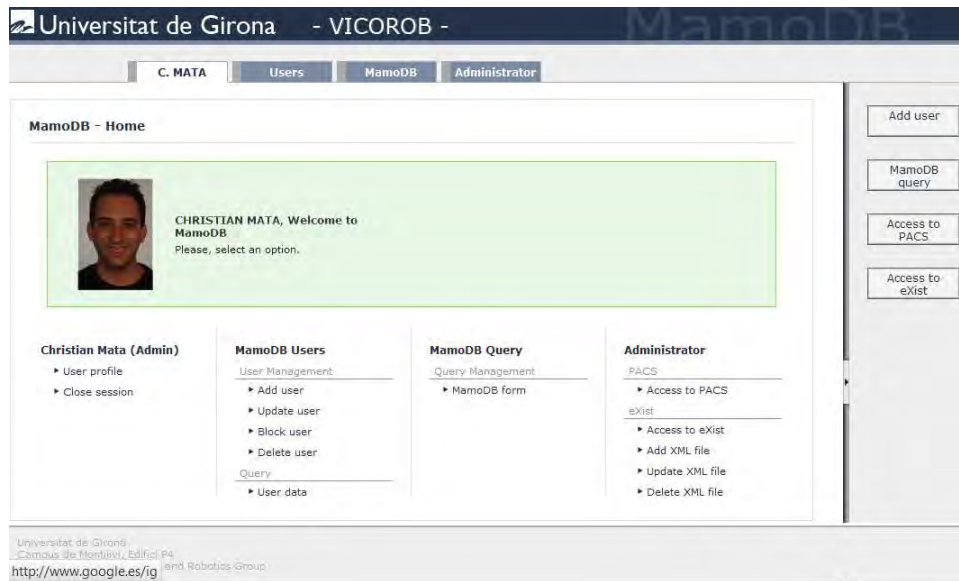
Figure 4.5: Example of the user profile form (level 3) of MamoDB

provide direct access to the most used features. The control panel also contains a personal information section where a user can edit such information as profile image, username and password.

Moreover, users have the ability to modify their profile data and their personal image. Using a query form, there is the option of changing personal information (profile) and information to access to MamoDB. Fields marked with an asterisk are required fields. Finally, the username and password can similarity be modified (see Figure 4.5).

4.3.3 Search engine

The most important part associated with the learning capabilities for training and guidance of users is shown in this section. Therefore, the management and training form, finding results, and characterization analysis of the mammographic findings are presented.

Management and training query form

The search engine uses queries to the MamoDB through a web form and provides a user with a list of clinical cases according to the query parameters, including both the image and annotation files. The first step required from the user for this training process is the use of the query MamoDB form. In order to obtain desirable results,

Figure 4.6: Example of a MamoDB query form.

the user should select appropriate criteria, as directed by the expert or medical staff. The aim of the search engine is to find mammographic studies according to the criteria used by experts. The search is performed in all the XML files stored in the eXist-db database and the list of results is shown in the web interface.

Figure 4.6 shows an example of a standard query form where a user can define his type of query (upper area) and select the information to be shown in the result (bottom area). Two different query options are defined: search all the mammographic examinations, or make an advanced search. In the example in Figure 4.6, a customized query search is performed. As can be seen, users can choose three different query types: searching for BI-RADS type, for study ID and for patient ID. Moreover, users can also select the information to show in the results list. This information is related with mammographic findings such as their number, location, shape, size, margins, density, or distribution.

Finding results

For a clinical use of the tool, the query forms allow one to search according to some specific medical criteria: reported diagnosis, BI-RADS classification, or presence of abnormalities (micro-calcifications, structural distortions, masses, etc.). Depending on the volume of records obtained from the results list, the application uses automatic paging group records in a more visual way as can be seen in Figure 4.7.

For every record obtained, the system allows four actions: see information on the DICOM image stored in the PACS server (see Figure 4.8), obtain the mammo-

Figure 4.7: Example of a MamoDB results list.

graphic information contained in the associated XML file (see Figure 4.9), display the mammographic image (see Figure 4.10), and download selected mammographic studies which will be explained in next subsection.

Characterization analysis of the mammographic findings

Three different actions are available once the records have been obtained:

- Display the information contained in the DICOM file (Figure 4.8).

- Get the diagnosis linked to the case, stored in an XML file (Figure 4.9).

- Display the digitized images contained in the DICOM file using MamoApplet (Figure 4.10).

When the list of results is obtained, the user can get information on all the studies and compare them in order to perform a classification and selection of the best results according to the preferences selected in the query form. MamoDB aids the diagnosis and the management of mammographic findings. Moreover, it offers an educational part as an e-learning tool for training, and is useful to novice radiologists. For example, for each case, a student can display acquisition information contained in the DICOM file, read annotations about the study included in the XML file, and display the images included in each study. Then, the student can compare his diagnosis with the corrected ones previously made by an expert.

Figure 4.8: Example of the information contained in the DICOM file.

Figure 4.8 shows an example of the DICOM image provided by PACS. Sometimes it is useful to know the details of the acquisition system machine in order to classify cases as well as patients. Notice that the user should have privileges to access the PACS server. So, it is another easy way to access the PACS.

On the other hand, in Figure 4.9 a user can obtain information from the XML files. As can be seen, this information is divided into three different parts. Firstly, the information related to the image file is shown (file name, patient, path, data acquisition and institution). In the second part the information related with the study is shown (Id study, name study, BI-RADS, view position, series number, description and number of overlays). In this example, the number of the contained overlays on the image is 2. For this reason, a third part shows information of the mass segmented areas by the doctor involved in this study (number of overlays, area, number of points, mass center, data and the username of the doctor). It is important to note that the information on the patient is anonymous, and only the patient ID is shown as it has no relation with the patient's identity. The main objective for users is to learn and to discuss the cases which will be useful for the doctors to do and support mammographic diagnosis.

Finally, in Figure 4.10 two examples of DICOM images associated with the selected study are displayed. In fact, it is an example of visualization of the Mamo Java applet tool (MamoApplet) which will be explained in next section. It also shows overlays marked by doctors and their individual annotations. This information is very important because users should compare and classify studies in order to download all the cases later.

Figure 4.9: Example of the information contained in the XML annotation file.

Downloading selected mammographic studies

The four options that MamoDB offers to the user is the possibility of downloading selected mammographic studies. An important feature in this system is that records can be visualized on screen or locally downloaded to the workplace selected by the user. This system offers the possibility of downloading a single record or using an automatic selection. Once selected, the DICOM files, together with their associated XML files, become part of the ZIP file downloaded by the user. As shown in Figure 4.11, a user can select the cases they want and start the downloading process immediately.

Searches on the web-application are performed through requests to the DICOM database using the DICOM protocol. The data obtained is then arranged according to the patient, studies, and series attributes, which later allows one to refine the search. Searches can be exhaustive as where the DICOM database returns a list of all patients who have undergone a study, or selective in which only a reduced subset of cases that fulfill the search condition is returned. The information concerning those studies is then related to the following DICOM attributes: studyID, manufacturer,
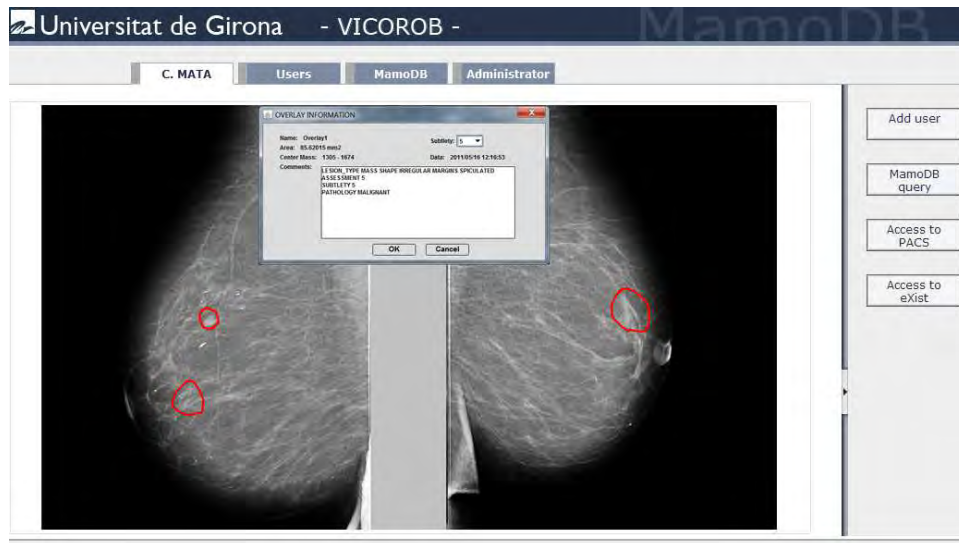
Figure 4.10: Display of DICOM images with their associated overlays. It is also possible to obtain a description of each overlay.
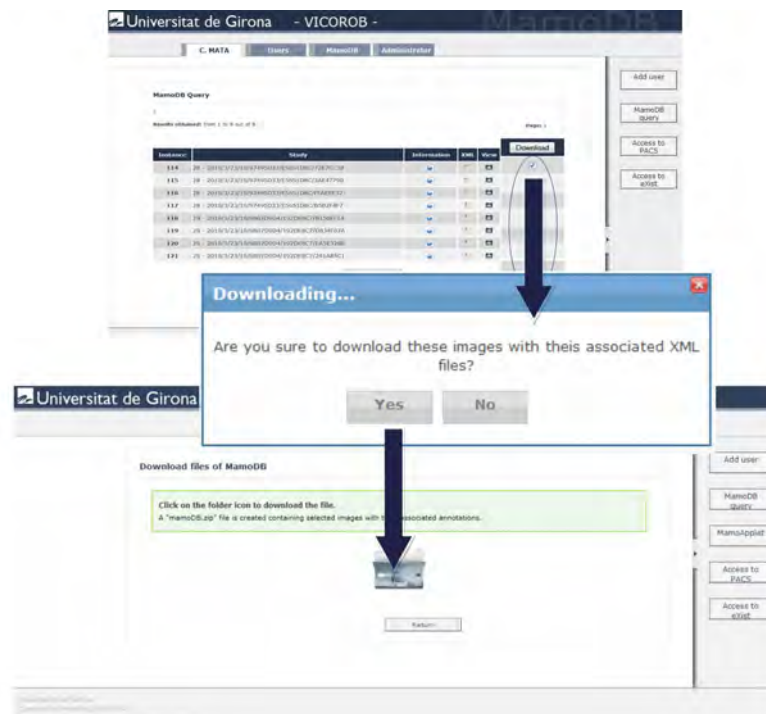
Figure 4.11: Example of creating a ZIP file containing a list of XML files and their corresponding DICOM mammographic images.

acquisition date, and the associated number of series and basic attributes.

Figure 4.12 shows an example of the search process in a sequence diagram. The user should be logged into the web-application to use the search engine. Once the user is authenticated, the search engine provides a data request form. Thus, the first step for users is to fill in the form according to the criteria query in order to generate and obtain a data subset of the mammographic images solicited. The server makes a petition to the eXist-db server by a previously defined user query using the query form. Then, a list of XML files are obtained and for each one the SOAP ID is obtained. This ID is the main key used to obtain DICOM image connected with the eXist-db server. The server then generates a list of XML files related to the DICOM images according to the parameters chosen. A ZIP file containing all these datais created using a MamoDB search engine, as depicted in Figure 4.11. Finally, the central server sends this ZIP file to the client which will download it for users.

4.4 MamoApplet: Interactive Java applet tool

A *MammoApplet* has been developed in Java, 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 computers and is independent of the operating system. This allows full portability of the application.

The next sections explain the architecture of the application 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.

4.4.1 System access

The interface is designed as an applet, hence the system needs a web server to maintain the webpage. This server also allows communication between users with both the PACS and the XML databases. A Java applet, for security reasons, can only establish communication with another application stored in the same machine as the web server where the applet is executed, and cannot store data in 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 is graphically shown in Figure 4.13. Notice that the XML database may be stored in the same PACS machine or in a different one[5].

---

[5]eXist-db official webpage. URL: `http://exist-db.org`
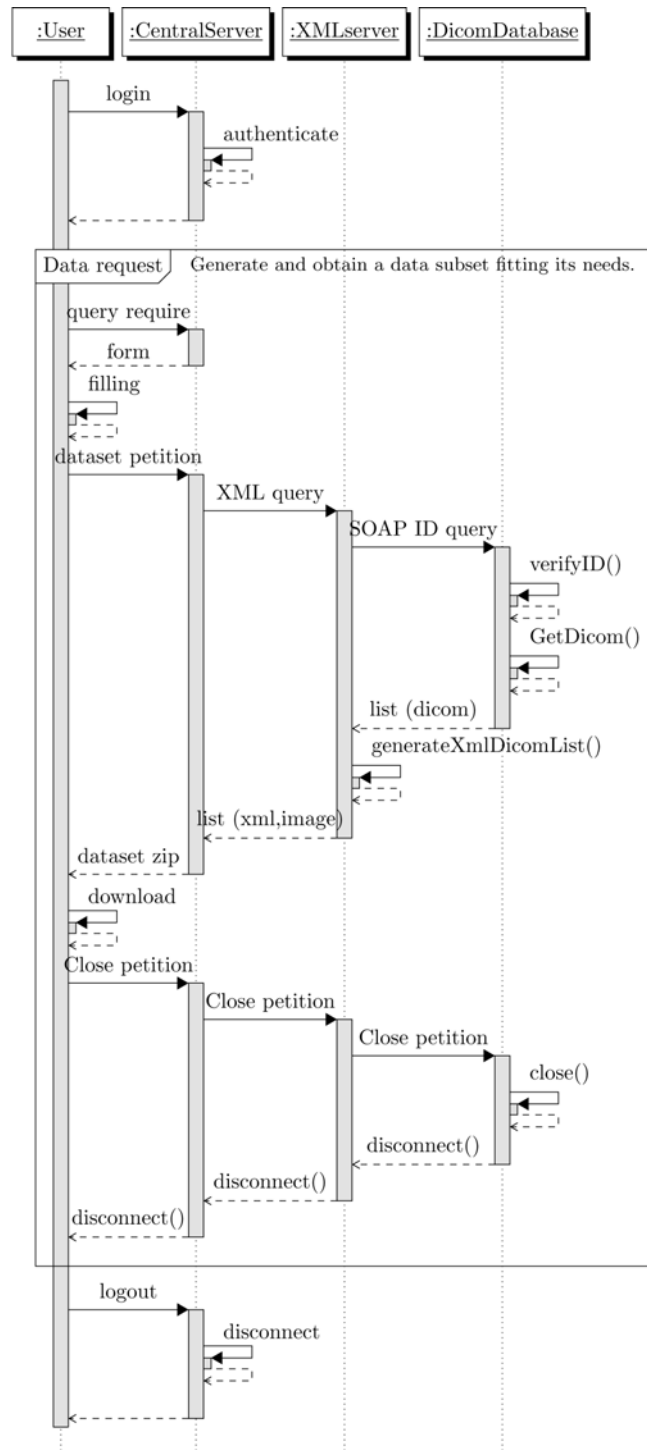
Figure 4.12: Sequence diagram of the internal downloading process.
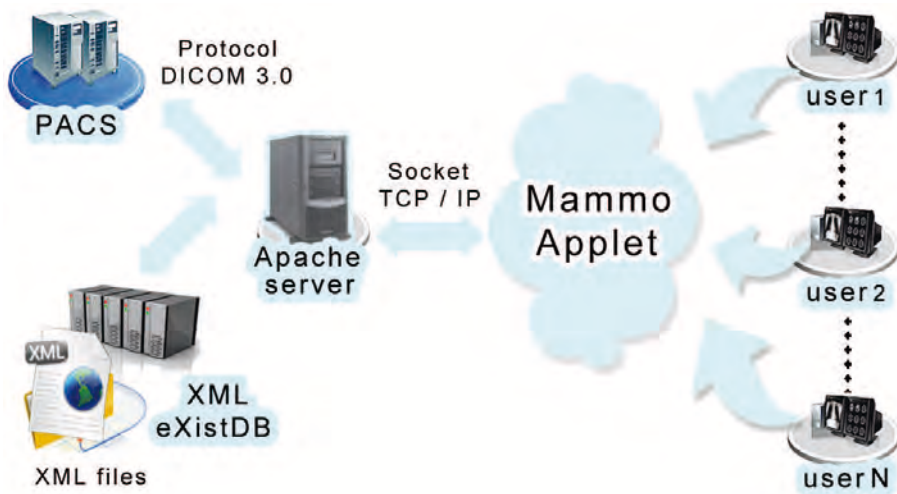
Figure 4.13: Interaction between PACS server, MammoApplet and XML database.

Also note 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.

As already described, the communication between the web server and the PACS uses the DICOM protocol. In this sense, a free distribution library PixelMed Java DICOM Toolkit (PixelMed) [32] is used, since this library also allows image processing and communication between different DICOM applications. Whatever the request, there are a number of attributes (IP address, name of PACS server, port,...) that both the PixelMed library and the PACS need to establish the connection.

Actually, the security of the system is done via username and password, requested by the Java applet. Once the user is validated, a home window is displayed to welcome to the user to the *MammoApplet* tool as shown in Figure 4.14.

4.4.2   MammoApplet engine

Facility in using the MamoDB search engine was explained in Section 4.3.3. In this subsection another *MammoApplet* search engine is presented. The main difference between the two is that only the first one obtains a list of mammographic examinations via a query form. Note that these mammographic examinations are made previously using the *MammoApplet* application. So, the aim is to visualize information related to the DICOM image and download a list of mammographic

Figure 4.14: First screenshot of the MamomDB applet, once the user has successfully logged in.

examinations (including DICOM images and XML files). Otherwise, the application only uses the search engine to open a complete examination in order to record mammographic findings.

Another difference of the MamoDB search engine is the option of finding examinations. In this case, the *MammoApplet* search engine offers the possibility of finding a patient ID, patient name, institution name and year of birth. The data obtained is then arranged according to the patient, studies, and series attributes, which later allows one to refine the search. Searches can be exhaustive, as when the PACS returns a list of all patients who have undergone a study, or selective, as when only a reduced subset of cases fulfilling the search conditions is returned. As can be seen, the DICOM query levels presented in subsection 1.1.2 of Chapter 1 were used. Then, an example of how a mammographic study is selected with the *MammoApplet* application is as follows:

- Find a patient via patient ID, patient name, institution name or year or birth. Then the patient or a list of patients appears.

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

- Once a study has been selected, *MammoApplet* returns all the series associ-

Figure 4.15: Example of a standard query provided by the search engine.

ated 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 4.15 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 displays the studies associated to the selected patient, and finally, the third corresponds to the series that belong to the selected study. When a series has been selected, the visualization panel is displayed on a new tab showing the DICOM images associated with the case along with the XML files containing the manually annotated overlays.

4.4.3   Visualization

The *MammoApplet* interface has been designed in a simple and intuitive way for radiologists and experts, with easy access to all the available tools. The interface panel is divided into two parts, as can be seen in Figure 4.16. The upper and larger part displays the DICOM images. In this example, a four part view containing MLO (left) and CC (right) images is visualized. The visualization tools offer different possibilities of arranging 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
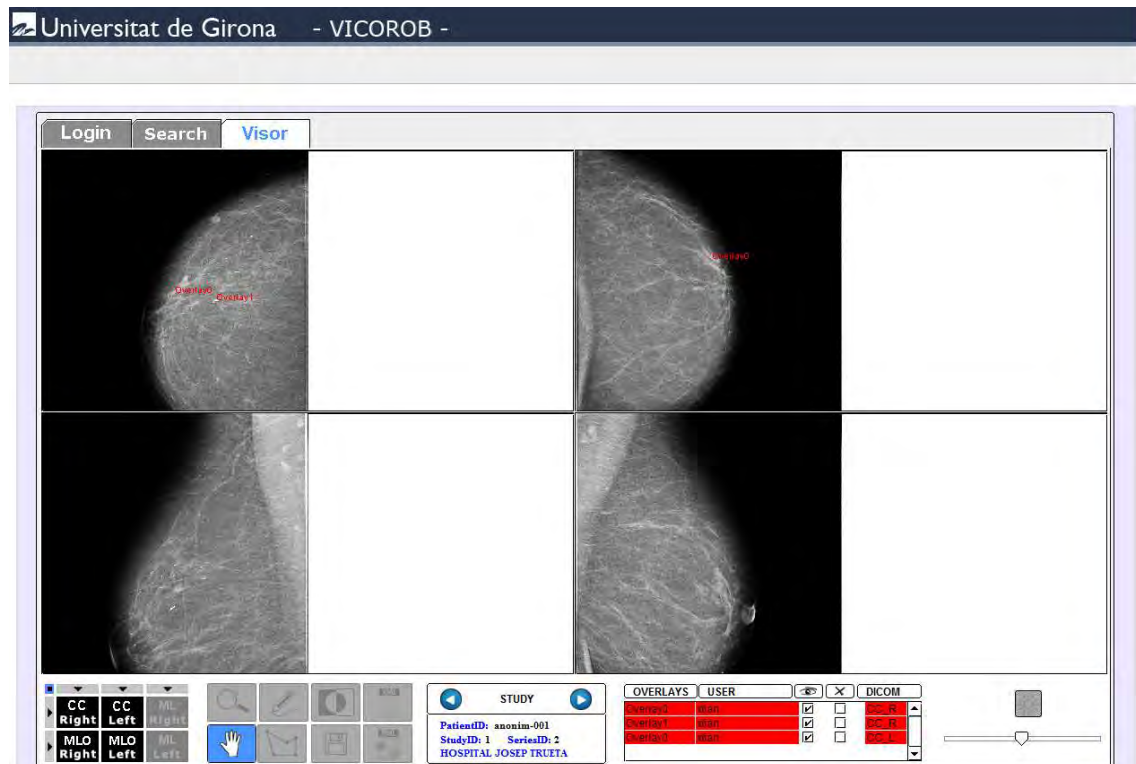
Figure 4.16: Example of a visualization of four mammographic images with their overlays. This example shows the MLO and CC view of right and left mammograms.

to the selected view of the breast. Users also have the possibility of visualizing just a single image. This latter way of visualization offers the radiologists the possibility of interacting with the image using a full resolution of the mammograms. Another visualization option is depicted in Figure 4.17 which shows an example of a visualization of the CC view of left and right mammograms. This view is useful to radiologists because it allows detecting breast asymmetries between the right and left breasts.

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 image processing tools offer a set of tools like zoom control, scroll contrast, inverse image, import the images, and scroll brightness. These tools are important for the users since they can see the image of the breast under different conditions. Finally, the bottom panel contains a 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 following studies of the same patient. This is very useful since the comparison of previous studies, i.e. the analysis of the breast
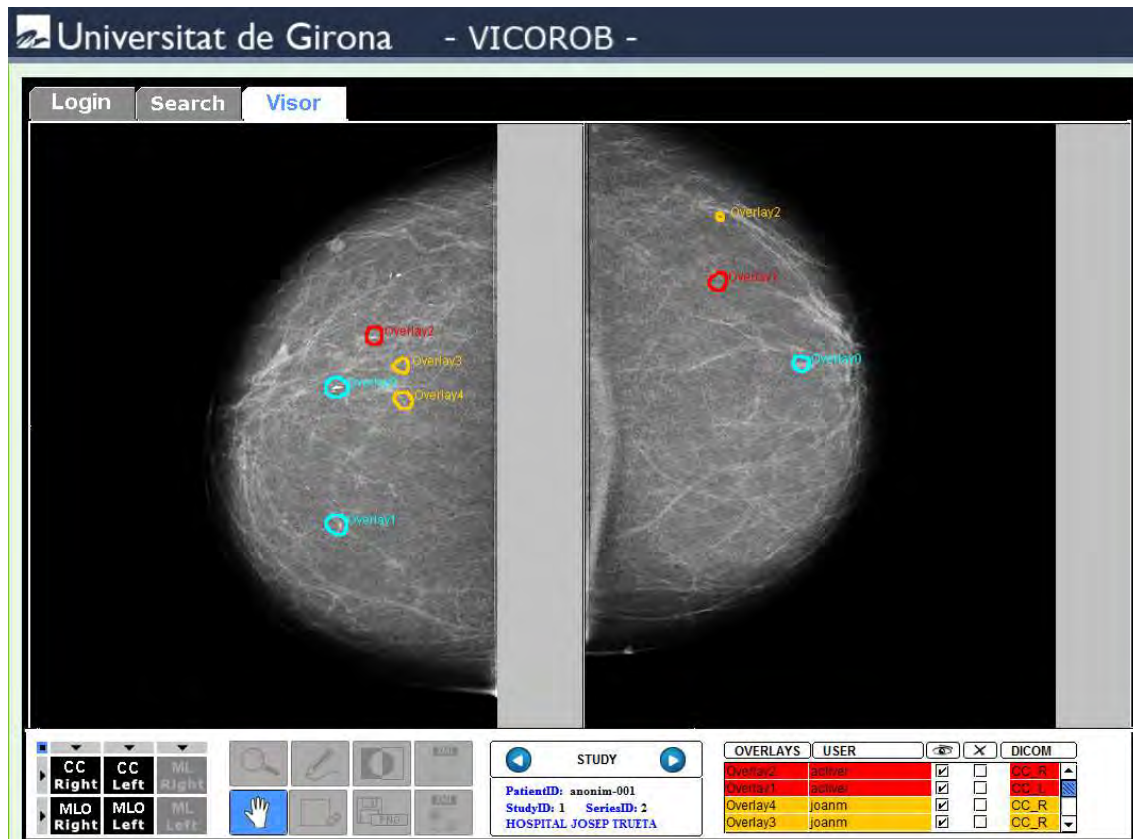
Figure 4.17: Example of a visualization of two mammographic images with their overlays. This example shows the CC view of the right and left mammograms.

evolution, is a procedure used by the experts to diagnose the patient and to perform longitudinal analysis.

4.4.4 Manual annotation

A manual annotation example is shown in Figure 4.18. In the control panel, (see bottom part of the figure) the user chose the manual segmentation tool and directly segments the Region Of Interest (ROI) area depicted in the example as a part "A". Then, a pop-up window containing information about the segmented overlay is displayed (shown in part "B"). As can be seen, the name of the overlay, the area, mass center, and data is generated. Also, the user can make comments concerning the overlay and select the type of subtlety. Then, the manual annotation process is finished after clicking the "Ok" button. Automatically, the new overlay appears in the overlays table.

Figure 4.18: An example of a manual annotation using MammoApplet application.

The visualization panel allows different actions related to the creation and manipulation of overlays. As can be seen in Figure 4.17, 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. In fact, making a collaborative work between users possible is one of the important goals of this thesis. This way it is easier for radiologists to know which overlays were made by themselves and, besides, it allows a fast comparison of evaluation made by different experts.

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

- Draw an overlay. Users can add new areas of interest in the mammograms by simply clicking left and right buttons on the mouse.

- Annotate an overlay. Once an overlay has been created, a pop-up window like the one shown in Figure 4.19(a) appears. This window displays information related to the annotation, like the area and the mass center location, which
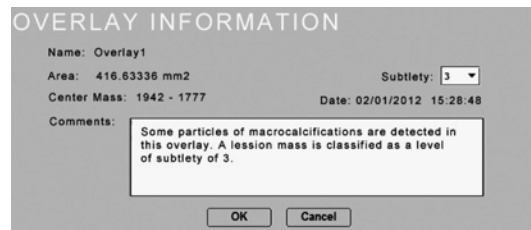
(a)
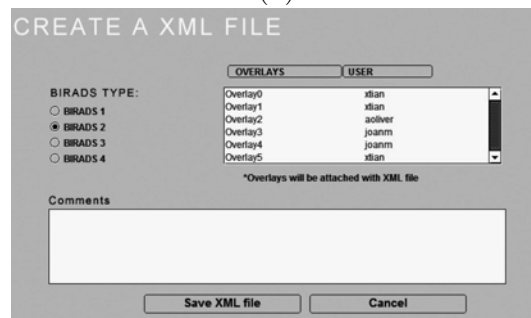
(b)

Figure 4.19: Examples of pop-up windows corresponding to (a) overlay information and (b) XML file.

are automatically computed, and also allows the expert to introduce extra information, like the subtlety of the lesion (degree of difficulty to find) or any additional comments. Also, this was shown in the previous figure.

- Delete an overlay. Users 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 4.19(b), is displayed. The user selects the breast density according to the BI-RADS lexicon and then may introduce comments regarding the final diagnosis.

All the information related to the annotations is also shown in the overlays table (right part of the bottom panel in Figure 4.17). This table shows the lists of overlays created for each case, specifying the username of the expert that created the overlay and the mammogram where the overlay is. It also allows visualizing or to deleting 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 from the table automatically displays its related information (the annotations made in the pop-up window by the expert).

4.4.5   An example of use

In this section, an indicative example of how to use a *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 shown in Figure 4.15. 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 drawn around the suspicious part in 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. Users can draw a polygonal area around the suspicious regions in the image, and associate these regions with the manual annotations through the pop-up window that is displayed when closing the polygon.

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

4.5   Design effort benchmarks

An exhaustive experimental study in order to evaluate the scalability, complexity and response speed is performed. Three analysis are performed: a) a MamoDB search engine, b) a MamoDB ZIP file generation and c) a MamoApplet application.

A) MamoDB search engine

A first analysis calculating the response time using a MamoDB search engine is performed. This analysis consists of evaluating the response time provided by an interval of user connections. In this sense, internal and external connections are used to evaluate a set of common actions: "Search all", "BIRADS", "Study" and "Patient". Table 4.3 shows the calculated response time using local connections. In order to facilitate the analysis, a number of user connections are divided into intervals (1-3 users, 5-8 users, 10 users, 15 users and 20 users) for each action. For

Response time obtained from internal connections (seconds)				
Features	1-3 users	5-8 users	10 users	15 users	20 users
Search all	9.04 ± 1.23	10.76 ± 1.35	13.28 ± 1.40	14.39 ± 1.37	16.25 ± 1.54
BIRADS	6.81 ± 1.17	8.53 ± 1.22	10.35 ± 1.31	12.46 ± 1.48	14.32 ± 1.48
Study	1.22 ± 1.29	2.84 ± 1.36	3.36 ± 1.28	4.07 ± 1.73	4.44 ± 2.11
Patient	1.14 ± 1.05	1.87 ± 1.12	2.42 ± 1.47	2.84 ± 1.87	3.27 ± 2.42

Table 4.3: Summary of the response time obtained using local connections by multiple users in MamoDB.
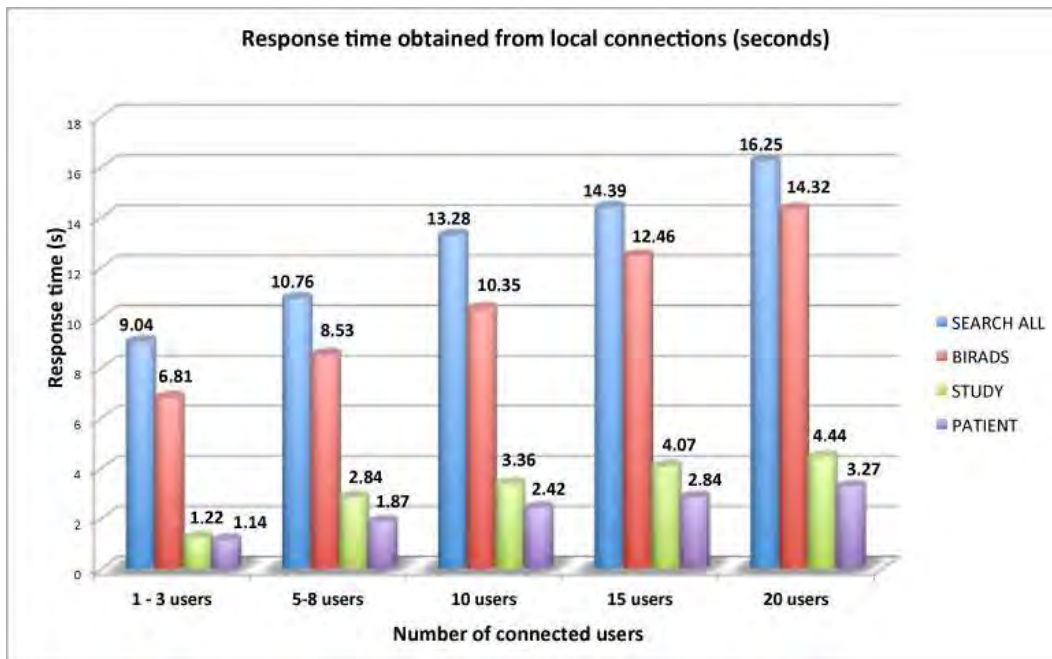
Figure 4.20: Bar diagram of the calculated response time of MamoDB using local connections.

each interval, the response time (in seconds.) and the mismatch error is calculated.

Figure 4.20 shows a representation of this response time using a bar diagram divided by intervals of user connections. As can be seen, the response time increases depending on the number of connected users. Concretely, "Search all" is the action that takes the longest time due to the requirements of compiling a full list of mammographic studies. Normally, users perform an advanced search such as "BIRADS", "Study" or "Patient" because it is more efficient obtaining an accurate list of results. The response time searching by "BIRADS" is close to "Search all" action because it also deals with a high volume of data but it is still less than "Search all". When searching with the "Study" or "Patient" action, the response time is decreased because the list of results is reduced.

	Response time obtained from external connections (seconds)				
Features	1-3 users	5-8 users	10 users	15 users	20 users
Search all	10.46 ± 1.34	12.38 ± 1.29	15.09 ± 1.10	17.68 ± 1.71	19.32 ± 2.13
BIRADS	7.93 ± 1.26	10.06 ± 1.18	12.04 ± 1.31	14.16 ± 1.81	16.14 ± 2.42
Study	2.33 ± 1.16	3.61 ± 1.39	4.86 ± 1.32	$6.37 \pm 1,76$	7.74 ± 2.84
Patient	2.28 ± 1.32	3.06 ± 1.25	4.24 ± 1.07	5.32 ± 1.92	6.51 ± 3.08

Table 4.4: Summary of the response time obtained using external connections by multiple users in MamoDB.
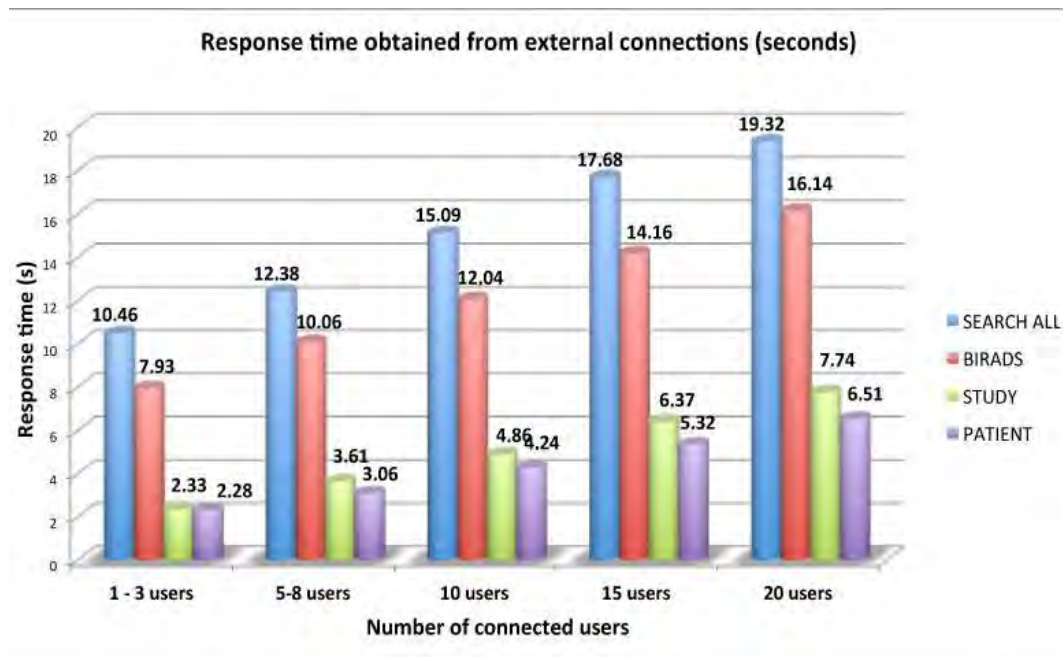
Figure 4.21: Bar diagram of the calculated response time of MamoDB using local connections.

In the same way, a second evaluation using external connections is also performed. Table 4.4 shows the response time calculated using the same interval of user connections and the same operations. Figure 4.21 shows a representation of this response time using a bar diagram divided by intervals of user connections. Comparing the results using internal connections with external connections the response time is doubled in all cases. In summary, a comparison of a time response between internal and external connection provided by different intervals of user connections using MamoDB is depicted in Figure 4.22.

In general terms, it is normal that when using external connections, the response time is higher compared with internal connections. When comparing the maximum point (20 users), the difference is around 3 seconds between internal and external. More than 20 user connections is considered unviable in our system's architecture
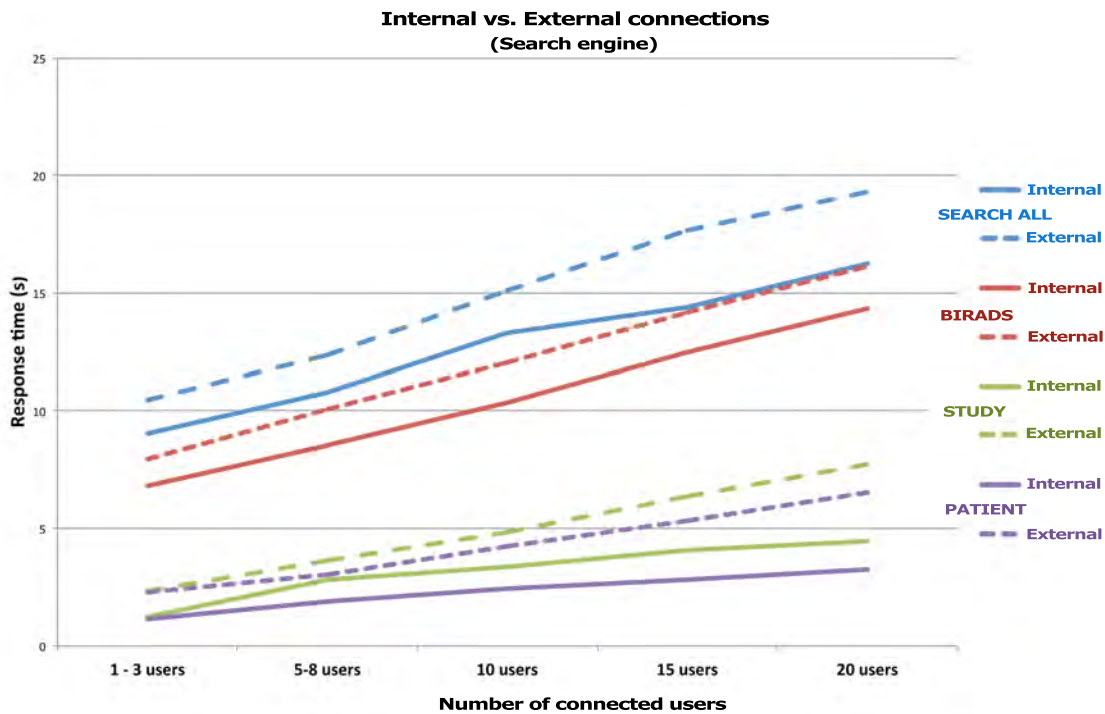
Figure 4.22: Comparison of internal and external connections using MamoDB.

because it could adversely affect the response time of the MamoDB. It is true that this difference it could be changed depending on the time gap caused by the users' internet connection. This time gap is described in Tables 4.3 and 4.4 respectively.

B) MamoDB ZIP file generation

A second analysis based on the evaluation of the response time to generate a ZIP file using the MamoDB search engine is performed. This analysis comprises a set of features to be evaluated again using internal and external connections. It is important to evaluate the response time according to the number of files in the search engine, i.e. the total size of these files and the size of the ZIP file containing all the data. Remember that this ZIP contains a set of images and XML files. Again, the number of user connections is divided into intervals (1-3 users, 5-8 users, 10 users, 15 users and 20 users) for each case.

Table 4.5 shows a summary of the response time obtained once a ZIP file is generated using internal connections. As can be seen, for a range of 2, 10, 20, 30, 40, 50 or 60 files, the response time is calculated also using an interval of user connections. The response time increases when the number of user's connections is

Response time of a ZIP file generation using internal connections							
Features			Response time (seconds)				
Files	Size	ZIP	1-3 users	5-8 users	10 users	15 users	20 users
2 files	26	10	6.26 ± 1.18	7.58 ± 1.32	8.95 ± 1.29	10.22 ± 1.47	11.54 ± 1.33
10 files	130	44	16,86 ± 1.06	18.42 ± 1.27	16.96 ± 1.40	21.53 ± 1.01	23.04 ± 1.42
20 files	240	80	21.63 ± 1.08	23.17 ± 1.28	29.57 ± 1.12	26.25 ± 1,46	27.79 ± 1.63
30 files	351	124	28.03 ± 1.02	29.57 ± 1.15	31.11 ± 1,27	32.65 ± 1.39	34.19 ± 2.03
40 files	481	167	31.06 ± 1,24	32.65 ± 1,13	34.19 ± 1.42	35.73 ± 1.62	37.27 ± 2.17
50 files	611	209	36.25 ± 1.42	37.79 ± 1.29	39.33 ± 1.16	40.87 ± 1.81	42.41 ± 2.26
60 files	741	259	52.31 ± 1.38	54.15 ± 1.33	55.69 ± 1.41	57.23 ± 1.77	58.77 ± 2.16

Table 4.5: Summary of the response time to generate a ZIP file using internal connections by multiple users in MamoDB. Sizes are given in MBytes.
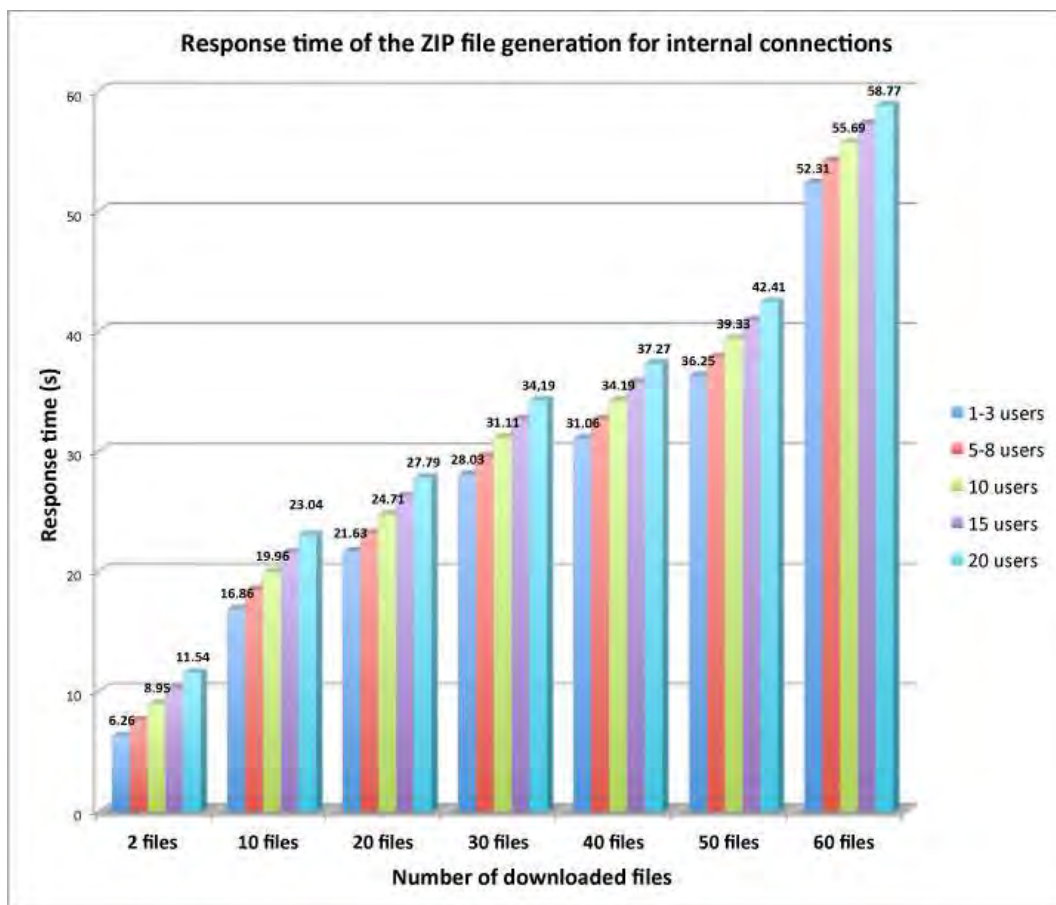
Figure 4.23: Bar diagram of the calculated response time to generate a ZIP file in MamoDB using local connections.

higher. Remember that users should select the number of files to be downloaded using the search engine. Then, MamoDB automatically generates the corresponding ZIP file. Figure 4.23 shows a representation of this response time using a bar diagram

Response time of a ZIP file generation using external connections							
Features			Response time (seconds)				
Files	Size	ZIP	1-3 users	5-8 users	10 users	15 users	20 users
2 files	26	10	7.53 ± 1.02	9.58 ± 1.29	11.63 ± 1.38	14.38 ± 1.77	19.33 ± 2.17
10 files	130	44	18.23 ± 1.14	20.28 ± 1.38	22.33 ± 1.32	25.08 ±1.81	30.03 ± 1.72
20 files	240	80	23.45 ± 1.05	25.51 ± 1.16	27.55 ± 1.23	30.32 ± 1,66	35.25 ± 1.83
30 files	351	124	30.12 ± 1.24	32.17 ± 1.11	34.22 +- 1.41	36.97 ± 1.69	41.92 ± 2.12
40 files	481	167	36.57 ± 1,12	38.62 ± 1,27	40.67 ± 1.68	43.42 ± 1.66	48.37 ± 2.35
50 files	611	209	38.21 ± 1.37	40.26 ± 1.33	42.31 ± 1.64	45.06 ± 2.08	50.01 ± 2.62
60 files	741	259	55.01 ± 1.29	57.06 ± 1.45	59.11 ± 2.34	61.86 ± 2.67	58.77 ± 2.35

Table 4.6: Summary of the response time to generate a ZIP file using external connections by multiple users in MamoDB. Sizes are given in MBytes.
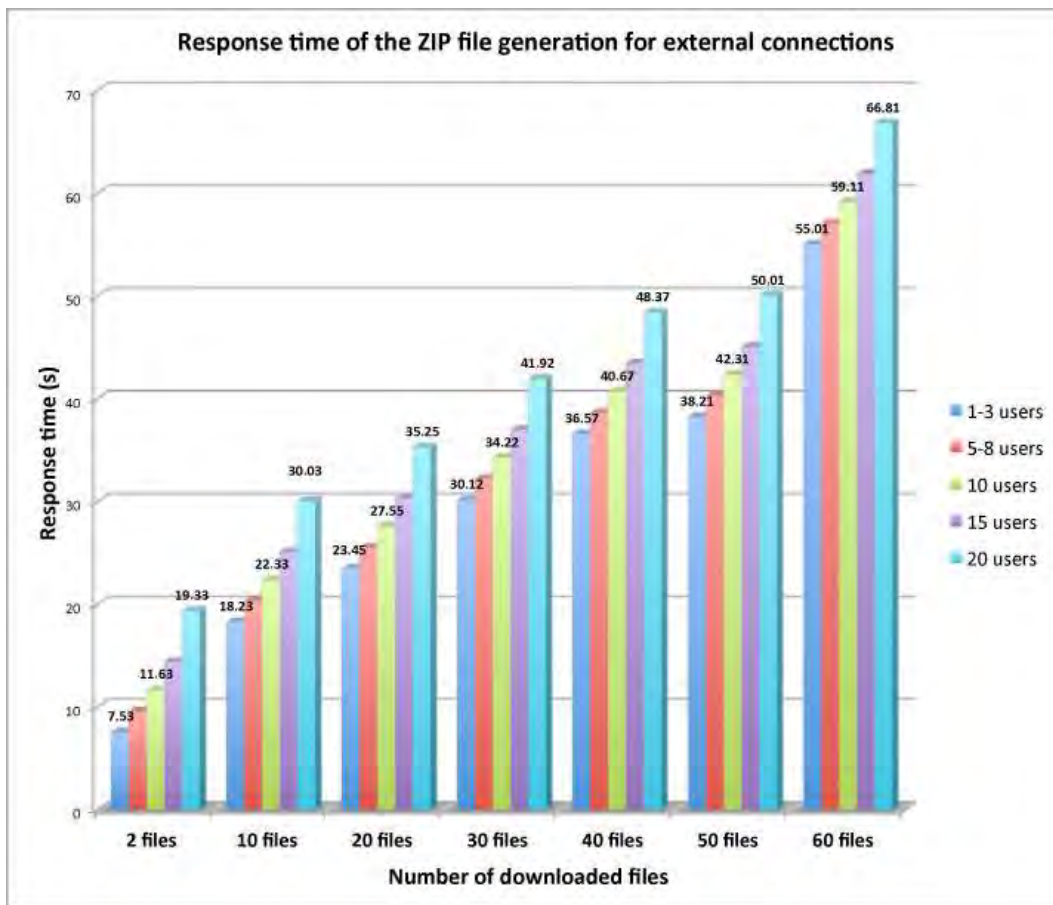
Figure 4.24: Bar diagram of the calculated response time to generate a ZIP file in MamoDB using external connections.

divided by intervals of user connections. In this case, the total of the files obtained from the MamoDB search engine is chosen as a reference point to evaluate the response time.
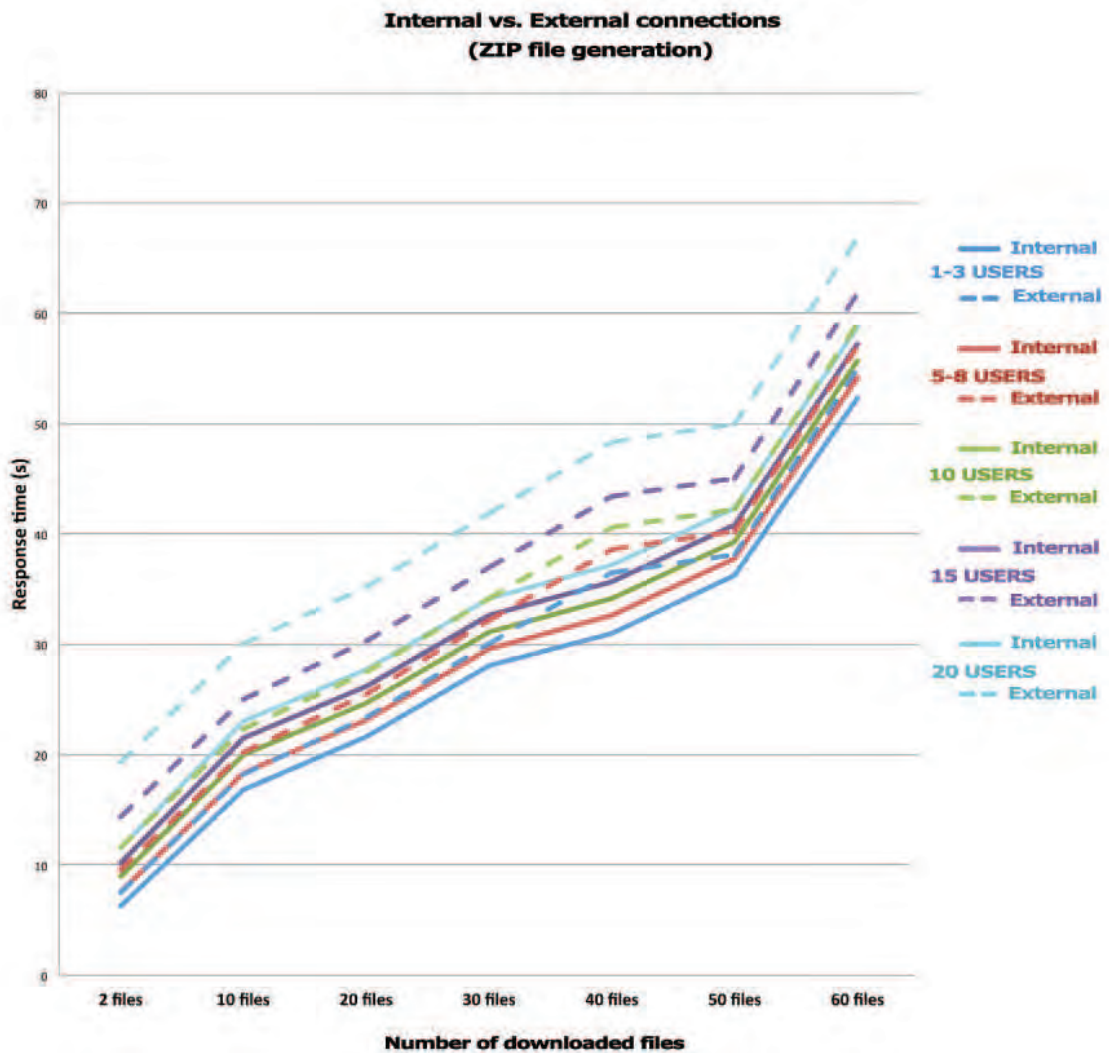
Figure 4.25: Comparison of internal and external connections generating a ZIP file.

A second evaluation using external connections is also performed. Table 4.6 shows a summary of the elapsed response time to generate a ZIP file using external connections by multiple users in MamoDB. When comparing the results obtained between Tables 4.5 and 4.6, the response time increases when using external connections. It is shown that after 15 simultaneous connections the response time also increases. Moreover, when using external connections, the mismatch error is between ± 2 or ± 3 seconds, as can be seen in Table 4.6. On other hand, Figure 4.24 shows a bar diagram of the calculated response time to generate a ZIP file in MamoDB using external connections.

In summary, a comparison of the internal and external connections generating a

	Response time obtained from internal connections (seconds)				
Features	1-3 users	5-8 users	10 users	15 users	20 users
Search all	12.49 ± 1.06	14.63 ± 1.21	16.77 ± 1.28	19.44 ± 1.86	21.58 ± 2.04
Search Advanc.	1.05 ± 1.10	3.19 ± 1.18	5.33 ± 1.32	7.47 ± 1.76	9.61 ± 1.97
Load files	5.53 ± 1.22	7.67 ± 1.38	9.81 ± 1.41	11.75 ± 1.67	14.14 ± 2.13
Create study	1.03 ± 1.34	2.59 ± 1.45	3.93 ± 1.29	4.97 ± 1.92	6.11 ± 2.17
Save study (create)	1.65 ± 1.04	3.38 ± 1.23	4.43 ± 1.44	6.07 ± 1.61	8.21 ± 1.87
Update study	1.03 ± 1.12	2.16 ± 1.39	3.45 ± 1.42	4.77 ± 2.19	5.91 ± 2.34
Save study (update)	1.54 ± 1.17	3.38 ± 1.21	4.28 ± 1.39	5.82 ± 2.07	8.35 ± 2.02

Table 4.7: Summary of the response time using internal connections by multiple users in *MamoApplet.*

ZIP file using 20 simultaneous connections is depicted in Figure 4.25. This was also chosen in a hypothetical case as a maximum response time for 60 files. The total size of the contained folder was 741 MB and the generated ZIP file was 259 MB. As can be seen, for each interval a difference of 7 or 8 seconds between external and internal connections is shown. In fact, 60 files is the maximum number of available files for which MamoDB can generate a ZIP file considering a maximum number of 20 connections as well.

C) MamoApplet application

A third analysis calculating the reponse time using the *MamoApplet* application is performed. *MamoApplet* is currently being used in the VICOROB department of the University of Girona (Catalunya, Spain). The present working database is composed of more than 300 patient datasets. It has been tested with the following common internet web browsers: Mozilla Firefox, Internet Explorer, Google Chrome, and Opera. In general terms, the most time-consuming part corresponds to loading a mammographic study. A typical study consists of a set of 4 images corresponding to different views (MLO and CC). Some cases can be loaded with a set of 5 or 6 images corresponding to LM or extended views.

Table 4.7 shows a summary of response times using internal connections by multiple users in *MamoApplet.* The response time is calculated for different actions using the *MamoApplet* application depending on the number of internal and external connections. These actions are: search all mammographic studies, advanced search, load files, create XML files, save a study after the creation of the XML files, update the XML files and save a study after the update of the XML files. In the same way as previous analysis, intervals of connections are used (1-3 users, 5-8 users, 10 users, 15 users and 20 users). A bar diagram of the calculated response time of *MamoApplet* using internal connections is depicted in Figure 4.26. In general terms, "search all" and "load files" are the two actions with the highest response time. In fact, both are the most important operations in the *MamoApplet* application.
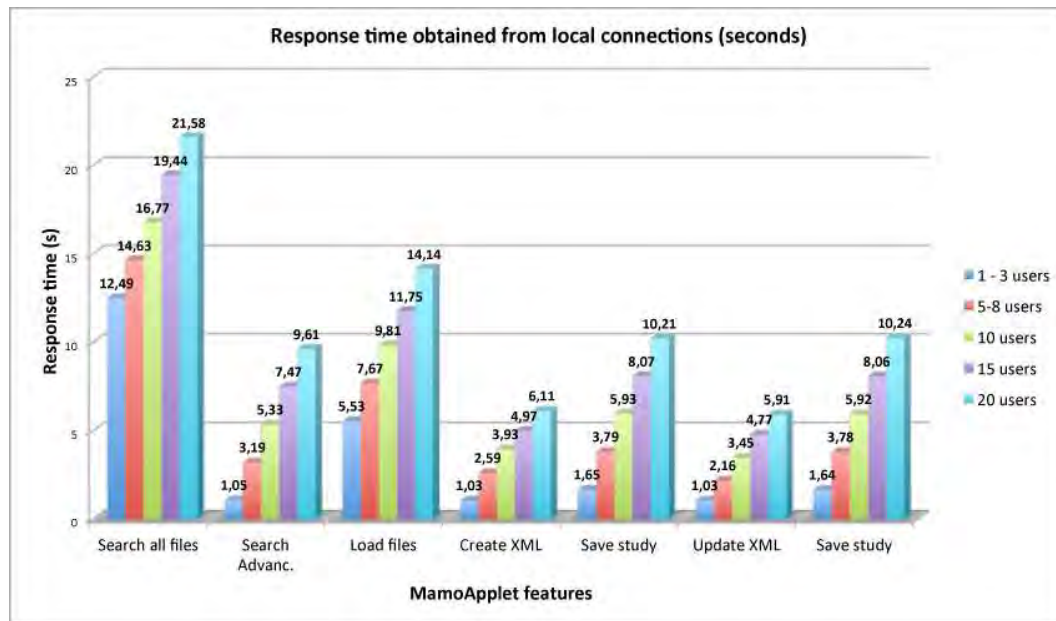
Figure 4.26: Bar diagram of the calculated response time of *MamoApplet* using internal connections.

	Response time obtained from internal connections (seconds)				
Features	1-3 users	5-8 users	10 users	15 users	20 users
Search all	18.22 ± 1.59	20.36 ± 1.67	22.51 ± 1.82	25.17 ± 2.67	28.16 ± 3.12
Search Advanc.	2.93 ± 1.61	5.07 ± 1.89	7.21 ± 1.91	9.35 ± 2.04	12.34 ± 2.78
Load files	13.71 ± 1.86	15.85 ± 1.72	17.99 ± 1.95	19.93 ± 2.14	22.92 ± 3.14
Create study	2.56 ± 1.75	4.72 ± 1.83	6.84 ± 2.01	9.38 ± 2.18	12.37 ± 2.89
Save study (create)	2.23 ± 1.58	4.37 ± 1.76	7.51 ± 2.14	11.25 ± 2.65	14.24 ± 3.08
Update study	2.44 ± 1.67	4.58 ± 1.72	6.72 ± 1.96	9.66 ± 2.87	12.65 ± 3.14
Save study (update)	2.26 ± 2.69	4.45 ± 2.65	8.14 ± 2.92	11.08 ± 3.57	14.46 ± 3.15

Table 4.8: Summary of the response time using external connections by multiple users in MamoApplet.

MamoApplet was accessed using external network connections from Dijon in France, several hundreds of kilometers from Girona. Table 4.8 shows a summary of the response time of common actions using external connections by multiple users. When comparing the results using internal connections, it is shown that the response time increase using external connections. Again, "search all" and "load files" are the two actions with the highest response time. Normally, the mismatch error is around ± 2 or ± 3 for all cases. However, the main difference is observed in the mismatch error when there are 20 user connections. A bar diagram of the calculated response time of *MamoApplet* using external connections is depicted in Figure 4.27.

A comparison between internal vs. external connections is depicted in Figure 4.28. In this case, the response time using 20 user connections for a internal
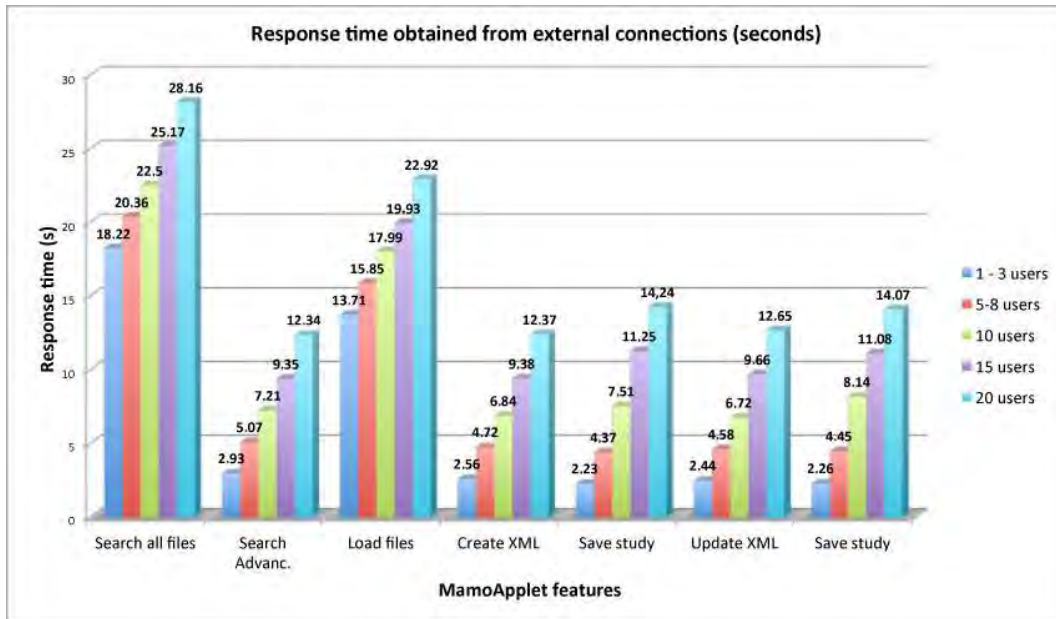
Figure 4.27: Bar diagram of the calculated response time of MamoApplet using external connections.

and external is compared. It is important to compare the difference between the response times for the "search all" and "load study" actions. This difference is around 7 or 8 seconds. Then, it is preferable not use more than 20 user connections in order to assure a good service to the users.

In fact, it has been noticed that the response time is proportionally increased when the number of connections also increase. The increment of the response time also affects to the throughput. Throughput is the rate at which a computer or network sends or receives data. The throughput and response time are linearly related and the influence is reciprocal of one another. If response time increases when the load is increased, the degradation of system performance is greater. At the moment, the web-based application is focused on a limited number of connections according to the maximum number of the users of the hospital (around 20 users). In the future, it is an important aspect to improve when a new users or medical centers wants to join in our architecture system. It is possible to improve the response time with a better hardware or an extended architecture using a combination of distributed database system. For example, using multiple server locations could be a solution due to the information is placed in few data centers to improve performance and availability. Also, using widely distributed servers and databases which the information could be served from a distributed network which has a set of small servers spread across many service providers. Of course, it should empower a system administrator to
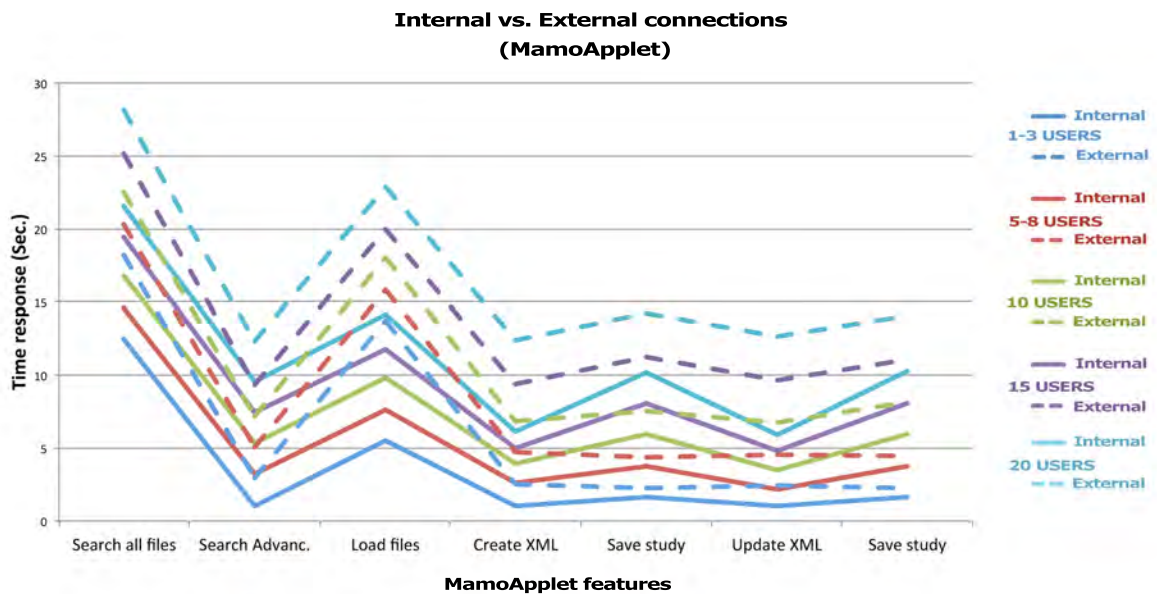
Figure 4.28: Comparison between internal and external connections using MamoApplet.

reason about the system's past behavior and estimate enhancements with a new hardware changes for the architecture system. In this sense, using a distributed databases system is more efficient and improves the throughput of the web-based applications and enables the real-time synchronization of data across any number of dispersed databases. A previous simulation should be done before to choose the best option to improve the actual architecture system.

4.6 Conclusions

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

Digital medical imaging technologies have become beneficial in modern medical practices and health care systems, providing powerful tools for diagnosis, treatment,

and surgery. The volume of medical images being generated has grown rapidly due to an increase in the number of clinical examinations performed in digital form and to the large range of available modalities. Therefore, the demand for online medical imaging systems that allow visualization and processing has increased significantly.

Chapter 5

ProstateAnalyzer: Web-based medical application for the management of prostate cancer from MR Imaging techniques and MR Spectroscopy

Attention to health is life's greatest hindrance.

Plato

In this chapter a new web-based medical tool for prostate cancer diagnosis called ProstateAnalyzer is presented. ProstateAnalyzer allows the visualization and analysis of Magnetic Resonance Imaging (MRI) and Magnetic Resonance Spectroscopy (MRS) in a single framework. A short presentation of Prostate Cancer (PCa), MR Imaging techniques and MR Spectroscopy, and the ProstateAnalyzer applet tool are presented. This chapter is an adaptation of a paper published in the journal IHSC[1].

---

[1][IHSC] ProstateAnalyzer: web-based medical application for the management of prostate cancer using multiparametric MR images, 2015 [117].

5.1 Background

5.1.1 Anatomy of the prostate

Prostate Cancer (PCa) has become a significant health care burden [164]. Early diagnosis and an active follow-up allow improved prognosis and prevent life-threatening conditions. Once the decision of treatment has been taken, having the most complete information for treatment and follow up is crucial. Among the techniques used to detect and diagnose PCa, MRI allows a non-invasive analysis of the anatomy and the metabolism in the entire prostate gland. The prostate is composed of the Peripheral Zone (PZ), Central Zone (CZ), Transition Zone (TZ) and Anterior Fibromuscular Tissue (AFT) (Figure 5.1). The PZ represents up to 70% of a normal prostate gland and around 75% of prostate cancers originate in this zone. The CZ represents about 25% of a normal healthy prostate gland in a young adult. Even though the frequency of cancers originating here is much lower, they tend to be of the more aggressive type [188].

The medical support systems used to assist in the diagnosis of prostate lesions generally focus on prostate segmentation [169, 37, 50, 55]. They rely on computerized techniques for prostate cancer detection applied to ultrasound, magnetic resonance and computed tomodensitometric images [56]. For example, Vos et al. [192] used 3D-T2 imaging to define a specific Region Of Interest (ROI), which was subsequently used on diffusion- and perfusion-weighted images to extract relevant features. The purpose was to train and classify the extracted set of features to calculate the likelihood of malignancy. Other related surveys have focused on magnetic resonance spectroscopic data [27]. The rapid growth of computer-based clinical examination has seen a significant increase in the number of online medical imaging systems [133, 111], some of which developed as web-based applications [101]. The main challenge is the access and interaction between the hospital database and distant users. Thus, the aim of our work was to develop a new architecture system allowing the inclusion of a web-based application connected to a prostate image database.

In the field of biomedical informatics, one of the long standing problems is finding a way to share medical data across a variety of media. Inherently, medical data is generated by a multitude of sources [114, 180]. XML has emerged as a leading facilitator. Although the XML is provided with predefined tags, one of its advantages is its extensible use. Over the last few years, a set of standards in the medical domain was developed, called Medical Markup Language (MML) to allow the exchange of medical data between different medical information providers [64, 63]. Therefore, an XML schema of tags can be defined for each individual case. Finally, the inclusion of XML databases facilitates the management of XML files by storing them in an
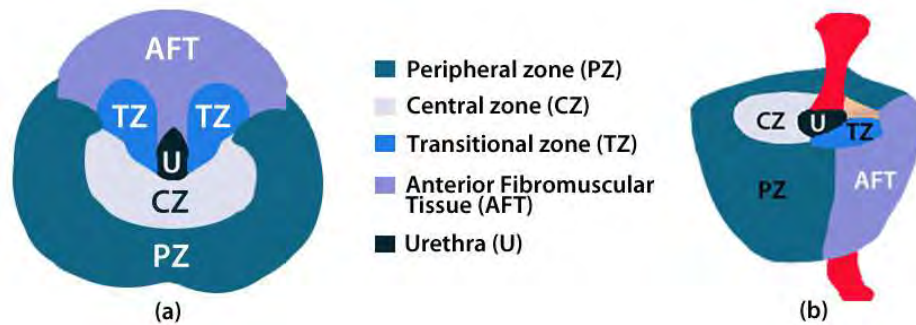
Figure 5.1: Anatomy of the prostate in (a) transversal and (b) sagittal planes.

efficient way [200]. Traditional object relational techniques, based on an XML model, are used to store XML files in an eXist-db database [119].

5.1.2 MR Imaging techniques and MR Spectroscopy

MRI has been established as the best imaging modality for the detection, localization and staging of PCa on account of its high resolution and excellent spontaneous contrast of soft tissues and the possibilities of multiplanar and multiparametric scanning [29]. As such, three MRI techniques (3D T2-Weighted Imaging (T2WI), Diffusion-Weighted Imaging (DWI) and Perfusion-weighted imaging (DCE)) and 3D Magnetic Resonance Spectroscopy (MRS) used in the diagnosis and follow up of PCa are presented in the next subsection.

3D T2-Weighted Imaging (T2WI)

A T2WI sequence [20] yields good contrast between PZ and CZ tissues. Also, it provides the best depiction of the prostate's zonal anatomy and capsule. T2WI is used for prostate cancer detection, localisation and staging. T2WI alone is not recommended because additional functional techniques improve both sensitivity and specificity. T2WI are obtained in 2–3 orientations. The axial (or pseudo-axial) T2WI sequence must cover the entire prostate and seminal vesicles, and are orthogonal to the rectum [17]. The phase encoding direction is left-to-right so that motion artefact does not overlap the prostate. Bowel motion artifacts should be reduced by administering an anti-peristaltic agent. The patient should be instructed about the importance of not moving during the image acquisition. An Endorectal Coil (ERC) is not an absolute requirement at either 1.5 T or 3 T, but a Pelvic Phased Array (PPA) coil with a minimum of 16 channels is required.

PCa typically manifests as a round or ill defined, low signal-intensity focus in PZ
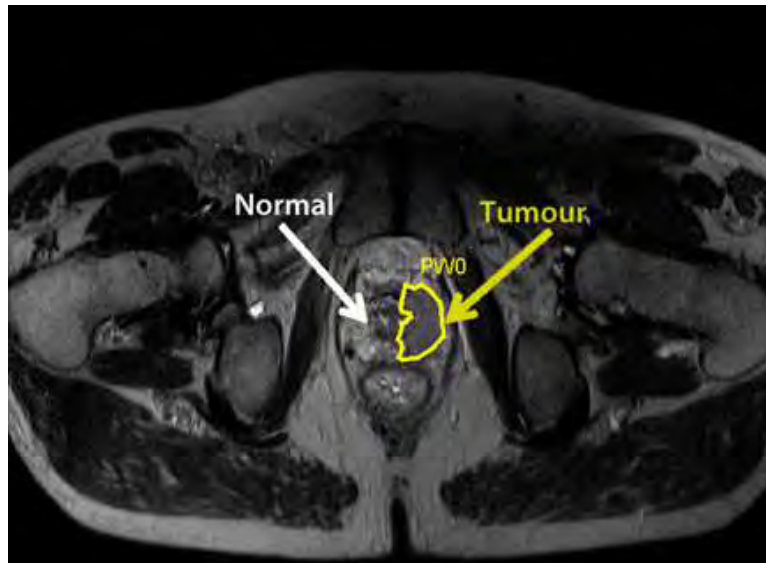
Figure 5.2: Example of a T2WI image from a prostate study.

(Figure 5.1 in T2WI). However, various conditions such as prostate intra-epithelial neoplasia, prostatitis, haemorrhage, atrophy, scars and post-treatment changes can mimic cancer in T2WI. Tumours located in the TZ are more challenging to detect, because of the signal intensity characteristics of the TZ. Also, a TZ tumour is often shown as a homogeneous signal mass with indistinct margins.

The interpretation of T2WI includes evaluation of the capsule, seminal vesicles and posterior bladder wall for extra-prostatic tumour invasion. Criteria for extra-capsular extension are abutment; irregularity and neurovascular bundle thickening; bulge, loss of capsule and capsular enhancement; measurable extra-capsular disease; and obliteration of the recto-prostatic angle. For seminal vesicle infiltration, the criteria are: expansion; low T2 signal intensity; and filling in of the prostate seminal vesicle angle. An example of a T2WI image is depicted in Figure 5.2 showing examples of normal and tumor tissues.

Diffusion-Weighted Imaging (DWI)

DWI is a powerful clinical tool, as it allows Apparent Diffusion Coefficient (ADC) maps to be calculated, enabling qualitative and quantitative assessment of prostate cancer's aggressiveness. AADC is a measure of the magnitude of diffusion (of water molecules) within tissue, and is commonly clinically calculated using MRI with DWI [2]. Cancer shows a lower ADC value than normal prostate tissue. Furthermore,

---

[2]Radiopaedia. URL: http://www.radiopaedia.org

ADC values correlate with Gleason scores [14].

DWI provides functional information on tissues such as cell organization, density and microstructure, and depends principally on the Brownian motion of water molecules [92]. The motion of water molecules is more restricted in tissues with a high cellular density and intact cell membranes and very low values are clearly indicative of cancer [151]. An example of a DWI image is depicted in Figure 5.3 (example of normal and tumor tissues).

Moreover, the image should be acquired in the axial plane with an echo planar imaging sequence employing parallel imaging. Motion probing gradients should be applied in three orthogonal directions [14]. The b-value identifies the measurement's sensitivity to diffusion and determines the strength and duration of the diffusion gradients. It combines the following physical factors into one b-value and is measured in s/mm [103]. According to [69], the minimal requirements are b-values of 0, 100, and 800–1000 s/mm2. The choice of these values enables calculation of diffusion sensitive ADC values. For optimal DWI, the b-values are: 0, 100, 500, and 800–1000 s/mm2.

ADC maps can be generated from the index DWI data on the MR console itself, and have to be analysed qualitatively and quantitatively. PCa demonstrates high signal intensity on DWI at high b-values and low signal intensity/value on ADC maps [17, 107]. For a qualitative assessment, high b-value (800–1000) s/mm2 DW images and ADC maps should be used. These should be evaluated in combination with T2WI for anatomical detail. However, some normal prostatic tissue, especially in the TZ, may reveal high signal intensity in DWI and low ADC, thus mimicking a tumour. This may be overcome by using very high b-values (>1000 s/mm2).

For quantitative assessments, ADC values are used. However, there is variability when using different field strengths, different b-values, and different models to fit the data [17]. Also, there is a considerable inter-patient variability. Thus absolute values should be used with care. It provides information about the tumour's aggressiveness, and improves specificity in prostate cancer detection compared with T2WI alone. DWI correlates well with tumour volume in the prostatic lesion index. It should, therefore, be part of routine assessments of patients with prostate cancer [67]. DWI is, however, affected by magnetic susceptibility effects resulting in spatial distortion and signal loss. Large b-values are required to suppress normal prostate tissue background signals and ADC maps should be used to minimise T2 shine-through [107].
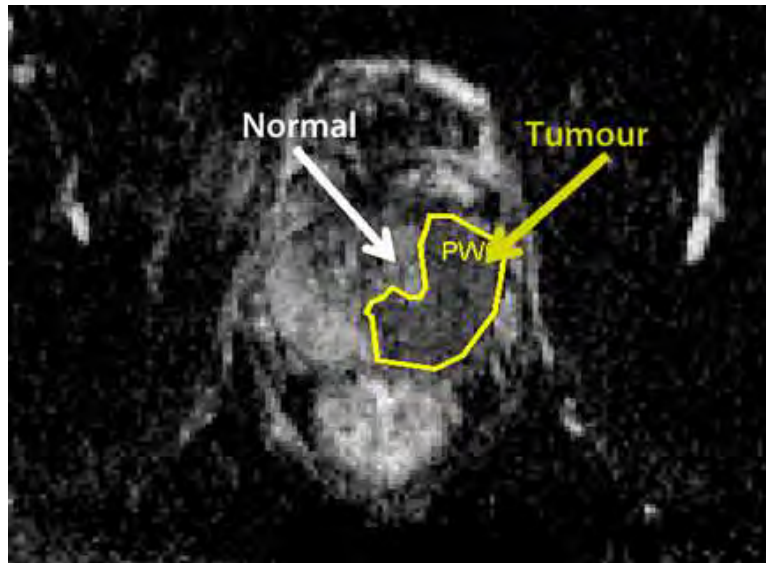
Figure 5.3: Example of a DWI image from a prostate study.

Perfusion-weighted imaging (DCE)

Perfusion imaging is based on Dynamic Contrast Enhancement (DCE) of the signal during the first pass of a contrast agent. The theoretic underpinnings of this vascular technique are based on tumor angiogenesis. In fact, there is a relationship between abnormal perfusion and neoangiogenesis in tumors [88, 13, 4].

DCE-MRI following the administration of a gadolinium-based contrast medium is the most common imaging method for evaluating a tumour's vascularity. As a normal prostate is also highly vascular, a comparison of pre- and post-gadolinium images is usually insufficient to discern prostate cancer [14]. DCE-MRI consists of a series of axial T1WI gradient echo sequences covering the entire prostate during a bolus injection (2–4 mL/s) of gadolinium-based contrast medium [14]. T1WI DCE-MRI imaging data can be assessed in three ways: qualitatively, semi-quantitatively or quantitatively.

DCE-MRI for prostate cancer detection, localisation, staging and recurrence detection is able to detect clinically important prostate cancer in 93% of cases [70]. Several studies have found that DCE-MRI is superior to T2WI for prostate cancer localisation. Although the literature is sparse, available data suggest that DCE may improve staging. Moreover, DCE-MRI is essential for the detection of post-prostatectomy and radiotherapy recurrences [67]. An example of a DCE image is depicted in Figure 5.4 (example of normal and tumor tissues).
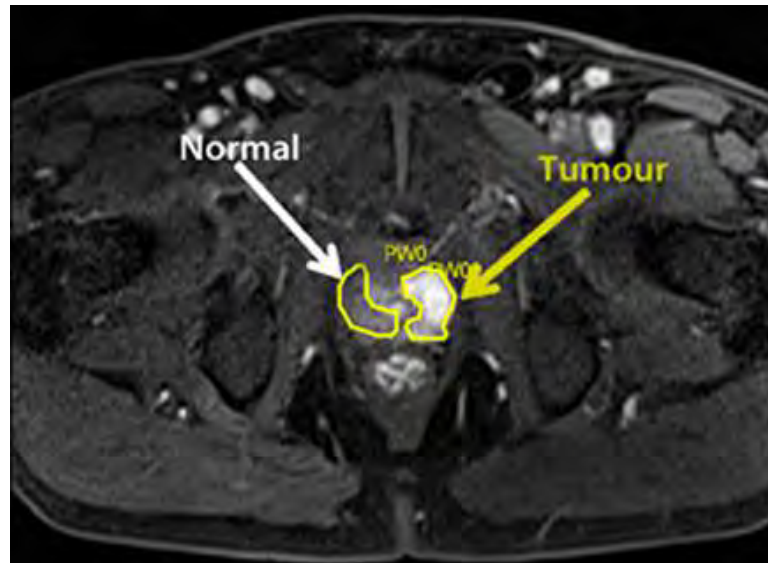
Figure 5.4: Example of a DCE image from a prostate study.

Magnetic Resonance Spectroscopy (MRS)

MRS is a technique that allows the study of metabolite concentrations by means of a 3D chemical shift imaging protocol [14]. This study is useful since healthy and cancerous tissues show different concentration levels. An LCModel is widely used for processing clinical single and multi-voxel spectroscopic data [146]. The LCModel method analyzes an in vivo spectrum as a linear combination of model in vitro spectra from individual metabolite solutions. Complete model spectra, rather than individual resonances, are used in order to incorporate maximum prior information into the analysis [145]. Specifically, PCa tissues show lower levels of citrate and higher levels of choline compared with healthy tissue [190, 34, 33, 138, 183]. Metabolic data are often presented in the form of concentration ratios eg. [Choline + Creatine]/Citrate. The exact ratio can vary with equipment and settings. For example, ratios at 3 T differ slightly from those at 1.5 T because of differences in the shape of the citrate spectrum [163]. However, it is generally accepted that peripheral zone voxels, in which the ratio of choline and creatine to citrate is at least 2 standard deviations higher than the average ratio in healthy tissues, are considered to represent possible cancer [189]. Voxels are considered highly suggestive of cancer if the ratio of choline and creatine to citrate is more than 3 standard deviations higher than the average ratio [89].

After post-processing, using commercially available software packages, spectral information is overlaid on T2WIs [14]. The relevant metabolites are citrate (marker of benign tissue), creatine (insignificant for diagnosis, but difficult to resolve from
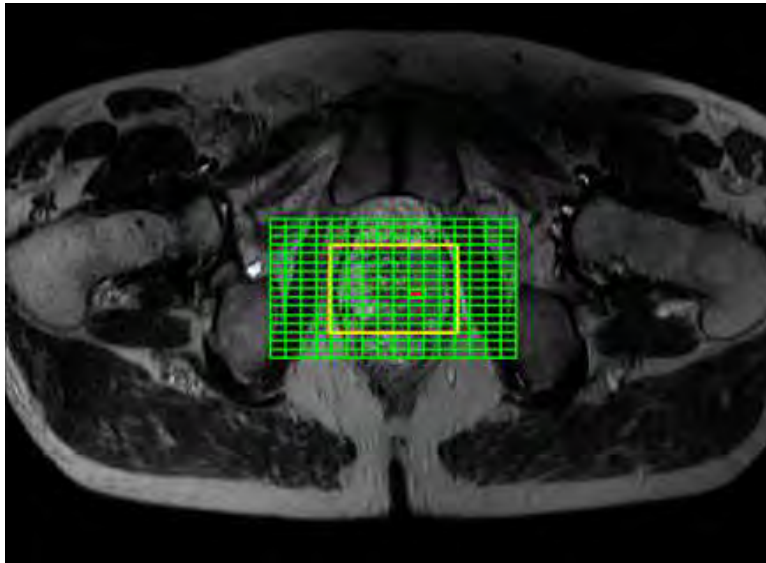
Figure 5.5: Example of a T2WI image with a spectroscopic grid surrounding the prostate gland as the region of interest.

choline), and choline (marker of malignant tissue). In quantitative analysis, the peak integrals of all metabolites are estimated by means of the choline-plus- creatine-to-citrate (CC/C) ratio. Cancer in PZ and TZ should have in at least two adjacent voxels a CC/C ratio exceeding respectively 2 and 3 standard deviations above the mean ratio [183]. In qualitative analysis, the peak heights of citrate and choline are visually compared.

Magnetic resonance spectroscopic imaging can be used to predict the presence or absence of cancer [14]. It also provides information about the lesion's aggressiveness, but does not give staging information owing to its poor spatial resolution. Thus, MRS is a valid tool for detecting cancer recurrence and monitoring therapy response [142].

An example of an MRS analysis is depicted in Figure 5.5 and 5.6. In the first example, a spectroscopic grid in a T2WI image is shown. The red square of the grid indicates the selected voxel to obtain the correspondent spectroscopic signal. It is a clear case of a cancerous tissue. The large amplitude of citrate resonances observed in the prostate tissue is due to an altered metabolism particular to this gland. Also, larger choline signals are associated with tumors in nearly all cancers [57].

(i) MRSI spectrum of healthy tissue   (ii) MRSI spectrum of cancerous tissue
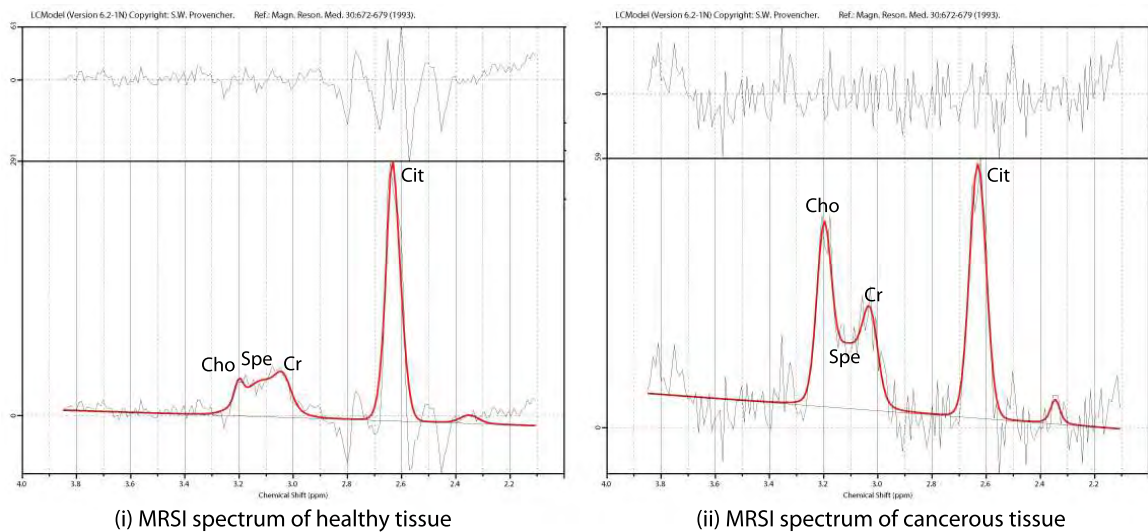
Figure 5.6: Example of a spectroscopic signal in a T2WI image in an MRS study. It shows a clear case of a detected tumor.

Multiparametric-MRI using T2WI in association with functional sequences such as DWI, DCE and/or MRS, has recently shown promising results in detecting and localizing prostate cancer [161, 78, 185, 167, 186]. In fact, recent studies which expand clinical interpretation and assess clinical performance in multiparametric MRI [75]. However, the difficulties in interpreting image, potentially discordant, from different MR pulse sequences are still a problem for a multiparametric-MRI. The main reason is because they still think that is not standardized enough to play a substantial role in patient management [161]. Also, the high number of possible combinations of shapes and signal abnormalities on the different MR sequences and the difficulty in interpreting them has discouraged some radiologists to use prostate multiparametric-MRI. Another limitation is that, by considering each sequence as simply positive or negative, it fails to account for variation in the degree to which a given sequence is abnormal, which can be substantial and may affect the potential significance of the lesion [195].

An overview and classification of the research related to Computer Aided Diagnosis (CAD) development for prostate cancer using multi-parametric MRI data is presented as a solution of some limitations [105]. The main problem is that no open multi-parametric dataset is currently available. Also, the use of standardized scoring systems such as PI-RADS still is not included in many research projects. PI-RADS goes beyond and provides a short verbal description of findings that correspond to each score from 1 through 5 for each individual sequence, there by taking a major step forward in addressing the issue of standardized interpretation. PI-RADS classification is presented in next sub-section [195].

5.1.3   PI-RADS classification

The increasing amount of data available to analyze a study has not been accompanied with the development of a single standardized way to report it. Often reports have been unstructured and written in a narrative way [38]. However, the European Society of Urogenital Radiology (ESUR) has recently proposed the Prostate Imaging-Reporting and Data System (PI-RADS) as the standard reporting structured scheme for prostate cancer [14]. Although other schemes exist, such as the Likert score [191], the use of PI-RADS is rapidly expanding [144, 156, 155, 150]. Therefore, we included the possibility to use this reporting scheme in our tool.

A scoring system similar to that successfully employed by breast radiologists (BI-RADS for X-ray mammography, breast ultrasound and MRI) should be used and prospectively validated for prostate MRI. According to ESUR prostate MR guidelines in 2012 [14] are being established as the common protocol in European countries [144, 156]. The PI-RADS definition is depicted in Figure 5.7.

5.2   Materials and data acquisition

The present working database is composed of anatomical, diffusion, and perfusion images and a set of spectroscopic data. In total, 10 patients (cases) with 640 images for anatomy, 140 images for diffusion, 6400 perfusion images, and around 10000 spectroscopic files for spectroscopy are used. It is important to note that the number of spectroscopic data files is very variable and depends on the prostate's size. The average age for all the patients is around 52 years.

Three-dimensional T2-weighted fast spin-echo (TR/TE/ETL: 3600 ms/143 ms/109, slice thickness: 1.25 mm) images (T2WI) were acquired with a sub-millimetric pixel resolution on an oblique axial plane.

DWI (TR/TE: 4200/101 ms) was performed using a pulsed gradient spin-echo technique with two b-values (100 and 800 sec/mm^2). The ADC map was generated from the raw data on a pixel-by-pixel basis.

The protocol was completed with DCE-MRI performed using a fat suppressed 3D T1 VIBE sequence (TR/TE/Flip angle: 3.25 ms/1.12 ms/10^0; temporal resolution: 6 sec/slab over approximately 5 minutes). A bolus injection of Gd-DTPA (Dotarem, Guerbet, Roissy, France) was administered at a dose of 0.2 ml Gd-DTPA/kg of body weight.

Three-dimensional MR spectroscopic data were acquired using a water and lipid-suppressed sequence. Data sets were composed of 16 x 16 x 16 phase-encoded spectral arrays giving a nominal spectral resolution of 0.14 cm^3. The spectral data
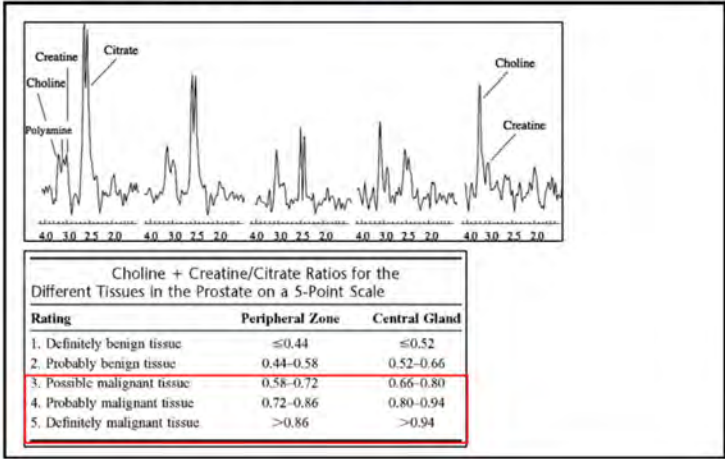
Score   Criteria

A1. T2WI for the peripheral zone (PZ)

1   Uniform high signal intensity (SI)
2   Linear, wedge shaped, or geographic areas of lower SI, usually not well demarcated
3   Intermediate appearances not in categories 1/2 or 4/5
4   Discrete, homogeneous low signal focus/mass confined to the prostate
5   Discrete, homogeneous low signal intensity focus with extra-capsular extension/invasive behaviour or mass effect on the capsule (bulging), or broad (>1.5 cm) contact with the surface

A2. T2WI for the transition zone (TZ)

1   Heterogeneous TZ adenoma with well-defined margins: "organised chaos"
2   Areas of more homogeneous low SI, however well marginated, originating from the TZ/BPH
3   Intermediate appearances not in categories 1/2 or 4/5
4   Areas of more homogeneous low SI, ill defined: "erased charcoal sign"
5   Same as 4, but involving the anterior fibromuscular stroma or the anterior horn of the PZ, usually lenticular or water-drop shaped.

B. Diffusion weighted imaging (DWI)

1   No reduction in ADC compared with normal glandular tissue. No increase in SI on any high b-value image ($\geq$b800)
2   Diffuse, hyper SI on $\geq$b800 image with low ADC; no focal features, however, linear, triangular or geographical features are allowed
3   Intermediate appearances not in categories 1/2 or 4/5
4   Focal area(s) of reduced ADC but iso-intense SI on high b-value images ($\geq$b800)
5   Focal area/mass of hyper SI on the high b-value images ($\geq$b800) with reduced ADC

C. Dynamic contrast enhanced (DCE)-MRI

1   Type 1 enhancement curve
2   Type 2 enhancement curve
3   Type 3 enhancement curve
+1   For focal enhancing lesion with curve type 2–3
+1   For asymmetric lesion or lesion at an unusual place with curve type 2–3

D1. Quantitative MRS for 1.5 T. Diagram references [50, 70]

Choline + Creatine/Citrate Ratios for the Different Tissues in the Prostate on a 5-Point Scale

Rating	Peripheral Zone	Central Gland
1. Definitely benign tissue	$\leq$0.44	$\leq$0.52
2. Probably benign tissue	0.44–0.58	0.52–0.66
3. Possible malignant tissue	0.58–0.72	0.66–0.80
4. Probably malignant tissue	0.72–0.86	0.80–0.94
5. Definitely malignant tissue	>0.86	>0.94

Score   Criteria

D2. Qualitative magnetic resonance spectroscopic imaging (MRSI)

1   Citrate peak height exceeds choline peak height >2 times
2   Citrate peak height exceeds choline peak height times >1, <2 times
3   Choline peak height equals citrate peak height
4   Choline peak height exceeds citrate peak height >1, <2 times
5   Choline peak height exceeds citrate peak height >2 times

Figure 5.7: A1 and A2. PI-RADS classification of T2WI

were processed using a modified version of LCModel (S. Provencher Inc.) for 1H spectroscopy of the prostate at 3T. Choline, creatine and citrate were all quantified individually.

5.3   ProstateAnalyzer framework

The ProstateAnalyzer framework [116] is a new medical tool in the form of a web-based medical application with the purpose of facilitating the inclusion of medical findings on existing prostate images using the combination of MRI techniques and MRS. In the literature, research works using some combination of different techniques focused on PCa are found. One of them used a combination between T2WI and DWI [67, 152], between T2WI and DCE [181], between T2WI and MRS [183], between DWI and DCE [99, 136], between DCE and MRS [51] or between T2WI, DWI, and DCE [102] among others. The novelty of our proposed framework is that it includes the use of the three common MRI techniques and MRS in a compact application. Generally, users have different tools to evaluate prostate images for MRI and MRS. However, it is necessary to compare regions of interest at the same time in order to determine pathologies or prostate lesions. For this reason, it is more efficient for each modality to provide a single tool that can easily manage all the data. In addition, a comprehensive list of clinical cases contained in the database is performed making a reported diagnosis using a PI-RADS classification with the presence of abnormalities.

Chapter 3 presented the generic system architecture framework. The proposed view model of the architecture was designed around a web-based application. Moreover, Chapter 4 presented the first framework also developed as a web-based tool within a web server and related to an XML database server used for data storage. The ProstateAnalyzer framework also uses this architecture system with some different particularities which will be detailed in this chapter. The ProstateAnalyzer also uses other digital databases of DICOM images, but in this case, containing prostate studies. Notice that a big difference between the MamoDB and ProstateAnalyzer is that the proposed DICOM database does not use a PACS system directly. A specific database contained in the server was created. The main reason for this is that, as different from the MamoDB, prostate studies have specific files related with different MR imaging techniques and Magnetic Resonance Spectroscopy (MRS) such as spectra information. In order to manage this information, experts and radiologists opted to make a new image database according to the MR sequence for each patient. Hence, the ProstateAnalyzer links this image database with the XML servers which in turn store prostate findings and reports using an XML database server. Also, the same eXist-db is used to store all the XML files related with the prostate studies.

The ProstateAnalyzer is a web-based platform that allows accurate overlays of prostate findings to be drawn and provides facilities to annotate other features related with lesions, like the subtlety of the lesions. Once the case has been fully annotated, the platform stores it in a database of annotations without modifying the original DICOM images.

5.3.1   Objectives

The main objective of the proposed system is to develop a new tool as a Java applet application, the ProstateAnalyzer, that encompasses the visualization and analysis of prostate MR images, as well as a new storage system of clinical diagnoses in a single package. The tool is able to characterize the morphological (location, shape, and size) and imaging features (mean and standard deviation of the signal, area and center of mass) of a region within the image, defined and annotated by an expert. Furthermore, it calculates the signal-time curve in perfusion studies and displays the LCModel signal spectrum for spectroscopic analysis [145].

Besides, it allows the analysis of the same image by different experts, an essential feature in order to obtain a robust evaluation. The final outcome of the analysis is summarized in terms of the PI-RADS protocol. With regard to the proposed architecture for storage of medical images, this should guarantee the connection and a secured access to the server, the DICOM database, and the system's storage of clinical diagnoses (XML database).

The main technological features when developing the ProstateAnalyzer are:

- Guarantee the connection and security access to the prostate image database using the DICOM standard format.

- Use of a search engine of prostate images based on their lexical information (patients, studies, series and/or images).

- Implement a Java applet tool, the ProstateAnalyzer, in order to manage the most common actions in image processing and visualization (zoom, contrast, PI-RADS lexicon, brightness,...).

- Implement a PI-RADS lexicon report for each prostate study.

- Allow the annotation of findings in the prostate images using manual segmentation tools.

- 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 prostate in a single XML file associated to the prostate study.

- Allow annotation by different experts doing a collaborative work.

The annotations are stored as XML files, one per prostate image and another one for each prostate study. An example of the XML structure of a prostate image is depicted in Figure 5.8. As can be seen, this XML contains information related to the prostate image and also with the list of overlays annotated by users. On the other hand, an example of the XML structure for each prostate study is depicted in Figure 5.9. In this case, global information related to the prostate study is generated by users. The most important part in this XML structure is that the information of a complete PI-RADS report is also saved. Comparing the two XML files, part of both structures is repeated for images of the same patient, but it allows a faster access to the overlay information.

5.3.2   ProstateAnalyzer interface

The ProstateAnalyzer was designed as a web application; hence the connection between the applet and the server can be both internal and external. The actual login process consists of an authentication measure with a unique username and password. One of the key elements in this application concerns the security of the data and the information across the network. In this sense, internal security measures to avoid vulnerabilities from the transfer of information are already included in the ZEND Server.

Figure 5.10 shows the control panel of the ProstateAnalyzer framework, which has been designed to take into account UI ergonomics and simplify access to the options in the system. This home page was designed with some of the functionalities of the MamoDB framework. The user can access all the functions in the system in a single click. There are two blocks of access elements: a drop-down menu bar and an expandable sidebar, which the user can make invisible. The latter's role is to provide direct access to the most used features. Moreover, the control panel contains a personal information section where a user can edit such information as profile image, username and password. Finally, the red squares depicted in Figure 5.10 is the direct access to the ProstateAnalyzer Java applet tool.

HTTP was originally designed for the transmission and reproduction of multimedia documents, but these encodings are not part of the HTTP standard. It is up to the applications themselves to break down and reassemble the information in order to transmit and receive it. Fortunately, Java applets are inherently based on the Internet Inter-ORB Protocol (IIOP). The ProstateAnalyzer uses this protocol based

```
<?xml>
<root>
  <patient>
     Information related to the patient
     <study>
        Information related to the study
        Information related to the prostate image
        <series>
          Information related to the series
          <list of overlays>
             <overlay 1>
                <label>User label</label>
                <user>Username</user>
                <listpoints>Coordinates points</listpoints>
                <centermass>Center mass coordinates</centermass>
                <area>Area of the overlay</area>
                <mean>Mean of the overlay</mean>
                <std>Standard deviation</std>
                <subtlety>Subtlety of the overlay</subtlety>
                <data>Data of the overlay</data>
             </overlay 1>
             ...
             <overlay N>
                Information of the overlay
             </overlay N>
          </list of overlays>
        </series>
     </study>
  </patient>
</root>
```

Figure 5.8: Example of the XML structure for a prostate image.

on the client/server computing model and a security algorithm implemented as a security mechanism to ensure the integrity of the medical data being transferred. In the literature, several architectures use the IIOP in their security applications [38, 137].

The most widely used regulations with respect to privacy and security are the Health Insurance Portability and Accountability Act (HIPAA) and the European Data Protection Directive 95/46/EC. Both regulations mandate health institutions

```
<?xml>
<root>
  <patient>
     Information related to the patient
     <study>
        Information related to the study
        <pirads>
            <piradsa1>Information related T2WI</piradsa1>
            <piradsa2>Information related T2WI</piradsa2>
            <piradsb>Information related DWI</piradsb>
            <piradc>Information related DCE</piradsc>
            <piradsd1>Information related MRS</piradsd1>
            <piradsd2>Information related MRS</piradsd2>
        </pirads>
        <comments>Information of the prostate study</comments>
     </study>
  </patient>
</root>
```

Figure 5.9: Example of the XML structure for a prostate study

to protect health information against unauthorized use or disclosure [202]. With the ProstateAnalyzer, the login/password signature allows knowledge of who is dealing with the individual data. Therefore, an audit-log for each study is produced, identifying the user who has read the data and thereby allowing one to generate audit trails on data access activities for any specific patient [104].

5.4   ProstateAnalyzer: Interactive Java applet tool

The ProstateAnalyzer 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 in the computer and is independent of the operating system. This allows full portability of the application.

The next sections explain the system's access and security policies, the search engine system, visualization options, and the main characteristics to annotate a prostate study and store it in the XML format.
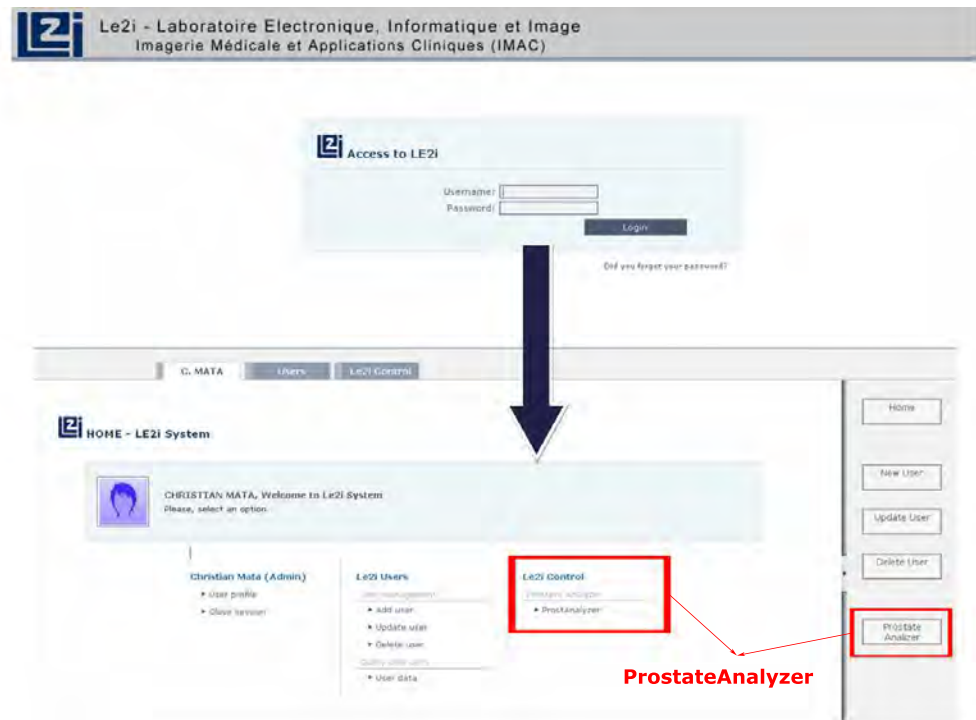
Figure 5.10: Example of the login access to the ProstateAnalyzer application (depicted in red)

5.4.1 System access

A system has been designed in order to link the ProstateAnalyzer application to a database of DICOM images and an XML server which stores the annotated files of the DICOM images. This architecture is supported by a Zend Server in order to allocate the ProstateAnalyzer application and an XML database provided by eXist-db [119]. This type of architecture has been used in previous projects for supporting other modalities in medical imaging databases such as PACS [116].

The ProstateAnalyzer application (Figure 5.11) communicates via socket TCP/IP with the Apache Server provided by Zend Server. This server is connected to a DICOM database, containing DICOM prostate images and with the eXist-db database, containing XML files. We opted for the creation of the XML annotation files in the database in order to preserve the integrity of the original DICOM database (usually stored in a PACS system).

Furthermore, in order to avoid vulnerabilities during the transfer of information, internal security measures have been implemented. Besides using control User ID private keys, the algorithms have systems that protect and encrypt the information
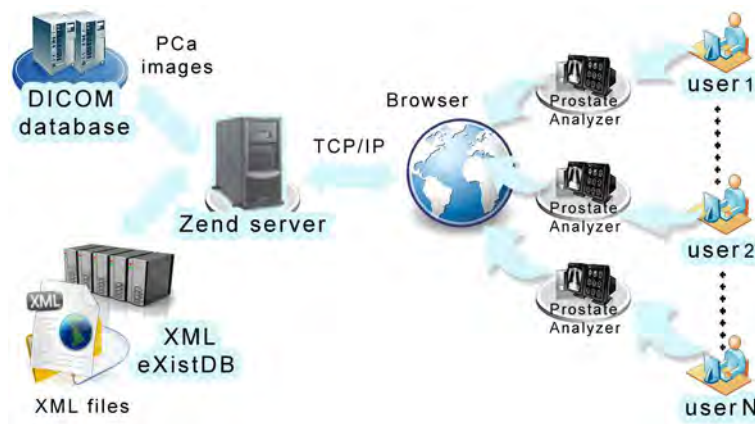
Figure 5.11: Network-based database system designed to implement the framework and the hardware used. The Zend server allows easy integration of the Apache server with database management tools. 423x234mm (72 x 72 DPI)

during its transmission. Currently, the SSL cryptographic protocol is used [178]. Finally, users can download a complete analysis made by one user in a PDF file such as a pre-medical report.

Actually, the security of the system is done via username and password, requested by the Java applet. Once the user is validated, a home window is displayed to welcome him to the ProstateAnalyzer tool as shown in Figure 5.12.

5.4.2 ProstateAnalyzer engine

The ProstateAnalyzer is related to the eXist-db manager, responsible for the XML database, incorporating its own server for access and management via web interface. We also opted for the creation of XML annotation files in the database. Again, all XML files are related to their corresponding DICOM images and contain the records provided by experts.

Figure 5.13 shows a scenario of the use of the eXist-db XML database. The database is configured at different levels in order to store the XML files associated with the studies, collections, status and images. Firstly, the ProstateAnalyzer application connects with the eXist-db to access a study, where each study contains a set of XML collections corresponding to T2WI, DWI, and DCE techniques. Consequently, for each dataset, XML files concerning prostate images are allocated. It is important to note that each XML file contains the annotations provided by all the experts. In this way, the main structure of the XML file is the same for the images of the same study, thus enabling faster access to the annotated information.
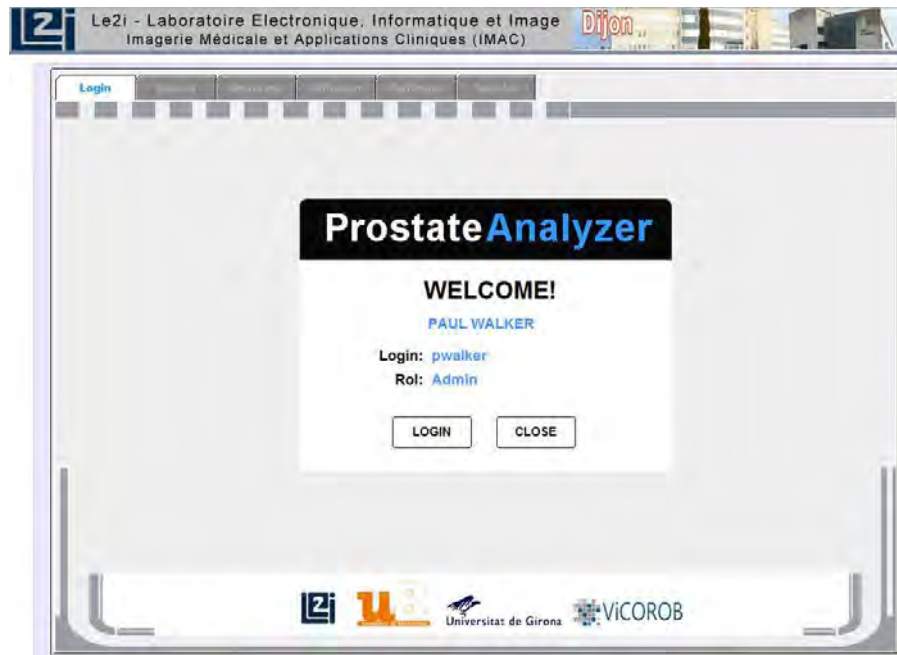
Figure 5.12: First screenshot of the ProstateAnalyzer applet, once the user has successfully logged in.

The search engine is the front-end for users once the ProstateAnalyzer has been accessed. An example of the search engine query page is depicted in Figure 5.14. There are two types of patient searches: search for all the patients or an advanced search. In the first option, the search engine returns a list with all the patients recorded in the database. In the second case, the system returns just the patients who share the same regular expression used as a query. To perform a search, an information concerning the four MR modalities described earlier is performed. For each item, the tool shows different attributes (patient name, comments, status) and the number of images provided by the concerned modality (anatomy, diffusion, perfusion, and spectroscopy). Moreover, each case is represented by a different color depending on the status of the study. When a state is color-coded as yellow, it means that the study has been partially analyzed ; a green color-coding indicates that the study has been completely analyzed and validated. Finally, white represents a study that has not yet been analyzed.

Moreover, at the status level for each study, an XML file associated with each user is saved. Consequently, once the user has logged into the ProstateAnalyzer, the corresponding XML file is obtained. It is important to note that the elements shown in the list are sorted by patient for the study number. For example, in Figure 5.14 the name of the patient appears (e.g. Patient 413) and also the number of the study
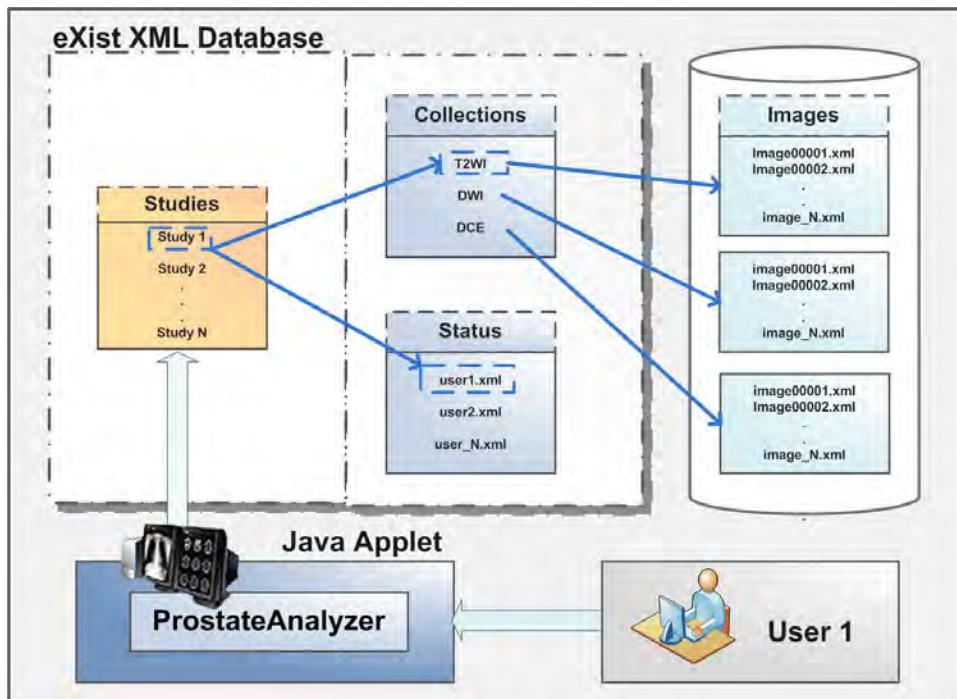
Figure 5.13: Connections to the eXist-db XML database within the ProstateAnalyzer and its interaction with the user.

"S1". Then a complete name is figured as "Patient 413 S1". It is a good way to identify the patient study number and to see it in a chronological way.

Then, an example of how a mammographic study is selected with the ProstateAnalyzer application is shown thus:

- Find a patient via the patient's name, comments (provided by users) or status (validated, partial or empty). Then the patient or a list of patients appears.

- Once a patient has been selected, the ProstateAnalyzer returns all the studies contained in the database.

- For each patient, the number of images contained in the database for anatomy, diffusion, perfusion and spectroscopy techniques appears.

5.4.3   Visualization

One of the main advantages of the ProstateAnalyzer application is that users can access a complete visualization of a prostate study by the four MR techniques.

Figure 5.14: Example of a standard query provided by the search engine.

Once a patient has been selected using the search engine, the application loads all the images and XML files and shows them in the visualization panel. It is easy to access for each visualization panel using the top menu (anatomy, diffusion, perfusion and spectroscopy). Moreover, for each panel and image from the selected technique, all the overlays annotated by users are depicted. Notice that these annotated ROIs are classified according to the prostate gland's region or tumour tissue detected. In order to simplify the name of the ROIs, the overlays are labelled as: Central zone (CZ), Peripheral zone (PZ), Transition Zone (TZ), Tumor (Tum1) and, in case two tumor regions appear, a Tumor2 (Tum2) is also labelled. In next sub-sections, an example of the four visualization modalities are presented.

Anatomy 3D-T2 Weighted Imaging (T2WI)

The ProstateAnalyzer offers a new visualization platform to provide annotations for different kinds of tissues (tumor number 1, tumor number 2, PZ, CZ, TZ, or others) obtained from T2WI studies. This visualization capability is combined with basic post-processing tasks such as zooming, gamma correction, user-specific ROI, and
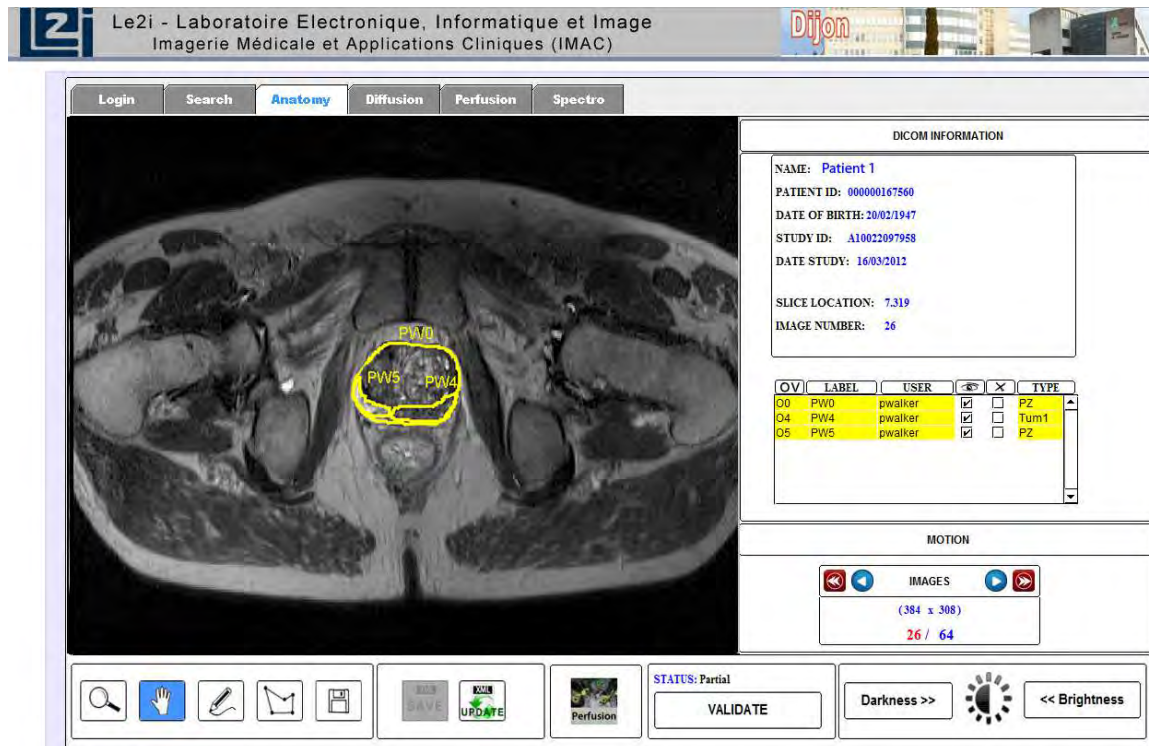
Figure 5.15: Shows an example of the anatomy (T2WI) panel visualization of the ProstateAnalyzer application with the ROIs annotated by users.

surface and volumetric measurements. In addition, the view provides information from the DICOM header, the status of the anatomical study (validated, partial, or empty) and the list of annotations added by different users.

Figure 5.15 shows an example of a visualization of the T2WI in a prostate cancer analysis. When an overlay is manually drawn, a label is automatically defined with the user's specified color. This label is defined using the first letter of the user's first name and surname followed by the number of annotations (in the figure, the first region of interest is identified by letters *AL0* in green). A table shows a list of color-coded overlays, specifying the username and the tissue type (indicated by the user). It also allows the visualization or to deletion of each overlay, although this latter option is only available to the expert who created it. Selecting an overlay of the table allows displaying its related information (the annotations filled in a pop-up window). Finally, after pressing the confirmation button from the pop-up window, the overlay is added to the table, displayed on the image and stored in the XML database.
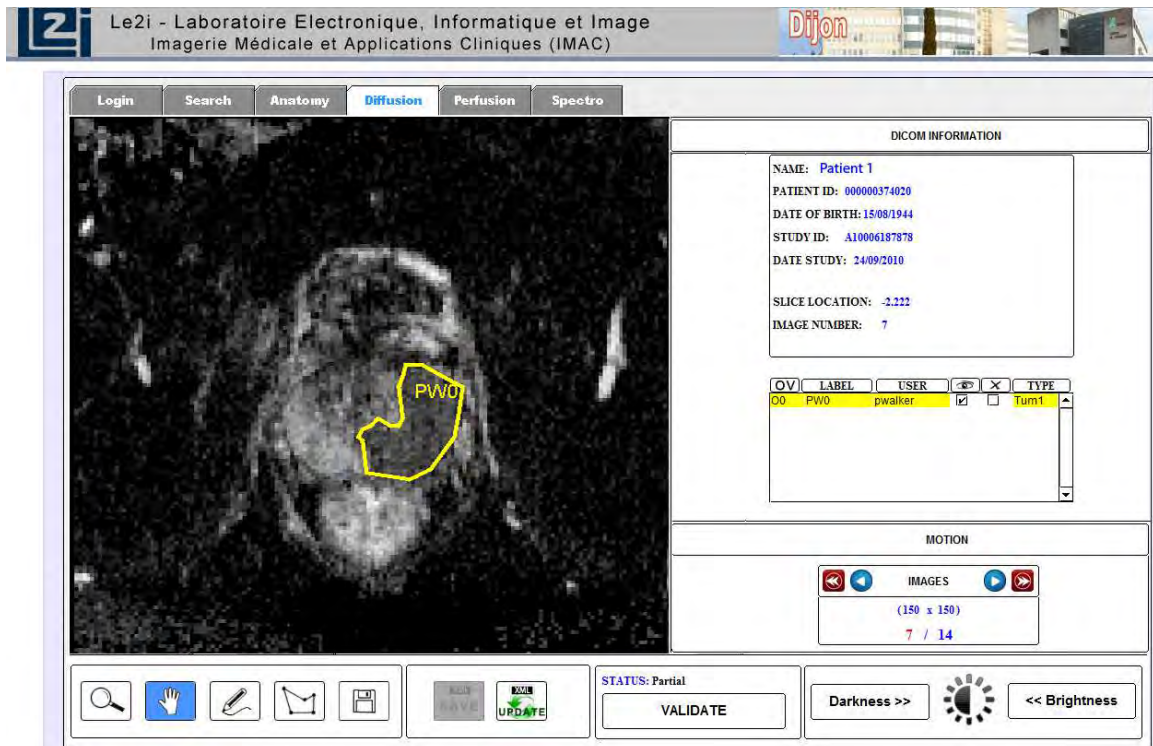
Figure 5.16: Shows an example of the diffusion (DWI) panel visualization of the ProstateAnalyzer application with the ROIs annotated by users.

Diffusion-Weighted Imaging (DWI)

The ProstateAnalyzer also provides a support tool for the DWI studies with similar basic post-processing tasks as previously presented for T2WI. Figure 5.16 shows an example of a diffusion prostate image with annotations provided by different users. In the example, a multiple user-specific ROI study is presented.

Perfusion Imaging Analysis (DCE)

Perfusion-weighted imaging (DCE) MRI is also catered for in the ProstateAnalyzer. This tool also contains basic post-processing tasks using the same features applied to T2WI and DWI. The most important difference with respect to other techniques is that perfusion images are displayed in time-space partitions. Furthermore, this viewer allows the display of the mean signal-time curve corresponding to the signal enhancement during the arrival of the contrast agent into the tissue.

Figure 5.17 shows an example of the ROI analysis in perfusion MR images. For each overlay, the mean signal from each ROI is calculated and displayed as
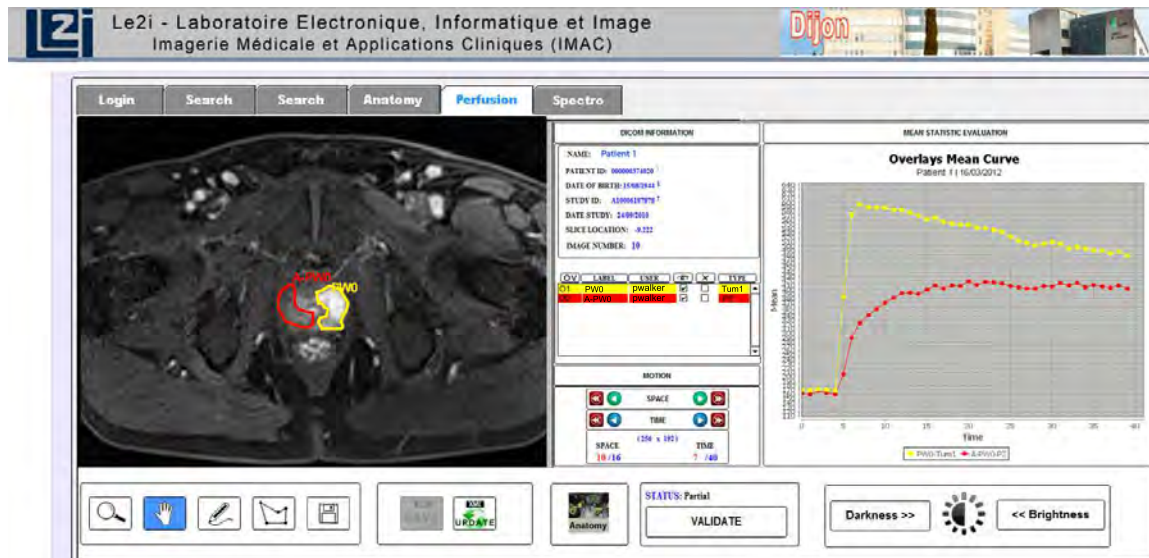
Figure 5.17: Shows an example of the perfusion (DCE) panel visualization of the ProstateAnalyzer application with the ROIs annotated by users.

the corresponding signal-time curve: the mean-curve is denoted with a different color according to its associated area. Two signal-time curves are depicted in the example. A healthy example is presented in the red signal-curve corresponding to the annotated region in the perfusion image (red overlay). In this case, the signal-curve presents a constant line meaning that the mean is more or less equal for all the images with a contrast agent concentration. On the other hand, the yellow curve corresponds to the yellow overlay annotated in the perfusion image. It is a clearly example of cancerous tissue. In difference from the healthy curve, the signal-curve presents a high peak (tumor detection) indicating the maximum mean value. Then, the line is decreases each time, finishing as a solid line.

Magnetic Resonance Spectroscopy (MRS)

The ProstateAnalyzer offers individual spectrum visualization from a 3D spectroscopic grid. Figure 5.18 presents an example of an MRS study with an LC Model-processed signal spectrum. This view displays a set of 3D-T2 images allowing the spatial location of the spectroscopy study. For each spectroscopic grid, there is a corresponding image. An example of a healthy spectrum tissue is presented. In order to obtain a specific spectrum, the user must click on the corresponding voxel (highlighted in red) within the grid. The LC Model is widely used for processing clinical single and multi-voxel spectroscopic data [146]. It allows individual and batch analyses of the main metabolites within the spectra of the prostate: citrate,
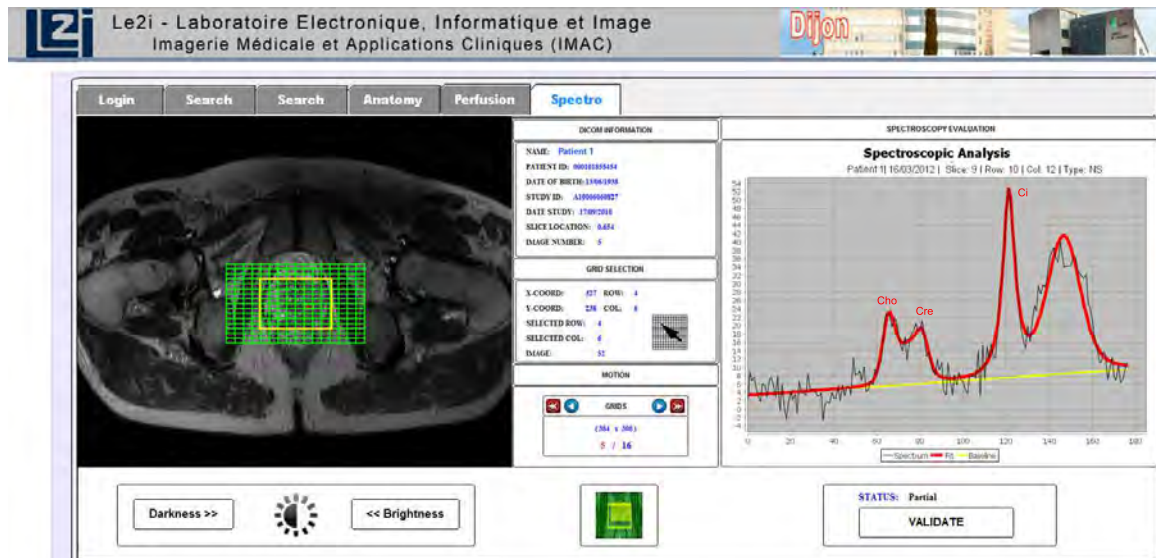
Figure 5.18: Shows an example of the spectroscopy (MRS) panel visualization of the ProstateAnalyzer application with the spectroscopic signal for a specifically selected voxel. In this case, healthy tissue is depicted in the spectroscopy signal.

choline, creatine, and spermine.

5.4.4 Interaction between MRI and MRS

One of the most important features of the ProstateAnalyzer is the interaction of the imaging modalities among themselves. One of the actual necessities is the analysis of different studies at the same time. The idea of making a new annotation in an MRI image and have it projected in another technique is an important advantage for the diagnosis. The ProstateAnalyzer, compared with other applications (discussed in Chapter 2), offers this possibility. One example of this case could be significant when an overlay is drawn on a 3D-T2 image, and how it can then be automatically annotated on the corresponding perfusion image or viceversa. Also, in the same way, how can the corresponding spectrum signal in the selected 3D-T2 image be obtained at the same time. It is important to notice that the modalities MRI, MRS and DCE are registered for the same patient. In this sense, two interactions are presented: a) interaction between MRS and MRI spectroscopy and anatomy and b) interaction between two MRI techniques such as anatomy and perfusion.
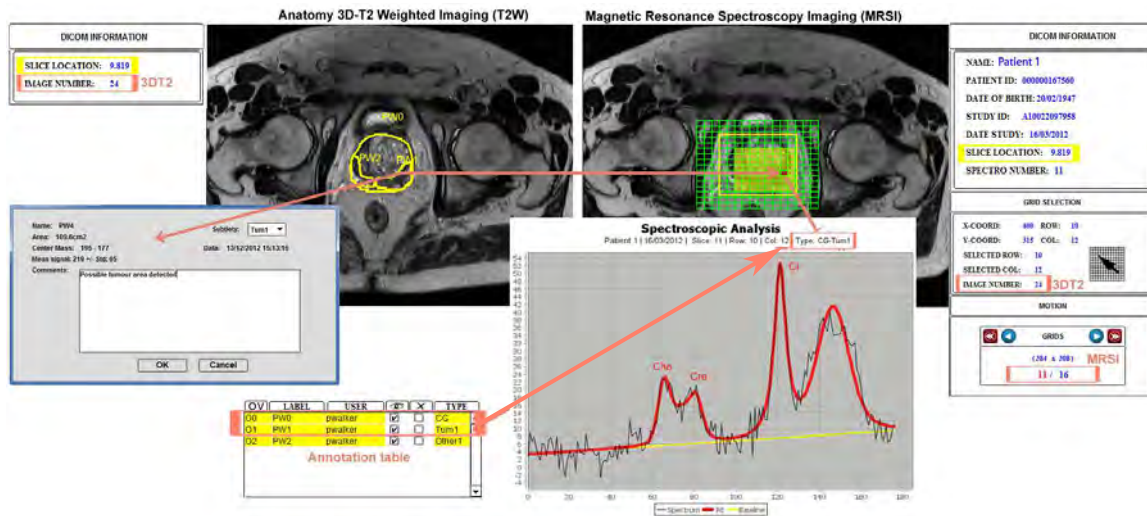
Figure 5.19: Example of the interaction between a 3D spectroscopic imaging study and T2W imaging.

A).- Interaction between anatomy and spectroscopy

The first interaction between anatomy and spectroscopy has been presented. As has been said previously, the general idea of this interaction is to explain how the corresponding spectrum signal once an overlay is annotated can be obtained in a 3D-T2 image. This interaction is especially important to apply regions in which a cancer tissue has been detected. For this reason, an example of a tumorous area is presented in order to show the usability of this interaction.

Figure 5.19 shows an example of the interaction between MRS and T2WI. The aim is to display the annotations obtained from the T2WI study panel on the spectrum illustrated in the graphical window. The way to link the processing is provided by the slice location (shown in the DICOM information panel of the MRS image, and boarded in yellow). In the MRS study panel, when the expert selects a voxel from the spectroscopic grid, the spectrum is displayed. The next step is to determine whether the selected voxel contains any annotations from the anatomical 3D-T2 panel. Notice that the T2-image used in MRS is provided to the same registration of 3D-T2 image. The way to identify the correspondence between T2WI and MRS is according the slice location. Then, it is loaded a spectroscopic data displaying the spectroscopic grid with the voxels in the image. If there is information, Prostate-Analyzer is displayed in the spectroscopic graphical area along with the spectrum. The example shows the selected voxel corresponding to an area (in the anatomic analysis) with two denominations for one tissue (named as *Tum1* and *CG* in the figure), and is also represented in the annotation table. This interaction is very
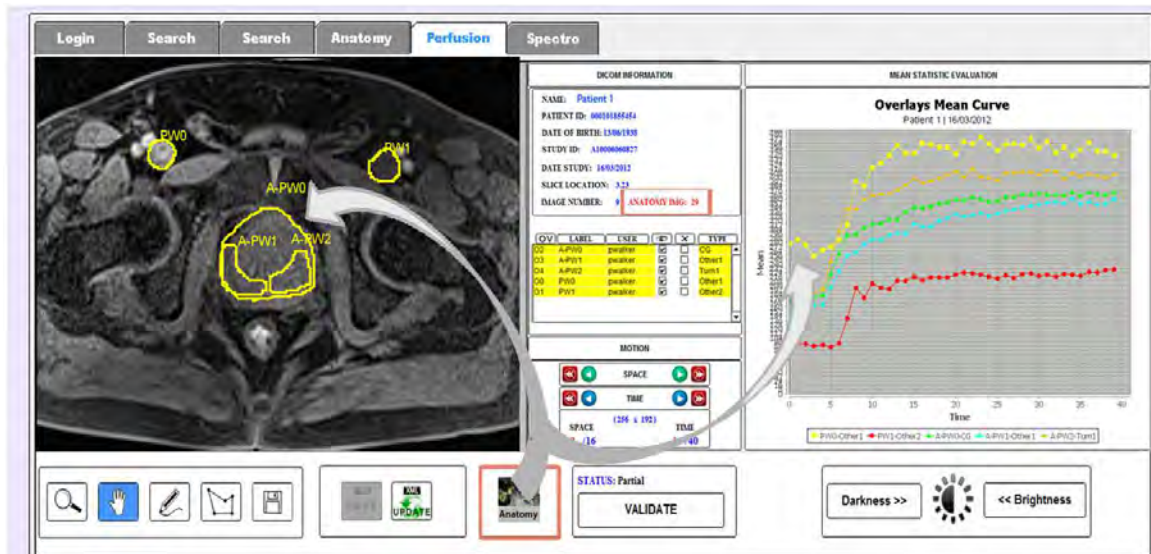
Figure 5.20: Example of the interaction between a perfusion slice and a 3DT2 image.

important because one can visualize the information of both techniques in the same representation.

B).- Interaction between anatomy and perfusion

In this case, the anatomic and perfusion studies do not use the same reference image due to differences in spatial resolution. ProstateAnalyzer application gets the correspondence between T2WI and DCE obtaining the slice location of the image which it is has been visualized in the MRI panel. In order to calculate the correspondence, the solution is to locate the image from a set of corresponding perfusion DCE-MRI images, as close as possible to the slice location of the selected anatomical 3D-T2 image.

Figure 5.20 shows an example of the analysis in a perfusion-weighted study using the ROI drawn on the corresponding 3D-T2 image. On the interface of the perfusion image analysis, when the "anatomy" button is pressed, all the annotations concerning the anatomy study are displayed. ProstateAnalyzer adapts the annotated regions into the correct position in the perfusion image, according to the spatial resolution (slice thickness) and pixel spacing. In order to distinguish the overlays between perfusion and anatomy, the viewer displays all the anatomy overlays and adds the prefix "A" before the label name.
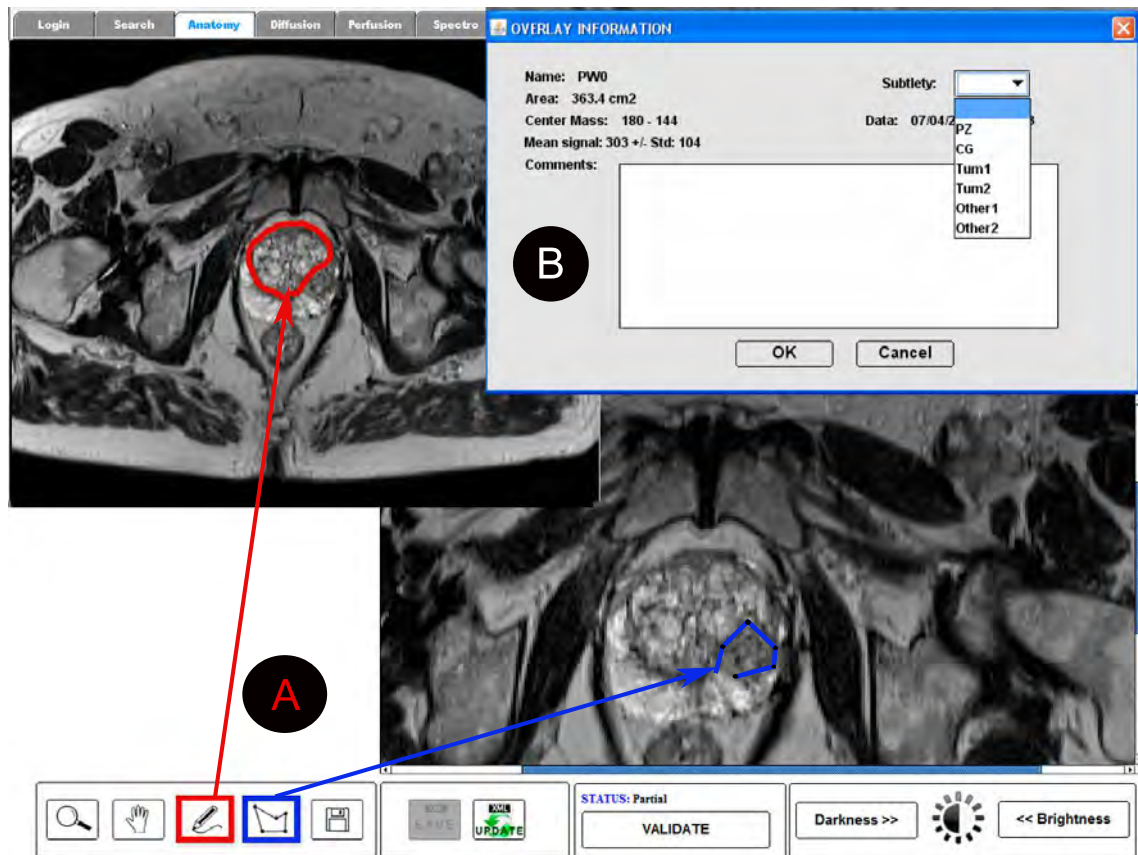
Figure 5.21: An example of manual annotation using ProstateAnalyzer application.

5.4.5   Manual annotation

A manual annotation example of the ProstateAnalyzer is shown in Figure 5.21. In the control panel, (see below in the figure) the user chose the manual segmentation tool and directly segments the ROI area depicted in the example as part "A". Then, a pop-up window containing information on the segmented overlay is displayed (shown in part "B"). As can be seen, the name of the overlay, area, mean, standard deviation, mass center, and data is generated. Also, the user can make comments concerning the overlay and select the type of subtlety (CZ, PZ, TZ, Tum1 or Tum2). Then, the manual annotation process finishes after a click on the "Ok" button. The new overlay appears automatically in the overlays table.

As can be seen in Figure 5.21, two different annotation options can be used by the user. An example of an overlay using a manual trace is shown in red (the button is shown in a red square). On the other hand, an example of an overlay using a polygonal trace is shown in blue (the button is shown in a blue square). Sometimes,

it is necessary to zoom the image and segment an area that requires a good accurate drawing. In this case, a polygonal trace is very useful (especially in a tumorous regions).

Once the clinical diagnosis has been performed, users have the option of creating an MR reporting, as depicted in Figure 5.22. As can be seen, a set of reporting information associated to each MRI and MRS modalities is displayed. Then, a "PI-RADS" button is available in the option panel. When it is pressed, a pop-up window appears where the user can report the findings for each technique according to the PI-RADS protocol.

The inclusion of the PI-RADS is carried out using a form with a drop-down list that includes the answers to the PI-RADS questions. Selecting the option, the corresponding number is assigned. The report is saved in an independent XML file and loaded into the XML database. The use of an independent XML file allows a faster search when looking just for similar cases in terms of PI-RADS scores.

Finally, the user has the possibility of changing the status of the prostate diagnosis. At the top of the pop-up window, a combo button is displayed to change the actual status from "validated" to "'partial" or viceversa.

In summation, we have implemented the following options to view and manipulate the overlays:

- Draw an overlay. Users can add new areas of interest in the prostate images by simply clicking the left and right buttons on the mouse.

- Annotate an overlay. Once an overlay has been created, a pop-up window like the one shown in Figure 5.21(b) appears. This window displays information related to the annotation, like the area and the mass center 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. This was shown in a previous figure.

- Delete an overlay. Users 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 4.19(b) is displayed.

- Make a PI-RADS report afterwards and introduce comments regarding the final diagnosis.
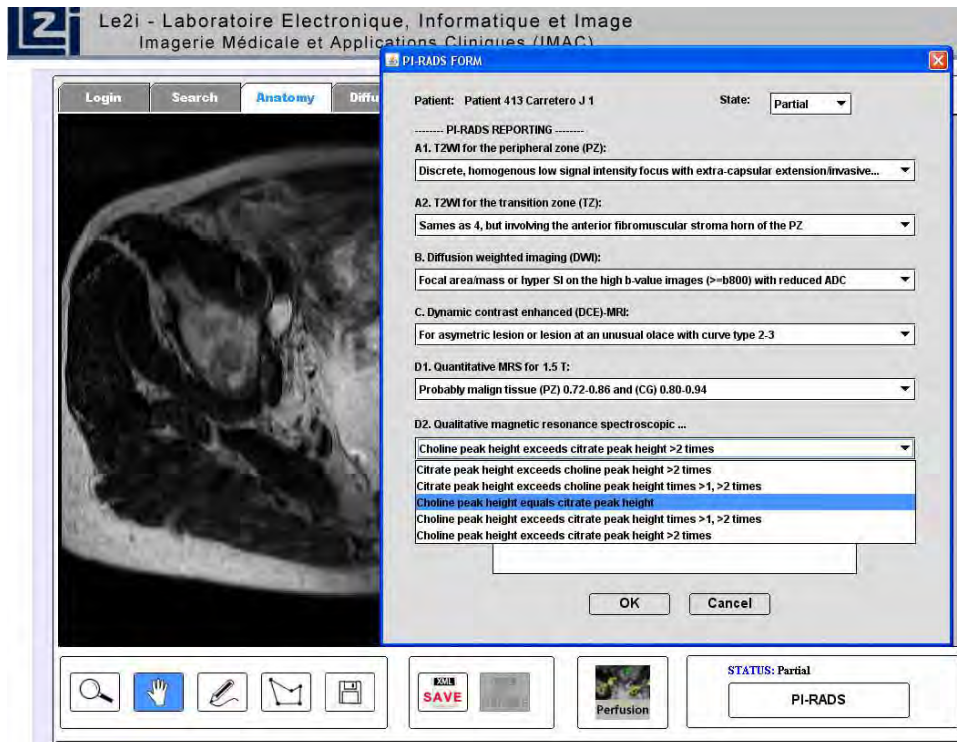
Figure 5.22: A pop-up window allows reporting the case according to the PI-RADS protocol.

5.5 Experimental results

5.5.1 An example of use

In this section a complete example of how to use the ProstateAnalyzer application is presented. One of the important features of ProstateAnalyzer is the multiparametric MRI examination (T2WI, DCE, DWI and MRS) of prostate cancer studies. The use of multi-parametric MRI prostate cancer imaging exams provides patients with more detailed information about their disease. In this sense, they can make the most informed treatment decision and/or understand whether the treatment they have received thus far has been effective. Also, it is able to use standardized scoring systems PI-RADS. In order to display the applications environment the results are fully detailed using the four visualizations platform. The main purpose is to demonstrate the usability of the ProstateAnalyzer application in the localization and analysis of a tumor.

The upper image in Figure 5.23 shows an example of a localized tumor displayed with T2WI. PCa usually shows a low signal intensity with T2WI that is well defined
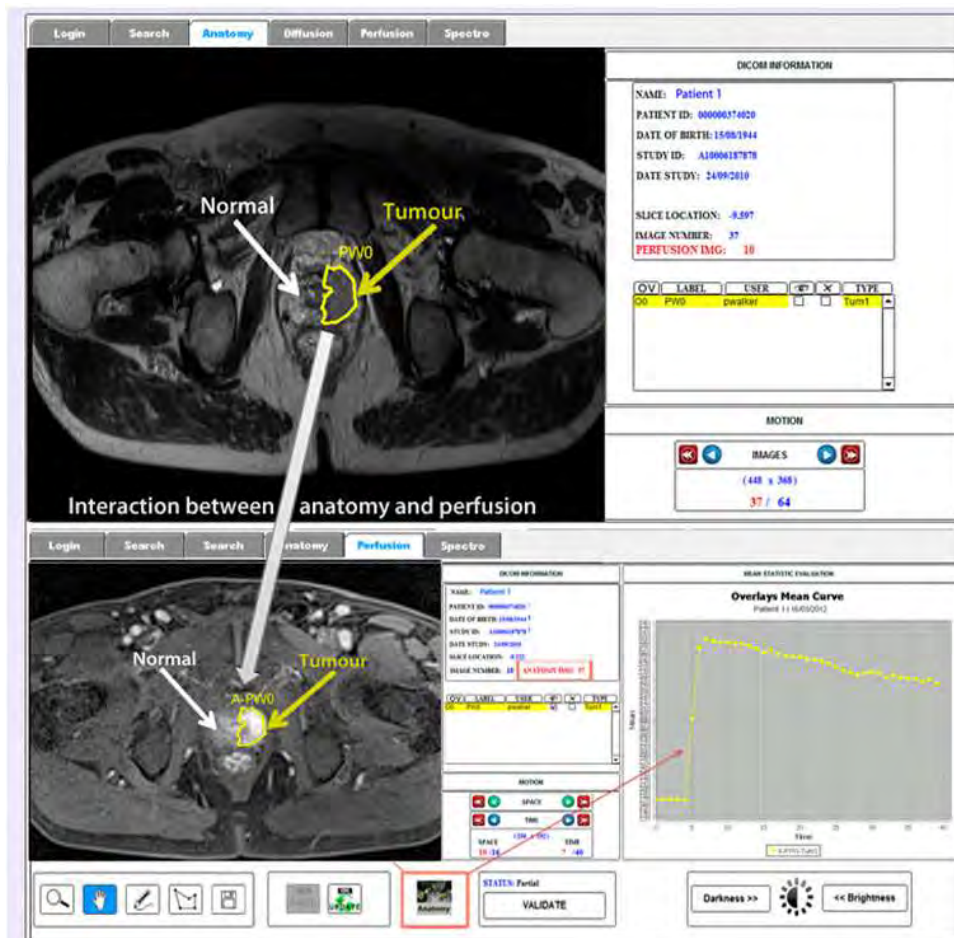
Figure 5.23: Interaction between T2WI and DCE.

with respect to normal peripheral zone tissue. However, it is not always easy to localize a pathologic area. One of the advantages of the ProstateAnalyzer application is it can analyze the same study using other techniques. Once a sequence in T2WI has been analyzed, the application offers an efficient solution to compare the same ROI in DCE, DWI and MRS. In DCE, the peak of the curve corresponds to the first pass of the contrast agent and indicates the rapid uptake of the contrast agent typically observed in cancer tissue.

For this case, we compare the mean curve of a representative tumor tissue with that of healthy tissue in Figure 5.24. The first curve (in yellow) is from the tumor tissue and the second from healthy tissue (in red). In this case, the first annotation corresponds to a tumorous region labeled "PW0" made in the perfusion panel. The second representing healthy tissue is annotated in the anatomical panel and displayed in the perfusion panel labeled "A-PWO". In the diffusion panel, the analysis
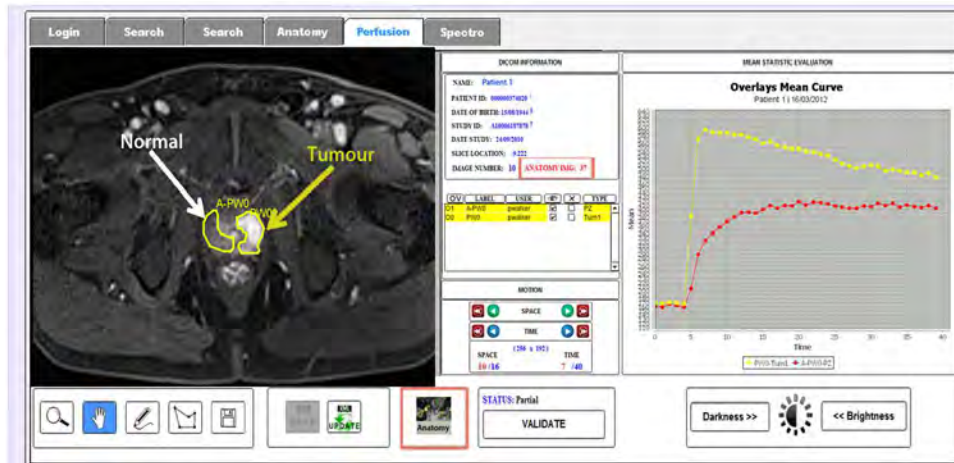
Figure 5.24: Example of two annotated regions (tumor and healthy tissue) represented in perfusion analysis. An illustration of the time-mean curve for DCE is also depicted in both cases (tumor in yellow, healthy area in red).
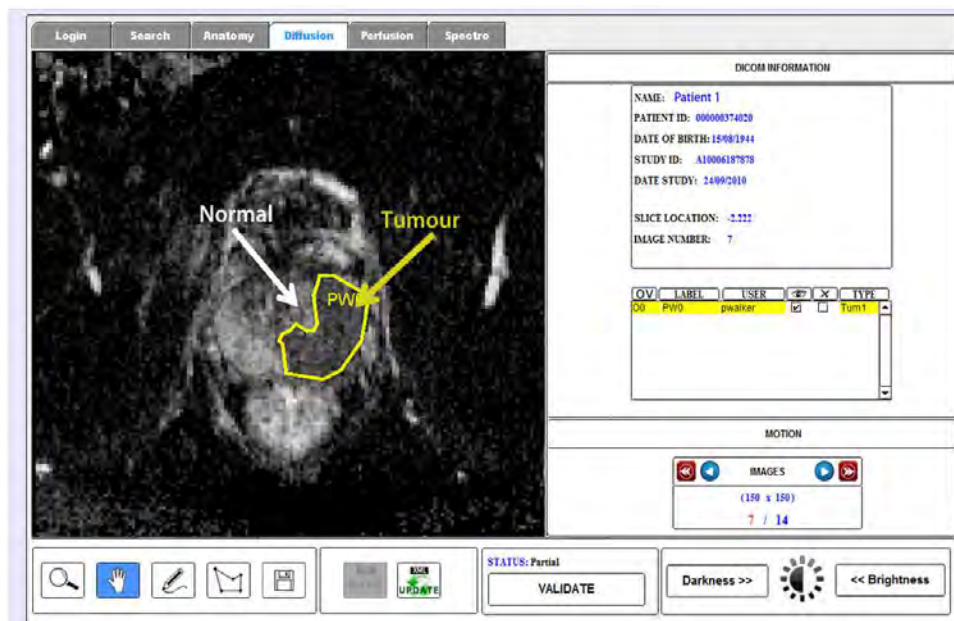
Figure 5.25: Example of a localized tumor in Diffusion-Weighted Imaging of the same prostate study.

also demonstrates the presence of a tumor (Figure 5.25). The tumor will correspond to the low signal in the DWI. It is important to notice that the user drew a different ROI from the one in Figure 5.23.

In the MRS panel (Figure 5.26), the example shows a selected voxel correspond-
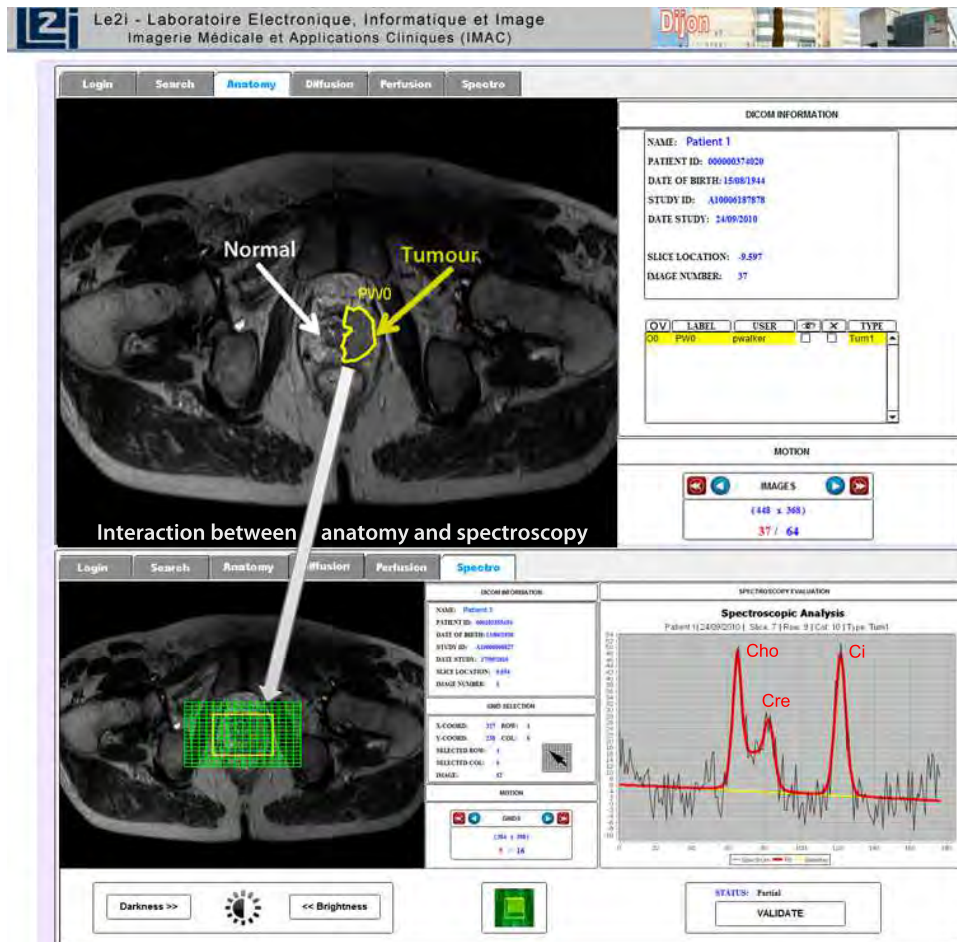
Figure 5.26: Example of the interaction between T2WI and MRS studies showing the spectra of the cancer voxel.

ing to an area of cancer tissue (Tum1). The LC Model-processed signal spectrum demonstrates how the signal is modeled in red from the noisy signal in black. Hence the different spectrum peaks are easily extractable. Graphically illustrating the levels of these metabolites is very important, since cancerous tissue is characterized in the spectrum as reduced citrate and elevated choline peaks. In the figure, the first two consecutive peaks correspond to choline and creatine, while the third one corresponds to citrate. It is clear that the relative levels of choline and creatine are very different from those observed in normal healthy tissue ((Figure 5.26)(d)), where higher levels of citrate and lower levels choline are observed.

Once the clinical diagnosis has been performed, the last step is to make the PI-RADS report saving or updating (depending on the case) the prostate study. In the same way as for a mammographic study (using BI-RADS), a PI-RADS report is

a very important point for the prostate diagnosis. An example of a PI-RADS report is shown in Figure 5.22 in Subsection 5.4.5.

5.5.2   Design effort benchmarks

The ProstateAnalyzer is currently being tested in the MR department of the University Hospital of Dijon (France). The present working database is composed of more than 1600 patient datasets. The time to access the Java Applet application (ProstateAnalyzer) by users is around $2 - 3$ sec in an intranet environment. It has been tested with the following common internet web browsers: Mozilla Firefox, Internet Explorer, Google Chrome, and Opera. It is preferable to use the Java Deployment Toolkit 6.0.160.1.1 or superior and to use the JavaTMPlatform SE 6 U16 6.0.160.1. The most time-consuming part corresponds to loading a prostate study. A typical study consists of a set of 64 anatomical images, 14 diffusion images, 640 perfusion images and a set of spectroscopic data containing up to 1000 files. It is important to note that the number of spectroscopic data files varies greatly and depends on the prostate's size (a complete spectrocopic data could be around 100MB). The average time to load and display a typical prostate study with previously presented data is around $60 - 70$ seconds.

The time necessary to record annotations into the database is around $3 - 5$ seconds for anatomical and diffusion images and around 12 seconds for perfusion images. In the case of the anatomical and diffusion images, the cost is calculated according to two steps: creating the XML file for each image and adding it to the eXist-db database. On the other hand, the computational cost is higher for perfusion images because of the calculation of the signal-time curve. In order to obtain a spectrum, the cost is around 5 seconds to display it in the graphic window. Around 100MB of spectroscopic data in a text file is loaded. The ProstateAnalyzer application loads this data and all the corresponding spectrum signal for each file is built.

With the new architecture, the ProstateAnalyzer can also be accessed using regular external network connections outside the intranet connection of the hospital. We tested the connection from Girona in Spain (several hundred kilometers from Dijon). Experiments were carried out by loading a large study consisting of a set of 64 anatomical images, 15 diffusion images, 720 perfusion images and a set of spectroscopic data containing up to 1200 files (corresponding to around 150 MB of the total transferred data). We loaded the case several times for a week, and the total elapsed time response varied between 100 and 130 seconds. When comparing the times provided by the internal and external connections using the same study, the difference was found to be around 40 to 70 sec. The difference was due to two main reasons. On the one hand, it depended on the instantaneous network load
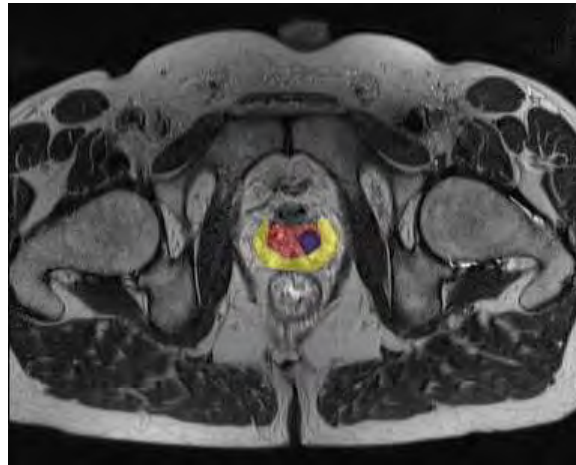
Figure 5.27: Example of an MRI prostate image from T2WI showing three anatomic areas: CZ (in red) and the PZ (in yellow) and the localization of a tumor (in blue).

and, on the other hand, on the number of users connected to the ProstateAnalizer server. The main conflict when multiple users are connected is in the downloading of cases, since the server must split the downloading process between the connected users. However, once the study has been loaded, the usability, visualization, and annotation times are independent with respect to the number of users.

5.5.3 Evaluation data analysis

The ProstateAnalyzer allows manual processing in regions of interest ROI, such as (CZ, PZ and tumor lesion), in prostate images using the combination of MRI techniques and MRS analysis. Due to the high volume of information present in the image, the purpose of the present study is to evaluate the variability of medical findings in prostate gland regions using T2WI between experts. Evaluation data analysis using the ProstateAnalyzer is described in this section.

We have chosen the T2WI because this provides the best depiction of the prostate's zonal anatomy and capsule. It is important to note that the TZ area is considered as part of the CZ zone in this evaluation. Experts draw prostate zones corresponding PZ and CZ but CZ is defined as a real anatomic (CZ+TZ). Then, TZ is considered as a part of the CZ because it marks the difference between two areas according to the T2WI images signal. Moreover, a T2WI sequence yields good contrast between PZ and CZ tissues [20].

Figure 5.27 depicts an example of the prostate gland showing the CZ (in red), PZ (in yellow) and tumor area (in blue). These represent anatomic areas used to make our evaluation. This study, made collaboratively, was performed using a pair

drawing or an independent drawing in two different ways: firstly, when a process is performed by a previous evaluation; and secondly, when this process is not evaluated. This methodology and some examples of the results obtained are also presented.

Evaluation procedure

The reason for making an evaluation in three different ways is due to the necessity of evaluating the variability of the medical findings in the different areas analyzed by different and independent experts. This evaluation procedure is preformed according to different experiments to analize the three main ROIs (CZ, PZ and tumor lesion) using the ProstateAnalyzer application. For each experiment, we asked the expert to draw ROIs if these areas are present in the image. A database encompassed by a total number of 640 prostate images acquired from 10 patients is used.

The first experiment ($E1$) consists of a prostate study evaluation provided by one expert. It consists of drawing ROIs in the prostate glands zones when they are required. For each ROI the area value is automatically calculated. Similarly, a second examination ($E2$) is carried out by a second expert in the same manner as $E1$. Finally, a third experiment is performed ($E3$) with the particularity that the ROIs made by the second expert are known. According to the procedure, pair drawing and independent drawing are chosen using the same ROIs. Indeed, the ProstateAnalyzer offers the possibility of showing all the ROIs drawn by every expert. The $E2$ is considered as the referent experiment. This means that the comparison procedure is done in two ways: firstly, between $E1$ and $E2$, and secondly between $E3$ and $E2$. Moreover, the volume of each considered structure (CZ, PZ and tumor) is considered. The volume is calculated as the surface of each area multiplied by the slice thickness.

Evaluation parameters

The correlation coefficient, the regression analysis and the Bland Altman [5, 22] plot were used to compare $E2$ with $E1$ and $E3$. Moreover, correlation estimation between the different experimentations was performed using a 2-sample t-test [154]. A p-value smaller than 0.05 was considered to represent statistical significance. Moreover, the contours from each experimentation were compared with each other. An edge-based approach using the Hausdorff distance [160] and region-based approach from the Dice index[60] were considered. The mean and the standard deviation of each parameter for the whole data set were calculated. Again, a 2-sample t-test was used to verify if there is any significant difference between the calculation of these parameters taking $E1$ and $E2$, and taking $E3$ and $E2$. Moreover, the number of cases in which one expert drew one area and not the other were counted and
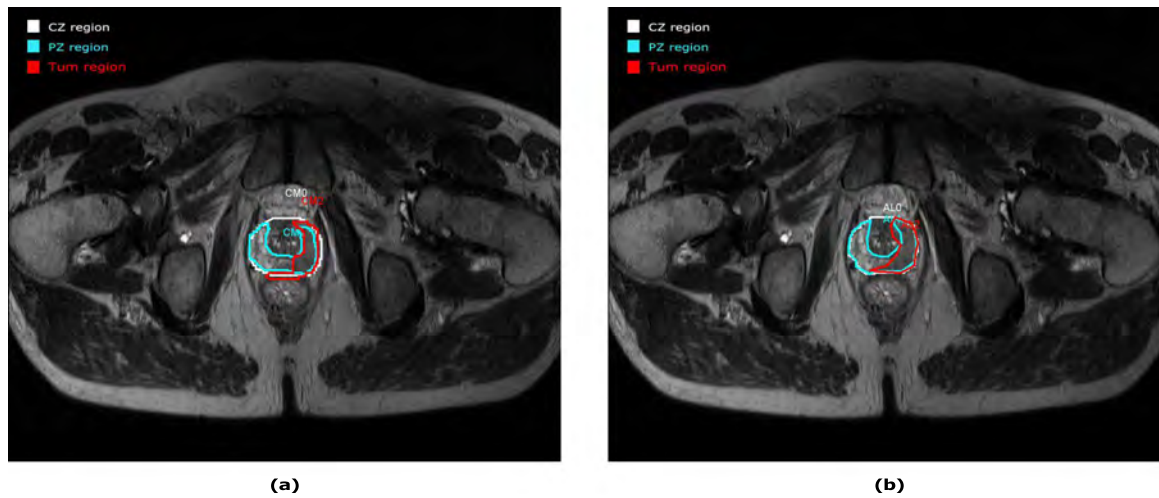
Figure 5.28: Example of a prostate study evaluation for E1 (left image) and E2 (right image).

presented as a percentage of the total number of cases.

Results

Two examples of a tumor analysis evaluation are presented. The first example is depicted in Figure 5.28. The left image in Figure 5.28 corresponds to an example of a drawing by the first expert and in the right image by the second expert.

Three ROIs are drawn in both images corresponding to CZ (white area), PZ (blue area) and Tumor (red area). When comparing the two drawings, a very good concordance between CZ and PZ areas can be observed. Concerning the tumorous area, a small deviation is appreciated but it could be considered relativity close between two experiments. However, not all the prostate studies were evaluated with good concordance between experiments. An example of this discordance is depicted in Figure 5.29. CZ and PZ have a good approximation between $E1$ and $E2$ but a big discordance is depicted in the tumor area. As can be seen, the tumor area (red region) corresponding to the $E1$ (left image) compared with the tumor area of $E2$ (right image) is different.

In this sense, it is possible to re-make the drawing in order to make a correction of previous prostate studies using the ProstateAnalyzer application. Figure 5.30 shows an example of a corrected case of $E2$. Now, the tumor area is approximately similar the same as the evaluation by $E1$.
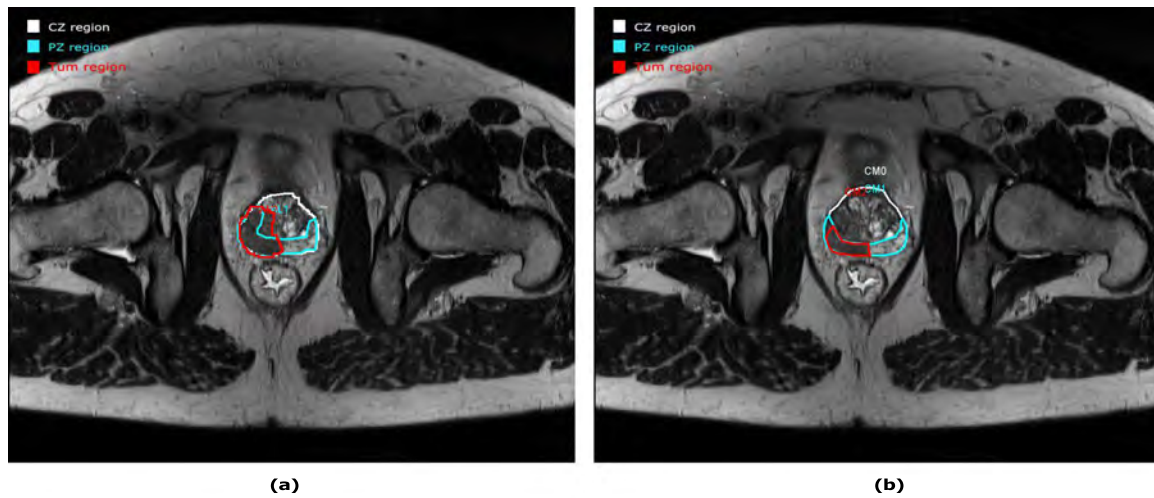
Figure 5.29: Example of a prostate study evaluation for E1 (left image) and E2 (right image). Discordance between both drawings for the tumor area is shown.
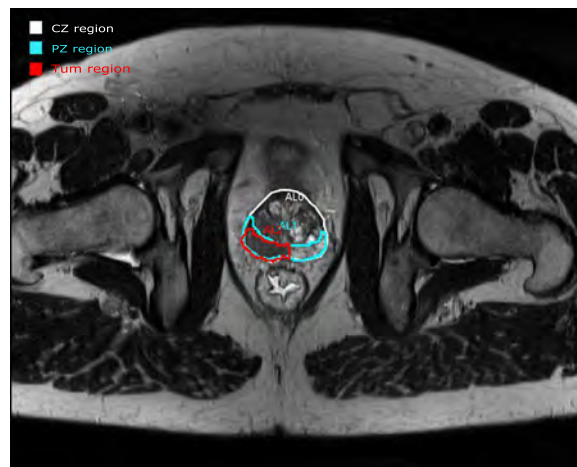
Figure 5.30: Example of a new evaluation of a prostate study for E3 concerning the tumor area.

Evaluation data

Several statistical analyses comparing the prostate gland's zones made during the examinations and according to the methodology of evaluation were performed. Two evaluation parts are performed using: a) anatomical parameters, and b) edge based and region-based parameters.

Patient	Processed slides	CZ		PZ		TUM	
		E1 vs. E2	E3 vs. E2	E1 vs. E2	E3 vs. E3	E1 vs. E2	E2 vs. E3
Patient 1	18	6%	6%	11%	0%	0%	0%
Patient 2	21	10%	10%	10%	10%	0%	0%
Patient 3	25	8%	0%	8%	0%	12%	4%
Patient 4	30	7%	0%	7%	0%	17%	0%
Patient 5	24	13%	0%	13%	8%	13%	0%
Patient 6	17	18%	0%	12%	0%	65%	0%
Patient 7	26	12%	4%	8%	0%	19%	0%
Patient 8	31	19%	10%	3%	6%	0%	0%
Patient 9	21	14%	0%	14%	0%	5%	0%
Patient 10	25	16%	0%	8%	0%	8%	0%

Table 5.1: Counting of the number of cases for each area (CZ, PZ and Tumor) that have not been evaluated by the two experts in each experiment.

Area	r		Regression line		Bland-Altman		t-test	
	E1 vs. E2	E3 vs. E2	E1 vs. E2	E3 vs. E2	E1 vs. E2	E3 vs. E3	E1 vs. E2	E2 vs. E3
CZ	0.95	0.98	y = 0.9x - 166	y = x - 12	-261.13 ± 168.20	-13.07 ± 118.09	0.01	0.36
PZ	0.91	0.94	y = 0.9x - 96	y = 0.9x + 21	-156.50 ± 95.71	-10.73 ± 84.60	0.01	0.32
TUM	0.96	0.98	y = 0.7x - 3	y = x + 3	-54.93 ± 64.34	-0.08 ± 27.13	0.02	0.47

Table 5.2: This table summarizes the analysis of CZ, PZ and the tumor area (in mm2) found in the prostate gland using the area as the anatomical parameter. $E2$ is the reference and is compared with $E1$ and $E3$. A P-value of < 0.05 between $E1$ and $E2$ for the area evaluation is present in all the cases.

A).- Anatomic parameters

The number of cases in which one expert drew one area, and not the other were counted and presented in Table 5.1 as a percentage of the total numbers of slices.

Considering all the patients, a percentage of the total mean for both experiments is calculated. For a CZ, this percentage is equal to 12% between $E1$ vs. $E2$ and 3% between $E3$ vs. $E2$. In the case of PZ, this percentage is equal to 9% between $E1$ vs. $E2$ and 3% between $E3$ vs. $E2$. Finally, for the tumor it is 13% between $E1$ vs. $E2$ and 0% between $E3$ vs. $E2$. It can be seen that in the second comparison between $E3$ vs. $E2$ the number of evaluated slices for each area is drastically reduced.

The correlation coefficient (r), regression line, Bland Altman and 2-sample t-test calculated for the three prostate gland zones are depicted in Table 5.2. In general, the results between $E3$ vs. $E2$ have improved. Concretely, the correlation coefficient of the area is improved between $E3$ vs. $E2$, being relativity close to one. Values of r close to 1 imply that there is a positive linear relationship between the data. Also, for every time that we have a positive correlation coefficient, the slope of the regression line is positive (as can be seen in the next figures 5.31, 5.32 and 5.33). In fact, it should be evident from this observation that there is definitely a connection between the sign of the correlation coefficient and the slope of the least squares line.

Figure 5.31 shows a regression analysis obtained from the comparisons between

Figure 5.31: Regression analysis obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of CZ.
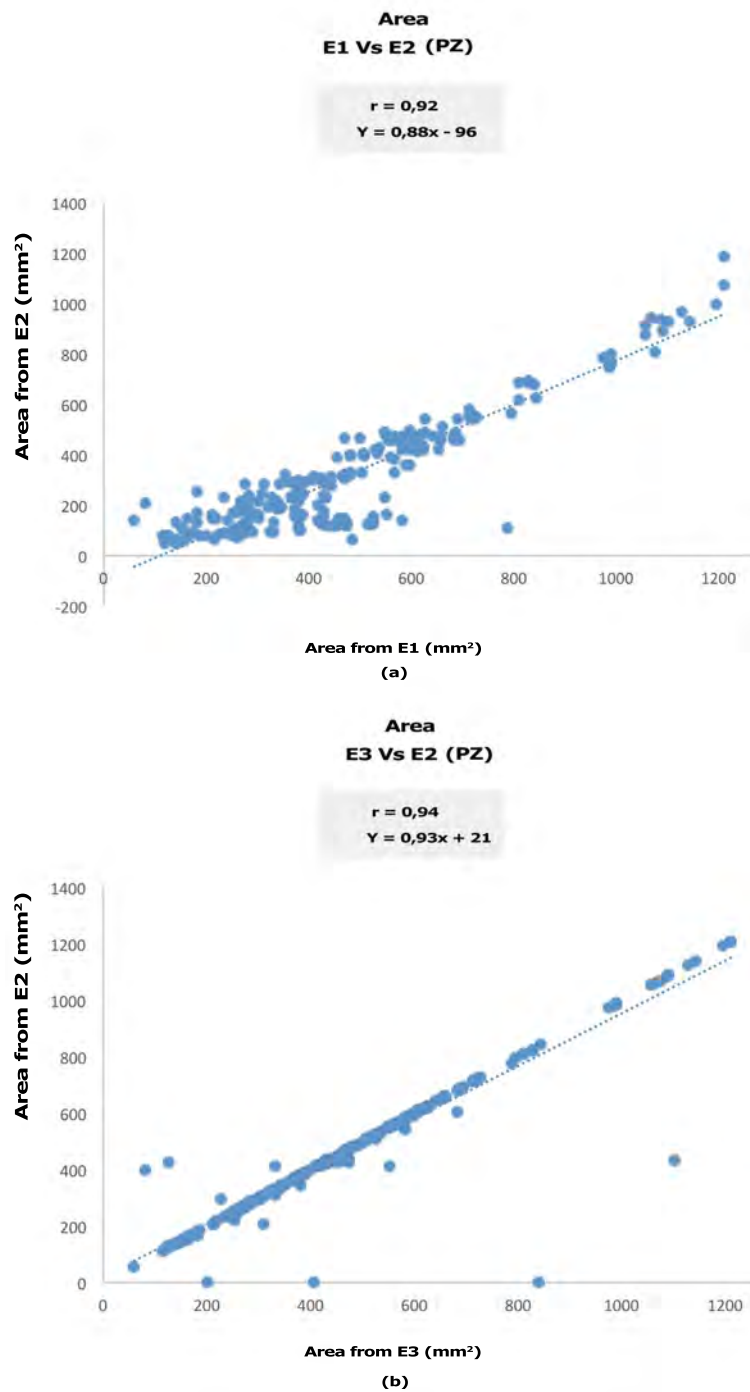
Figure 5.32: Regression analysis obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of PZ.

$E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of CZ. Also, Figure 5.32 regression analysis obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of PZ. In both cases, the best results are obtained concerning the comparison between $E3$ vs. $E2$. It is agreed that this is due to the slope of the regression line being close to 1. Moreover, our interest area is focused on the tumor area. In this sense, Figure 5.33 details the linear regression analysis concerning the evaluation of the tumor's area. A tumor area has been chosen due to its importance and because this area is more difficult to analyze and provides more variations among experts. When comparing the two obtained regression lines, an improvement is noted in Figure 5.33(b) compared with Figure 5.33(a). Firstly, a direct relationship exists in both cases because the slope of the line is positive and is closest to 1. Also, the concentrations of points in Figure 5.33(b) lie close to the regression line compared with Figure 5.33(a).

On the other hand, Bland-Altman plots are calculated for each area. Figure 5.34 shows a Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of CZ. In the same way, Figure 5.35 shows a Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of PZ. Both examples show agreement between $E3$ vs. $E2$. Again, our interest area is focused on the tumor area. In this sense, Figure 5.36 details the Bland-Altman plot concerning the evaluation of the tumor area. The mean of the difference between $E3$ vs. $E2$ is close to zero, meaning that there is no bias between the two measurements.

In the same way as the previous study, a tumor area has been chosen due to it being the most import case, more difficult to analyze and provides more variations. Figure 5.36 details the corresponding Bland-Altman plot concerning the evaluation of the tumor area. A visual assessment of agreement is made using a scatterplot of the two experiments, overlaying a line of equality. This agreement is depicted in Figure 5.36(b).

Volume is the second anatomic parameter to be evaluated. Again, in this evaluation the correlation coefficient, regression line, Bland-Altman plot and a 2-sample t-test are also calculated and described in Table 5.3.

The coefficient correlation (r) of the volume calculated for the CZ equals 0.98 between $E1$ vs. $E2$ and 0.99 between $E3$ vs. $E2$ respectively, PZ is 0.78 between $E1$ vs. $E2$ and 0.95 between $E2$ and $E3$, and for the tumor it is 0.94 between $E1$ vs. $E2$ and 0.99 between $E2$ vs. $E3$. On the other hand, Bland-Altman plot is also calculated for the volume. For CZ it is 40 ± 17 between $E1$ vs $E2$ and -0.9 ± 3 between $E3$ vs. $E2$, for PZ it is 20 ± 13 between $E1$ vs. $E2$ and 3 ± 12 between $E3$ vs. $E2$, and for the tumor it is 7 ± 6 between $E1$ vs. $E2$ and 0.4 ± 0.9
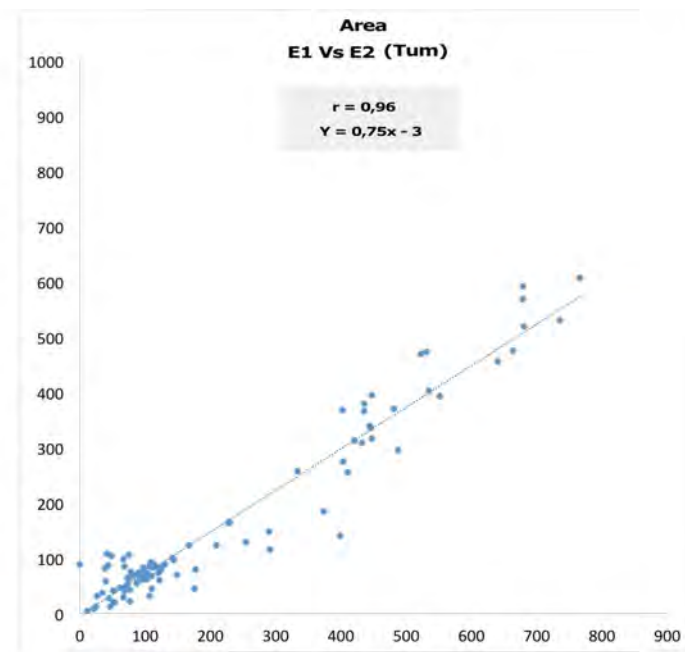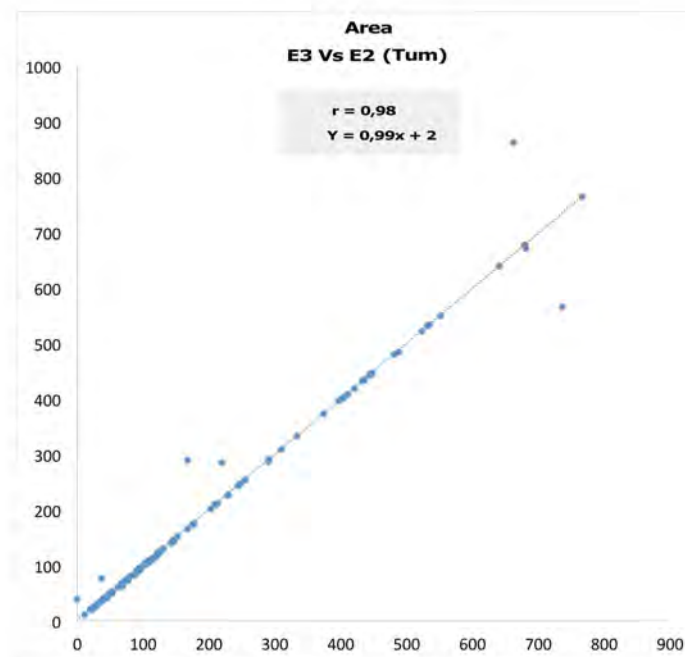
Figure 5.33: Regression analysis obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of a tumor.

(a)

(b)

Figure 5.34: Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of CZ. For the $E1$ vs. $E2$: Mean = -261; SD = 168 and for $E3$ vs $E2$: Mean= -13; SD=118.

Figure 5.35: Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of PZ. For the $E1$ vs. $E2$: Mean = -156; SD = 96 and for $E3$ vs $E2$: Mean= -11; SD=85.
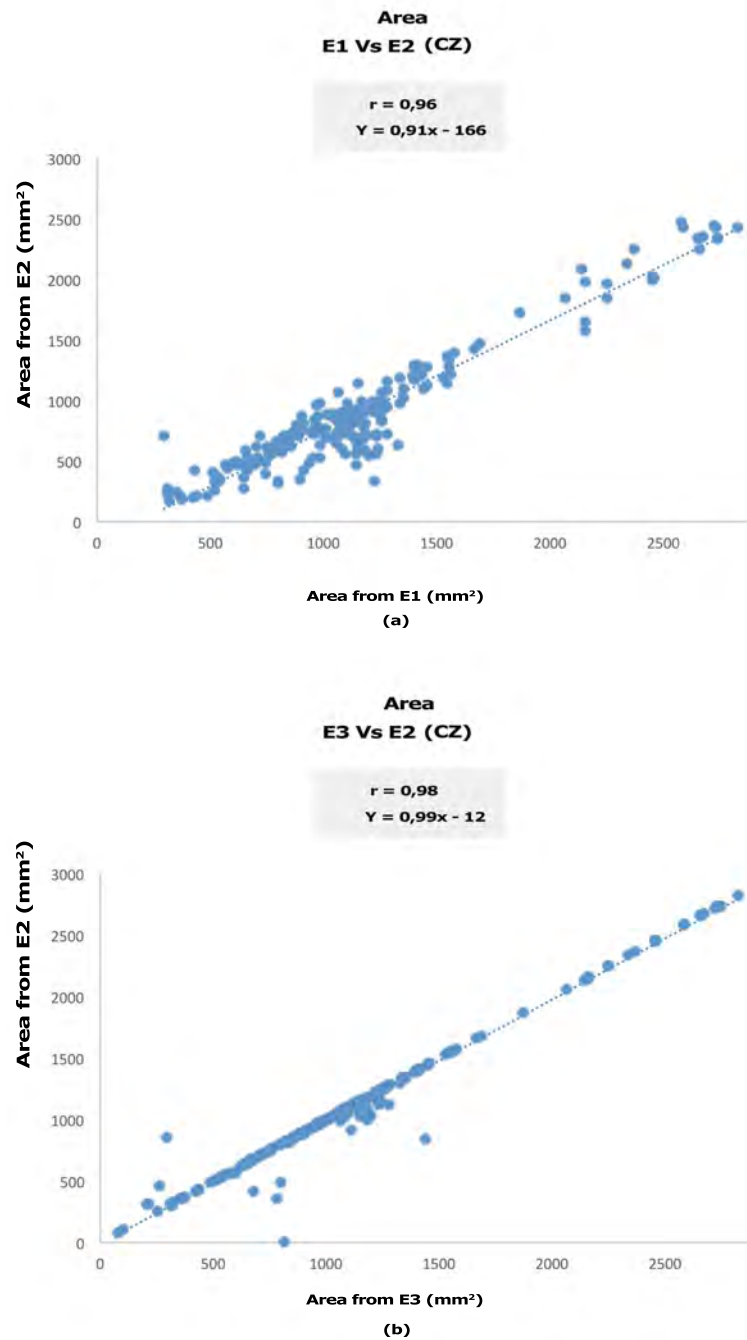
(a)

(b)

Figure 5.36: Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of a tumor. For the $E1$ vs. $E2$: Mean = -55; SD = 64 and for $E3$ vs $E2$: Mean= -0.1; SD=27.

	r		Regression line		Bland-Altman		t-test	
Volume	E1 vs. E2	E3 vs. E2	E1 vs. E2	E3 vs. E2	E1 vs. E2	E3 vs. E3	E1 vs. E2	E2 vs. E3
CZ	0.98	0.99	y = 1.6x - 20	y = 0.95x + 6	39.9 ± 19.43	-0.9 ± 3.07	0.02	0.96
PZ	0.78	0.95	y = 0.8x + 25	Y = 0.7x + 16	19.91 ± 3.07	2.91 ± 12.09	0.04	0.80
TUM	0.94	0.99	y = 1.5 + 1	y = x - 0.6	6.87 ± 6.40	0.4 + 0.9	0.24	0.47

Table 5.3: This table summarizes the analysis of the tumor area found in the prostate gland using the volume as anatomical parameter. $E2$ is the reference, and compare with E1 and E3. A P-value of < 0.05 between $E1$ and $E2$ for the area evaluation is present in all the cases.

	Hausdorff distance		Dice Index	
Areas	E1 vs. E2	E3 vs. E2	E1 vs. E2	E3 vs. E2
CZ	8 ± 3	4 ± 1	0,7 ± 0,2	0,9 ± 0,1
PZ	11 ± 5	5 ± 2	0,6 ± 0,2	0,9 ± 0,1
TUM	10 +- 4	8 ± 11	0,7 ± 0,1	0.9 ± 0,1

Table 5.4: This table summarizes the analysis of Hausdorff distance and Dice Index for the CZ, PZ and TUM area. A P of $< 0,05$ each time between $E1$ vs $E2$ and $E3$ vs $E2$ is present in all the cases.

between $E3$ vs. $E2$. However, the regression line for the volume is not good due to too little representation of the data and some influential points. In this sense, an influential point is an outlier that greatly affects the slope of the regression line. It has a disproportionate effect on the slope of the regression equation. For this reason, only the area has been analyzed as the most representative case of the anatomical parameter.

B).- Edge based and region-based parameters

The Hausdorff distance and the Dice index between the different drawings are presented in Table 5.4. Mean and standard deviation were calculated in order to compare between $E1$ vs. $E2$, and between $E3$ vs. $E2$. They are used to evaluate the same experiments. Again, between $E3$ vs. $E2$ an improvement is observed with respect to the results obtained between $E1$ vs. $E2$. The mean Hausdorff distance is reduced using the second comparison of experiments in all the cases. In the same way, the analysis of the Dice Index is around 0.9 for the same experiments providing a good result between $E3$ vs. $E2$, compared with $E3$ vs. $E2$.
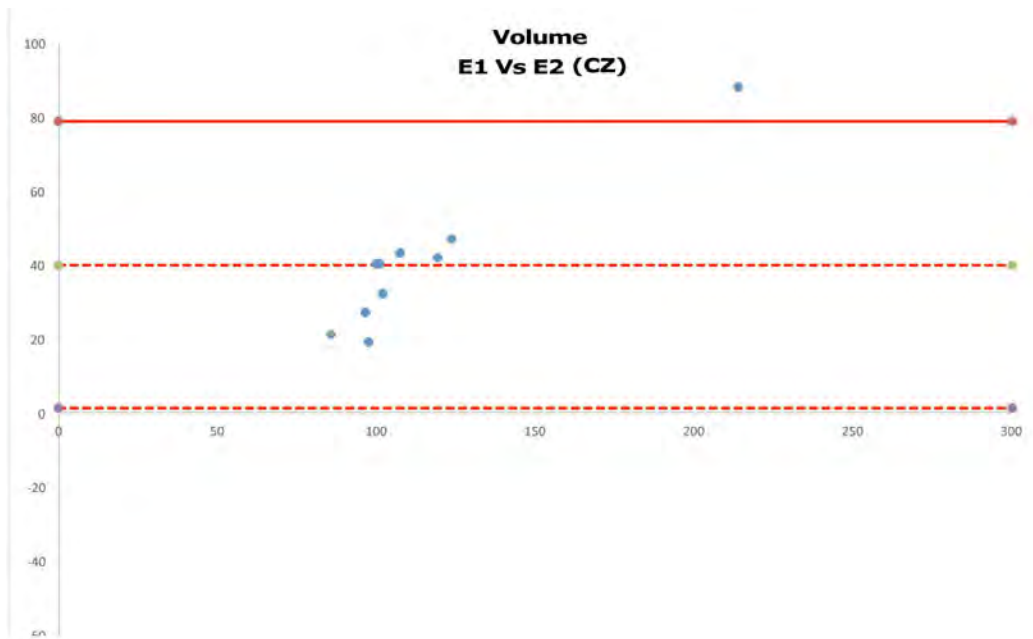
(a)
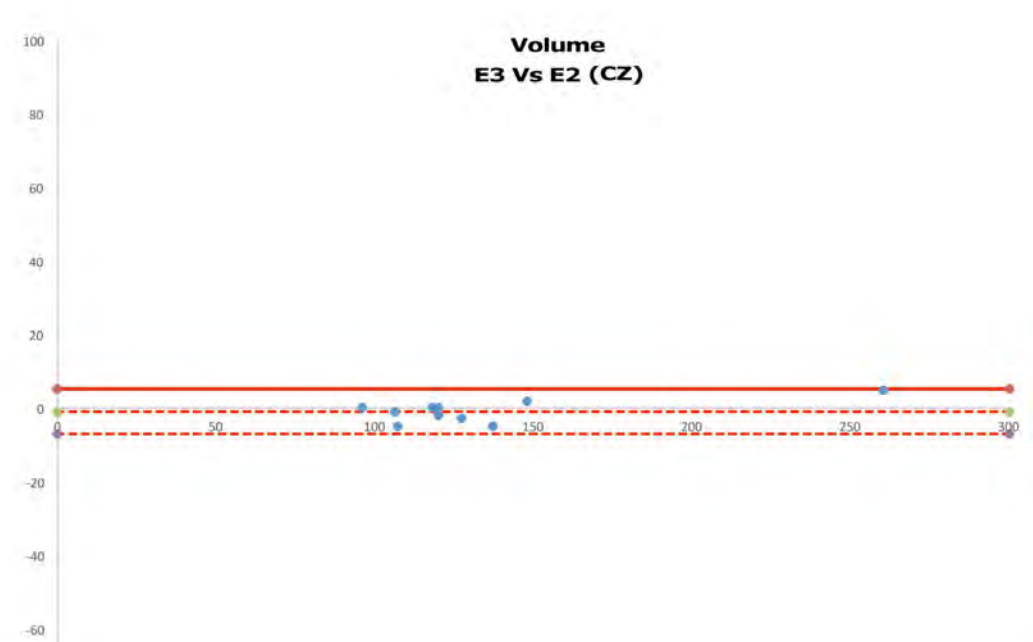
(b)

Figure 5.37: Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of CZ. For the $E1$ vs. $E2$: Mean $= 40$; SD $= 19$ and for $E3$ vs $E2$: Mean$= -0.9$; SD$=-3$.

(a)

(b)
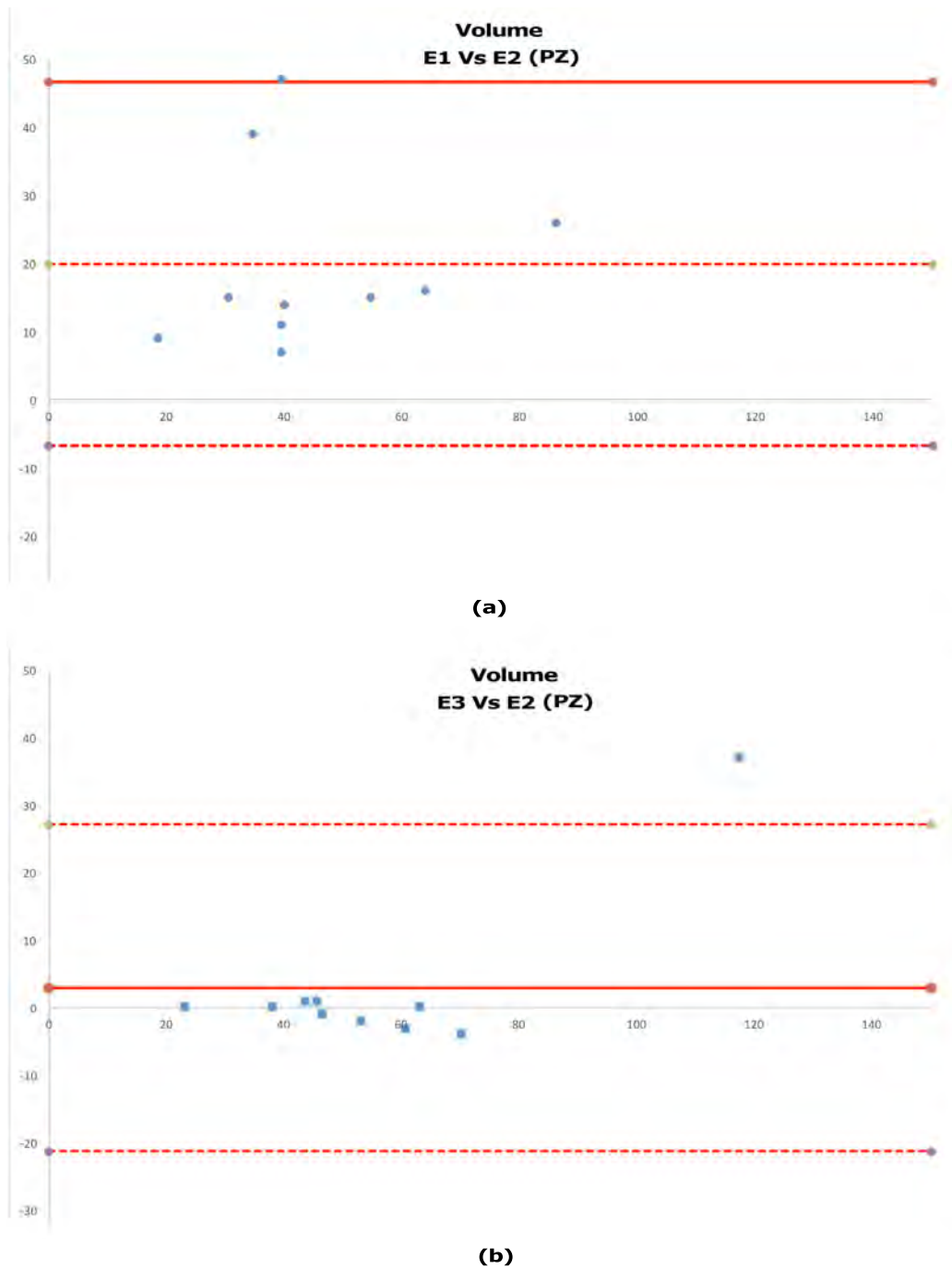
Figure 5.38: Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of PZ. For the $E1$ vs. $E2$: Mean = 20; SD = 13 and for $E3$ vs $E2$: Mean= 3; SD=12.
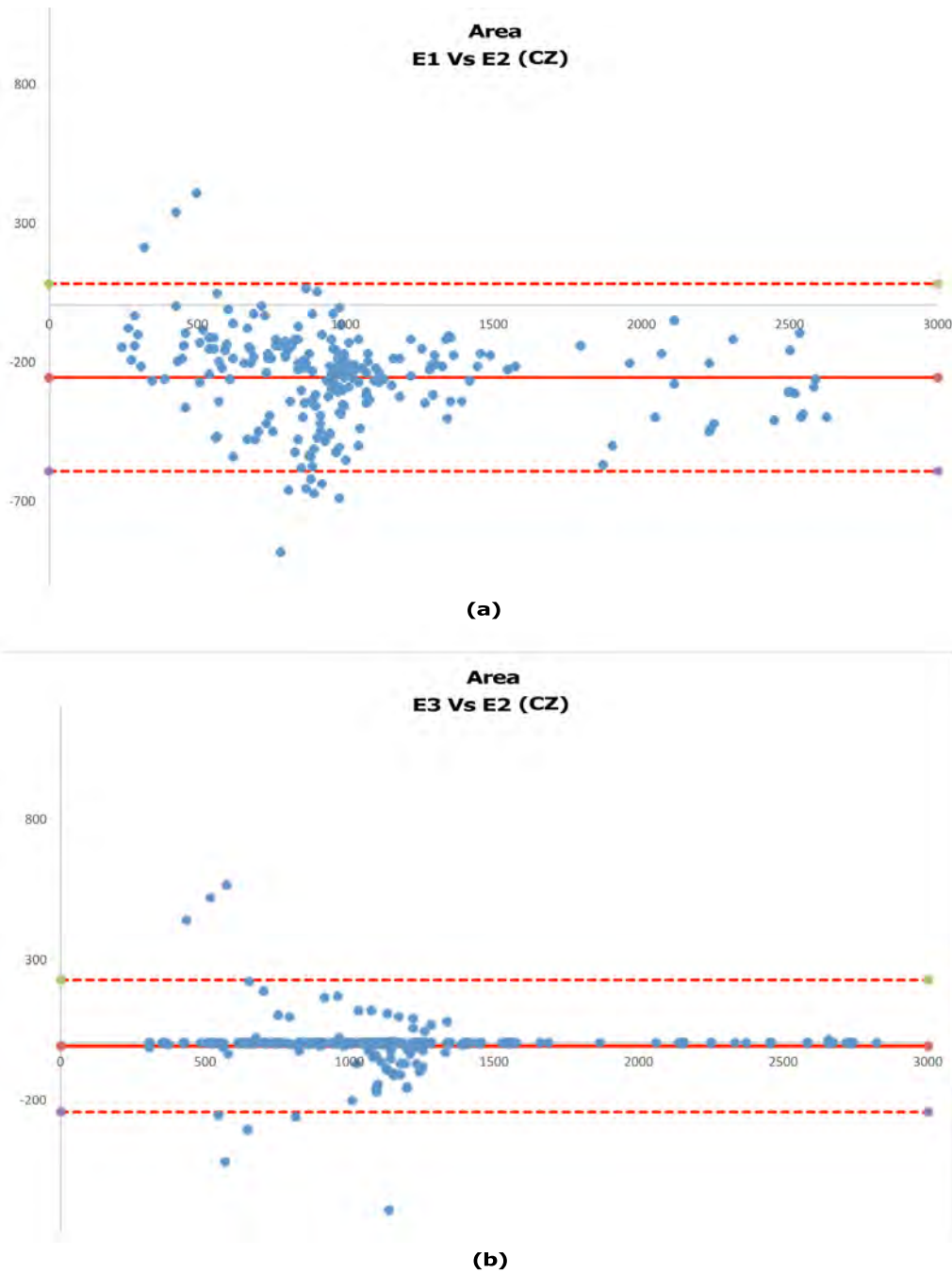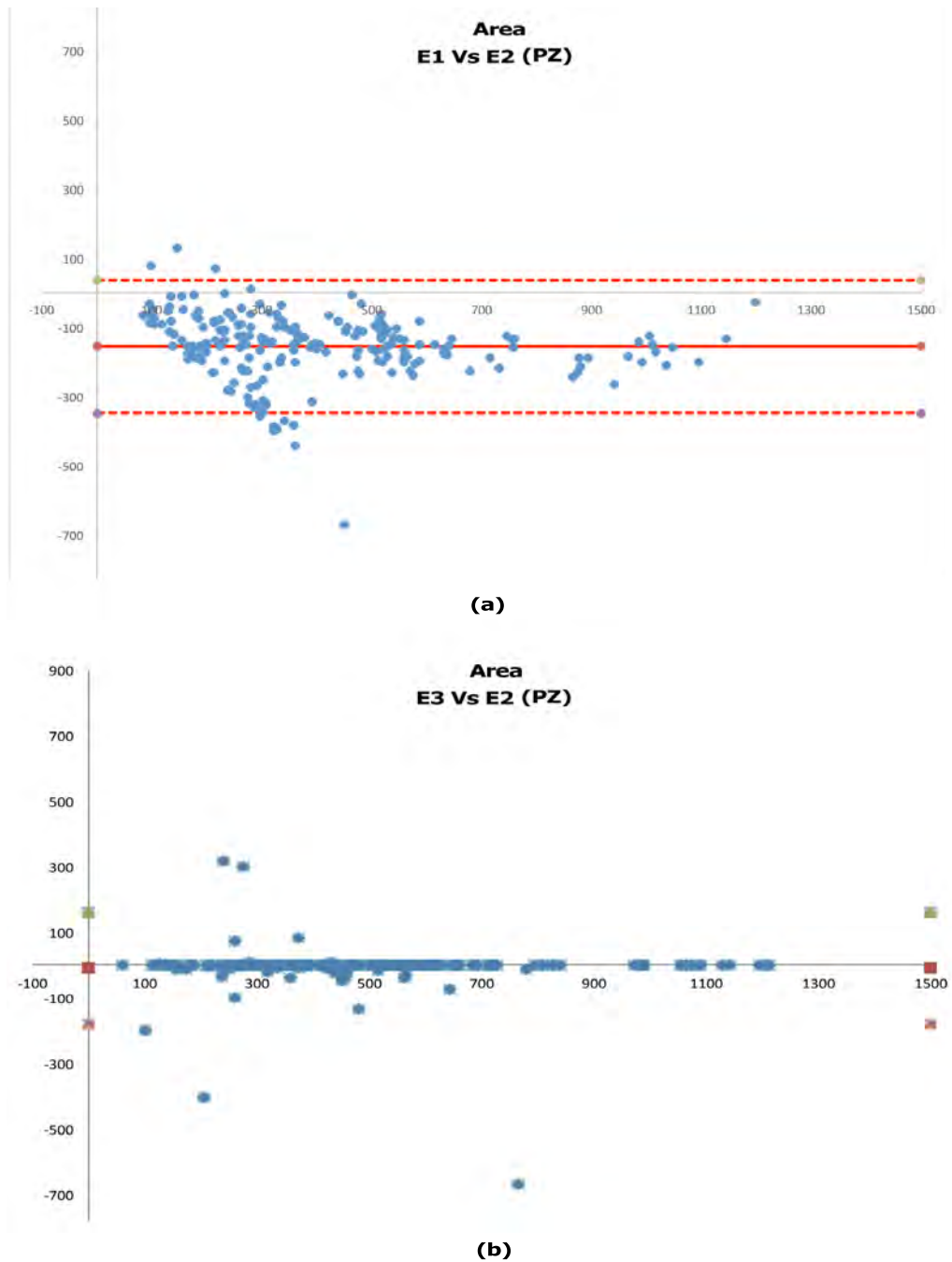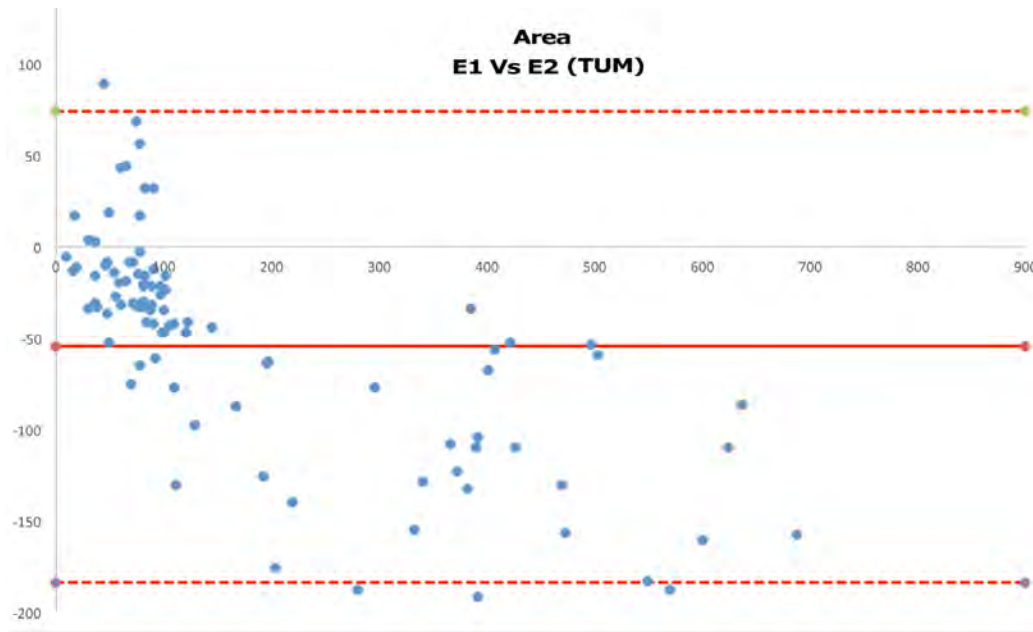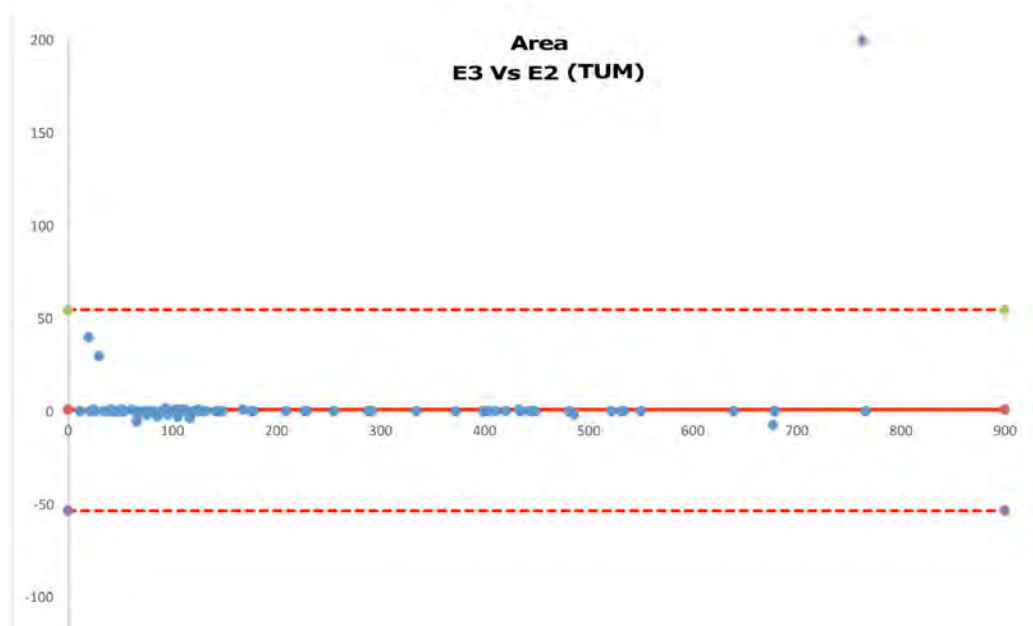
(a)

(b)
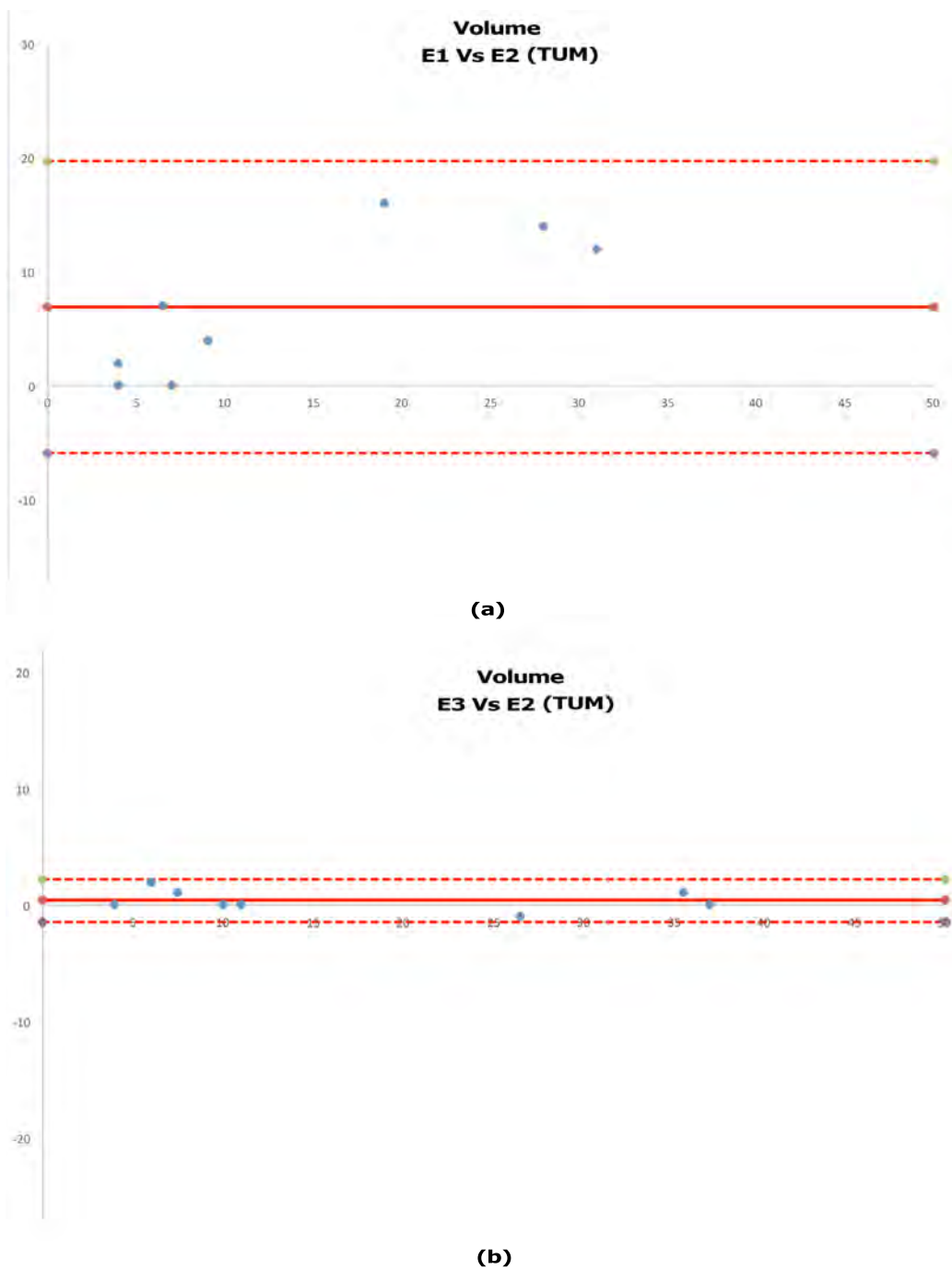
Figure 5.39: Bland-Altman plot obtained from the comparison between $E1$ vs. $E2$ (a), and $E3$ vs. $E2$ (b) concerning the calculation of the surface of a tumor. For the $E1$ vs. $E2$: Mean = 7; SD = 6 and for $E3$ vs $E2$: Mean= 0.4; SD=0.9.

Conclusion

In our study, an evaluation of anatomical, edge based and region-based parameters are performed and discussed. Three experiments are performed and comparisons between them focused on anatomical and edge based and region-based parameters. Then, a set of results using different analysis such as the correlation coefficient, regression line analysis, a Bland-Altman plot, and a 2-sample t-test are presented. In this case, the area was chosen as the anatomical parameter. A second analysis for the edge based and region-based parameters based on Hausdorff distance and Dice Index is also performed.

In general, a positive assessment is depicted in the analysis of the comparison between $E3$ vs. $E2$. The proposed analysis tries to compare different experiments of prostate gland regions using a set of parameters. It has been demonstrated that there is good improvement when there is less variability between the users using the ProstateAnalyzer tool. There is an important decrease in variability when the $E2$ evaluation is used (always used as a reference point). Hence, the advantage in knowing the processing of other experts is important in order to improve this evaluation.

5.5.4   Discussion

The ProstateAnalyzer is a web-based application for the analysis of prostate images using four MR modalities: T2WI, DWI, DCE and MRS. One of the most evident advantages of ProstateAnalyzer is that it allows simultaneous analysis of a prostate study using the different imaging modalities. It also provides interaction among them and the MR spectroscopy.

The application is currently dedicated to our acquisition protocol, but can be easily adapted to other protocols. In fact, the implementation of the the Prostate-Analyzer offers the possibility of extending and adapting it to other MR modalities. As a proof of concept, this project is an adaptation of previous work based on mammography [116] and demonstrates that our approach is easily portable to other sources of medical images.

The principal problem encountered in the diagnosis of a prostate study is the localization of a ROI containing the tumor tissue. Generally, experts use different tools to validate the diagnoses using different software and making many annotations in different files. This is not a practical solution to managing abundant medical data. the ProstateAnalyzer offers a solution because it allows experts to analyze prostate examinations on T2WI, DWI, DCE and MRS panels in the same application.

The development of the proposed interface has been made in such a way that it is

simple and intuitive to use. The interface is divided into four panels with the purpose of visualizing a patient study, simultaneously, through the different techniques. As has been previously said, it is important to note that the main objective was to provide useful tools for experts to manage examinations with different types of images and data. The ProstateAnalyzer allows the annotation of findings provided by different experts in prostate examinations. All annotations are saved in XML files associated with the prostate images. Although a prostate study can be shared among many users, they can still validate or modify their own diagnosis in individual cases.

Once the study has been analyzed, the findings need to be reported, if possible in a standardized way. The ProstateAnalyzer includes the PI-RADS structured reporting scheme, according to ESUR prostate MR guidelines in 2012 [14], which is becoming established as the common protocol in European countries [144, 156]. The use of a standardized graphic reporting scheme facilitates the communication between referring colleagues, and increases the quality and diagnostic value of prostate analyses.

The design effort benchmarks to estimate the accessing response time is also described. In general terms, the response time is acceptable considering the high volume (images and data) to be loaded and displayed. Depending on the user's connections the response time be get over the calculated average. But the timeout limit that the ProstateAnalyzer requires to load a complete prostate study is around 3 minutes maximum. After this time, the prostate study will be aborted and it is necessary to start again. It is important to define a timeout in order not to saturate the server and not be affected by the connection of the users.

On the other hand, a prostate study that allows collaborative work in the evaluation of cancer tissues is also presented. In this study, the variability of medical findings in different areas analyzed by different and independent expert is evaluated. Three experiments between them focused on anatomical and edge based and region-based parameters are performed and compared. The evaluation procedure used to make an evaluation with significant results is presented. In general, it has been demonstrated that there is good improvement when there is little or no a variability between users using the ProstateAnalyzer tool. The advantage of knowing the process of other experts is important in order to improve the evaluation.

Finally, our tool offers the possibility of working remotely via the web and represents an improvement in data management. Access to the ProstateAnalyzer should be available to multiple users in order to make both local and external connections easier. In order to facilitate this task, the application is implemented as a Java applet tool. Thus, it is not necessary to install any program on computers to run our interface especially it users have a browser that supports Java technology. Moreover,

applets can be executed from any operating system (Windows, Linux and MAC) because they are running in a Web browser. The ProstateAnalyzer is integrated in a server to manage medical images stored in a prostate database.

5.6   Conclusions

In this chapter, we have presented The ProstateAnalyzer, a new medical tool that allows the evaluation of PCa in an effective way. The ProstateAnalyzer visualizes the different MRI techniques (anatomy, diffusion, and perfusion) together with MRS, and allows of making and displaying medical findings (annotations). In addition, the ProstateAnalyzer also includes the PI-RADS reporting protocol, thereby offering the possibility of fully reporting the prostate study in a standardized way.

It has been implemented as an interactive Java applet application with the purpose of facilitating the inclusion of medical findings on existing prostate images, using the combination of MRI techniques and MR Spectroscopic analysis. Furthermore, a new architecture is presented to store the medical records in an XML database which stores a set of annotated files.

Chapter 6

Conclusions and further work

> There is no real ending. It's just the place
> where you stop the story.
>
> Frank Herbert

This final chapter summarizes the proposed system architecture and the development of several applications presented in this thesis. A description of the contributions presented in this manuscript and to the research area of web-based medical applications. Finally, it also highlights possible future improvements and directions of research related to the work that has been presented.

6.1   Conclusions of the thesis

The importance of developing database models and architecture systems in order to store medical images and aid medical diagnosing have increased over the last decade. This increase in medical data supposes a new paradigm to develop new system architectures and digital databases to store patient information along with medical images. Thus, our interest focused on the development of a generic system architecture, a model for medical applications to be used by the medical community with the support of medical staff and hospitals for the diagnoses of cancer. In order to demonstrate how the proposed generic medical tool can be adapted to different image modalities, we have focused our endeavors on breast and prostate cancer analyses. Furthermore, the support of the University of Girona (working in breast cancer) and the University of Burgundy (working in prostate cancer) has been key in the development of this project with a cotutelle program.

The first chapter introduced medical imaging management and data acquisition

focused on X-ray radiology and MRI techniques. Concretely, the acquisition of X-ray radiology focuses on mammographic images and MRI on prostate images. These images are acquired using the DICOM protocol which was also described. This introductory chapter presented an overview of PACS as a common DICOM database describing the usability and main features. Finally, the context, motivation, main objectives and aims of this thesis were also explained.

The second chapter was devoted to a thorough survey of the state-of-the-art in web-based medical applications. This chapter reviewed the web-based applications applied to the medical imaging field, analyzing how applications store and exchange huge amounts of information. Indexing tools based on the XML protocol, allowing a fast retrieval of the information, and security issues were also studied. The purpose of this chapter was to highlight the main advantages and drawbacks of the projects reviewed.

The third chapter focused on the proposed system architecture. This system architecture framework covers the relationship between image processing tools and medical databases, with the integration of web-based medical applications. An introduction to the background of our purpose, concepts used to design the proposed architecture, a proposed view model, and the advantages and drawbacks of the proposed system architecture framework compared to others were presented in this chapter. Finally, one of the most important advantages of our proposed architecture is to foster collaborative work, by sharing valuable information and promoting the aid of medical diagnosis by experts.

The last chapters of this thesis explained in detail two different ways to develop a web-based application using the proposed system architecture. The aim was to demonstrate the capabilities, advantages and facilities of implementing a web-based application in different medical areas and modalities, by using the proposed architecture framework model. X-ray radiology was applied to mammography (chapter four) to aid in the diagnosis of breast cancer; while MRI was used in prostate studies to aid the diagnosis of prostate cancer(chapter five). These two modalities were chosen due to the fact that they offered the possibility of collaborative work between different universities and hospitals.

6.1.1 Contributions

A number of research projects published in recent literature using web-based applications in medical informatics have been studied. However, there are still major problems to solve. The main problem is access and information sharing between different centers, and the interaction between hospitals and experts to assist in medical diagnosis. From this point of view, the main objective of this thesis is to develop

and design an architecture model in order to maintain a system management and diagnostic oriented to a web-based medical imaging application. Hence, a centralized database of digitally acquired medical images was performed.

The aim was to provide a useful tool to the medical and scientific communities in order to manage this image database, including the associated diagnoses. Moreover, the goal was to develop a medical tool to help the medical community in the early detection of cancer. The main contributions of this thesis to clinical community are:

- A new database of medical cases building a new system architecture is developed. Database queries are performed through a web form and provide the user with a list of clinical cases according to the query parameters, including both the image and the annotation files. This architecture links both the PACS and the XML server which store the clinical cases (DICOM and annotation files). The advantages of the proposed architecture are to aid in the medical diagnosis and allow a collaborative work.

- A novel web-based application to aid in the diagnosis of breast cancer, based on the proposed system architecture: MamoDB. Firstly, a search engine for the clinical community is available to aid in the diagnosis of cancer. The main contribution is to help radiologists to diagnose breast cancer in digital mammography more effectively, by providing the user with other, similarly diagnosed cases. Moreover, MamoDB offers the possibility of downloading selected mammographic studies. The system also offers the possibility of downloading a single record or use an automated selection. A Java applet application is implemented to visualize and annotate the mammographic findings by the users. Finally, it is possible creates a BI-RADS report using the application.

- A novel web-based application to aid in the diagnosis of prostate cancer based on the proposed system architecture: the ProstateAnalyzer. The application allows one to create, recover and update prostate diagnosis provided by different experts. It represents an efficient way to share and save information concerning the state of the diagnoses. The principal problem encountered in the diagnosis of a prostate study is the localization of a ROI containing tumor tissue. Generally, experts use different tools to validate the diagnoses using different software and make many annotations in different files. This is not a practical solution to managing great quantities of medical data. The ProstateAnalyzer offers a solution because it allows experts to analyze the prostate examination on T2WI, DWI, DCE and MRS panels in the same application.

- In general terms, although public databases of mammographic images are available, many of them have been acquired by analogical acquisition with subsequent digitization of the film. For breast imaging, this process entails an

inherent loss of quality depending on the digitizer system, loss of information acquisition and non-standardization of image formats. The proposed system architecture is used to manage and centralize a public database of digitally acquired mammographic images. Furthermore, by using the two proposed web-based applications, a user can have access to information both locally and remotely.

- One of the most important contributions concerning the ProstateAnalyzer is the interaction between different imaging modalities supported by the tool . One of the immediate necessities is the ability of analyze different studies at the same time. In this sense, two interactions are presented: a) the interaction between anatomy as with MRI technique and MRS spectroscopy, and b) the interaction between two MRI techniques such as anatomy and perfusion. Finally, the ability to create a PI-RADS report using the application is also available.

- The interest in collaborative work in the evaluation of cancer issues using the ProstateAnalyzer is presented. Exhaustive evaluations of the medical findings in different regions of the prostate gland (CZ, PZ and Tum) for several experiments using the ProstateAnalyzer application are evaluated and discussed in this work.

6.2   Further work

There has been a constant technological evolution in computer science and network field recently, and medical imaging and medical applications can benefit from it. In this sense, it means that our proposal could be improved by adapting it to new technologies and developments. Further work related to our work could focus on improvements in such things as management control for servers, visualization and post-processing tools, backup systems or security protocols, among others. Independently, improvements related to the usability of the proposed web-based applications could be improved according to the necessities and requirements of medical staff.

Hence, future directions are presented divided into two categories: those related to improving the reliability of our proposal in the short-term, and future research lines departing from this thesis.

6.2.1   Short-term proposal improvements

Some potential improvements could be made in the near future. In regard to the proposed system architecture, short-term improvements could focus on security pro-

tocols. At the moment, all the packages implemented (such as the ZEND server, PACS database and eXist-db used to develop our proposed architecture model) provide a security system. However, the managed data is typically confidential and therefore, improvements such as the implementation of an encryption or anonymization algorithms could be made. Moreover, in order to preserve and save the data, a backup system could be implemented by taking advantage of the actual technologies and tools such as cloud systems (using cloud computing).

Now, some proposal improvements concerning the two web-based applications (MamoDB and ProstateAnalyzer) are proposed. Concerning the MamoDB application, the definition of a standard XML file format for the medical community could prove extremely useful and would correspond to the annotations of the images associated with mammography. Moreover, we envision the definition of custom search forms that ensure optimal results, and are tailored to the needs of the user. Further improvements could be investigated, in terms of data transfer, and the use of new algorithms to improve the transfer of data in a more efficient way. One solution could be to add more options to render the search more accurate with respect to mammographic findings retrieved from the database.

Concerning the ProstateAnalyzer application, it allows downloading PDF files containing annotations and graphical results (mean signal-time curve and signal spectra). It could also be useful to generate a pre-report with all the graphic results, annotations, and include the PI-RADS report in a PDF file. However, experience has revealed several limitations of the PI-RADS standard, in part due to rapid progress in the field. In an effort to make PI-RADS standardization more globally acceptable, the ACR [1], European Society of Urogenital Radiology (ESUR) [2] and the AdMeTech Foundation [3] established a steering committee to build upon, update and improve upon the foundation of PI-RADS v1. A second version of the score has been recently launched [187]. In this sense, ProstateAnalyzer should be updated including the new version of PI-RADS reporting.

In terms of response time, it has been evaluated and it is considered acceptable considering the high volume (images and data) to be loaded and displayed. Depending on the user's connections, the time response can rise to over the calculated average. In this sense, a limit of time has been assigned to the ProstateAnalyzer application in order to load prostate studies. One improvement could be finding new alternatives for loading large quantities of data (normally for a spectroscopic dataset). For example, a pre-script could be implemented to save this data in a

---

[1]American College of Radiology (ACR). URL: http://www.acr.org

[2]European Society of Urogenital Radiology (ESUR). URL: http://www.esur.org

[3]AdMeTech Foundation is a 501(c)(3) non-profit organization providing global leadership in transforming early detection and treatment of life-threatening diseases. URL: http://www.admetech.org

database in order to reduce the number of text files with spectroscopic data. Also, in prostate studies with a number of associated prostate images, the computational cost could increase a lot. In particular, this problem could be affected by perfusion studies. A normal prostate study contains 640 images but in some cases could have 720 images. One solution could be to load single frames only on the basis of their spatial correspondence instead of the whole dataset.

6.2.2   Future research lines

In the long term, there are several new research lines departing from this thesis that could be studied.

This web-based tool has been developed using X-ray radiology to aid in the diagnosis of breast cancer, and using MRI in the case of prostate cancer diagnoses. However, it is easily portable to other organs (brain, heart, etc.). One of the advantages of our proposed work is the use of standard protocols such as DICOM and PACS databases. This is an important factor because it is possible to implement new applications focusing on different image modalities. For example, it will be possible to adapt it to an Echography with Ultra Sound (US) also used in mammography and it can be applied to other pathologies/diseases.

Concerning the database architecture, an improvement to add extra information related to the calculation of metrics other than the mean value of the signal in a given ROI could be implemented. It has been noticed that the tendency of a recent research works published in the literature indicates that quantitative parameters such as the 10th percentile of the ADC distribution, or the skewness or kurtosis of the distribution of the ADC value or the T2-weighted signal could be interesting in characterizing prostate focal lesions seen on MRI [157, 139]. In this sense, besides the implementation of these metrics analysis into the web-based application, the design or adaptation of the actual XML structure in order to incorporate the metrics data into the XML files. Alternatively, another solution could be adding an extension tool into the web-based applications to export the ROIs information using a standard format to be downloaded by users in own the workstations. In this sense, it will be useful in order to make own calculations and analysis supposing a benefit for radiologists.

Finally, we also propose the possibility of displaying 2D and 3D metabolite maps in the ProstateAnalyzer application. For this task, two main problems must be solved. First, specialized spatial and spectral data processing methods for which several sources are not commonly available are needed (e.g. morphological analysis and spectral characteristics of the observed metabolites). Secondly, a manual pre-processing stage is needed in order to co-register the spectroscopic grid and the

different image sequences.

Appendix A

Analytical tools

In this appendix the main tools used to estimate a series of value are shown. Also, these tools are used in order to compare two series of data and tests to evaluate segmentation. The main common analytica tools used in this thesis are: Mean, Standard Deviation (SD), Student t-test, Pearson correlation coefficient, regression analysis, Bland-Altman analysis, Hausdorff Distance (HD), and Dice Similarity Coefficient (DSC), all of which will be explained in the next sub-sections.

A.1 Mean

The Mean, generally represented by μ, is the average of the numbers: a calculated "central" value of a set of numbers. The arithmetic mean of a sample being:

$$X_n = \{x_1, x_2, ..., x_n\}. \tag{A.1}$$

Mean is the sum of the sampled values divided by the number of items and is defined by equation A.2:

$$\mu = \frac{\sum_{i=1}^{N} X_n}{N}. \tag{A.2}$$

A.2 Standard deviation

The Standard Deviation (SD)), generally represented as σ, is a measure of the statistical dispersion. It is a measure of statistical amount of variation or dispersion from the average. A low standard deviation indicates that the data points tend to

be very close to the mean, and a high SD indicates that the data points are spread out over a large range of values [73].

The calculation of the SD can be summarized as follows [174]:

- Work out the average (mean value) of your set of numbers.

- Work out the difference between each number and the mean.

- Square the differences.

- Add up the square of all the differences.

- Divide this by one less than the number of numbers in your set. This is called the variance.

- Take the square root of the variance and you've got the standard deviation.

Formally, SD is defined by equation A.3

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2. \tag{A.3}$$

In the alternative notation this can be expressed by equation A.4

$$\sigma = \sqrt{\frac{\sum_{i=0}^{N-1}(x_i - \mu)^2}{N-1}}. \tag{A.4}$$

A.3 Student t-test

A t-test is any statistical hypothesis test in which the test statistic follows a Student's t distribution if the null hypothesis is supported. A hypothesis is a statement designed to be proven or disproven. This null hypothesis will usually stipulate that there is no significant difference between the means of the two data sets. The t-test has been used to compare two sample means, and the calculation of the t-statistic in this case is given by the following expression A.5:

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}}. \tag{A.5}$$

where SD is the standard deviation of the sample.
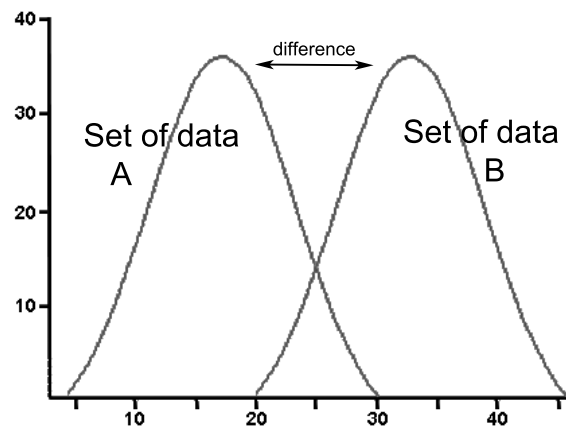
Figure A.1: Example of a Student's t-test distribution applied to a two mean sample.

It is best to try and determine whether there is a difference between two independent sample groups. For the test to be applicable, the sample groups must be completely independent, and it is best used when the sample size is too small to use more advanced methods.[154]. Figure A.1 shows an example of two sets of data, *A* and *B* applying Student's test analysis.

The criteria of a t-test should fill the following requirements [1]:

- The number of points in each data set must be the same, and they must be organized in pairs, in which there is a definite relationship between each pair of data points.

- If the data were taken as random samples, you must use the independent test even if the number of data points in each set is the same.

- Even if data are related in pairs, sometimes the paired t is still inappropriate.

- Here's a simple rule to determine if the paired t must not be used - if a given data point in group one could be paired with any data point in group two, you cannot use a paired t test.

---

[1]David R. Caprette. Student Test. Experimental Biosciences. URL: `http://www.ruf.rice.edu/~bioslabs/tools/stats/pairedttest.html`

A.4 The Pearson correlation coefficient

Pearson's correlation coefficient or Correlation coefficient, generally represented by γ, is a measure of the linear correlation (dependence) between two variables X and Y, giving a value between $+1$ and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. It is widely used in the sciences as a measure of the degree of linear dependence between two variables. It was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s [2].

The Correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations [16]. Pearson's correlation coefficient when applied to a sample is commonly represented by the letter r and may be referred to as the sample correlation coefficient or the sample Pearson correlation coefficient.

Correlation coefficient between two random variables X and Y is defined as A.6:

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{SD}(X)\mathbf{SD}(Y)}}, \tag{A.6}$$

where SD is Standard Deviation and "Cov" is the Covariance. The covariance is a measure between two jointly distributed real-valued random variables X and Y with finite second moments is defined in A.7:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] \tag{A.7}$$

where E[X] and E[Y] are the expected value of X and Y, also known as the mean of X and Y respectively.

The main features of the Correlation coefficient can be defined as the following [3]:

- The **value of** r is such that $-1 < r < +1$. The $+$ and $-$ signs are used for positive linear correlations and negative linear correlations, respectively.

- **Positive correlation**: If x and y have a strong positive linear correlation, r is close to $+1$. An r value of exactly $+1$ indicates a perfect positive fit. Positive values indicate a relationship between x and y variables so that as values for x increases, values for y also increase.

---

[2]Boundless Statistics. Coefficient of correlation. Boundless Statistics. Boundless, 02 Jul. 2014. Retrieved 06 Apr. 2015. URL: `https://www.boundless.com/statistics/textbooks/`

[3]MathBits.com. Statistics 2: Correlation coefficient, 2015. URL: `http://mathbits.com/MathBits/TISection/Statistics2/correlation.htm`
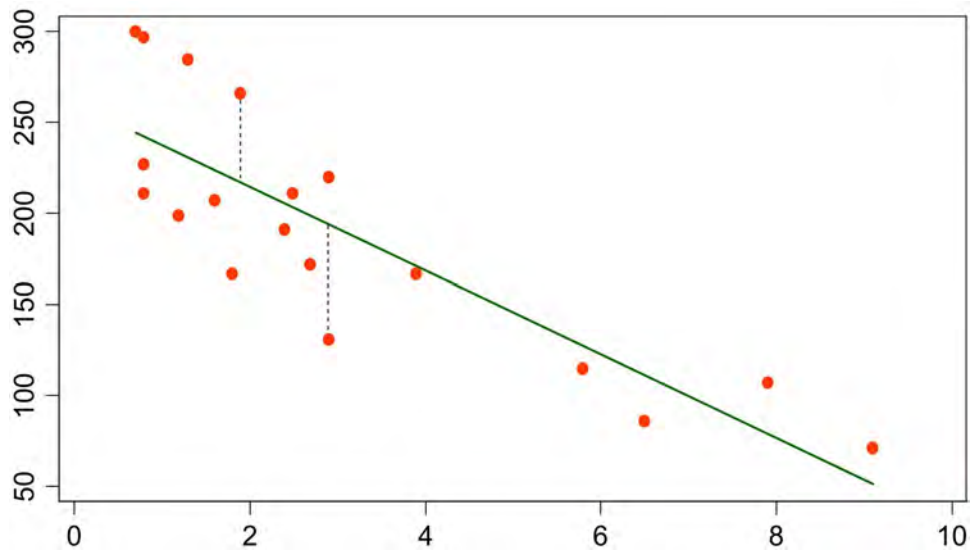
Figure A.2: Examples of scatter diagrams with different values of correlation coefficient (r).

- **Negative correlation**: If x and y have a strong negative linear correlation, r is close to -1. An r value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between x and y so that, as values for x increase, values for y decrease.

- **No correlation**: If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables.

- **A perfect correlation** of $\pm\,1$ occurs only when the data points all lie exactly on a straight line. If $r = +1$, the slope of this line is positive. If $r = -1$, the slope of this line is negative.

Examples of scatter diagrams with different values of correlation coefficient (r) is depicted in Figure A.2.

A.5 Regression analysis

Simple linear regression analysis is a technique in parametric statistics that is commonly used for analyzing the mean response of a variable Y which changes according to the magnitude of a variable X. It forms the basis of one of the more important
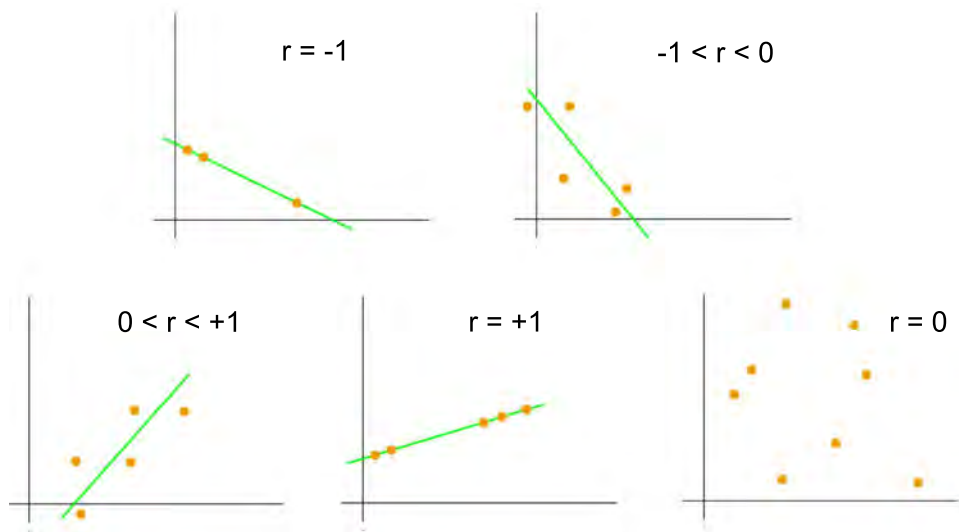
Figure A.3: Example of a regression line in green. Vertical distances from least-squares lines are residuals, and a good line for prediction makes these distances small.[5]

forms of inferential statistical analysis [4].

As with correlation, regression analysis is used to analyze the relation between two continuous (scale) variables. However, regression analysis is better suited to studying functional dependencies. The term functional dependency implies that X (partially) determines the level of Y. In fact, there is an independent variable (X) and a dependent variable (Y). For example, there is a function dependency between age and blood pressure since, as one ages, blood pressure increases. In contrast, there is no functional dependency between arm length and leg length since increasing the length of an arm will have no effect on the leg's length (or vice versa) [5].

A line is identified by its slope (the angle of the line describing the change in Y per unit X) and intercept (where the line crosses the Y axis). Regression describes the relation between X and Y with just such a line defined in the corresponding equation A.8:

$$\hat{y} = a + bx. \tag{A.8}$$

When discussing our line, let:

---

[4]Oxford Journals. Research Methods II: Simple Linear Regression. URL: http://www.oxfordjournals.org

[5]B. Burt Gerstman. 15. Regression, 2004. URL: http://www.sjsu.edu/faculty/gerstman/StatPrimer/

$\hat{y}$ represent the predicted value of Y,

a represent the intercept of the best fitting line, and

b represent the slope of the line (also called the regression coefficient).

However, because statistical data has random scatter, identifying a good line is not a trivial matter. The random scatter around the line is identified as the distance of each point from the predicted line. These distances, called **residuals**, are shown as dotted lines in Figure A.3.

The goal it to determine a line that minimizes the sum of the squared residuals. This line is called **the least squares line** [153]. **The slope(b)** of the least squares line is given by equation A.9:

$$\hat{y} = a + bx\frac{SS_{xy}}{SS_{xx}}, \tag{A.9}$$

where SS_{xy} is the sum of the cross-products and SS_{XX} is the sum of the squares for variable X. Then, the **intercept** of the least squares line is given by equation A.10:

$$a = \bar{y} - b\bar{x}, \tag{A.10}$$

where $\bar{y}$ is the average value of Y, b is the slope, and $\bar{x}$ is the average value of X. If we find that the slope of the regression line is significantly different from zero. In fact, when slope "b" is different to one means that is a case of bias. In conclusion, there is a significant relationship between the independent and dependent variables.

A.6 Bland-Altman analysis

The Bland-Altman analysis is the most popular method of an agreement approach. D.G. Altman and J.M. Bland first published this approach in 1983 in a statistical journal [5] and later in Lancet [22]). Assessing agreement between two observers or measurement methods can be done by using Bland and Altman plots and calculating limits of agreement. In a Bland and Altman plot, the difference between the two measurements per subject is plotted against the mean of the two measurements [47]. An example of a Bland-Altman plot is depicted in Figure A.4.

The basic concept of Bland-Altman's approach is the visualization of the difference of the measurements made by the two methods, then plotting the differences or the bias (Y-axis) versus the mean of the two readings (X-axis). A good agreement is when the mean is close to zero with a low SD. In addition, additional reference
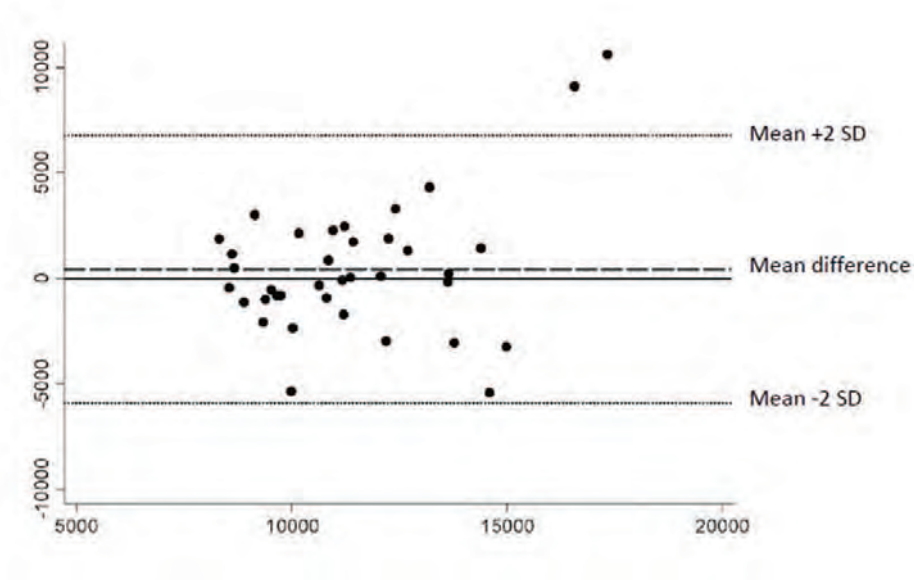
Figure A.4: Example of a Bland-Altman plot [24].

lines such as the zero bias line and 95% upper (0 + 1.96 SD) and 95% lower (0 - 1.96 SD) are also overlaid on the same scatter plot. When there is no systematic bias, the plot of the differences are symmetrical around zero [6].

A.7   Hausdorff Distance (HD)

The Hausdorff Distance (HD) is the maximum distance from one of set points to the nearest point in the other set [160]. But in this thesis, this set of points is used to define a contour. More formally, HD from set

$$A = \{a_1, a_2, ..., a_n\} \tag{A.11}$$

to set

$$B = \{b_1, b_2, ..., b_n\} \tag{A.12}$$

---

[6]Fernández, F. and Fernández G. Validating the Bland-Altman Method of Agreement, University of Nevada-Las Vegas. URL: `http://www.lexjansen.com/wuss/2009/pos/POS-Fernandez.pdf`
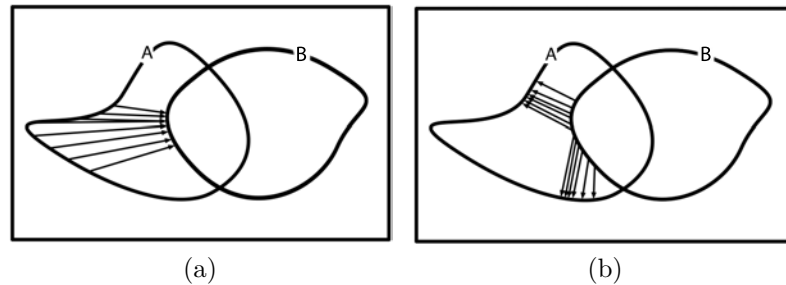
Figure A.5: Illustration of the non-symmetry property of the Minimum Distance (MD) metric. (a) $\text{MD}(a_i, B)$, (b) $\text{MD}(b_i, A)$

is a maximin function. The Hausdorff Distance (HD) measures the worst possible discrepancy between the two delineations A and B defined in equation A.13 [7].

$$\text{HD}(A, B) = \max \left\{ \max_{a_i \in A} \text{MD}(a_i, B), \max_{b_i \in B} \text{MD}(b_i, A) \right\}. \qquad (A.13)$$

The HD as defined in equation A.13 has been used for assessing the segmentation results.

$$\text{MD}(a_i, B) = \min_{b_j \in B} \|a_i - b_j\|. \qquad (A.14)$$

The Minimum Distance (MD) is defined in equation A.14 and corresponds to the minimum distance [44] between a particular point a_i in the contour A (so that $a_i \in A$) and any other point within delineation B.

Notice that it is calculated as the maximum of the worst discrepancy between (A, B) and (B, A) since MD is not a symmetric measure [8], as can be observed in figure A.5.

A.8   Evaluation criteria: region-based approaches

When analyzing and comparing the areas described by segmented regions A and B, B is defined as the reference area (see fig. A.6b), 4 areas are described: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN); corresponding to the regions of the confusion matrix in figure A.6a.

---

[7] Waitser, M. Mathematical Expression Recognition Using the Hausdorff Distance, 2004. URL: http://books.google.es/books?id=ZmwaOgAACAAJ

[8]   Máximo,   C.O.   Hausdorff   Distance,   2010.   VDM   Publishing.   URL: http://books.google.es/books?id=pgUcygAACAAJ

Segmentation Ground Truth (GT) (reference)

		Positive	Negative
Segmentation Outcome (prediction)	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

(a)

(b)
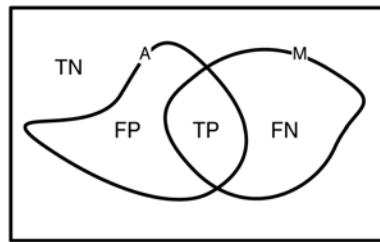
Figure A.6: (a) Statistical hypothesis test errors confusion matrix. (b) Graphic representation of the statistical hypothesis test errors for assessing the performance in terms of area.

True Positive (TP) is found as the area in common $(A \wedge B)$ between the two delineations A, B.

True Negative (TN) is found as the area $(\overline{A} \wedge \overline{B})$ not belonging to either of the delineations A nor B .

False Positive (FP) is found as the area $(A \wedge \overline{B})$ belonging to the assessing segmentation A and not as a part of the reference delineation B.

False Negative (FN) is found as the area $(\overline{A} \wedge B)$ corresponding to the reference delineation B but not as a part of the assessing segmentation A.

Area metrics (or indexes) for assessing the segmentation are defined as a dimensionless quotient relating the 4 regions (TP, FP, FN and TN) described by the segmentation outcome being assessed (denoted A in fig:A.6a) and the reference segmentation (denoted B). Most of the indexes are defined in the interval $[0, 1]$ and some projects report their results as a percentage.

A.9 Dice Similarity Coefficient (DSC)

Before explaining the definition of the Dice Similarity Coefficient (DSC) as also known as the Dice Index, it is important to define first the concept of Area Overlap [**Massich14**]. Area Overlap (AOV), also known as the overlap ratio, the Jaccard Similarity Coefficient (JSC) [52] or Similarity Index (SI) [170], represents the percentage or amount of area common to the assessed delineation A and the reference delineation B divided by the union according to equation A.15.

$$AOV = \frac{TP}{TP + FP + FN} = \frac{|A \cap B|}{|A \cup B|} \qquad \in [0, 1]. \tag{A.15}$$

The Dice Similarity Coefficient (DSC), which can have a value ranging from zero to one, is defined as two times the volume of the intersection between two segmentations divided by the sum of the volumes of the two segmentations. The difference between DSC and AOV is that DSC takes into account the TP area twice, one for each delineation. The DSC index is given by equation A.16 and the relation between AOV or JSC and the DSC similarity indexes is expressed by equation A.17. Notice that the DSC similiarity index is expected to be greater than the AOV index [60]. Finally, when the index is close to 1, then a good overlap (or a good concordance) is represented, otherwise when it is close to 0 there is discordance.

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2|A \wedge B|}{|A| + |B|} \qquad \in [0, 1]. \tag{A.16}$$

$$DSC = \frac{2 \cdot AOV}{1 + AOV}. \tag{A.17}$$

Bibliography

[1]   R. Agarwal et al. "If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging". In: *J. Med Internet Res.* 15.2 (2013), e43.

[2]   C.B. Akgül et al. "Content-Based Image Retrieval in Radiology: Current Status and Future Directions". In: *J. Digit. Imaging* 24.2 (2011), pp. 208–222.

[3]   A. Aljabre. "Cloud Computing for Increased Business Value". In: *Business & Society* 3.1 (2012).

[4]   R. Alonzi, A. Padhani, and C. Allen. "Dynamic contrast enhanced MRI in prostate cancer". In: *Eur. J. Radiol.* 63.3 (2007), pp. 335–350.

[5]   D. G. Altman and J. M. Bland. "Measurement in Medicine: The Analysis of Method Comparison Studies". In: *The Statistician* 32.3 (1983), pp. 307–317.

[6]   American College of Radiology. *Breast Imaging Reporting and Data System.* 2nd Edition. 1995.

[7]   American College of Radiology. *Illustrated breast imaging reporting and data system birads.* 3rd edition. 1998.

[8]   M. Angell, J.P. Kassirer, and A.S. Relman. "Looking back on the millennium in medicine [editorial]". In: *New Engl. J. Med.* 342.1 (2000), pp. 42–49.

[9]   C.W. Arnold et al. "Informatics in radiology: A prototype Web-based reporting system for onsite-offsite clinician communication". In: *Radiographics.* 27.4 (2007), pp. 1201–1211.

[10] National Electrical Manufacturers Association. *Digital Imaging and Communications in Medicine (DICOM).* 3rd Edition. National Electrical Manufacturers Association, 2006.

[11] National Electrical Manufacturers Association. "Digital Imaging and Communications in Medicine (DICOM)". In: 2011.

[12] A. Baltasar and A. González-Sistal. "Design of a Web-Tool for Diagnostic Clinical Trials Handling Medical Imaging Research". In: *J. Digit. Imaging* 24.2 (2011), pp. 196–202.

[13] R.L. Bard. *Dynamic Contrast-Enhanced MRI Atlas of Prostate Cancer.* Springer, 2009. ISBN: 978-3-540-78423-4.

[14] J.O. Barentsz et al. "European Society of Urogenital Radiology. ESUR prostate MR guidelines 2012". In: *Eur. Radiol.* 22.1 (2012), pp. 746–757.

[15] L.W. Bassett et al. "Film-screen mammography: comparison of views". In: *J. Natl. Med. Assoc.* 81.4 (1989), pp. 391–394.

[16] J. Benesty et al. *Pearson Correlation Coefficient.* Vol. 2. Springer Berlin Heidelberg, 2009, pp. 1–4.

[17] A. Bhavsar and S. Verma. "Anatomic Imaging of the Prostate". In: *Hindawi Publishing Corporation BioMed Research International* 1 (2014), pp. 1–9.

[18] W.D. Bidgood et al. "Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging". In: *J Am Med Inform Assoc.* 4.3 (1997), pp. 199–212. ISSN: 05315131.

[19] R. Bird, T. Wallace, and B. Yankaskas. "Analysis of cancers missed at screening mammography". In: *Radiology.* 184.3 (1992), pp. 613–617.

[20] R. Bitar et al. "MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask". In: *Radiographics.* 26.2 (2006), pp. 513–537.

[21] A.G. Biternas, G.T.P. Charnock, and I. Kuprov. "A standard format and a graphical user interface for spin system specification". In: *J. Magn. Reson. Imaging* 240.0 (2014), pp. 124–131.

[22] J.M. Bland and D.G. Altman. "Statistical methods for assessing agreement between two methods of clinical measurement". In: *Lancet* 327.8476 (1986), pp. 307–310.

[23] R.R. Bond et al. "A review of ECG storage formats". In: *Int J Med Inform.* 80.10 (2011), pp. 681–697.

[24] S.E. Bonn et al. "Active-Q: Validation of the Web-Based Physical Activity Questionnaire Using Doubly Labeled Water". In: *J. Med Internet Res.* 14.1 (2012), e29.

[25] N.F. Boyd et al. "Mammographic density and the risk and detection of breast cancer". In: *New Engl. J. Med.* 356.3 (2007), pp. 227–236.

[26] I. Carrión, JL. Fernández-Alemán, and A. Toval. "Are personal health records safe? A review of free web-accessible personal health record privacy policies". In: *J. Med Internet Res.* 14.4 (2012), e114.

[27] P.R. Carroll, F.V. Coakley, and J. Kurhanewicz. "Magnetic Resonance Imaging and Spectroscopy of Prostate Cancer". In: *Rev. Urol.* 8.1 (2006), S4 –S10.

[28] V. Castelli and L.D. Bergman. *Image Databases: Search and Retrieval of Digital Imagery*. Vol. 1. ISBN 978-0-471-32116-3. John Wiley & Sons, inc., 2002.

[29] M. Chen et al. "Prostate cancer detection: comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined". In: *Acta Radiol.* 49.5 (2008), pp. 602–610.

[30] J. Chipman et al. "Providing access to risk prediction tools via the HL7 XML-formatted risk web service". In: *Breast Cancer Res. Treat.* 140.1 (2014), pp. 187–193.

[31] J. Choe and S.K. Sun K. Yoo. "Web-based secure access from multiple patient repositories". In: *Int J Med Inform.* 77.4 (2008), pp. 242–248.

[32] D.A. Clunie. *DICOM Structured Reporting*. 1st Edition. PixelMed Publishing, 2000. ISBN: 0-9701369-0-0.

[33] G. Crehange et al. "Early choline levels from 3-tesla MR spectroscopy after exclusive radiation therapy in patients with clinically localized prostate cancer are predictive of plasmatic levels of PSA at 1 year". In: *Int J Radiat. Oncol. Biol. Phys.* 81.407–413 (2011).

[34] G. Créhange et al. "Tumor volume and metabolism of prostate cancer determined by proton magnetic resonance spectroscopic imaging at 3T without endorectal coil reveal potential clinical implications in the context of radiation oncology". In: *Int J Radiat. Oncol. Biol. Phys.* 80 (2011), pp. 1087–1094.

[35] S. Deftereos et al. "A Java-based electronic healthcare record software for beta-thalassaemia". In: *J. Med Internet Res.* 3.4 (2001), e33.

[36] M. Delisle. *Mastering phpMyAdmin 3.4 for Effective MySQL Management*. PACKT publishing, 2012. ISBN: 9781849517782.

[37] L. Dickinson et al. "Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting". In: *Eur. J. Neurol.* 59.4 (2011), pp. 477 –494.

[38] L. Dickinson et al. "Scoring systems used for the interpretation and reporting of multiparametric MRI for prostate cancer detection, localization, and characterization: could standardization lead to improved utilization of imaging within the diagnostic pathway?" In: *J. Magn. Reson. Imaging* 37.48–58 (2013).

[39] B.H Do et al. "Informatics in radiology: RADTF: A Semantic Search-enabled, Natural Language Processor-generated Radiology Teaching File". In: *Radiographics.* 30.7 (2010), pp. 2039–2048.

[40] K. Doi. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential". In: *Int J Med Inform.* 31.5 (2007), pp. 198–211.

[41] D. Dragan and D. Ivetic. "Architectures of DICOM based PACS for JPEG2000 Medical Image Streaming". In: *Comp. Science. Infrom. Syst..* 6.1 (2009), pp. 185–203.

[42] K. Dreyer, A. Mehta, and J. Thrall. *PACS: a guide to the digital revolution.* 2nd Edition. Springer, 2002. ISBN: 978-0387260105.

[43] M.M. Eberl et al. "BI-RADS classification for management of abnormal mammograms". In: *J Am Board Fam Med* 19.2 (2006), pp. 161–164.

[44] G.A. Edgar. *Measure, Topology, and Fractal Geometry.* Springer UTM, 1995. ISBN: 978-0387972725.

[45] A.D. Elster. "An index system for comparative parameter weighting in MR imaging". In: *J Comput Assist Tomogr* 12 (1988), pp. 130–134.

[46] A.D.A.M Medical Encyclopedia. "A service of the National Library of Medicine". In: *PubMed Health* (2013).

[47] A.M. Euser, F.W. Dekker, and S. Le Cessie. *Mathematical Statistics and Data Analysis.* Vol. 61. JCLEP, 2008, pp. 978–982.

[48] J. Fernàndez-Bayó et al. "Distributing medical images with internet technologies: a DICOM web server and a DICOM java viewer". In: *Radiographics.* 20.2 (2000), pp. 581–590.

[49] A.E. Flanders. "Next generation web search: augmenting information access for radiologists". In: *Radiographics.* 27.5 (2007), pp. 1519–1521.

[50] V. Fradet et al. "Prostate cancer managed with active surveillance: role of anatomic MR imaging and MR spectroscopic imaging". In: *Radiographics.* 256.1 (2010), pp. 176 –183.

[51] Jurgen J. Fütterer et al. "Prostate cancer localization with dynamic contrast-enhanced MR imaging and proton MR spectroscopic imaging." In: *Radiographics.* 241.2 (2006), pp. 449–458. ISSN: 0033-8419.

[52] L. Gao, X. Liu, and W. Chen. "Phase- and GVF-Based Level Set Segmentation of Ultrasonic Breast Tumors". In: *J. Applied Mathematics* 2012 (2012), pp. 1–22.

[53] X. Gao, Y. Qian, and R. Hui. "The State of the Art of Medical Imaging Technology: from Creation to Archive and Back". In: *Int J Med Inform.* 5 (2011), pp. 73–85.

[54] A. Gerstmair et al. "Intelligent image retrieval based on radiology reports". In: *Eur. Radiol.* 22.12 (2012), pp. 2750–2758.

[55] S. Ghose et al. "A supervised learning framework of statistical shape and probability priors for automatic prostate segmentation in ultrasound images". In: *Med. Image Anal.* 17.6 (2013), pp. 587–600.

[56] S. Ghose et al. "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images". In: *Comput. Meth. Prog. Biomed.* 108.1 (2012), pp. 262 –287.

[57] K. Glunde, Z.M. Bhujwalla, and S.M. Ronen. "Choline metabolism in malignant trans formation". In: *Nat Rev Cancer* 11.2 (2011), pp. 835–848.

[58] R.G. Grainger and J. Allison. "Grainger & Allison's Diagnostic Radiology: A Textbook of Medical Imaging. 4th ed, 3 vols [book review]". In: *Radiographics.* 226.3 (2003), pp. 706–706.

[59] D. Gritzalis and C. Lambrinoudakis. "A security architecture for interconnecting health information systems". In: *Int J Med Inform.* 73.3 (2004), pp. 305–309.

[60] y. Guang-Zhong and J. Tianzi. *Medical imaging and augmented reality.* Second International Workshop, MIAR. Springer-Verlag Berlin Heidelberg, 2004. ISBN: 9783540228776.

[61] M. Güld et al. "A generic concept for the implementation of medical image retrieval systems". In: *Int J Med Inform.* 76 (2005), pp. 252–259.

[62] M.O. Güld et al. "Quality of DICOM header information for image categorization". In: *Proc. SPIE.* Vol. 4685. Medical Imaging. 2002, pp. 280–287.

[63] J. Guo et al. "The development of MML (Medical Markup Language) version 3.0 as a medical document exchange format for HL7 messages". In: *J Med Syst.* 28.6 (2004), pp. 523 –533.

[64] J. Guo et al. "The latest MML (Medical Markup Language) version 2.3– XML-based standard for medical data exchange/storage". In: *J Med Syst.* 27.4 (2003), pp. 357 –366.

[65] H. van der Haak et al. "Data security and protection in cross-institutional electronic patient records". In: *Int J Med Inform.* 70.3 (2003), pp. 117–130.

[66] S. Haas et al. "Aspects of privacy for electronic health records". In: *Int J Med Inform.* 80.2 (2011), e26–e31.

[67] Masoom A. Haider et al. "Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer." In: *Am. J. Roentgenol.* 189.2 (2007), pp. 323–328.

[68] J.D. Halamka, C. Osterland, and C. Safran. "CareWeb, a web-based medical record for an integrated health care delivery system". In: *Int J Med Inform.* 54.1 (1999), pp. 1–8.

[69] T. Hambrock et al. "Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer". In: *Radiographics.* 259 (2011), pp. 453–461.

[70] N. Hara et al. "Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer". In: *Am. J. Roentgenol.* 62 (2005), pp. 140–147.

[71] D. Haraguchi et al. "A web-based image viewer for multiple PET-CT follow-up studies". In: *IEEE Eng. Med. Biol. Soc.* (2011), pp. 5279–5282.

[72] S. Hastings et al. "A Grid-Based Image Archival and Analysis System". In: *J Am Med Inform Assoc.* 12.3 (2005), pp. 286–295.

[73] M. Hazewinkel. *Quadratic deviation.* Encyclopedia of Mathematics, Springer, 2001. ISBN: 978-1-55608-010-4.

[74] K. Heckman and T.J. Schultz. *Picture Archiving and Communication System (PACS).* Vol. 29. ISBN 978-3-540-34350-9. Springer New York, 2006, pp. 249–267.

[75] J.V. Hegde et al. "Multiparametric MRI of Prostate Cancer: An Update on State-of-the-Art Techniques and Their Performance in Detecting and Localizing Prostate Cancer". In: *J. Magn. Reson. Imaging* 37.5 (2013), pp. 1035–1054.

[76] P. de Heras Ciechomski et al. "Development and Implementation of a Web-Enabled 3D Consultation Tool for Breast Augmentation Surgery Based on 3D-Image Reconstruction of 2D Pictures". In: *J. Med Internet Res.* 14.1 (2012), e21.

[77] H.H. H.K. Huang. "Short history of PACS. Part I: USA". In: *Eur. J. Radiol.* 78.2 (2011), pp. 163–176.

[78] C.M. Hoeks, J.O. Barentsz, and T. et al. Hambrock. "Prostate cancer: multi-parametric MR imaging for detection, localization, and staging". In: *Radiographics.* 261.1 (2011), pp. 46–66.

[79] J. Honeyman-Buck. "Biomedical Image Processing: Springer-Verlag Berlin Heidelberg". In: *J. Digit. Imaging* 25.6 (2011), pp. 689–691.

[80]   W. Hsu et al. "SPIRS: a Web-based image retrieval system for large biomedical databases". In: *Int J Med Inform.* 78.1 (2009), s13 –s24.

[81]   S. Hualei et al. "MIAPS: A web-based system for remotely accessing and presenting medical images". In: *Comput. Meth. Prog. Biomed.* 113.1 (2014), pp. 266–283.

[82]   H. K. Huang. *PACS and Imaging Informatics: Basic Principles and Applications.* 2nd. edition. John Wiley & Sons, Inc., 2005. ISBN: 9780471654780.

[83]   H.K. Huang. "Medical imaging, PACS, and imaging informatics: retrospective". In: *Rad. Phys. Tech.* 7.1 (2014), pp. 5–24.

[84]   K.H. Hwang, H. Lee, and D. Choi. "Medical Image Retrieval: Past and Present". In: *Healthc Inform Res.* 18.1 (2012), pp. 3–9.

[85]   K.H. Hwang, H. Lee, and D. Choi. "Merits of usability testing for {PACS} selection". In: *Int J Med Inform.* 83.1 (2014), pp. 27–36.

[86]   Digital Imaging and NEMA Publications Communications in Medicine (DICOM). "DICOM strategic document". In: The Association of Electrical Equipment and Medical Imaging Manufacturers (NEMA), 2008.

[87]   SearchHealth IT. "Picture Archiving and Communication System (PACS)". In: 2011.

[88]   A.S. Jackson et al. "Dynamic contrast-enhanced MRI for prostate cancer localization". In: *Brit. J. Radiol.* 82.974 (2009), pp. 148–156.

[89]   J.A. Jung et al. "Prostate depiction at endorectal MR spectroscopic imaging: investigation of a standardized evaluation system". In: *Radiographics.* 233.701–708 (2004).

[90]   George C. Kagadis et al. "Cloud computing in medical imaging". In: 40.7, 070901 (2013).

[91]   F.J. Kammerer et al. "MyCases – A Portable Application for Radiologic Case Collections". In: *J. Digit. Imaging* 27.5 (2014), pp. 557–562.

[92]   O. Kiliçkesmez et al. "Diffusion-weighted MRI of urinary bladder and prostate cancers". In: *Diagn. Interv. Radiol.* 15.2 (2009), pp. 104–110.

[93]   Jinman Kim, David D. Feng, and Tom W. Cai. "A Web Based Medical Image Data Processing and Management System". In: *Selected Papers from the Pan-Sydney Workshop on Visualisation - Volume 2.* VIP '00. Australian Computer Society, Inc., 2001, pp. 89–91. ISBN: 0-909-92580-1.

[94]   M. Kimura et al. "MERIT-9: a patient information exchange guideline using MML, HL7 and DICOM". In: *Int J Med Inform.* 51.1 (1998), pp. 59–68.

[95] E.H. Kluge. "Informed consent and the security of the electronic health record (EHR): some policy considerations". In: *Int J Med Inform.* 73.3 (2004), pp. 229–234.

[96] D. Korenblum et al. "Managing Biomedical Image Metadata for Search and Retrieval of Similar Images". In: *J. Digit. Imaging* 24.4 (2011), pp. 739–748.

[97] G.V. Koutelakis and D.K. Lymperopoulos. "PACS through web compatible with DICOM standard and WADO service: advantages and implementation". In: *IEEE Eng. Med. Biol. Mag. Soc.* 1 (2006), pp. 2601–2605.

[98] G.V. Koutelakis and D.K. Lymperopoulos. "WADA Service: An Extension of DICOM WADO Service". In: *IEEE Trans. Inform. Technol. Biomed.* 13.1 (2009), pp. 121–130.

[99] Piotr Kozlowski et al. "Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosisÂă-ÂăCorrelation with biopsy and histopathology". In: *J. Magn. Reson. Imaging* 24.1 (2006), pp. 108–113.

[100] W. Kuijpers et al. "A Systematic Review of Web-Based Interventions for Patient Empowerment and Physical Activity in Chronic Diseases Relevance for Cancer Survivors". In: *J. Med Internet Res.* 15.2 (2013), e37.

[101] T. Kurc et al. "An XML-based System for Synthesis of Data from Disparate Databases". In: *J Am Med Inform Assoc.* 13.3 (2006), pp. 289–301.

[102] Deanna L. Langer et al. "Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI". In: *J. Magn. Reson. Imaging* 30.2 (2009), pp. 327–334. ISSN: 1522-2586.

[103] D. Le Bihan et al. "MR imaging of intravoxel incoherent motions; application to diffusion and Perfusion in neurologic disorders". In: *Radiographics.* 161 (1986), pp. 401–407.

[104] W.B. Lee, C.D. Lee, and A. Ho KI. "HIPAA-compliant key management scheme with revocation of authorization". In: *Comput. Meth. Prog. Biomed.* 113.3 (2014), pp. 809–814.

[105] G Lemaitre et al. "Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review". In: *Comput. Biol. Med.* 60.0 (2015), pp. 8–31.

[106] H.U. Lemke. "Short history of PACS. Part II: Europe". In: *Eur. J. Radiol.* 78.2 (2011), pp. 177–183. ISSN: 0720-048X.

[107] H.K. Lim, J.K. Kim, and K.S. Kim K.A. Cho. "Prostate cancer: apparent diffusion coefficient map with T2-weighted images for detection–a multireader study". In: *Radiographics.* 250 (2009), pp. 141–151.

[108] Y. Lim et al. "A Web-based Collaborative System for Medical Image Analysis and Diagnosis". In: *Selected Papers from the Pan-Sydney Workshop on Visualisation - Volume 2*. VIP '00. Australian Computer Society, Inc., 2001, pp. 93–95. ISBN: 0-909-92580-1.

[109] H. van der Linden et al. "Inter-organizational future proof {EHR} systems: A review of the security and privacy related issues". In: *Int J Med Inform.* 78.3 (2009), pp. 141–160.

[110] L.R. Long et al. "Content-Based Image Retrieval in Medicine: Retrospective Assessment, State of the Art, and Future Directions". In: *Int J Health Inf. Syst. Inform.* 4.1 (2009), pp. 1–16.

[111] A. Madabhushi et al. *Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions.* ISBN 978-3-642-23943-4. International Workshop Held in Conjunction with MICCAI 2011. Springer, 2011.

[112] S. Mahmoudi et al. "Web-based interactive 2D/3D medical image processing and visualization software". In: *Comput. Meth. Prog. Biomed.* 98.2 (2010), pp. 172–182.

[113] B. Mansoori, K.K. Erhard, and J.L. Sunshine. "Picture Archiving and Communication System (PACS) Implementation, Integration and Benefits in an Integrated Health System". In: *Acad. Radiol.* 19.2 (2012), pp. 229–235.

[114] E. Marcos et al. "A database for medical image management". In: *Comput. Meth. Prog. Biomed.* 86.3 (2007), pp. 255–269.

[115] J. Martí et al. "MamoDB: a web-based tool for training radiologists in the diagnosis of digital mammography". In: *International Conference on Education and New Learning Technologies* (2011), pp. 2359–2367.

[116] C. Mata et al. "MammoApplet: an interactive Java applet tool for manual annotation in medical imaging". In: *IEEE Int. Conf. Bioinf. Bioeng.* (2012), pp. 34–39.

[117] C Mata et al. "ProstateAnalyzer: web-based medical application for the management of prostate cancer using multiparametric MR images". In: *Inform. Health Soc. Care* (2015), p. 0.

[118] G.K. Matsopoulos et al. "MITIS: a WWW-based medical system for managing and processing gynecological-obstetrical-radiological data". In: *Comput. Meth. Prog. Biomed.* 76.1 (2004), pp. 53–71.

[119] W. Meier. *eXist: An Open Source Native XML Database.* Vol. 2593. A.B. Chaudhri et al. (Eds.) Web Databases and Web Services. Springer-Verlag, 2003, pp. 169 –183.

[120] P. Mildenberger, M. Eichelberg, and E. Martin. "Introduction to the DICOM standard". In: *Epidemiol. Rev.* 12.4 (2002), pp. 920–927. ISSN: 05315131.

[121] M. Miller. *Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online.* 1st edition. Que publishing, 2008. ISBN: 0-7897-3803-1.

[122] D.G. Mitchell and M. Cohen. *MRI principles.* 2nd edition. Saunders, 2004. ISBN: 978-0721600246.

[123] J.M. Monguet et al. "E-Therapy-Learning". In: *International Conference on Education and New Learning Technologies* (2010), pp. 2166–2175.

[124] S.A. Moorhead et al. "A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication". In: *J. Med Internet Res.* 15.4 (2013), e85.

[125] M. Moskowitz et al. "Evaluation of new imaging procedures for breast cancer: proper process". In: *Am. J. Roentgenol.* 140.3 (1983), pp. 591–594.

[126] H. Müller et al. "A reference data set for the evaluation of medical image retrieval systems". In: *J. Comp. Medical Img. Graph.* 28.6 (2004), pp. 295–305.

[127] H. Müller et al. "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions". In: *Int J Med Inform.* 73.1 (2003), pp. 1–23.

[128] H. Münch et al. "Web-based distribution of radiological images from PACS to EPR". In: *Int. Congress Series* 1256 (2003), pp. 873–879. ISSN: 05315131.

[129] M. Mustra, K. Delac, and M. Grgic. "Overview of the DICOM standard". In: *ELMAR, 2008. 50th International Symposium.* Vol. 1. 2008, pp. 39–44.

[130] T. Neubauer and J. Heurix. "A methodology for the pseudonymization of medical data". In: *Int J Med Inform.* 80.3 (2011), pp. 190–204.

[131] J.E. de Oliveira, A. de Albuquerque Araújo, and T.M. Deserno. "Content-based image retrieval applied to BI-RADS tissue classification in screening mammography". In: *World J Radiol.* 3.1 (2011), pp. 24–31.

[132] A. Oliver. "Automatic mass segmentation in mammographic images". In: 2007.

[133] A. Oppelt. *Systems for Medical Diagnostics: Fundamentals, Technical Solutions and Applications for Systems Applying Ionizing Radiation.* Germany: Publicis Publishing, 2005.

[134] M. Ortega et al. "Sirius: A web-based system for retinal image analysis". In: *Int J Med Inform.* 79.10 (2010), pp. 722–732.

[135] S. Oster et al. "caGrid 1.0: an enterprise Grid infrastructure for biomedical research". In: *J Am Med Inform Assoc.* 15.2 (2008), pp. 138–149.

[136] A. Oto et al. "Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis." In: *Am. J. Roentgenol.* 197.6 (2011), pp. 1382–1390. ISSN: 1546-3141.

[137] I. Papadakis, V. Chrissikopoulos, and D. Polemi. "Secure medical digital libraries". In: *Int J Med Inform.* 64.3 (2001), pp. 1–8.

[138] S. Parfait et al. "Classification of prostate magnetic resonance spectra using Support Vector Machine". In: *Biomed. Sig. Proc. Contr.* 7.5 (2012), pp. 499–508. ISSN: 1746-8094.

[139] Y. Peng et al. "Quantitative Analysis of Multiparametric Prostate MR Images: Differentiation between Prostate Cancer and Normal Tissue and Correlation with Gleason Score-A Computer-aided Diagnosis Development Study". In: *Radiographics.* 267.3 (2013), pp. 787–796.

[140] J. Philbin, F. Prior, and P. Nagy. "Will the Next Generation of PACS Be Sitting on a Cloud?" In: *J. Digit. Imaging* 2.24 (2011), pp. 179–183.

[141] O.S. Pianykh. *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide.* 2nd ed. Springer, 2012. ISBN: 978-3-642-10850-1.

[142] B. Pickett et al. "Use of MRI and spectroscopy in evaluation of external beam radiotherapy for prostate cancer". In: *Int J Radiat. Oncol. Biol. Phys.* 60 (2004), pp. 1047–1055.

[143] D.B. Plewes. "Contrast mechanisms in spin-echo MR imaging". In: *Radiographics.* 14 (1994), pp. 1389–1404.

[144] D. Portalez et al. "Validation of the European Society of Urogenital Radiology scoring system for prostate cancer diagnosis on multiparametric magnetic resonance imaging in a cohort of repeat biopsy patients". In: *Eur Urol.* 62.6 (2012), pp. 986–996.

[145] S.W. Provencher. "Estimation of metabolite concentrations from localized in vivo proton NMR spectra". In: *Magn. Reson. Imaging* 30.6 (1993), pp. 672–679.

[146] W. Provencher. "Automatic quantitation of localized in vivo 1H spectra with LCModel". In: *NMR Biomed.* 14 (2001), pp. 260–264.

[147] T. Rajala et al. "Development of a Research Dedicated Archival System (TARAS) in a University Hospital". In: *J. Digit. Imaging* 24 (2011), pp. 864–873.

[148] A.M. Rassinoux et al. "Architecture for Knowledge-Based and Federated Search of Online Clinical Evidence". In: *J. Med Internet Res.* 7.5 (2005), e52.

[149] A.M. Rassinoux et al. "Xml as standard for communicating in a document-based electronic patient record: a 3 years experiment". In: *Int J Med Inform.* 70.2 (2003), pp. 109–115.

[150] L.A. Reisæter et al. "1.5-T multiparametric MRI using PI-RADS: a region by region analysis to localize the index-tumor of prostate cancer in patients undergoing prostatectomy". In: *Acta Radiol.* (2014).

[151] J. Ren et al. "Diffusion-weighted imaging in normal prostate and differential diagnosis of prostate diseases". In: *Abdom. Imaging.* 33.6 (2008), pp. 724–728.

[152] J. Ren et al. "Seminal vesicle invasion in prostate cancer: prediction with combined T2-weighted and diffusion-weighted MR imaging". In: *Eur. Radiol.* 19.10 (2009), pp. 2481–2486. ISSN: 0938-7994.

[153] A.C. Rencher and W.F. Christensen. *Chapter 10, Multivariate regression - Section 10.1, Introduction.* Vol. 2593. Methods of Multivariate Analysis, Wiley Series in Probability and Statistics 709 (3rd ed.), John Wiley & Sons, 2012, p. 19.

[154] J.A. Rice. *Mathematical Statistics and Data Analysis.* Third Edition, Duxbury Advanced, 2006.

[155] M.C. Roethke et al. "Evaluation of the ESUR PI-RADS scoring system for multiparametric MRI of the prostate with targeted MR/TRUS fusion-guided biopsy at 3.0 Tesla". In: *Eur. Radiol.* 24.2 (2014), pp. 344–352.

[156] A.B. Rosenkrantz et al. "Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales". In: *Radiographics.* 269.2 (2013), pp. 482–492.

[157] A.B. Rosenkrantz et al. "Utility of Diffusional Kurtosis Imaging as a Marker of Adverse Pathologic Outcomes Among Prostate Cancer Active Surveillance Candidates Undergoing Radical Prostatectomy". In: *Am. J. Roentgenol.* 201.4 (2013), pp. 840–846.

[158] A. Rosenthal et al. "Cloud computing: a new business paradigm for biomedical information sharing". In: *J. Biomed. Inform.* 43 (2010), pp. 342–353.

[159] A. Rosset et al. "Integration of a Multimedia Teaching and Reference Database in a PACS Environment". In: *Radiographics.* 22 (2002), pp. 1567–1577.

[160] G. Rote. "Computing the minimum Hausdorff distance between two point sets on a line under translation". In: *Inform. Proces. Letters* (1991), pp. 123–127.

[161] O. Rouvière et al. "Is it possible to model the risk of malignancy of focal abnormalities found at prostate multiparametric MRI?" In: *Eur. Radiol.* 22.5 (2012), pp. 1149–1157.

[162] D.L. Rubin. "Creating and curating a terminology for radiology: ontology modeling and analysis". In: *J. Digit. Imaging* 21.4 (2008), pp. 355–362.

[163] T.W. Scheenen et al. "Three-dimensional proton MR spectroscopy of human prostate at 3 T without endorectal coil: feasibility". In: *Radiographics.* 245.507–516 (2007).

[164] F.H. Schröder et al. "Prostate-Cancer Mortality at 11 Years of Follow-up". In: *New Engl. J. Med.* 366.11 (2012), pp. 981–990.

[165] D. Schweiger, Z. Trajanoski, and S. Pabinger. "SPARQLGraph: a web-based platform for graphically querying biological Semantic Web databases". In: *BMC Bioinformatics* 15.279 (2014), pp. 1–5.

[166] G. Schwitzer. "A Review of Features in Internet Consumer Health Decision-support Tools". In: *J. Med Internet Res.* 4.2 (2002), e11.

[167] KM. Selnæs et al. "Peripheral zone prostate cancer localization by multiparametric magnetic resonance at 3 T: unbiased cancer identification by matching to histopathology". In: *Invest. Radiol.* 47.11 (2012), pp. 624–633.

[168] N. Sethi et al. "Disaster easily averted? – Data confidentiality and the hospital desktop computer". In: *Int J Med Inform.* 83.5 (2014), pp. 385–391.

[169] V. Shah et al. "Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging". In: *Med. Phys.* 39.7 (2012), pp. 4093 –103.

[170] J. Shan, H.D. Cheng, and Y. Wang. "Completely Automated Segmentation Approach for Breast Ultrasound Images Using Multiple-Domain Features". In: *Ultrasound Med. Biol.* 38.2 (2012), pp. 262–275.

[171] R. Siegel et al. "Cancer statistics, 2014". In: *CA: Cancer J. Clin* 64.9–29 (2014).

[172] Luís Bastião A. Silva, Carlos Costa, and José Luis L. Oliveira. "A PACS archive architecture supported on cloud services." In: *Comput. Aided Rad. Surg.* (2011), pp. 1–10. ISSN: 1861-6429.

[173] U. Sinha et al. "A review of medical imaging informatics". In: *Ann N Y Acad Sci.* 980.1 (2002), pp. 168–197.

[174] S.W. Smith. *The Scientist & Engineer's Guide to Digital Signal Processing.* DSP Guide, 1997. ISBN: 978-0966017632.

[175] Blausen.com staff. "Blausen gallery 2014". In: 2014. ISBN: 20018762.

[176]   N. Strickland. "PACS (picture archiving and communication systems): film-less radiology". In: *Arch. Dis. Child.* 83.1 (2000), pp. 82–86.

[177]   S. Sucurovic. "Implementing security in a distributed web-based EHCR". In: *Int J Med Inform.* 76.6 (2007), pp. 491–496.

[178]   S. Sumathi and S.N. Sivanandam. *Introduction to Data Mining and its Applications.* Vol. 29. ISBN 978-3-540-34350-9. Springer, Studies in Computational Intelligence (SCI), 2006.

[179]   Chia-Chi T. et al. "A medical image archive solution in the cloud". In: *IEEE international conference on software engineering and ser- vice sciences (IC-SESS 2010)* (2010), pp. 431–434.

[180]   D. Tahmoush and H. Samet. "A new database for medical images and information". In: *in Proceedings of SPIE - Medical Imaging 2007: Image Processing.* 2007.

[181]   A. Tanimoto et al. "Prostate cancer screening: The clinical value of diffusion-weighted imaging and dynamic MR imaging in combination with T2-weighted imaging". In: *J. Magn. Reson. Imaging* 25.1 (2006), pp. 146–152.

[182]   W.M. Tellis and K.P. Andriole. "Finding the optimal picture archvingand communciation system(PACS)architecture: A comparison of three PACS designs". In: *J. Digit. Imaging* 14.1 (2001), pp. 72–76.

[183]   C. Testa et al. "Prostate cancer: sextant localization with MR imaging, MR spectroscopy, and 11C-choline PET/CT." In: *Radiographics.* 244.3 (2007), pp. 797–806.

[184]   M. Tortajada et al. "Breast peripheral area correction in digital mammograms". In: *Comput. Biol. Med.* 50 (2014), pp. 32–40.

[185]   B. Turkbey et al. "Integrating multiparametric prostate MRI into clinical practice". In: *Cancer Imaging* 11.Spec No A (2011), pp. 27–37.

[186]   B. Turkbey et al. "Prostate Cancer: Value of Multiparametric MR Imaging at 3 T for Detection-Histopathologifc Correlation". In: *Radiographics.* 255.1 (2010), pp. 89–99.

[187]   T. Vaché et al. "Characterization of Prostate Lesions as Benign or Malignant at Multiparametric MR Imaging: Comparison of Three Scoring Systems in Patients Treated with Radical Prostatectomy". In: *Radiographics.* 272.2 (2014), pp. 446–455.

[188]   S. Verma and A. Rajesh. "A clinically relevant approach to imaging prostate cancer: review". In: *Am. J. Roentgenol.* 196.3 (2011), S1–10.

[189]   S. Verma et al. "Prostate MRI and 3D MR Spectroscopy: How We Do It." In: *Am. J. Roentgenol.* 194.6 (2010), pp. 1414–1426.

[190] G.M Villeirs et al. "A qualitative approach to combined magnetic resonance imaging and spectroscopy in the diagnosis of prostate cancer". In: *Eur. Radiol.* 73.352–356 (2010).

[191] A. Villers et al. "Dynamic contrast enhanced, pelvic phased array magnetic resonance imaging of localized prostate cancer for predicting tumor volume: correlation with radical prostatectomy findings". In: *J Urol.* 176.2432–2437 (2006).

[192] P.C. Vos et al. "Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI". In: *Phys. Med. Biol.* 55.6 (2010), pp. 1719 –1734.

[193] C. Wang, F. Ritter, and A. Smedby. "Making the PACS workstation a browser of image processing software: a feasibility study using inter-process communication techniques". In: *Int. J. Comput. Assist. Radiol. Surg.* 5.4 (2010), pp. 411–419.

[194] P. Welter et al. "Bridging the integration gap between imaging and information systems: a uniform data concept for content-based image retrieval in computer-aided diagnosis". In: *J Am Med Inform Assoc.* 18.4 (2011), pp. 506–10.

[195] A. Westphalen and A. Rosenkrantz. "Prostate Imaging Reporting and Data System (PI-RADS): Reflections on Early Experience With a Standardized Interpretation Scheme for Multiparametric Prostate MRI". In: *Am. J. Roentgenol.* 202.1 (2014), pp. 121–123.

[196] R. Wetering and R. Batenburg. "Towards a Theory of PACS Deployment: An Integrative PACS Maturity Framework". In: *J. Digit. Imaging* (2014), pp. 1–14.

[197] L.S. Wilkinson and N.T.F Ridley. "The Practical Application of the UK 5-Point Scoring System for Breast Imaging: How Standardisation of Reporting Supports the Multidisciplinary Team." In: *Brit. J. Radiol.* 84.1007 (2011), pp. 965–966.

[198] S. Wong. *Medical image databases.* Vol. 465. The Springer International Series in Engineering and Computer Science, 1998. ISBN: 978-3-540-34350-9.

[199] P. Wunderbaldinger et al. "World Wide Web and Internet: applications for radiologists". In: *Eur. Radiol.* 9.6 (1999), pp. 1170–1182. ISSN: 0938-7994.

[200] *XML Databases and Biomedical Informatics.* Lindsay, J. School of Engineering. UCONN CSE 300, University of Connecticut, 2008.

[201] J. Zhang, J. Sun, and J.N. Stahl. "PACS and Web-based image distribution and display". In: *J. Comp. Medical Img. Graph.* 27.2-âĂŞ3 (2003), pp. 197–206. ISSN: 0895–6111.

[202] Z. Zhou and B.J. Liu. "HIPAA compliant auditing system for medical images". In: *J. Comp. Medical Img. Graph.* 29.2 (2005), pp. 235–241.