



UNIVERSIDADE DO ALGARVE

Diversifying medical imaging of breast lesions

Master thesis

Riyadh Mohammed Al-Tam

2015

Supervisors:

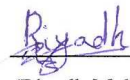
Prof^{ta} Doutora Maria Margarida da Cruz Silva Andrade Madeira e Carvalho de Moura
Prof.^a Dr^a Teresa Leonor Isabel Dulce Fiel de Vasconcelos Figueiredo

Statement of Originality

Diversifying medical imaging of breast lesions

Statement of authorship: The work presented in this thesis is, to the best of my knowledge and belief, original, except as acknowledged in the text. The material has not been submitted, either in whole or in part, for a degree at this or any other university.

Candidate:



(Riyadh Mohammed Al-Tam)

Copyright © Riyadh Mohammed Al-Tam. A Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científico e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Abstract

Breast cancer is the most frequent cancer in women and remains the main cause of woman mortality, having also registered an increased incidence worldwide. Early detection of breast cancer is of major importance to increase the survival rate. Ultrasound, mammography and magnetic resonance modalities are used to increase the efficiency and accuracy of diagnose. The goal of the work described here is the implementation of a repository of medical images of the breast regardless of origin that can be used by medical doctors, researchers or in a medical faculty, to collect, anonymize, share, annotate and classify. Available standard and tools were used and integrated with the special-purpose ones developed. DICOM, an international standard for medical digital image format and communications protocol, was used. dcm4chee was selected to implement a Picture Archiving and Communication System (PACS) and was integrated within a MySQL server to provide the required functionalities of the data repository. In order to increase the security of the whole system, VPN (IPsec) is used to create a secure channel between the MySQL server and the client solutions. Patient privacy is guaranteed by the deletion of identifiers and use of anonymisation algorithms before upload. Annotation is achieved using the image application FIJI complemented with a modified “DICOM Exporter” class from the Tudor application. Interaction with the repository is facilitated by client side solutions that integrate some tools and extend them by implementing new desired functionalities. These solutions, integrating FIJI, connectors to dcm4chee and MySQL and using VPN (IPsec), support the activities of each of the user roles considered, namely, main doctor, hospital doctor, teacher, student, researcher and admin. Testing was done with more than one hundred files uploaded to the repository.

Keywords: *Annotation, Anonymisation, Security, Dcm4chee, FIJI, Breast lesion, Classification, Network.*

Resumo

O Cancro da Mama é o cancro mais frequente no sexo feminino e a permanece com a principal causa de mortalidade para as mulheres, tendo também registado um aumento da incidência em todo o mundo. A deteção precoce do cancro da mama é muito importante para aumentar a probabilidade de sobrevivência. As modalidades de ultrassom, mamografia, ressonância magnética são utilizadas a fim de aumentar a eficiência e a eficácia de um diagnóstico. O objetivo do trabalho aqui descrito é a implementação de um repositório de imagens médicas que, independentemente da origem dessas imagens e usado por médicos, investigadores ou em contexto de ensino, seja usado em recolha, anonimização, partilha, anotação e classificação. Standards e ferramentas disponíveis foram usadas e integradas com umas de propósito particular desenvolvidas. DICOM, um standard internacional para o formato de imagens médicas e protocolo de comunicação, foi usado. A aplicação dcm4chee foi selecionada para implementar um PACS (do inglês, Picture Archiving and Communication System, um sistema de arquivo e comunicação) e foi integrada num servidor MySQL para prover às funcionalidades requeridas do repositório de dados. Para aumentar a segurança do sistema, VPN (IPsec) foi usado para criar um canal seguro entre o servidor e as soluções clientes. A privacidade dos pacientes é garantida pela eliminação de identificadores e uso de algoritmos de anonimização antes do carregamento para o repositório. A anotação é conseguida com a aplicação de imagem FIJI complementada com a modificação de uma classe “DCOM Exporter” proveniente da aplicação Tudor. A interação com o repositório é facilitada por soluções clientes que integram algumas ferramentas e as estendem implementando as funcionalidades desejadas. Estas soluções, integrando FIJI, conetores a dcm4chee e MySQL e usando VPN (IPsec), suportam as atividades de cada papel de utilizador considerado, nomeadamente médico principal, médico hospitalar, professor, aluno, investigador e administrador. O teste foi feito com mais de uma centena de ficheiros no repositório.

Palavras-chave: *Anotação, anonimização, Segurança, dcm4chee, Fiji, câncer de mama, de classificação, de rede*

Acknowledgments

I would like to express my gratitude to my supervisors Prof^a Doutora Maria Margarida Madeira e Moura and Dr^a Teresa Figueiredo for the useful comments, remarks and engagement through this work. Prof^a Doutora Margarida helped me challenge myself and push towards higher goals. She taught me the right approach of looking at problems, formulating a good design and grounding myself in the core tenets of self-belief, consistency and coherence. This thesis would have been impossible to complete without her support, encouragement, and insight.

Special thanks for all teachers of the University of Algarve who taught me different topics, which in turn helped me to complete this thesis.

Lastly, I thank my ever-awesome family for supporting and encouraging me through the entire period and bearing with me for periods of minimal communication. Their endless love, unconditional support and encouragement have given me the strength to accomplish this thesis.

Contents

| | |
|--|----|
| Abstract | 3 |
| Resumo..... | 4 |
| Acknowledgments..... | 5 |
| 1 Introduction..... | 10 |
| 1.1 Motivation | 10 |
| 1.2 Goal | 11 |
| 1.3 Thesis outline | 11 |
| 2 Background..... | 13 |
| 2.1 Breast Imaging..... | 13 |
| 2.1.1 Breast Cancer | 13 |
| 2.1.2 Breast Cancer Causes | 13 |
| 2.1.3 Breast Anatomy..... | 14 |
| 2.1.4 Breast Cancer Types | 15 |
| 2.1.5 Mammography..... | 15 |
| 2.1.6 Ultrasound..... | 16 |
| 2.1.7 Elastography..... | 16 |
| 2.1.8 Picture Archiving and Communication System | 16 |
| 2.2 Selected Issues on Medical Informatics | 17 |
| 2.2.1 PACS Servers..... | 17 |
| 2.2.2 DICOM file format | 25 |
| 2.2.3 PACS viewers | 31 |
| 2.2.4 Dcm4chee..... | 32 |
| 2.2.5 MySQL server overview..... | 36 |
| 2.2.6 Overview of gluing technologies | 37 |
| 3 System Analysis and Requirements..... | 39 |
| 3.1 Introduction..... | 39 |
| 3.2 Glossary of Terms | 40 |
| 3.3 Functional requirements | 43 |
| 3.3.1 Use case diagram..... | 43 |
| 3.3.2 Use case descriptions | 46 |
| 3.4 Data base diagram..... | 51 |
| 4 Medical Applications | 54 |
| 4.1 Introduction..... | 54 |

| | | |
|-------|---|----|
| 4.2 | MAA (Medical Admin Application) | 56 |
| 4.3 | MMA (Medical Main Application) | 57 |
| 4.4 | MDA (Medical Doctor Application) | 57 |
| 4.5 | MTA (Medical Teacher Application) | 57 |
| 4.6 | MSA (Medical Student Application) | 58 |
| 4.7 | MRA (Medical Researcher Application) | 58 |
| 4.8 | Methodology | 58 |
| 4.8.1 | Ordering DICOM files | 58 |
| 4.8.2 | Anonymization | 60 |
| 4.8.3 | Grouping files | 65 |
| 4.8.4 | Security | 65 |
| 5 | Discussion, conclusion, and future work | 68 |
| 5.1 | Discussion | 68 |
| 5.2 | Conclusion | 69 |
| 5.3 | Future work | 70 |
| 6 | User Manuals | 71 |
| 6.1 | MedDoctor plugin User Manual | 71 |
| 6.2 | Server side | 71 |
| 6.2.1 | The installation steps | 71 |
| 6.2.2 | The configuration steps for dcm4chee | 73 |
| 6.2.3 | The usage steps for server side | 74 |
| 6.3 | Client side | 80 |
| 6.3.1 | The installation steps for a doctor, a student, a teacher, a hospital doctor, or a researcher | 80 |
| 6.3.2 | The usage steps for a doctor, a student, a teacher, a hospital doctor, or a researcher | 82 |
| | Bibliography | 97 |

List of Figures

| | |
|---|----|
| Figure 2.1: Breast anatomy (National Institutes of Health, 2014)..... | 15 |
| Figure 2.2: MRIdb components (Woodbridge, 2013) | 18 |
| Figure 2.3: Dicoogle components and interfaces (Costa, 2011) | 20 |
| Figure 2.4: DICOM model (Evans, 2008)..... | 27 |
| Figure 2.5: DICOM element (NEMA, 2014) | 28 |
| Figure 2.6: Interlaced way..... | 30 |
| Figure 2.7: Separated way | 30 |
| Figure 2.8: Dcm4chee architecture (Evans, 2008) | 34 |
| Figure 2.9: Dcm4chee workflow..... | 35 |
| Figure 3.1: System requirements | 40 |
| Figure 3.2: The system registration..... | 43 |
| Figure 3.3: A main user tasks | 44 |
| Figure 3.4: Sending/retrieving DICOM files to/from the archive..... | 44 |
| Figure 3.5: The workflow between a teacher and a student | 45 |
| Figure 3.6: Basic tasks of doctor and researcher | 45 |
| Figure 3.7: Classify a patient study | 46 |
| Figure 3.8: The proposed database to keep classifications, reports, DICOM files parameters, methods, and groups. | 53 |
| Figure 4.1: The architecture of the system | 56 |
| Figure 4.2: Anonymize DICOM header | 62 |
| Figure 4.3: An example of the proposed views to give limited access to some rows in a table..... | 67 |
| Figure 6.1: An admin login interface | 74 |
| Figure 6.2: Add users and configure the connection interface..... | 75 |
| Figure 6.3: The interface of inserting a user to the system. | 76 |
| Figure 6.4: Using the default configuration of dcm4chee..... | 76 |
| Figure 6.5: Add new role in decm4chee..... | 78 |
| Figure 6.6: Assign a role to the admin user..... | 78 |
| Figure 6.7: Receive and run commands of the researcher application at the server. | 79 |
| Figure 6.8: A web site for download FIJI for windows 7 (32bit)..... | 80 |
| Figure 6.9: Put a Zipped FIJI file into a folder | 81 |
| Figure 6.10: FIJI extracting..... | 81 |
| Figure 6.11: FIJI application path | 82 |
| Figure 6.12: FIJI interface | 82 |
| Figure 6.13: Login interface of a doctor | 83 |
| Figure 6.14: Anonymize and send DICOM files and reports to dcm4chee | 84 |
| Figure 6.15: The Criteria of retrieve DICOM files from dcm4chee..... | 84 |
| Figure 6.16: Classifying DICOM files | 85 |
| Figure 6.17: Drawing shapes on DICOM images | 86 |
| Figure 6.18: Classify a study date for a patient | 87 |
| Figure 6.19: Create a group..... | 88 |
| Figure 6.20: Add elements to a group..... | 88 |
| Figure 6.21: Assign a group to another group..... | 89 |
| Figure 6.22: Load single names to the system database..... | 89 |

| | |
|--|----|
| Figure 6.23 : Review the classification of students and put grades..... | 90 |
| Figure 6.24 : Export students' grades to csv files | 91 |
| Figure 6.25 : Modified batch processor to apply a group of plugins on DICOM files..... | 92 |
| Figure 6.26: Retrieve a group of DICOM files parameters. | 94 |
| Figure 6.27: Apply a group of plugins to a group of DICOM files..... | 95 |
| Figure 6.28: Read a csv file and send its lines to PACS server..... | 96 |

List of Tables

| | |
|---|----|
| Table 2.1: Summary of PACS servers' pros and cons | 21 |
| Table 2.2: Different transfer syntaxes (NEMA, 2014):..... | 29 |
| Table 3.1: Anonymization procedure | 46 |
| Table 3.2: Classify a DICOM file | 47 |
| Table 3.3: Classify a patient study | 47 |
| Table 3.4: Send or retrieve a DICOM file from/to PACS | 48 |
| Table 3.5: A group of users, files, or reports..... | 49 |
| Table 3.6: Putting a grade of student..... | 49 |
| Table 3.7: Assign a group to another group | 50 |
| Table 3.8: Apply a method on a DICOM image | 50 |

1 Introduction

Breast cancer is a disease that afflicts a significant number of women worldwide. Whenever a suspicion is raised or above a certain age, periodical exams usually including mammography and sonography are recommended.

In the last decade, the technological advances have introduced into the practice Elastography, a complementary imaging technique using ultrasound, which produces real-time qualitative images representing the relative stiffness of the tissue. In spite of being promising, the use of elastography is not widespread and, due to its relative novelty, elastography is not yet a mature technology. Other imaging technologies, such as mammography or MRI, are used to evaluate the medical conditions in some particular cases.

Although several viewers are available, from the manufacturers or from the medical informatics community, these viewers lack the capability to present images from different sources. The set of available solutions is small if, beyond viewing, one wants to add marks and classification to the files.

Many medical images can be collected for long-terms and may be for short-terms, so in general, PACS systems are used to collect such images and allowing for high-performance and high-speed access.

The ability to diagnose is enhanced by the knowledge that comes from experience. The opportunity to work on many cases and to be able to know how accurate the diagnose was is an important tool for students and young doctors. The available databases of medical images are few, and fewer have the classification of the lesions.

1.1 Motivation

The main idea behind this project is to create a repository that keeps all the medical images of patients with different cases of breast lesions. It can be used at a hospital or at the medical faculty. It not only helps doctors or radiologists to collect images of patients that have been taken from mammography, ultrasound, or MRI modalities, but also allows them to annotate, anonymize, send, retrieve, share and classify these files. Different medical files might be classified and reviewed by different doctors or radiologists, which helps to suggest suitable treatments for patients. Many medical images can be accessed by students to be classified, which it increases the knowledge of students. To support teachers at the faculty of medicine with a real medical images with different cases of breast lesions to be used for teaching purposes and to create tutorials. Developers cannot only get a group of medical images to apply

the technique of image processing to detect breast lesions, but also compare their results with the correct classification of doctors at a hospital.

1.2 Goal

Nowadays, there are a lot of tools that prepared to detect and diagnose diseases in the medical area. For this dissertation, we want to create a system that supports radiologists, doctors, teachers, students, or researchers for classifying breast lesions. Medical files will be collected as data repository and can be accessed through the Internet. The ability to annotate DICOM files should be available. The personal data inside medical files must be anonymized. Different users will be available; some of them work at a hospital, others work at a medicine faculty, while the rest work outside of medical area such as developers. This system must work on different operating systems and be able to integrate methods and tools previously developed by other authors aiming image pre-processing, segmentation, and image conversion.

1.3 Thesis outline

The organization of this thesis is uncommon. It consists of six chapters, including the current one. The current chapter introduces the work done, presents the goal and motivation of this project and the thesis outline.

Chapter 2 presents the background of the work, namely in what concerns breast imaging and some selected issues in medical informatics. Included is general medical information about breast cancer, its types and causes, and the changes that occur in the breast of a female after puberty and some imaging modalities used in the early diagnose of breast cancer such as Ultrasound, Mammography, and Elastography. In what concerns medical informatics, an introduction to a PACS system and its uses and the reviewing of advantages and disadvantages of PACS servers and DICOM files viewers. The DICOM standard and the components of its DICOM file are discussed in more detail to allow a developer to get a general idea about how to deal correctly with a DICOM file, especially in reading or writing data. The selected PACS server (dcm4chee) is shown with its components tools for retrieving or sending DICOM files. A small brief review about the gluing techniques of this system is presented.

System analysis and requirements are discussed in Chapter 3, introducing the extra requirements that should be considered beyond those supported by the chosen tools, FIJI and Dcm4chee, to achieve the requirements of the work presented here.

Chapter 4 presents how the proposed client solutions work together. The advantages of each is presented and the approaches studied to the issues of anonymization and security are also discussed.

In Chapter 5, a discussion of alternative approaches is presented and some concluding remarks and suggestions of future work are presented.

The uncommon choice of including the user manuals as the Chapter 6 was deemed necessary to convey the work done on the development of this project. The user manuals contain the necessary information to install, configure, and use the developed plugins.

2 Background

2.1 Breast Imaging

2.1.1 Breast Cancer

Breast cancer occurs when some cells of the breast start to be a malignant tumour, usually in the ducts and lobules (Radiological Society of North America, Inc, 2013). It is one of the leading causes of cancer mortality among women in the United States (Radiological Society of North America, Inc, 2013). During the life of a woman, the approximately of developing invasive breast cancer is one in eight (Radiological Society of North America, Inc, 2013). Many studies show that early detection of suspicious lesions is crucial for the diagnosis of a patient. The tumour typically grows into nearby tissues and enters the bloodstream or lymphatic system, which can spread out to other organs. The mammography and ultrasonography (USG) are used to depict the breast and to detect breast cancer. There are several types of breast cancers. Ductal carcinoma in situ (DCIS) is the most frequent type of breast cancer eight (Radiological Society of North America, Inc, 2013). DCIS has frequent markers called clusters of micro-calcification. The mammography equipment is not enough to diagnose the breast lesions for all cases, especially when a large amount of glandular tissue exists, because the produced breast image will be very bright, which significantly decreases the visibility of micro-calcification (Radiological Society of North America, Inc, 2013). There are many studies show that mammography cannot give a clear screen image for all types of breast lesions, so ultrasound and magnetic resonance imaging (MRI) are used to help supplement mammography to give better results (Radiological Society of North America, Inc, 2013). MRI or ultrasound screening is not appropriate for all women, so a doctor or a radiologist determines which one is suitable.

2.1.2 Breast Cancer Causes

There are many factors that might influence the appearance of breast cancer, but until now the scientific community cannot know exactly which factors cause atypical cells to become cancer. Certain inherited DNA mutations, breast density, and hormones in many cases can increase the risk for the appearance of cancer (American Cancer Society, Inc., 2013). This disease can appear in men, however it is about a hundred times more common in women than men. The age can increase the possibility of breast cancer, for example, in women younger than 45 there is about one out of eight invasive breast cancers, while in women's age 55 or older there are about two of three invasive breast cancers (American Cancer Society, Inc., 2013).

Under 45 years, the breast cancer is more common in African-American women than Asian, Hispanic, and Native-American women. Women with dense breasts can develop the cancer more than women with less dense breasts. Drinking alcohol increases the risk of developing breast cancer. The breastfeeding at a young age reduce breast cancer. Walking for a period between 1.25 to 2.5 hours per week reduces the risk of developing breast cancer by 18% cancers (American Cancer Society, Inc., 2013).

2.1.3 Breast Anatomy

The breasts are located in the left and right sides of the upper abdominal region of the trunk. The base of the breast extends from the lateral border of the sternum to the midaxillary line and from the 2nd to the 6th ribs (Hall, 2012). Figure 2.1 shows an anatomical image of a female breast and a profile image of the mammary gland (Gray, 2014). The mammary gland is responsible for producing milk. Both males and females have the glandular tissue in the breast, but after puberty, the glandular tissue begins to develop in response to the release of estrogen and progesterone in females (Gray, 2014). The surface of each breast is convex and has a small conical nipple on the top (Gray, 2014). The nipple's base is surrounded by an areola, which has a slightly rough surface due to the presence of rudimentary mammary glands and areolar glands. The breast composes not only of the mammary gland tissue, but also fibrous and fatty tissue, ducts, blood and lymph vessels, nerves, and 15 – 20 lobes (Gray, 2014). Each lobe consists of a group of lobules, which hold bulbs that actually produce breast milk. (University of Connection, 2014). Lactiferous ducts enlarge during lactation to form a small lactiferous sinus, which carry milk from the lobes of each breast to the nipple. The fibrous tissue connects lobes together and extend over the entire surface of the breast. The surface of the gland tissue covers with the fatty tissue, except for the areola. The breast shape and size depends on the fatty tissue density. The weight and dimension of breast vary from person to another and at different periods of life (Gray, 2014). The development of breast is increased during pregnancy, when oestrogen level is raised, or when mammary glands are secreting milk for lactation. In the old age, the breast becomes with more lipomatous tissue. The breast in children consists mainly of ducts with scattered alveoli, and these ducts in females and males are the same. The breast is made up mostly of fibrous tissue and glands in adolescents, but in adults, the fat replaces some of the fibrous and glandular tissue.

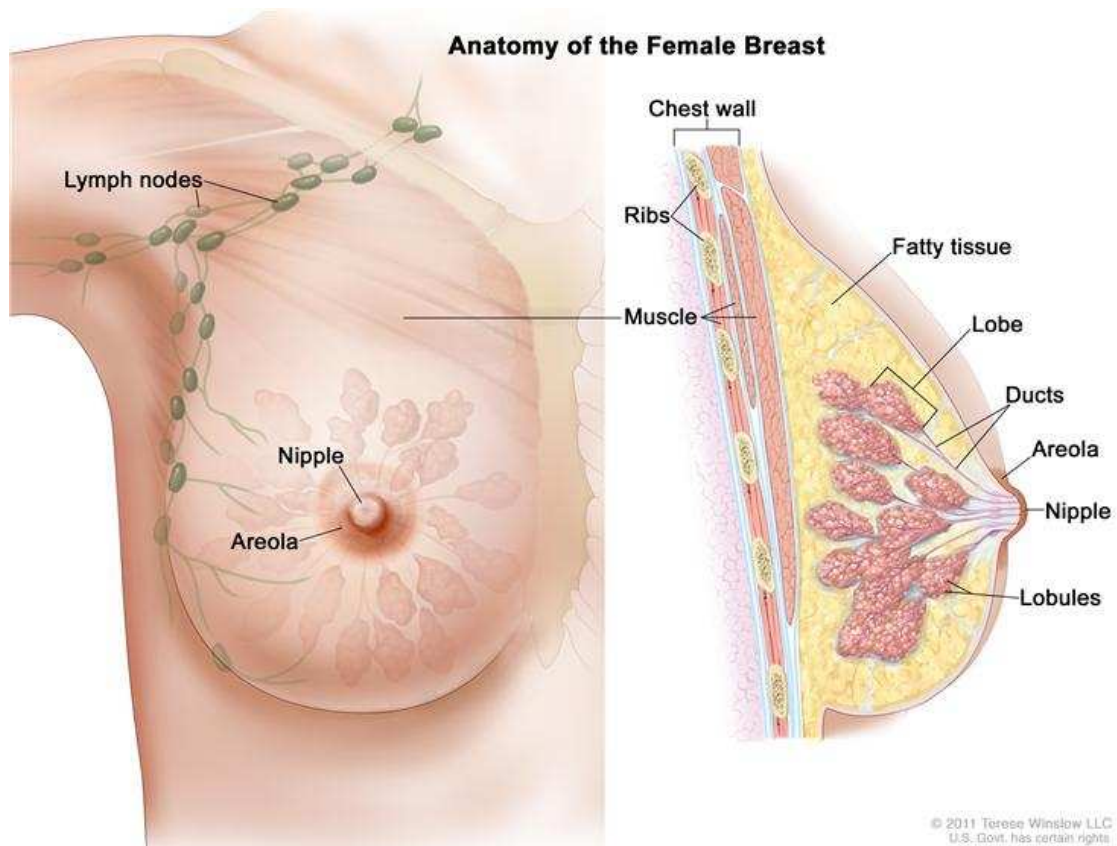


Figure 2.1: Breast anatomy (National Institutes of Health, 2014)

2.1.4 Breast Cancer Types

There are many types of breast cancer, but some of them are rare. The most common types are Lobular carcinoma in situ, Ductal carcinoma in situ (DCIS), Invasive (or infiltrating) ductal carcinoma, and Invasive (or infiltrating) lobular carcinoma (American Cancer Society, Inc., 2013). The majority of breast tumours detected by mammography are benign. There are some less common types of breast cancer, such as inflammatory breast cancer, Triple-negative breast cancer, Paget disease of the nipple, Phyllodes tumour, and Angiosarcoma. The cancer is called non-invasive if the malignant cells have not gone through the basal membrane, but is contained entirely in the lobule of the ducts. And, called invasive if the cancer has broken through the basal membrane and spread into the surrounding tissue.

2.1.5 Mammography

Mammography is an imaging technology that allows to evaluate the breast tissue. (National Library of Medicine, 2013). Mammography is performed to screen a woman breast to detect an early breast cancer, which helps for a better treatment and early diagnosis. The

result of mammography is a mammogram to be used by a radiologist or a doctor to diagnose the breast disease (National Library of Medicine, 2013). It is recommended for women at age 40 or older to do this test once or twice a year. A woman should consider yearly mammograms if she has a mother or sister who had breast cancer at a younger age. Mammography is also used to follow a woman who has had an abnormal mammogram and symptoms of breast cancer. There are different symptoms such as a lump, breast pain, nipple discharge, changes in the nipple, dimpling of the skin on the breast, or other symptoms (National Library of Medicine, 2013).

2.1.6 Ultrasound

It is increasingly used for diagnostic purposes by using sound waves to create images of the organs and structures inside the human body (Leighton, 2007). Ultrasound is used in many different fields such as veterinary and human medicine. Ultrasound instrumentation allows monitoring of the moving structures inside the body, for example, foetal movement and blood flow. It is a sound wave higher than 20 kHz and this means that its frequencies are above the range of human audible frequencies.

2.1.7 Elastography

It is a dynamic technique that uses ultrasound to noninvasively assess the mechanical stiffness of tissue by measuring tissue distortion in response to external compression (Analogic, 2015). Ultrasound Elasticity Imaging or elastography is used for breast examination with a number of high-resolution linear transducers to differentiate between benign from malignant breast lesions. It helps to reduce the number of biopsies.

2.1.8 Picture Archiving and Communication System

Nowadays, a PACS (Picture archiving and communication system) is used for storing, sending, and retrieving medical images. It can directly acquire the medical images from different imaging instruments and be accessed from workstations for viewing and reporting. In general, PACS systems have both high-performance and high-speed requirements. Most of PACS systems use databases to keep some information of medical images (information of patients, doctors, images, and so on.) to be retrieved or viewed later, and also keep the original medical images in special folders on the server. A lot of PACS systems were prepared to work internally not in the Internet. Therefore, some of them use TLS (Transport Layer Security) certificates to make sure a user is authorized to send or retrieve medical images through the

Internet, while other use a user name and password for the same function. A user is able to send/retrieve medical images through some tools that establish a connection to the PACS. To view the medical images, the user needs to use a web browser or install any DICOM viewer that is able to retrieve files from a PACS server.

2.2 Selected Issues on Medical Informatics

2.2.1 PACS Servers

Because of the urgent need for a system to provide a convenient and efficient access to medical images from different modalities for doctors or radiologists, PACS systems were developed. This project needs a PACS system to collect medical files, so a survey for the best PACS server with characteristics obeying to the goals of this work was done. A lot of PACS servers are available with different working mechanisms: Dcm4chee, CDMEDIC, MRIdb, Orthanc, Dicoogle, OSPACS, OpenSourcePACS, ClearCanvas, Conquest DICOM, DCMTK. Some of these servers are working with relational databases, while the rest are not, but all servers have the same mechanism in collecting DICOM files in folders inside operating systems. The following part discusses advantages and disadvantages for each PACS server.

Dcm4chee is an open source tool that collects some applications for supplying the healthcare enterprise (Morate, 2012). These applications are written in Java. The portability, performance, deployment, and implementation of DICOM files are supported. Many robust services are available for a user such as storing and querying DICOM files, web access to DICOM files, make commitment, notification, and write DICOM files to CD media. Recently, some software are created based on this tool such as CDMEDIC and MRIdb servers.

MRIdb is an open source software, which consists of server and client components. These components are written in Java. The server side works in Linux, while the client works in a web browser (Woodbridge, 2013). The source and binary version are freely available from the MRIdb website (Woodbridge, 2013). MRIdb was designed based on a number of other software, including dcm4chee DICOM server, a storage system, an authentication service and a relational database system, as shown in Figure 2.2. A user-friendly interface is provided for converting, searching, and viewing imported images. This software was developed by Mark Woodbridge at Imperial College London (Woodbridge, 2013). This system stores some information about DICOM files in PostgreSQL database and allows a user to view a DICOM image on Weasis DICOM viewer or on ImageJ, send DICOM files to the server archive, or download them with automatically anonymization from the archive (Mwoodbri, 2014). Since

this server just only do automatically anonymization during the download operation, consequently, a DICOM file with original data exist at the archive and another DICOM file with anonymisation data will be available after downloading. This operation has two drawbacks: first, personal data still in DICOM files at the archive because this server do only anonymise DICOM files after downloading but this is not acceptable for health privacy especially if an attacker or a developer can view these data. Second, if a user wants to classify a DICOM file for detecting breast lesion, first the DICOM file should be downloaded, and then classify the DICOM file and send it back to the archive. Here a problem will occur, two DICOM files with different information are in the archive and both of them are related to the same patient. In this situation, researchers, or doctors cannot follow different classifying methods or compare results for the same patient.

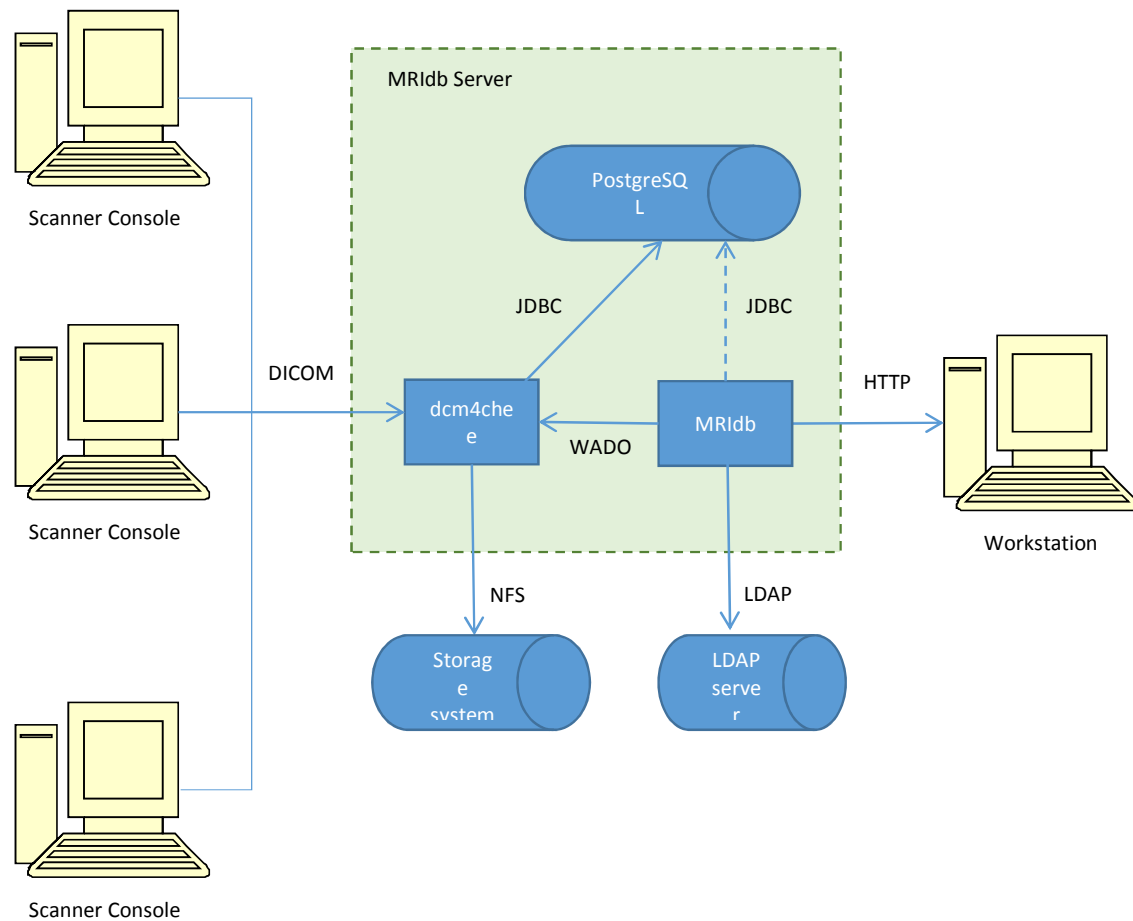


Figure 2.2: MRIdb components (Woodbridge, 2013)

DCMTK is an open source software written in ANSI C and C++. Many libraries and applications are collected to perform large parts of the DICOM standard such as examining, constructing and converting DICOM files, handling offline media, sending and receiving the

DICOM files (OFFIS, 2014). DCMTK's components are executed by the command line. Since this software was written in C and C++, it is platform-dependant. Based on this server, Orthanc and CDMEDIC are created. The first server is Orthanc. It is a free and open source DICOM server software written in C++ by Sebastien Jodogne (Jodogne, 2013). It works as standalone and runs in different platforms at least works on Windows or Linux. Since the architecture of this software is designed to be compliant with the DICOM standard, the operations of sending, storing, and retrieving of DICOM files are available. Clients can use a graphical interface called Orthanc explorer to view and interact with the content of Orthanc remote server (Jodogne, 2013). The second server is CDMEDIC, which it is created based on DCMTK, dem4chee, and MySQL (M.D., 2007). It is an open source software written in Java and works as a server-client application. Where the server side works on Debian, Mac OSX, or Ubuntu, while the web browser acts as the client side to send or retrieve DICOM files. This server does not work on windows platforms.

ClearCanvas is an open source software written in C#, JavaScript, Ruby (Synaptive Medical, 2014). A user can store, view, and manage DICOM files. It is designed to be easy to use, easy to access, highly customizable de-identification, collaborate and share with others, whether you are an administrator, a clinician, a researcher or a commercial innovator (Synaptive Medical, 2014). This server is good at first glance but the facilities that are given to users are too limited, and a user may have access to all DICOM files or not, i.e. the system does not allow some of DICOM be accessible to a special user (ClearCanvas Inc., 2009). Besides, annotation on DICOM files are not supported. The anonymization procedure is prepared to change the original information of a patient manually and also it cannot change the patient information in the pixels of DICOM images (ClearCanvas Inc, 2013).

Open Source Picture Archiving and Communication System (OSPACS) was designed to store and display DICOM files. It is an open source tool written in C#. Currently it has been used to store more than 100,000 ultrasound images by the Institute of Women's Health (University College London) (Medfloss.org, 2014). The last version of this tool was version 2 in May 23, 2007 by WillStott and only support windows operating system (Medfloss.org, 2014).

Conquest DICOM is an open source software developed by Mark Oskin. Many features are supported since it has been developed based on the public domain UCDCM DICOM code (Marcel van Herk, 2013). UCDCM allows storing, verifying, sending and retrieving by means of the programmable SQL database (Marcel van Herk, 2013). This server allows a user to

convert, view, select images. The annotation and classification on DICOM images are not supported.

OpenSourcePACS is an open source software designed to store and retrieve images. Now, it is retired, no longer supported, and not available for download. (Bui, 2007)

Dicoogle is a PACS archive supported by a document-based indexing system and by peer-to-peer (P2P) techniques (Costa, 2011). This system does not use the traditional approach of relational databases to collect information, rather than it gathers and indexes data from files over a set of distributed repositories (Costa, 2011). Since this system keeps medical imaging with their information, the operation of query and retrieval will be fast (Costa, 2011). It could also complement the traditional centralized database with more agile indexing and retrieving mechanism (Costa, 2011). In this way, there is no need to create tables, relations, fields because it collects DICOM textual data elements. FTP or email is used to access to indexed medical data on the target system as shown in Figure 2.3.

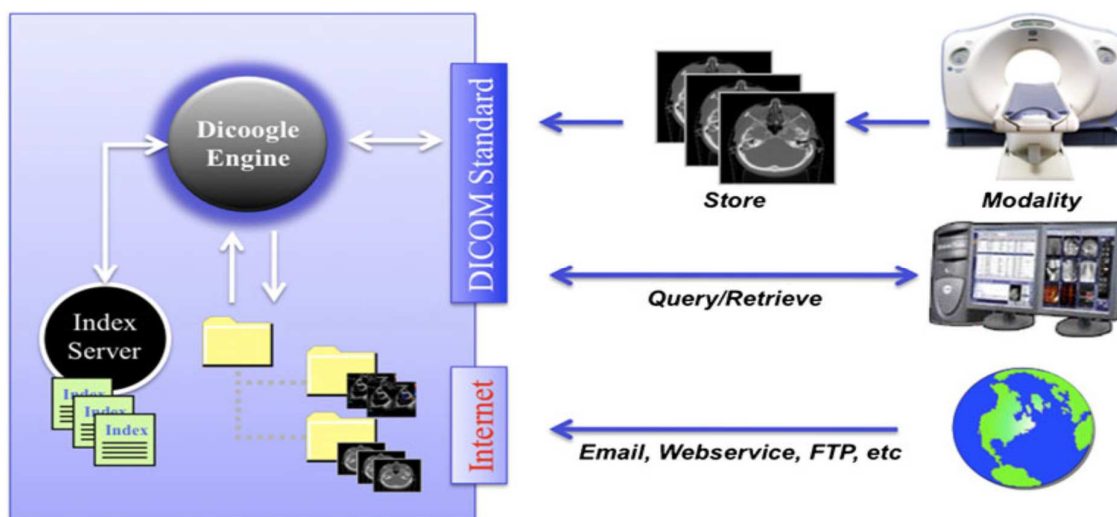


Figure 2.3: Dicoogle components and interfaces (Costa, 2011)

This system is suitable for small or medium-sized organizations and it can be used in regular clinical workflow, research, or teaching (Costa, 2011). Institutions of PASC can also be installed in parallel with this system and indexing all exist image information. This system does not care either about restricting access to DICOM files or creating different users with different privileges. Since this system uses p2p architecture, it difficult to manage DICOM files and also viruses might be transformed from a computer to another. The search in this application needs to remember patients' names in DICOM files because it asks a user to enter

the patient name manually. Also, this system is not suitable for a lot of users or a big organization.

Table 2.1 summarizes the previous considerations (2.2.1) and presents the pros and cons for each PACS server.

Table 2.1: Summary of PACS servers' pros and cons

| PACS Server | Pros | Cons |
|-----------------|--|---|
| Dcm4chee | 1- An open source 2- Written in Java 3- Support multi-platform 4- It was created to be compliant with DICOM standard. 5- Other servers are created based on it such as CDMEDIC and MRIdb. 6- It supports different image types such as US, CT, MR, SC, DX, XA, VL, RT. 7- It is supplied with command lines tools for sending or retrieving files. 8- Hanging Protocol, Storage Commitment are available. 9- DICOM viewer: Weasis, ImageJ, TUDOR DICOM, FIJI, and web browser acts as viewer to connect and retrieve files . 10- Supported database(s) engines: PostgreSQL 8.1+, MySQL 4.1+, Oracle 9i+, SQL Server 2000+, DB2 8.1+, and Firebird 2.1+. 11- Transport Layer Security (TLS) encryption is used to create a secure connection. | 1- Annotation on DICOM images is not supported. 2- Restrict access to users is limited. 3- Anonymization is not included. 4- Security through untrusted network is not enough. 5- Classify DICOM files are not available. |
| MRIdb | 1- An open source 2- Written in Java | 1- Not work in multi-platform. |

| | | |
|-----------------|--|---|
| | <ul style="list-style-type: none"> 3- It is created based on dcm4chee DICOM server, a storage system, an authentication service and a relational database system. 4- The server works on Linux, while the web browser acts as the client side. 5- It uses PostgreSQL database to store some data of DICOM files. 6- DICOM viewers can connect to this server such as Weasis, ImageJ, TUDOR DICOM, and FIJI to retrieve or send files. 7- Automatically anonymization occur during the download files. | <ul style="list-style-type: none"> 2- Annotate and classify DICOM files are not supported. 3- Restrict access to users is limited. 4- Anonymization happed only during downloading a file, which make the classification is difficult to be followed by users. 5- Security through untrusted network is not enough. 6- Classify DICOM files are not available. |
| DCMTK | <ul style="list-style-type: none"> 1- An open source. 2- Written in ANSI C and C++. 3- Many libraries and applications are collected to perform large parts of the DICOM standard. 4- Orthanic and CDMEDIC are created based on this server. 5- Support different image types such as US, CT, MR, SC, DX, XA, VL, RT. 6- Transport Layer Security (TLS) encryption is supported. 7- DICOM viewer: DICOMscope | <ul style="list-style-type: none"> 1- Platform-dependant. 2- Annotation on DICOM images is not supported. 3- Restrict access to users is limited. 4- Anonymization is not included. 5- Security through untrusted network is not enough. 6- Classify DICOM files are not available. 7- No Hanging Protocol support. |
| Orthanic | <p>Has the same pros of DCMTK but a client uses a graphical interface called</p> | <p>Has the same cons of DCMTK server.</p> |

| | | |
|--------------------|--|--|
| | Orthanc explorer to retrieve or send files. | |
| CDMEDIC | It is an open source software written in Java. It is created based on DCMTK, dem4chee, and MySQL. The server side works on Debian, Mac OSX, or Ubuntu, while the web browser acts as the client side. | This server does not work on windows platforms. Also it has the same cons of DCMTK. |
| ClearCanvas | <ol style="list-style-type: none"> 1- An open source software written in C#, JavaScript, Ruby. 2- It is designed to be easy to use, easy to access, highly customizable de-identification, and it is able to collaborate and share with others. 3- It uses MS SQL Express 2008. | <ol style="list-style-type: none"> 1- The facilities that are given to users are too limited, 2- The annotation on DICOM files are not supported. 3- The anonymization procedure is prepared to change the original information of a patient manually. 4- The anonymization procedure cannot change the patient information in the pixels of DICOM images. 5- Classification of DICOM files is not available. |
| OSPACS | <ol style="list-style-type: none"> 1- An open source tool written in C# using Visual Studio 2005 tools (Stott, 2008). 2- It has been used to store more than 100,000 ultrasound images by the Institute of Women's Health. 3- Microsoft SQL Server database is used. | <ol style="list-style-type: none"> 1- The last version of this tool was version 2 in May 23, 2007. 2- It only supports windows operating system. 3- The annotation uses white pixels |

| | | |
|-----------------------|---|--|
| | | <p>4- The classification of a DICOM file is not supported.</p> <p>5- It removes all patient-identifying information from the DICOM header data</p> <p>6- Documentation is not enough.</p> |
| Conquest DICOM | <p>1- Written in C/C++.</p> <p>2- It has been developed based on the public domain UCDCM DICOM code (Marcel van Herk, 2013). UCDCM allows storing, verifying, sending and retrieving by means of the programmable SQL database.</p> <p>3- This server allows a user to convert, view, select images.</p> | <p>1- Only Windows, UNIX are supported.</p> <p>2- Only support US, CT, MR, SC, DX, XA images.</p> <p>3- The annotation and classification on DICOM images are not supported.</p> <p>4- No Hanging Protocol support.</p> |
| OpenSourcePACS | <p>It an open source software designed to store and retrieve images.</p> | <p>It is retired, no longer supported, and not available for download.</p> |
| Dicoogle | <p>1- Supported by a document-based indexing system and by peer-to-peer (P2P) techniques.</p> <p>2- It gathers and indexes data from files over a set of distributed repositories.</p> <p>3- The query operations will be faster through free text or metadata.</p> <p>4- FTP or email is used to access to indexed medical data on the target system.</p> <p>5- This system is suitable for small or medium-sized organizations and it</p> | <p>1- This system does not care either about restricting access to DICOM files or creating different users with different privileges.</p> <p>2- Since this system uses p2p architecture, it is difficult to manage DICOM files and also viruses might be transformed from a computer to another.</p> |

| | | |
|--|--|---|
| | can be used in regular clinical workflow, research, or teaching. | <p>3- The search in this application needs to remember patients' names in DICOM files because it asks a user to enter the patient name manually.</p> <p>4- Also, this system is not suitable for a lot of users or a big organization.</p> <p>5- Annotation and classification on DICOM files are not supported</p> |
|--|--|---|

2.2.2 DICOM file format

2.2.2.1 Introduction

Recently, most of the modalities in a hospital generate medical images, handled using the DICOM (Digital Imaging and Communications in Medicine) files standard. DICOM is the international standard for digital image format, and biomedical images' file structure and related information (Bidgood, Horii, Prior, & Van Syckle, 1997). For clinical use, it uses a special file format and a network communications protocol with necessary data and quality. It is implemented in many devices such as X-ray, CT, MRI, ultrasound, and increasingly used in ophthalmology and dentistry devices. Therefore, tens of thousands of imaging devices implement this standard, so it is one of the important standards for healthcare in the world (NEMA, 2015). This standard allows many devices (scanners, servers, workstations, printers, and network hardware.) to integrate and communicate from different manufacturers using a picture archiving and communication system (PACS). Information of patients, procedures, doctors, and images will be supported for shared management. The standard addresses some areas such as network image, network image interpretation, network print, imaging procedure, and off-line storage media management. The essence of DICOM is a file format and a networking protocol. A DICOM viewer will be necessary to display DICOM images, to diagnose and classify them, and maybe to read or modify some elements of the DICOM header. The medical imaging applications in the hospital network use the DICOM protocol to exchange

DICOM images, and retrieve some information of DICOM file from the PACS to a workstation in order to display or modify them.

2.2.2.2 DICOM File Components

DICOM file contains a file header portion, File Meta Information portion, and a single SOP instance. This header consists of a 128 byte preamble available for a special using such as enabling a multi-media application to randomly access images stored in a DICOM data set; followed by a 4 byte DICOM prefix, followed by the File Meta Elements (data set) (Evans, A Very Basic DICOM Introduction - dcm4che-2.x - Confluence, 2008), (NEMA, 2013). The preamble should be initialized with zeros if it is not used. This header shall be present in every DICOM file. Figure 2.4 shows a DICOM file (image IOD (Information Object Definition)), which consists of a list of attributes. Some DICOM files contain image and other not. DICOM files contain more than just images. Every DICOM file holds patient information (name, ID, sex and birth date), important acquisition data (e.g., type of equipment used and its settings), and general study (date, time, referring physician, accession number). A PACS system uses some of DICOM file header elements to create indexing data to be used for queries by DICOM viewers.

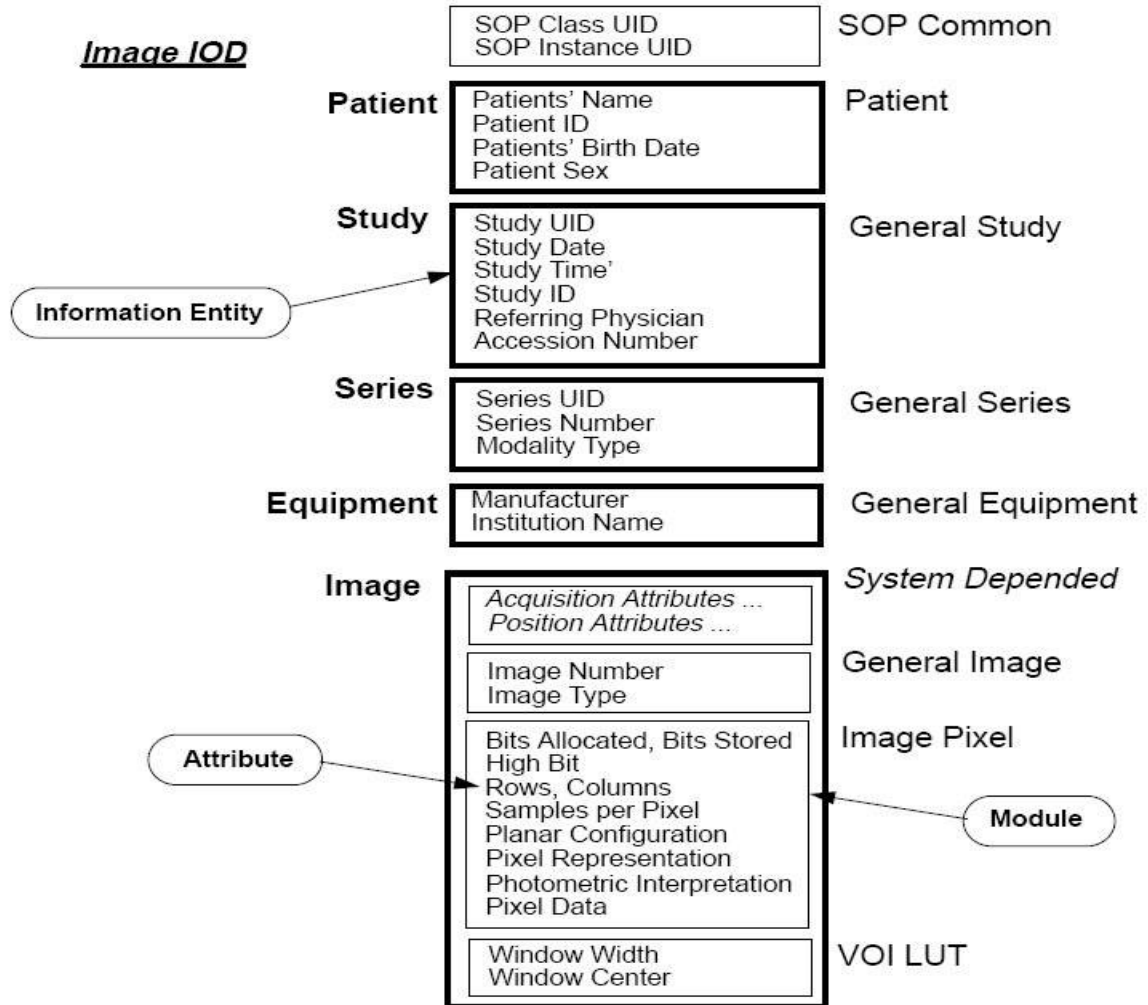


Figure 2.4: DICOM model (Evans, 2008)

2.2.2.3 DICOM Elements or Attributes

In reality, a DICOM object is comprised of DICOM elements or DICOM attributes. Each DICOM element has a tag, VR (value representation), length and value as shown in Figure 2.5 (NEMA, 2014). Each tag uniquely defines the element and its properties and has a special form (gggg,eeee) where gggg is a group value and eeee is an element value, and both of them start from 0000 to 9999. For example, (0010,0010) is the element that specified for patient's name. The second component of DICOM element is VR (value representation), which defines the data type of element. There are different data types such as UI which is abbreviation of unique identifier data type, US for unsigned short data type, CS for Coded String and OB for other byte (NEMA, 2014). All elements have length and this length should be even, but there are some elements that have a single value such as patient's sex, it might be a 'M' value for male or 'F' value for female or 'O' value for other. In this situation, the element length should be 2 (even number) and this value will be padded by a space (ASCII 0x20) value. This padding is for string types like CS or UI, but if the data type is binary like US, the padding should be null (0x0) value.

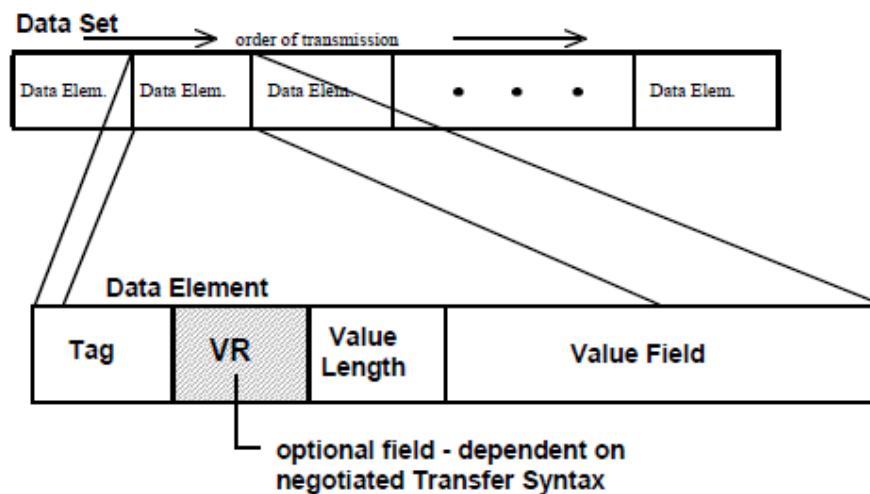


Figure 2.5: DICOM element (NEMA, 2014)

2.2.2.4 Important DICOM Elements

There are thousands of DICOM elements and just some of them will be appeared in a DICOM file according to the necessary information. Therefore, the elements of DICOM files from one DICOM file to another might be different. In this work, we will focus in the important elements according to the requirements of this project. The DICOM elements have many groups but we will use some of the important groups: patient, study, series, equipment, and image

groups as depicted in figure 2.4. Each of these groups have UID (unique identifiers) elements and also can help in the query/retrieve operations. For example, patient’s ID uses to identify the patient group, study instance UID for study group, series instance UID for series, and sop(service object pair) instance UID for an image or for a DICOM file (NEMA, 2014). Transfer syntax is an important element inside DICOM file, which uses to describe the format of a DICOM file and the network transfer methods. Many transfer syntaxes are available such as Implicit VR Little-endian, Explicit VR Little-endian, Explicit VR Big-endian, or JPEG Lossless (Medical Connections Ltd, 2014) as shown in Table 2.2. Transfer syntax is required during creating a new DICOM file or convert a file (jpeg, png, pdf, etc.) into a DICOM file.

Table 2.2: Different transfer syntaxes (NEMA, 2014):

| Number | Transfer syntax | Value |
|--------|---------------------------|---------------------------|
| 1 | Implicit VR Little-endian | 1.2.840.10008.1.2 |
| 2 | Explicit VR Little-endian | 1.2.840.10008.1.2.1 |
| 3 | Explicit VR Big-endian | 1.2.840.10008.1.2.2 |
| 4 | JPEG Lossless | 1.2.840.10008.1.2.4.57 |
| 5 | JPEG Lossless First Order | 1.2.840.10008.1.2.4.70 |
| 6 | RLE Lossless | 1.2.840.10008.1.2.5 |
| 7 | JPEG 2000 (Lossless) | 1.2.840.10008.1.2.4.90 |
| 8 | JPEG-LS (Lossless) | 1.2.840.10008.1.2.4.80 |
| 9 | JPEG Baseline | 1.2.840.10008.1.2.4.50 |
| 10 | JPEG Extended | 1.2.840.10008.1.2.4.51 |
| 11 | JPEG 2000 (lossy) | 1.2.840.10008.1.2.4.91 |
| 12 | JPEG-LS (Lossy) | 1.2.840.10008.1.2.4.81 |
| 13 | MPEG-2 | 1.2.840.10008.1.2.4.100 & |
| 14 | MPEG-4 | 1.2.840.10008.1.2.4.102 & |
| 15 | Deflate | 1.2.840.10008.1.2.1.99 |
| 16 | JPIP | 1.2.840.10008.1.2.4.94 |
| 17 | JPIP-Deflate | 1.2.840.10008.1.2.4.95 |

As we mentioned above, there are some DICOM files that contain images and others are not, so if a DICOM file contains an image, the image elements should be included in the DICOM header such as sample per pixel, photometric interpretation, planar configuration, rows, columns, bits allocated. These elements describe the pixels of the image of DICOM file. The image pixels store in the pixel data element (7FE0,0010). The size of the image can be defined by the height element (0028,0010) and the width element (0028,0011). Samples per pixel element (0028,0002) defines the number of colour channels. Two values represent the colour in an image, 1 for the single channel (black/white) and 3 for the three colour channels (red, green and blue). The photometric interpretation (0028,0004) element defines what every

colour channel holds. It can refer to the colour space that used to encode an image and show how the image should be displayed (Medical Connections Ltd, 2014). This element has different values: MONOCHROME2, MONOCHROME1, PALETTE COLOUR, RGB, YBR_FULL, YBR_FULL_422, YBR_PARTIAL_422, YBR_RCT, or YBR_ICT (Medical Connections Ltd, 2014). RGB is the most colour format and should start with red, green and blue respectively. MONOCHROME2 or MONOCHROME1 are usually used in grayscale images (like CT, MR, or fluoroscopic images). "YBR_FULL" and "YBR_FULL_422" are commonly used in JPEG. Planar configuration element (0028,0006) uses with colourful images to show how the colour channels are arranged in the pixel data. It has two values: either '0' means the channels are interlaced where the order of colour is one by one, first red, then green, and then blue to get the colour of one pixel as depicted in Figure 2.6; '1' means the channels are separated, where first all the reds, then all the greens and then all the blues as shown in Figure 2.7. The separated way is rather rare and it usually used with RLE(Run-Length encoding) compression.

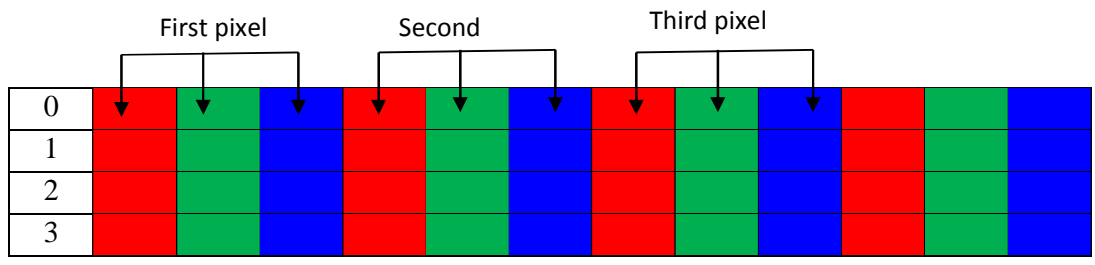


Figure 2.6: Interlaced way

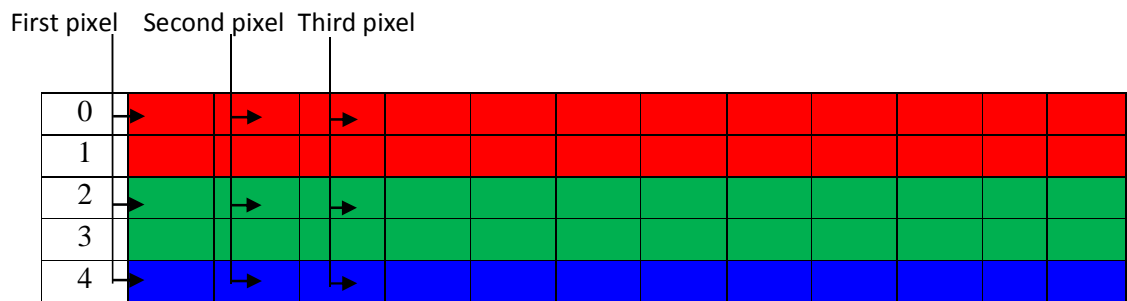


Figure 2.7: Separated way

The pixel structure defined by Bits Allocated (0028,0100), Bits Stored (0028,0101), and high bit (0028,0102) elements. Bits Allocated (0028,0100) element is responsible for determining the allocated space or the size of a pixel cell for every sample in bits. For example, every channel is encoded in 8 bits in 24 bit RGB image. Bits Stored (0028,0101) determines a number of allocated bits that are actually used. While high bit (0028,0102) specifies how stored bits are aligned inside the allocated bits (NEMA, 2014).

Pixel Representation (0028,0103) element has two values, either '0' which refers to unsigned values or '1' which refers to signed values. The default is unsigned values. All attributes of group 0028 will be encoded in signed short (SS) values or unsigned short (US) values according to the value of this element. Number of frames (0028,0008) element determines the number of frames in an image. Usually if there's only one frame, this element will be omitted, but in case of multi-frames for a DICOM file, the value of this element should be appeared and set.

2.2.3 PACS viewers

A user wants to view medical images from a local or remote system to classify, analysis, or whatever she/he wants, therefore, a DICOM viewer is necessary now. Some viewers work as a standalone application, while the rest work based on a web browser as mentioned above in PACS server 2.1.1 section. There are some DICOM servers that use a web browser as a client interface for a user to connect and display DICOM files such as Cleanconvan, Dcm4chee, MRIdb, Conquest DICOM, Dicoogle, Orthanc or CDMEDIC. While other servers use a command line tool, or a standalone application to connect and send, retrieve, or view DICOM files such as DCMTK, OSPACS, or dcm4chee. Recently, there are plenty of free tools that work as a standalone viewer for DICOM files.

TUDOR DICOM is an open source Java application, contains interfaces to perform some DICOM operations such as anonymous one or a group of DICOM files' headers, send/retrieve DICOM files from/to a PACS server, read/write DICOM files from/to a disc (Hermena, 2008), (Hermen, 2009). Also other features of showing DICOM images are supported such as windowing, zooming, shifting, and measuring. It has two versions: one works as standalone and another works as a plugin in ImageJ. It was created based on dcm4che version 2, Java Advanced Imaging (JAI) API, and ImageJ.

ImageJ is an open source package written in Java for image processing field (Rasband 1997). At least, it needs Java 1.4 virtual machine or later to run on different operating systems such as Windows, Mac OS, Mac OS X and Linux. A friendly-interface is supported with menus and tool bars to view, modify, analyse, and process images. Different commands are available to support researcher or student from different areas like segmentation, contrast manipulation, sharpening, smoothing, edge detection, median filtering, etc. Many image formats can be supported like TIFF, GIF, JPEG, BMP, DICOM, FITS. It is useful for graphic designers because it helps in calculating the area and pixel values, measuring distances angles, creating

density histograms, or lining profile plots (Rasband, 2012). ImageJ is able to read and annotate a DICOM file but it keeps the modification with different formats such as JPEG, TIFF format.

FIJI is an open source software written in Java to support image processing operations. It is created based on ImageJ (MediaWIKI, 2013). Many image processing procedures can be used specially for image registration, image segmentation, 3D reconstruction, and 3D visualization. A developer can read, add, modify, and rebuild all the internal libraries and plugins. Not only are the previous plugins of ImageJ available in FIJI, but also new plugins too. Recently, different fields are using FIJI like biology, genetics, image processing, life science, and material science. A friendly interface is supplied with menu bar and tools for processing and analysing data. Fiji works on many platforms such as Windows, Linux, MacOSX, Intel 32-bit or 64-bit.

Matlab is a high-level language and interactive environment for developing algorithms, analysing and viewing data, and performing numerical computation (TutorialsPoint, 2015). It is supported with a group of tools for building applications. External applications that are written in Java, C++, or .net can be integrated with Matlab by some Matlab functions (TutorialsPoint, 2015). Many areas use Matlab such as mathematics, chemistry, physics, and image processing. DICOM files are supported by a special tool in Matlab called “Image Processing Toolbox”, which is responsible to read and write data to/from a DICOM file. Unfortunately, “Image Processing Toolbox” does not support dealing with DICOM files through a network (MathWorks, Inc., 2015).

2.2.4 Dcm4chee

2.2.4.1 Introduction

In recent decades, development occurred in the area of digital imaging and rapid adoption of picture archiving and communication systems (PACS), so the exchange of information within the radiology departments has become very crucial and influential. And, since the DICOM has become the standard for medical imaging, it is necessary to have tools that manipulate and store DICOM files. According to these requirements, JDICOM was created to deal with DICOM files, is a server written in Java. After that, DCM4CHEE was appeared, which is an archive system for DICOM files with the possibility of modifying, collecting, and retrieving of DICOM objects. The following part explains the advantages of the selected PACS server (dcm4chee) with its components.

2.2.4.2 Dcm4chee's architecture

The architecture of dcm4chee was designed as a modular for supporting different services such as a web-based user interface, DICOM interface, HL7(healthcare messaging protocol), web access to DICOM object (WADO), Audit Record Repository(ARR), media creation, and XDS/XDS-I(Image document sharing protocol) (Evans, Home - dcm4chee-2.x - Confluence, 2012) as shown in Figure 2.8. Dcm4chee uses the web interface of JMX console to enable or disable each of these services. There are many features of dcm4chee, the following are some of them:

1. C-ECHO SCP/SCU (ping)
2. C-STORE SCP/SCU (storage)
3. C-FIND SCP (query)
4. C-MOVE SCP (retrieve)
5. N-ACTION SCP/SCU (storage commitment)
6. Create new users
7. Grant or revoke the users some authorities to access to the DICOM files
8. Export DICOM files

Where C-ECHO is a command to check if the DICOM server is listening in a special port or not, C-STORE is a command to store DICOM files in dcm4chee, C-FIND is a command to find images or data in dcm4chee, C-MOVE is a command from client (SCU) to dcm4chee to retrieve a DICOM file, and finally N-ACTION is a command to check if a DICOM file is stored successfully or not.

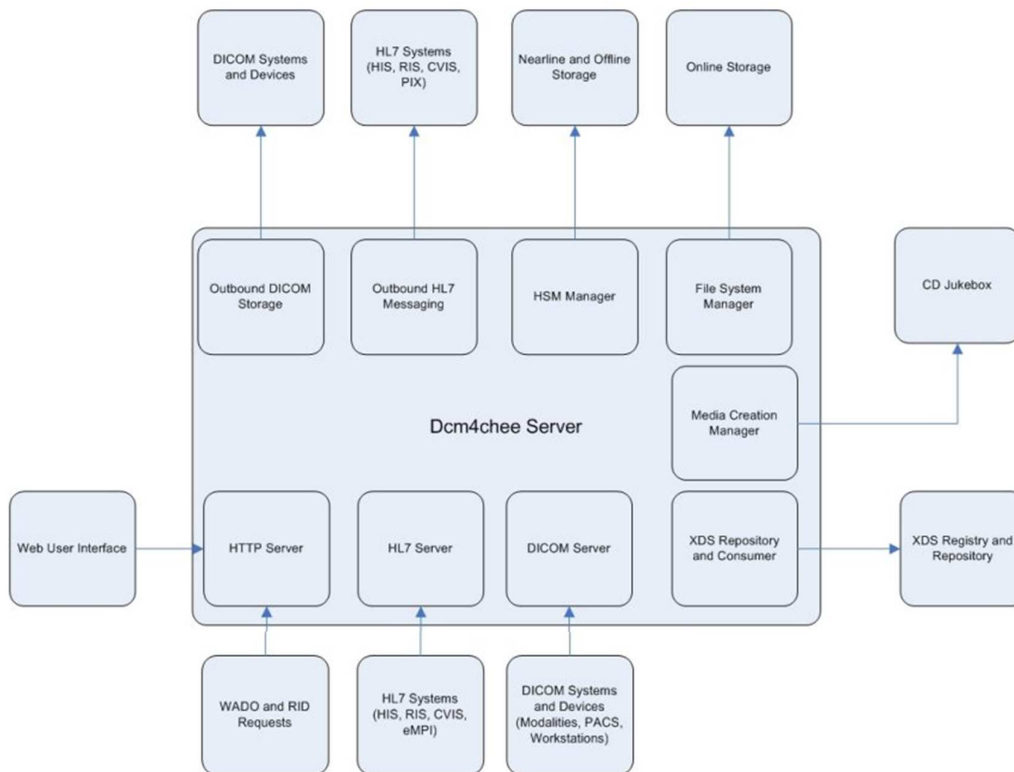


Figure 2.8: Dcm4chee architecture (Evans, 2008)

2.2.4.3 Installation

Preferably, install this software on a computer with High specifications. This system needs at least 512 MB RAM, 100 MB hard disk space, and 400 MHz CPU or better. A database to store some DICOM information for a web-based interface is compulsory. Many databases engines are supported such as PostgreSQL 8.1.x, MySQL 4.1+, Oracle 9i/10g, SQL Server, and DB2 8.1+. In this work, MySQL 5.5 and dcm4chee 2.0.18 are implemented. For more information about the steps of installation with the necessary packages, please review the appendix.

2.2.4.4 The Technology

Dcm4chee is supported with a set of command line tools that allow sending, retrieving manipulating DICOM objects. Dcm4che implements apache maven project to perform the basic functions of managements (monitoring dependencies, package building, etc.). Maven is a management and comprehensive tool based on the concept of a project object model (POM) (Apache Software Foundation, 2002-2014). It is responsible for managing and building any

java-based projects, reports, and documentations. Additionally, Maven uses different projects to supports users with easily management processes such as Ant, Archetype, Doxia, JXR, Plugin Tools, SCM, and Wagon (Apache Software Foundation, 2002-2014). For modifying and building the source code of dcm4chee, maven is necessary. The binary version of dcm4chee is used in this work, not the source code.

2.2.4.5 Dcm4chee's Tools

Since dcm4chee is an archive system with collection of open source applications and tools for the healthcare enterprise, the integration of the work between these tools is very important. Figure 2.9 shows the workflow of dcm4chee with different tools to support sending, retrieving, and manipulating DICOM files.

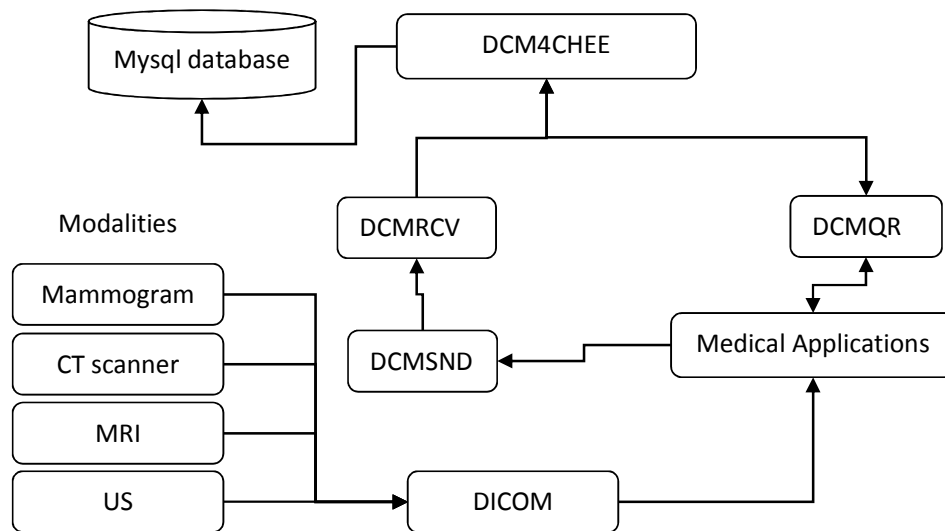


Figure 2.9: Dcm4chee workflow

2.2.4.5.1 DCMRCV utility

Is one of the supporting command line tools, it is responsible for listening on a TCP port and determine the name of application entity title (AET) to listen for incoming DICOM files from the senders, and to store them in a folder (Evans, Dcmrcv - dcm4che-2.x - Confluence, 2006). This tool is an open source written in Java ®.

2.2.4.5.2 DCMSND

An open source command line tool that used to send one or more DCIOM files to the listening port of dcmrcv utility. It is possible to send a group of DICOM files in a folder one time to dcm4chee by using this tool (Evans, Dcmsnd - dcm4che-2.x - Confluence, 2008).

2.2.4.5.3 DCMQR

Is used to create a query to ask dcm4chee for retrieving some information of DICOM header or retrieving a DICOM file (Chetan Uberoy, 2010). Two commands are available to retrieve DICOM image: c-get and c-move. C-move allows to send images from one AET to other AETs, while c-get cannot do that. C-move SCP (Service Class Provider) or the server needs to establish a TCP connection to the end point of SCU (Service Class User) or the client, but this is impossible in some cases such as if SCU uses dynamic IP (DHCP), or if there is intervening firewalls. C-get is able to retrieve data from dcm4chee to special folder at SCU. Therefore, c-get is used to retrieve data from dcm4chee.

2.2.4.5.4 Weasis

Weasis is a standalone application responsible for viewing DICOM files from dcm4chee data repository. It is an open source package written in Java ®. A user can search the archive without the need of in-depth DICOM knowledge. A number of different operations are available to users, including showing some information as patient's name, age, study date, series date, and import and export DICOM files.

2.2.5 MySQL server overview

A free, open source tool and compliant with the standard form of SQL data language (Tutorialspoint, 2015). It supports a user with database management system with ability to connect and query any desired data. Each database has one or more tables and each table has one or more rows and columns to store data in. It can handle and deal with large databases reliably and quickly. The ability to work in different operating systems is available. The connectivity, speed, security of MySQL server makes it suitable for accessing databases and make it the most popular database to be used by a lot of applications (Tutorialspoint, 2015) , (Oracle Corporation, 2015).

2.2.6 Overview of gluing technologies

2.2.6.1 JAVA

Java is a high level, object-oriented, platform independent language, developed by Sun Microsystems in 1995 (Tutorialspoint, 2013). It was originally designed for interactive television, but it was too advanced for the digital cable television industry at that time. It is robust and secure, architecture-neutral and portable, interpreted, threaded, and dynamic programming language (Tutorialspoint, 2013). The flexibility of Java is high; a lot of applications and applets are created based on Java. Unlike some other languages such as C and C++, a created program can be run on various operating systems without having to rewrite or recompile the code because of the Java run-time environment, which interprets the Java code and tells the operating system what to do (Tutorialspoint, 2013). The syntax of Java is similar to the C/C++ syntax, so programmers of C/C++ can learn it fast. Since Java is one of the most popular programming languages in use, different applications are created such as client-server web, game, mobile, and TV applications (Tutorialspoint, 2013). Basically, Java virtual machine (JVM) is necessary to run the byte codes (class files) of the compiled programs for running on any computer regardless of the computer architecture.

2.2.6.2 Python

Python is one of the high-level programming languages, used for general purposes and used to deal with a wide range of problems (Python Software Foundation, 2013). An object oriented programming, multiple programming paradigms, and imperative and functional programming are supported. Because of its high-level data structures, interpreted nature, dynamic typing, and elegant syntax, it can be used in rapid application development, large applications, and web applications. Program maintenance cost is reduced because its simplicity, readable and learn its syntax easily (Python Software Foundation, 2013). Programmers can write smaller code than some other languages like C. The execution time for running is slower than to some other languages such as C/C++, because it needs an interpreter to interpret a code, while C and C++ need a compiler. Therefore, it can be easily extended with C/C++ to make the run time of the applications almost at the same speed of C/C++ applications. Programmers usually have a steep learning curve because of the increased productivity that provides. The extensive standard libraries and interpreter of Python are free and available in binary or source form for the most important platforms and can be freely distributed. The cycle of edit-test-debug is quick because of there is no compilation step in this programming language. An

exception will be raised when the interpreter catches an error, so the python's debugging programs will never cause a segmentation fault if there is a bug or bad input. The debugger allows evaluation of arbitrary expressions, inspection of the local and global variables, breakpoints setting, and line stepping through a code.

Pydicom is a python package, which allows to modify or read DICOM files. Python uses this package for manipulating data elements inside DICOM files. The main disadvantage of the current version is, the compressed pixel data cannot easily altered as in uncompressed pixel data (Mason, 2012).

Enthought Canopy is an appropriate python environment for scientific and analytic computing (Enthought, Inc., 2013). The Canopy provides tools and interfaces for data management and analysis, scripting, and testing. Its text editor provides auto completion and error checking for the code. It has IPython window that lets us easily test the code and see the results after running the code. The IPython console allows to execute the entire script or selected lines. Plenty of python packages are available to support the technical and scientific computing, such as Traits, TraitsUI, Pyface, BlockCanvas, GraphCanvas, and SciMath.

2.2.6.3 OpenCV

OpenCV is a library that collects a lot of algorithms to be used by industry and academia for computer vision applications and researches (Itseez, 2013). It is an open source written in C and C++. Many interfaces such as C++, C, Python, Matlab, Java, and other languages are supported and it can be run on different operating systems such as Windows, Linux, Mac OS, iOS and Android operating systems. It was designed for computational efficiency, real-time image processing, and computer vision. The library takes the advantage of multi-core processing because it is written in optimized C/C++. A lot of algorithms are supported in many areas in vision, factory product inspection, medical imaging, security, user interface, camera calibration, stereo vision, and robotics (Bradski, 2008).

3 System Analysis and Requirements

3.1 Introduction

Actually, the selected tools at Chapter 2 are not enough to accomplish all requirements of this thesis, so this part will analyse the missing requirements to be created and integrated with the selected tools: dcm4chee, FIJI, and MySQL server. In order to facilitate the understanding of the working mechanism of this system, Figure 3.1 shows the general requirements of the system, which consists from different levels. Three levels are proposed in the system: input, processing, and output. The input level is responsible for collecting reports and DICOM files to the system archive. These reports can be images or pdf files. All DICOM files and reports should be anonymized before sending them to the archive. In the processing level, some DICOM files can be retrieved from the archive to apply pre-processing or segmentation methods by a researcher, while others can be retrieved to annotate or classify them manually by a doctor, teacher, or student. Finally, in the output level, different classification results for different users (doctors, researchers) can be shared. Besides, the classifications will be graded by a teacher and viewed by students. This means that, many users such as doctors at hospital, students or teachers at medicine faculty, or researchers can interact with this system. Many users can classify a DICOM file, so many derived DICOM files will be generated for each original DICOM file.

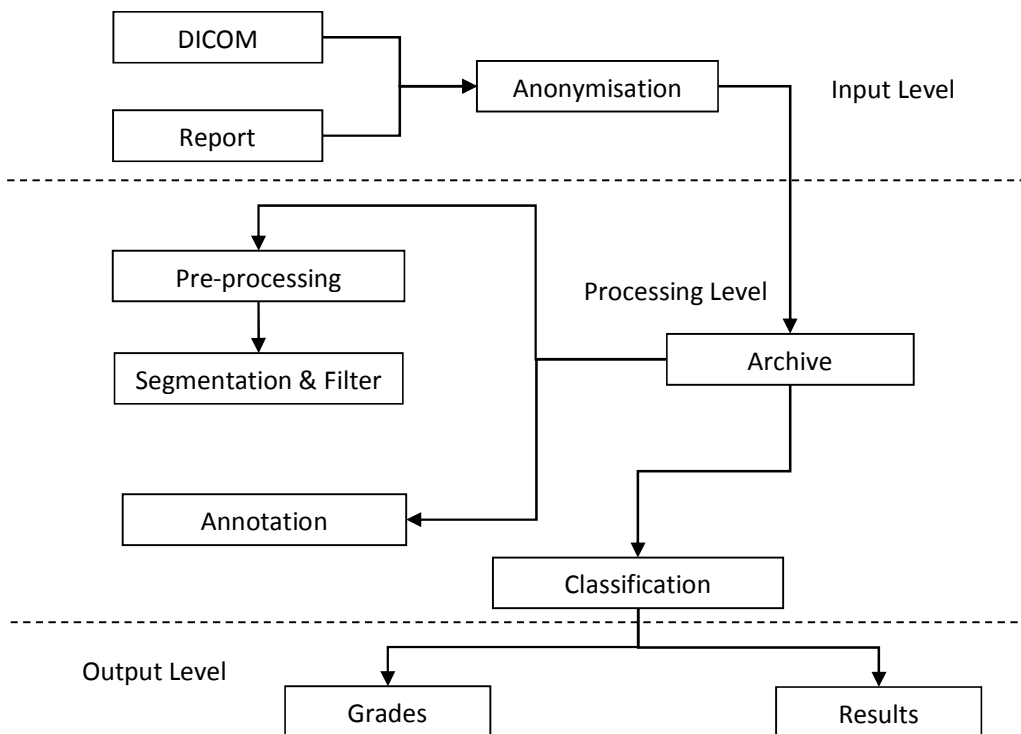


Figure 3.1: System requirements

The system analysis process in this work is divided into the glossary of terms, functional requirements, and database diagram to show all stakeholders of the system and expected inputs and outputs.

3.2 Glossary of Terms

- **Admin user:** is a root user that is responsible for creating users' accounts with different privileges. Different types of users privileges will be supported such as doctors, students, teachers, or researchers privileges. This user is able to grant other users a limited access to DICOM files and reports of the system. The root user is also able to change a user privilege from type into another one, *e.g.* change the privileges of a user from a teacher's privileges into a doctor's privileges or vice versa .
- **Main user:** is the realization of a real main doctor in the domain of this system. This user is responsible for exporting a group of DICOM files from modalities or devices to a special folder. And then, import a group of DICOM files or reports from the folder to be anonymized and sent to the archive. the main user is also able to share all DICOM files and reports in the archive to be used later. The ability to put annotations (drawing texts or shapes) on DICOM files should be available. Assign a group of DICOM files or reports to a group of users should be supported. All classifications of other users can be seen by this user, since this user is the main one and has the full control permission of DICOM files, reports, or groups in the entire system.
- **Student:** an actor is able to get a group of DICOM files as tasks, start classifying them, and send the classification results back to the archive. This user can show classifications' grades that belong to her/him. A student can get feedback by a teacher about the mistakes in classifications.
- **Teacher:** an actor that supervises a group of students by teach them different status of breast lesions, assign tasks, and grades them. This user must get a group of DICOM files and a group of students by a main user. A teacher is able to create different groups from the assigned groups by the main user. A group of files can be assigned to a group of students as tasks.
- **Doctor:** this actor can do exactly what a student do but without grading, and classifications of this user will keep for long time, while student classifications will keep for a limited time. In addition, this user is able to classify a group of DICOM files that belong to a patient in

specific study date as a global classification, while a teacher and a student cannot do that. A doctor at a hospital represents this user. It is possible to classify one or more DICOM files at once.

- **Researcher:** this user can get a group of DICOM files with ability to apply one or more methods to help in detecting breast lesions. The Modified DICOM files will be sent back to the archive. In addition, she/he is able to run a batch process to a group of DICOM files: it is a process that allows executing a group of methods (plugins) to a group of files such as FIJI batch process (MediaWiki, 2015).
- **DICOM file:** is a file that contains different information that are related to a doctor who made diagnosis, patient, study, modality, and image of a DICOM file. It is the backbone of the system. This file has two parts: the header and the pixel of image.
- **BI-RAD:** a standard that defines classes with their meaning and are used by radiologists to summarise the diagnosis. It starts from 0 to 6. The value 0 means a mammogram study is still incomplete (The Johns Hopkins University, 2014). This score needs to be reviewed later, therefore, further information is needed to make a final assessment and assign the true BI-RADS score (The Johns Hopkins University, 2014). The score 1 means no evident signs of cancer were found (The Johns Hopkins University, 2014). The score 2 means there is no apparent cancer in the current study. The score 3 means a study is probably normal but there is an approximately two percent chance of cancer (probably benign) (The Johns Hopkins University, 2014). The score 4 means that a patient mammogram has an approximately 20 percent to 35 percent chance of the breast cancer (The Johns Hopkins University, 2014). The score 5 means a mammogram results are highly suspicious with a 95 percent chance of breast cancer (The Johns Hopkins University, 2014). Finally, the score 6 means that a diagnosis of a patient mammogram has confirmed with a breast cancer (The Johns Hopkins University, 2014).
- **Patient:** for each DICOM file, information of the patient should be included. A patient can have one or more DICOM files in different dates with some information such as name, ID, or birthdate inside the header of each DICOM file. In some cases, the patient's information might be appeared in the pixels of DICOM images.
- **Study:** A study includes all the information gathered about a patient at specific date. One study might have a group of DICOM files. Classification can be applied to studies, and a study might be classified by one or more doctors. The information of study is also included in the header of a DICOM file.

- **Report:** is a file that can be an image (jpeg) or a pdf file, which consists of some data that are related to a patient and a doctor with the diagnosis results. For one or more DICOM files, one or more reports can be created.
- **DICOM element:** an attribute that determine an element in a DICOM file header, such as study date, patient's name, etc. Each element has tag, value, and type as mentioned at the chapter 2.
- **Registration:** each user should register in the system to have access to some functionalities of the system. A user name, password, and IP are the necessary inputs to be used to register a user in this system.
- **Lesion:** is an area of abnormal tissue change. Users can attribute characteristics of a lesion. Such characteristics may include, but are not restricted to, contour, inner structure or diagnose.
- **Group:** a set of objects or entities that are manipulated as items. A group may be a set of users, of DICOM files, or reports.
- **Tag:** is an attribute that represents a relevant characteristic of part of an image, of the total image, or of a collection of images. Each tag consists of a tag name and a range of values. For example, a tag may be attributed to a lesion and lesion type or BI-RADS and respective score values (0-6), or any value that describes these parts or the collection of images. This tag differs from the element tag inside DICOM header, where the element tag is used to uniquely identifying each element inside the DICOM header.
- **Classification:** a process of assigning tags to objects (DICOM files) to help finding early solutions for breast lesions. Different tags can be added to describe a classification. The classification can be applied with/without annotation: is the process to draw a shape(or a set of shapes) or texts on a DICOM image. Each classification contains one or more tags.
- **Anonymisation:** is a process that changes original data with other data for the health privacy rights. Each file that has personal information in should be anonymized.
- **Privilege:** is the process that gives a user ability or authority to access the system and its files. Different privileges are supported such as those of the student, teacher, doctor, main user. Only one of them will be assigned to a user.
- **Grade:** a teacher is able to grade the work done by students according to theirs classifications. Students can see the grades, but each student is only able to view her/his grades.

- **Method:** A developer or a researcher can apply one or more methods to one or more DICOM files to detect breast lesions. A method can be internal, which appears as plugin inside the whole system; or it can be external, which executes outside of the system.

3.3 Functional requirements

In this part, the functional requirements will be presented as use case diagrams with a group of tables to show what the system should do. Eclipse software is used to create the use case diagrams.

3.3.1 Use case diagram

A use case defines a goal-oriented set of interactions between external actors and the system under consideration. Actors are parties from outside the system that interact with the system. In the following part some diagrams show the interaction between actors and the system.

A registration in this system is managed by a root or an admin user who has full access to the archive and can create all privileges of other users. First, to register a new user, this user will send the admin user an email or file with his name and email. After that, the admin user will create an account for the user with random password and determine the user privileges. The created user can change his password later as shown in Figure 3.2.

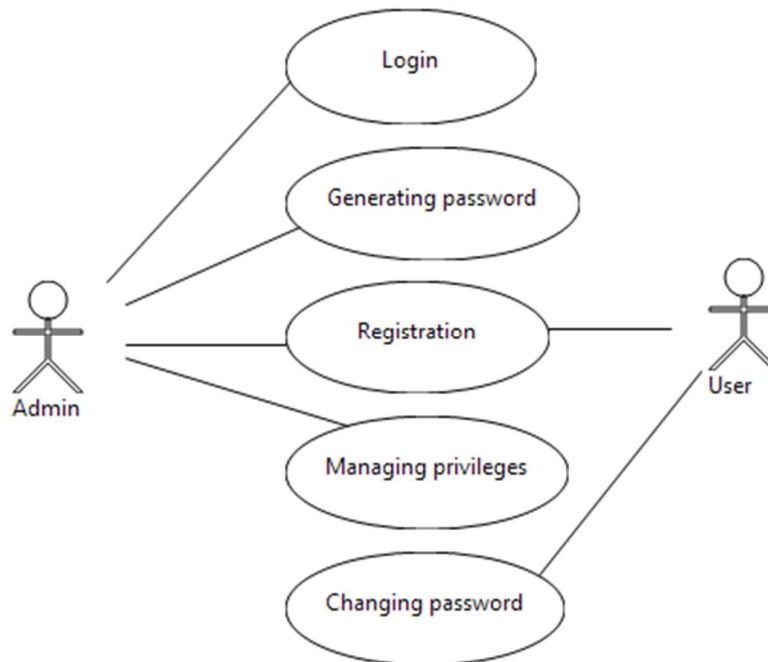


Figure 3.2: The system registration

After registration, a user has different tasks according to his privileges. An anonymisation procedure can be applied only by a main user for all DICOM files and reports before sending them to the archive. Figure 3.3 shows some tasks of the main user the system.

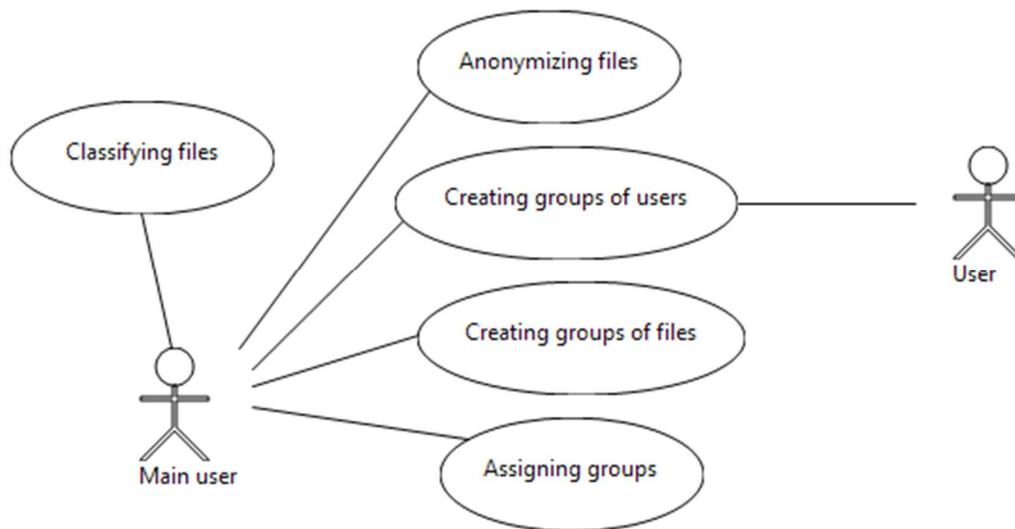


Figure 3.3: A main user tasks

In general, to send or retrieve files, a user should login to the system with a user name, password, and IP as shown in Figure 3.4. All users are able to send or retrieve files except the admin of this system.

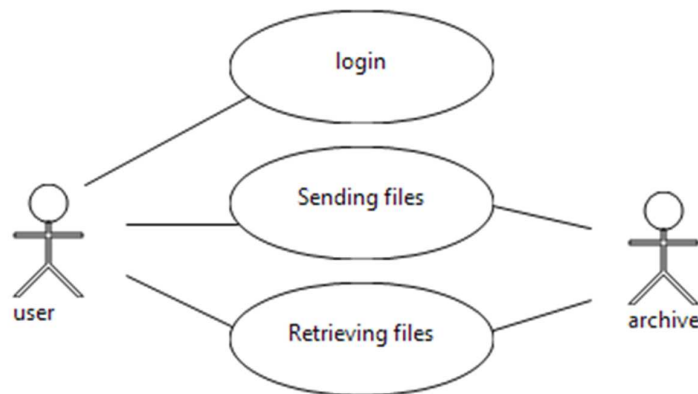


Figure 3.4: Sending/retrieving DICOM files to/from the archive

This system will be applied at different areas. In the medicine faculty, a teacher can create different groups of students and also different groups of files to be assigned to the groups of students as tasks. The teacher can grade the work done by the students. Each student is able to view only his grades as shown in figure 3.5. There are a number of similarities tasks between

the teacher and the student such as classifying, sending, or retrieving files. The student is not able to create groups, while the teacher can.

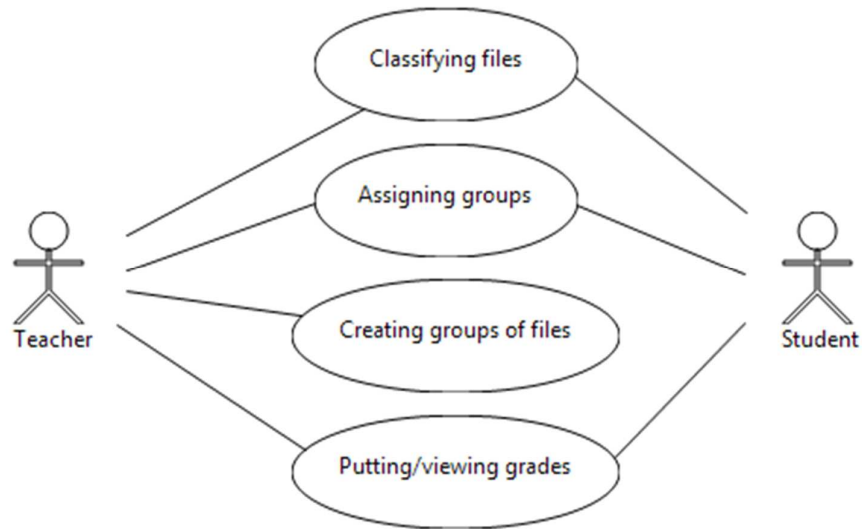


Figure 3.5: The workflow between a teacher and a student

At a hospital, a doctor can classify one or more DICOM files and share results with a researcher as shown in figure 3.6, while the researcher is able to apply one or more methods to one or more DICOM files.

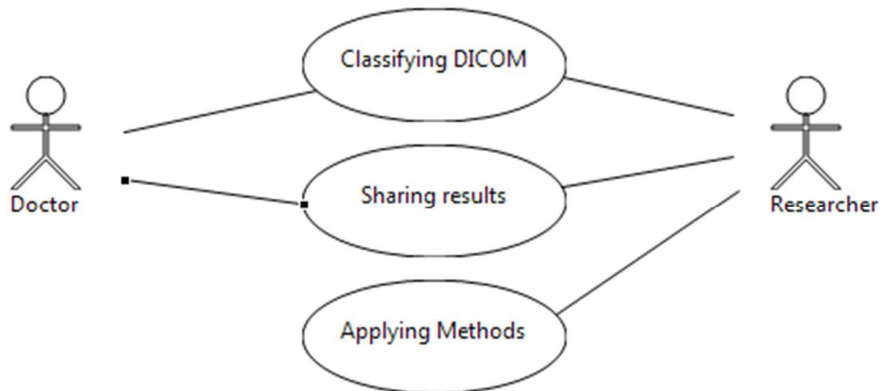


Figure 3.6: Basic tasks of doctor and researcher

A patient may have one more studies and one or more doctors can classify each one as shown in Figure 3.7. Each study may have one or more classified DICOM files. The results of classification can be shared between doctors and might be used as general classification for the patient.

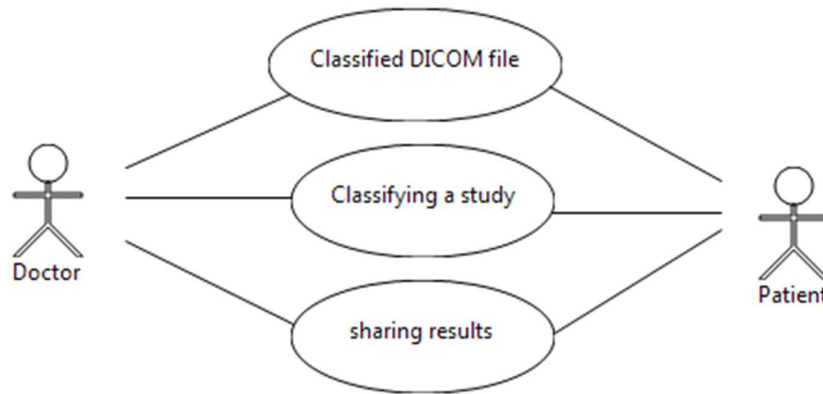


Figure 3.7: Classify a patient study

3.3.2 Use case descriptions

Due to space considerations and for sake of clarity, some of use cases are presented by tables to represent the relevant activities of this system. Each table consists of nine rows and two columns. For example in Table 3.1, the “Description” value in the second row is used to describe the use case “Anonymization” in general, while the “Source” value is used to determine a user(or a set of users) or system who should do the “Anonymization”. The “Destination” is used to determine a target user (or a set of users) or system that should use the “Anonymization” later to access to the system. The “Input” is used to identify the system inputs or what are the requirement inputs for the “Source” to do the “Anonymization”, while “Output” is identified the results that the “Destination” will get. How the system will behave in the “Anonymization” operation is defined by “Action”. Finally “Pre-condition” and “Post-condition” values are used to determine the system constrains that should be done before or after applying the “Anonymization”.

Table 3.1: Anonymization procedure

| Use Case | Anonymization |
|-------------|---|
| Description | This operation does not only anonymize the data in the DICOM file header, but also image pixels must be anonymized if the image contains personal information. If a group of DICOM file have the same patient name, so they should have the same anonymized name and ID. By the same way, a report can be anonymized but if the report is pdf file, some texts will be removed, while if the report is an image file, the top and bottom parts will be removed. |
| Source | A main user |

| | |
|-----------------------|--|
| Destination | The system archive (dcm4chee) will be the destination, if the input is a DICOM file. Otherwise the system database will. To be clear, this system is used dcm4chee as archive system to store DICOM files, while other files will be store at a database, we will see the proposed database later in this chapter. |
| Input | DICOM file headers, DICOM image pixels, or associated reports of DICOM files. |
| Output | Anonymized DICOM files or reports |
| Action | System checks the personal data in DICOM files to be replaced with random values, while in reports, the system will remove some parts. |
| Pre-condition | The system will check if there is other files that have the same patient information, if so, the system will collect them to be anonymized with the same patient name and ID. |
| Post-condition | None |

Table 3.2: Classify a DICOM file

| Use Case | Classify DICOM file |
|-----------------------|---|
| Description | A user is able to classify a DICOM file by using tags and relevant values. The classification does not only use tags, but also allows to draw some texts or shapes to mark or annotate DICOM images to determine the place that tags explains. Therefore, the classification can be a group of tags associated with annotated DICOM images, or can be only tags without annotation. |
| Source | A doctor, a student, a teacher or a researcher |
| Destination | A file will be stored at the system archive (dcm4chee), while the classification tags will be kept at the system database. |
| Input | A DICOM file. |
| Output | Classified DICOM file |
| Action | System will keep annotation on an image of the DICOM file and allows a user to add some tags. |
| Pre-condition | A user should login to the system and retrieve the DICOM file from the system archive to her/his machine. |
| Post-condition | None |

Table 3.3: Classify a patient study

| Use Case | Classify patient studies |
|-----------------------|--|
| Description | A doctor can classify patient studies in different study dates. A classification can be used as a general classification for patients. Also, this classification uses one or more tags with relevant values to describe the current study. This classification is inserted to the system without annotation. |
| Source | A doctor or a main user |
| Destination | The system database |
| Input | Classified DICOM files, patient name, doctor name, study date |
| Output | Classified patient study |
| Action | The system will generate a general study classification for one or more classified DICOM files. It shows a message to check if the operation is done successfully or not. |
| Pre-condition | DICOM files that are related to a current study should be classified. |
| Post-condition | None |

Table 3.4: Send or retrieve a DICOM file from/to PACS

| Use Case | Send/retrieve file |
|----------------------|--|
| Description | All DICOM files and reports will be collected in the system, so a user can send them through a network. The user is able to retrieve DICOM files from the archive to his local machine. |
| Source | A doctor, a teacher, a student, a researcher, or a main user |
| Destination | The system archive and the system database. |
| Input | A DICOM file or a report |
| Output | New DICOM file or report is added to the archive or to the system database. |
| Action | The system checks a listening port at a server, start sending a DICOM file or a report packet-by-packet to the archive through the listening port, and send a successful message back to a user application if the operation is done successfully or shows an error message. |
| Pre-condition | A DICOM file or a report does not exist in the archive or in the system database and also the format of them are supported (*.dcm, *.pdf, *.jpg) by the system. |

| | |
|-----------------------|------|
| Post-condition | None |
|-----------------------|------|

Table 3.5: A group of users, files, or reports

| Use Case | Create a group |
|-----------------------|--|
| Description | DICOM files, reports, or users can be collected in different groups. Each group name should be uniquely identified. |
| Source | A teacher or a main user |
| Destination | The system database |
| Input | New group name with a list of users, a list of DICOM files, or a list of reports. |
| Output | A group of users, DICOM files, or reports is created |
| Action | The system will ask to insert a unique group name and select a set of files or users to be added to the group. And then, the system will show successful message if the operation is done, otherwise, the system will show an error message. |
| Pre-condition | A group name does not exist in the system database. |
| Post-condition | None |

Table 3.6: Putting a grade of student

| Use Case | Put a grade |
|--------------------|---|
| Description | <p>A teacher can review a classification of a student to put a grade for her/his classification. For each classified DICOM file, the teacher can put a grade. The teacher can put any grade value such as 0-20, 0-100, or any number value can be added. The system calculates the grades of students by using this equation:</p> <p>The final grade for a student</p> $= \frac{\text{The total sumation grades of classified DICOM files}}{\text{The count of the files in the task}} \quad (3.1)$ |
| Source | A teacher |
| Destination | the system database |
| Input | A DICOM files, classification of the student, and the student name |
| Output | Grades of students |

| | |
|-----------------------|--|
| Action | The teacher can insert a grade for a student classification manually, after that, the system inserts the grade for the student to the system database and show successful message. |
| Pre-condition | A grade should be a number. |
| Post-condition | The system will put zero value as a grade for a student if she/he does not classify any DICOM file in the task that assigned to the user. |

Table 3.7: Assign a group to another group

| Use Case | Assign a group |
|-----------------------|--|
| Description | The system allows a user to assign a group of DICOM files or reports to a group of users. It is possible also to assign a group of users to another group of users. For example, consider a class with one teacher and many students, to allow the teacher to manage a group of students; first, the teacher should be put in a group and also the students also should be put in another group, and then assign the group of students to the group of teacher. Now, the teacher has access to all students at the second group. By using this way, it is possible to create different groups of files to be accessed and shared by different groups of users. Also the classification results can be shared by putting the classified DICOM files and reports in groups (source groups) and assign this groups to the desired groups of users (destination groups). |
| Source | A main user or a teacher |
| Destination | The system database |
| Input | A source group name of DICOM files, reports, or users; and a destination group name of users. |
| Output | A group of users have access to a group of DICOM files, reports, or users. |
| Action | Assign a source group to another destination group. |
| Pre-condition | A user cannot assign a group to the same group. |
| Post-condition | None |

Table 3.8: Apply a method on a DICOM image

| Use Case | Apply Method |
|-----------------|---------------------|
|-----------------|---------------------|

| | |
|-----------------------|---|
| Description | A method in this work is a plugin and it can be applied on one or more DICOM images to detect breast lesions. A researcher is responsible to create her/his plugin inside FIJI to be implemented on these images. Otherwise, he/she can retrieve DICOM files from the system archive to her/his local device, and then he/she uses FIJI plugins or external tools such as Matlab to detect breast lesions or to apply any other purposes. Not only a group of methods can be run at the client machine, but also a group of methods can be run at the remote server machine by sending commands from the researcher application (client) with names of methods and desired files as parameters to ask the server to retrieve these files locally and apply these methods on the retrieved files to create the modified DICOM files. |
| Source | A researcher |
| Destination | the system archive and database |
| Input | A list of DICOM files |
| Output | Modified DICOM files |
| Action | Retrieve the selected DICOM files to the user machine or to the server machine. If a user want to apply these methods on her/his local machine, so the system applies the selected methods on the retrieved DICOM files and keeps the modifications on the same DICOM files. The same thing will happen if a user want to apply the selected methods on the retrieved DICOM files at the server machine. |
| Pre-condition | A DICOM file(s) must be in the system archive and retrieved to a folder at the server machine or at the user machine. |
| Post-condition | None |

3.4 Data base diagram

According to the previous requirements, a new database is created with necessary tables. MySQL Workbench 6.2 CE is used to create the database diagram. All reports will be kept in this database, while only DICOM files parameters will be kept to be used by the system applications (doctor, student, teacher, and researcher applications) to retrieve files. This parameters are necessary because the system archive (dcm4chee) needs an identifier of DICOM

file (SOP class UID) and a patient name or ID and a transfer syntax of the current DICOM file as parameters to retrieve any DICOM file. Users' classifications with/without applied methods will be kept in this database as well, as shown in Figure 3.8. In the following database, new tables are appeared such as criteria, referenced name, and configuration tables. Where "referenced name" table helps to keep a list of single names to be chosen to create a random names to be replaced with the original patients' names of DICOM files during the anonymization process. While, "configuration" table is used to determine the defaults configuration of the archive to send or retrieve DICOM files. Finally, "criteria" table is used to keep the criteria of search of classified DICOM files to be used later by a researcher.

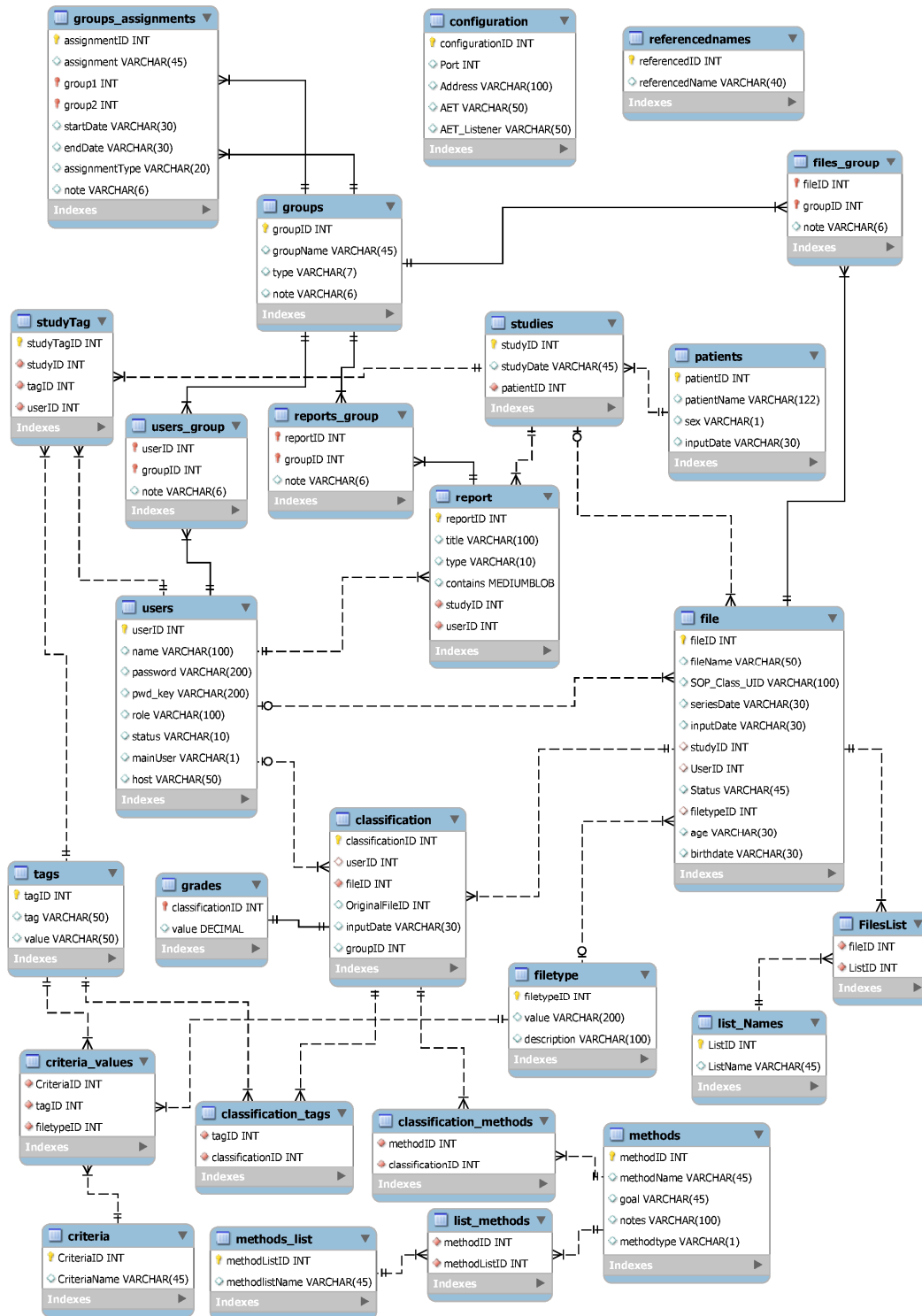


Figure 3.8: The proposed database to keep classifications, reports, DICOM files parameters, methods, and groups.

4 Medical Applications

4.1 Introduction

As mentioned in the background chapter, there are many applications were prepared to support DICOM standard, but according to the goals of this thesis, no one of them performs all the requirements that should be done. Therefore, we will try to integrate different tools with new applications to achieve the requirements of this work.

Dcm4chee is used to collect DICOM files as the system archive because his characteristics, please review Chapter 2 for more details. For sending DIOCM files to dcm4chee, dcmsnd tool is used, while for retrieving DICOM files from dcm4chee, dcmqr is used. To annotate (to draw texts, shapes on a DICOM image) or to do pre-processing and segmentation, FIJI tool is used with a new plugin to keep the modified image as DICOM file format. For Classifying DICOM files, some interfaces are created, which appear as plugins in FIJI. This system has six Java applications, admin, main user, doctor, student, teacher and researcher applications. The consistency and the exchange of information between the system tools are shown in the Figure 4.1. The main user, doctor, student, teacher, and researcher applications use the source code of dcmsnd and dcmqr tools to send or retrieve DICOM files to/from dcm4chee. Dcmsnd and dcmqr are encapsulated in the system applications as classes to do their functions automatically. This system uses also different complementary tools such as JDBC, MySQL server, SSL, TLS, IPsec, FIJI items, dcmrcv or dcmsnd, for connecting, encrypting, annotating, and listening tasks. SSL and TLS are supported by MySQL and dcm4chee but in this work, only VPN (IPsec) is applied. SSL or TLS is included in Figure 4.1 as other options if a user want to apply other types of security. In this chapter, the six applications of the system are discussed with more details. Basically, each application asks for a user name, password, and IP to log in. Different functions will be available such as automatically hide the personal data inside the headers of DICOM files, classify the DICOM files, and share the results among users. A user can select a folder with a group of DICOM files and reports in to upload them to dcm4chee PACS. The system will automatically use dcmsnd class to upload the files of the selected folder, so, a user does not care about dcmsnd parameters (Evans, Dcmsnd - dcm4che-2.x - Confluence, 2008). In the same way, the system retrieves DICOM files from dcm4chee by using dcmqr class but a user will determine the parameter of retrieving such as patient name and study date in the system interfaces. Since some interfaces of the system use “Image Plus” class and “IJ” library of FIJI, most FIJI’s commands or plugins that analyse, process, annotate, or segment images can be applied to DICOM images of this

system. This system has two database: dicom_db and pacs_db. The pacs_db database is used by dcm4chee to keep some information of DICOM files such as doctors, patients, modalities, and images information. While, users' information, privileges, reports, results of classifications will be kept in dicom_db. "dicom_db" is the system database that we had been suggested in the Chapter 3.

In the implementation stage of the system, NetBeans 8.0.1 software is used to build and create all the system applications. Since we need some of these applications to work as plugins in FIJI, the source code of them were copied into a folder inside FIJI tool to build them again using FIJI script or editor.

It is possible to write a code directly in FIJI editor, but this editor is not good compared to the facilities found in the NetBeans IDE. NetBeans editor allows a user to manage projects and build them very easily. Debugging and tracing code is very simple. For example, it is a common practice to watch variables during execution, which is nicely provided in NetBeans.

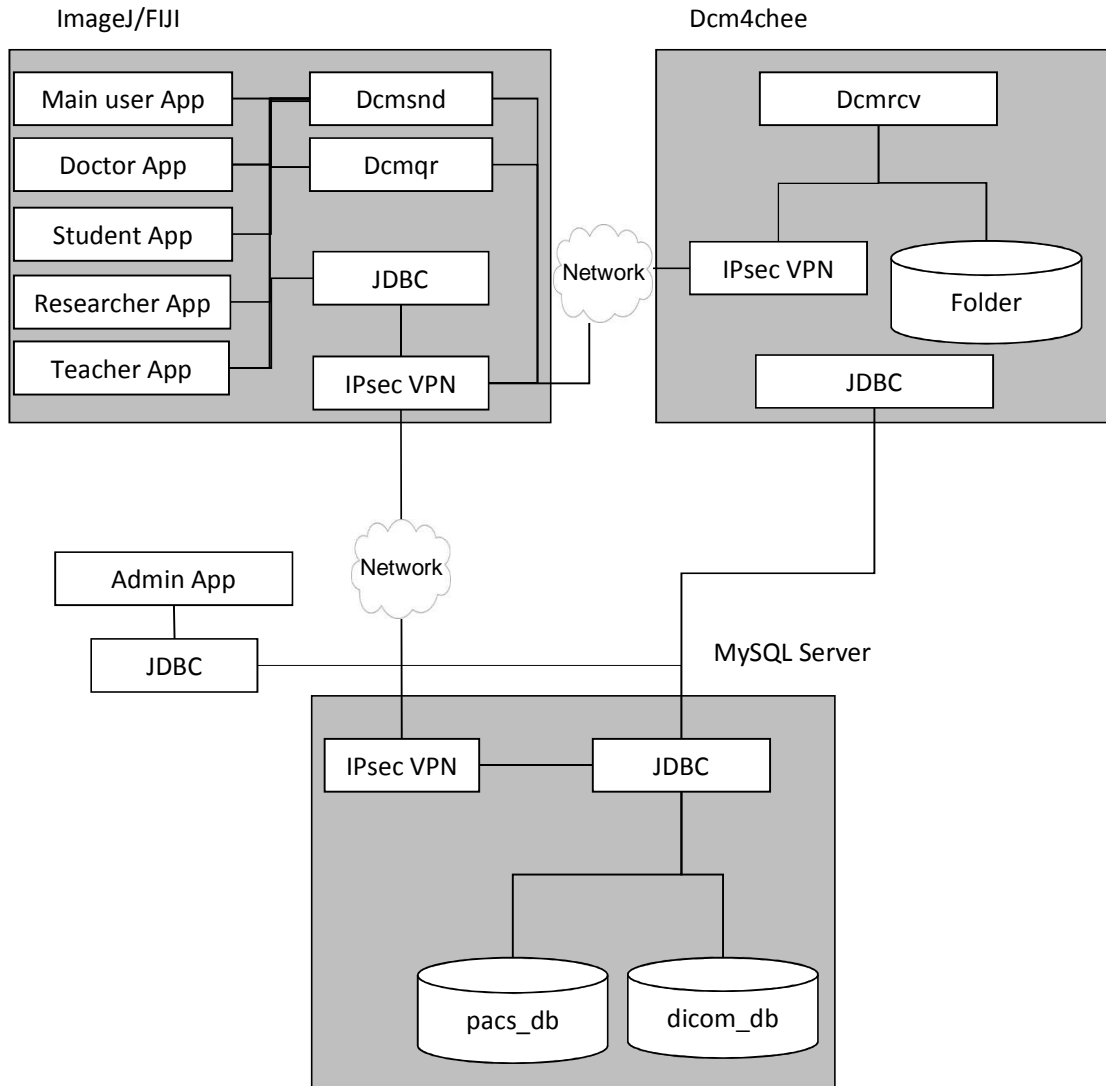


Figure 4.1: The architecture of the system

4.2 MAA (Medical Admin Application)

This application is prepared only for a root user in MySQL server, who is responsible to create users with different privileges. It works as a standalone application with easy interfaces to manage users. Different privileges for users are available such as doctor, main user, researcher, teacher, and student privileges. Each user can be only one of these privileges. Privileges for each one will be automatically assigned to each user according to selected privileges during the user creation. That means, the privileges that will be assigned to a user is different from one to another if they are using two different user privileges. It is possible to modify the privilege of a user by using grant or revoke commands in MySQL editor but to ease

managing of the users of the whole system, this application is created. During the creation of user, the root user can determine if a current user have access to dcm4chee files or not. Please see appendix for more details about the usage and other characteristics of this application.

4.3 MMA (Medical Main Application)

MMA application allows a user to manage a group of users, DICOM files, or reports by collecting them into groups. For the privacy rights of private health information, the personal information inside files will be anonymized before send them to dcm4chee. Through this application, the user can create new groups with different types such as users, files, or reports with the ability to assign a group of files, reports, or users to a group of users. MMA helps to see all the classifications that done by any user in the whole system. It is able to retrieve DICOM files from dcm4chee and import images of these files into FIJI for annotating. A standard or global classification can be added for a patient at specific study date to be used as the best classification result for the patient, which it enhances a diagnosis and helps in early detection for breast lesions. It appears as a plugin in FIJI, please see the appendix for more details about how to install and use this application.

4.4 MDA (Medical Doctor Application)

This application is prepared for a doctor who works at a hospital. It allows the doctor to retrieve some DICOM files with the ability to classify them and send the modified DICOM files back to dcm4chee. It also allows to add a global classification for a group of DICOM files at a specific patient study date. The classification can be created with/without annotations. Almost the functionalities provided by this system are similar to the functionalities provided by Medical Main Application except this application cannot create groups and cannot load one-word names to the system to be used for anonymization procedure. Please review the appendix for more details.

4.5 MTA (Medical Teacher Application)

A teacher at a medicine faculty can use this application with ability to have access to a group of DICOM files, reports, or users that assigned by MMA user to her/his. Creating different groups and assign a group of files to a group of students are supported. The classifications of students can be reviewed with ability to grade each classification and export the final grades to CSV files.

4.6 MSA (Medical Student Application)

A student is able to get a group of DICOM files as a task from a teacher to perform classifications. Students will increase their knowledge by classifying different cases of DICOM files with different breast cancer lesions. This application appears as plugin inside FIJI, so the student is able to annotate DICOM images. Each student can only view her/his grades.

4.7 MRA (Medical Researcher Application)

MRA allows a user to get a group of DICOM files to perform classifications. The user can create new plugins inside FIJI to be applied to one or more DICOM files to detect breast lesions. This application is supported with some plugins to keep modification on the DICOM files and execute one or more plugins to one or more DICOM files such as “DICOM Save” and “Batch Processor” plugins. “DICOM Save” plugin is responsible for keeping the annotation directly on a DICOM file without log in to the system, while “Batch Processor” is a modified version of FIJI batch process, but with ability to run one or more plugins and save the results on one or more DICOM files without log in to the system.

4.8 Methodology

Different methodologies were applied to anonymize DICOM files. All DICOM files will be sorted according to the patients’ names. And then, two different ways were suggested to anonymize the personal data inside the DICOM files. The first methodology is to apply the encryption to anonymize the DICOM files. The second one is to apply new method to replace the personal data with random values inside the DICOM files. The following part explains these methodologies with more details and show why this system choose the second methodology.

4.8.1 Ordering DICOM files

The necessity for this order, is to collect all DICOM files that belong to the same patient name and patient ID in a group of folders, to anonymize them with the same random string value (the patient name will be replaced with this value) and the same sequential integer value (patient ID will be replaced by this value) and so on. The system uses the merge sort as shown **Merge_Sort** Algorithm to sort a group of DICOM files by patient ID. There are a lot of sorting algorithms but this one was chosen because it works better than others such as bubble and insertion sort in medium size set of elements (Cormen, Leiserson, Rivest, & Stein, 2001).

Merge (Input, Tmp, Left, Right, RightEnd)

```

1- LeftEnd := Right - 1
2- K := Left
3- Num := RightEnd - Left + 1
4- While (Left <= LeftEnd and Right <= RightEnd)
5-     If (Input [Left] <= Input [Right])
6-         Tmp [K] := Input [Left]
7-         K:=K+1
8-         Left:=Left+1
9-     Else
10-        Tmp [K] := Input [Right]
11-        K:=K+1
12-        Right:=Right+1
13-    End if
14- End while
15- While (Left <= LeftEnd) // Copy rest of first half
16-     Tmp [k] = Input[Left]
17-     K := K + 1
18-     Left := Left + 1
19- While (Right <= RightEnd) // Copy rest of right half
20-     Tmp[K] = Input [Right]
21-     K := K + 1
22-     Right:= Right + 1
23- For(int i = 0; i < Num; i++, rightEnd--) // Copy Tmp back
24-     Input [RightEnd] = Tmp [RightEnd]

```

Merge_Sort (Input, Tmp, Left, Right)

```

1- If (Left < Right) then
2-   Center:= (Left + Right) / 2
3-   Merge_Sort(Input, Tmp, Left, Center)
4-   Merge_Sort(Input, Tmp, Center+1, Right)
5-   Merge(Input, Tmp, Left, Center+1, Right)

```

Merge_Sort (Input)

```

1- Create array Tmp [Input.length]
2- Merge_Sort (Input, Tmp ,0, Input.length-1)

```

4.8.2 Anonymization

Anonymization is a procedure to change data so it can no longer be traced back to an individual, for privacy sake. The system will anonymize DICOM files and reports before send them to the dcm4chee server. The anonymization procedure should anonymize DICOM files for a patient with the same name and ID because different doctors will classify these files to follow the patient situation and suggest a suitable treatment. A lot of algorithms and tools are created for anonymization and applied with different programming languages. One of these tools is DICOM cleaner tool (PixelMed, 2015), which allows to change or remove some of the original data of DICOM files, but this tool has one drawback, it anonymizes all DICOM files in a folder and sub-folders with the same data even if they are belong to different patients. If we use DICOM cleaner tool, we must to anonymize DICOM files folder-by-folder and must to make sure that all DICOM files in a folder follow to the same patient. This way will take a lot of time especially if the number of patients are too big. Other tools such as DICOM Anonymizer (Slashdot Media, 2014), Anonymize IJ DICOM (Cooper, 2009), (O'Dell, 2003) are available but no one of them is enough to do the anonymization procedure as the requirements of this work. Therefore, we suggest to use anonymization procedure that grantees the anonymized data cannot return back again. To do that, two ways are applied: use Deletion of 18 identifiers function with encryption, or use Deletion of 18 identifiers function with a new procedure to replace the original values with random values.

Deletion of 18 Identifiers function is a de-identifying way that be used by health organizations to determine the important elements that have personal data in a medical file (UC Davis Health System, 2014). It shows 18 personal data inside medical files that should be removed or anonymized such as names, elements of date (date of birth, admission date, etc.), phone number, fax number, Electronic mail addresses, Social security numbers, All information that determine the location (State, street address, city, county, precinct, zip code), medical record numbers, health plan beneficiary numbers, account numbers, certificate/license numbers, vehicle identifiers and serial numbers, including license plate numbers, device identifiers and serial numbers, web universal resource locators (URLs), internet protocol (IP) address numbers, biometric identifiers (finger and voice prints), full face photographic images and any comparable images, and any other unique identifying number, characteristic, or code that can identify the individuals.

First, doing the anonymization using encryption. There are a lot of algorithms to do encryption procedure such RES or DES, but these types of algorithms can encrypt data with

keys and also use the same keys to decrypt the encrypted data. Therefore, these algorithms are not good to do anonymization because the encrypted data might be returned back to the original data especially if a user get these keys. Therefore, it is better to use encryption with one way functions. Hash function is a one way function that turns original texts into a fixed string of digits (MathWorld Team, 2014). This technique can be used for creating digital signature, which in turn protect messages from disclosure. Hash function is not reversible, and this means it is infeasible to generate the original strings from output and also it is infeasible to find two different inputs with the same output (MathWorld Team, 2014). There are a lot of hash functions such as MD5, SHA-0, SHA-1, SHA-2, etc., but only SHA-2 was used in this work. SHA-2 is abbreviation of Secure Hash Algorithm #2, which consists from a set of cryptographic hash functions such as SHA-224,SHA256,SHA-384,SHA-512 (FIPS, 2002). SHA-256 is chosen because it does not have any collision outputs till now and grantee more security (FIPS, 2002). Two part inside DICOM files should be anonymized: the header elements and image of DICOM. Since each DICOM header contains a group of elements and each element contains a tag, type, and value, the value of the element should be changed to something else, please see Chapter 2 for more information about important DICOM elements. Figure 4.2 explains the proposed way to anonymize the important element values inside DICOM header. Deletion of 18 Identifiers function is used to determine the important elements that should encrypted by SHA-2 with a random secure salt.

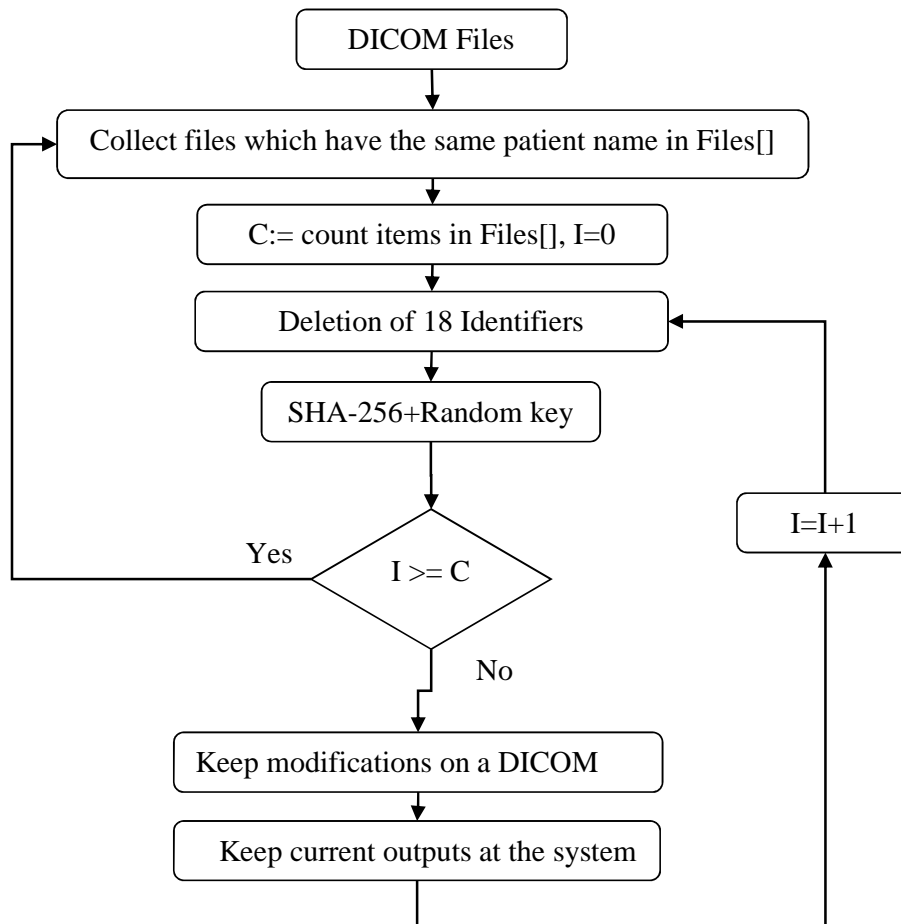


Figure 4.2: Anonymize DICOM header

Now, to anonymize the image of a DICOM file, a rectangle with limited size with black fill out colour will be drawn on the image.

Unfortunately, using Deletion of 18 Identifiers function with encryption has one big drawback, it is difficult to memorize patient names after encryption with long encryption outputs or even if with short outputs. The second option is used since it is easier to apply and can be memorized easily.

Secondly, doing the anonymization using Deletion of 18 Identifiers function with a new procedure to replace the important data with random values. Since we need to replace the original patients' names with names that can be easily memorized, more than 3000 different single names were kept in a special table in dicom_db database of this system to be used as list to create anonymized patient's names. For any important number value inside a DICOM header, '0' value will be used to replace it, while an important string value will be replaced with empty value. Another way is applied for date values. As we know the date contains of three part: day,

month, and year. Therefore, a day always will be changed into constant value “5”, while a month will be rounded to the nearest number as shown in Algorithm 2.

Algorithm 2 anonymize month of date.

input: month of a date

output: anonymized month value

```
1- begin
2-     Month:= month of the date, new_month=0
3-     if month <= 3 then new_month=02
4-     else if month > 3 and month <= 6 then new_month=05
5-     else if month > 6 and month <= 9 then new_month=08
6-     else new_month=11
7-     end if
8- end
```

For anonymizing a year, the following equation is used:

$$\text{Year} = ((\text{int}) (\text{Math.round}(\text{Integer.parseInt}(\text{year})/10.0) * 10)) \quad (4.1)$$

Where Math.round() is a function in Java to round a number to the nearest integer number, while Integer.praseInt() is a function in Java to cast string values into number values. The purpose of this equation is, to round the year to the nearest tens, for example, values from 2010 till 2015 will be changed into 2010, while values from 2016 till 2020 will be changed into 2020. Algorithm 3 shows the whole steps to anonymize any important element inside DICOM files.

Algorithm 3 anonymize DICOM element

input: get a DICOM file element

output: anonymized element value

```
1- begin
2-     if (element is a number)
3-         New_element:= 0
4-     else if (element is a string)
5-         New_element:= ""
6-     else if (element is a date)
7-         begin
8-             Day=05
9-             Month:= use algorithm 2 (month of the date)
10-            Year:= ((int) (Math.round(Integer.parseInt(year)/10.0) * 10))
```

```

11-             New_element:=Day+Month+Year
12-         end
13-     end if
14- end

```

The whole procedure to open a group of DICOM files, doing anonymization, and send DICOM files is represented by Algorithm 4:

Algorithm 4 open DICOM files, doing anonymization, send them

input: FilesList[] := get a group DICOM file element, ElementList= get important elements that should be anonymized in FilesList[], NamesList[]= get a group of single names

output: Anonymized DICOM files

```

1- begin
2-     Merge_Sort (FilesList.PatientIDs) for sorting the list by patient ID
3-     group := 0, ID := 1
4-     for I:=0 to FilesList.length
5-         begin
6-             do
7-                 I++
8-                 while (FilesList[I-1].PatientID = FilesList[I].PatientID)
9-                     Patient_Name := ""
10-                    for Z=0 to 3
11-                        begin
12-                            RandomIndex := 1 + (Integer) * Math.random() *
                                NamesList.length
13-                            Patient_Name := Patient_Name +
                                NamesList[RandomIndex]+'_'
14-                        end
15-                    end if
16-                    for w= group to I
17-                        begin
18-                            FilesList[w].PatientID := ID
19-                            FilesList[w].PatientName := Patient_Name
20-                            for d=0 to FilesList[w].Tags.getSize()
21-                                begin

```

```

22-         if ( FilesList[w].Tags.getTag(d) <> Tag.PatientID
           or
           FilesList[w].Tags.getTag(d) <> Tag.PatientID )
23-         if(FilesList[w].Tags.getTag(d) in ElementList)
24-             Use algorithm 3 with parameter
                (FilesList[w].Tags.getTag(d))
25-         end if
26-     end if
27-     end
28- end for
29-     send FilesList[w] to dcm4chee (PACS server)
30- end
31- end for
32-     group := I + 1, ID := ID + 1
33- end
34- end for
35- end

```

Different reports from a hospital were be reviewed to determine the location of important data to be anonymized, we find that the headers and trailers for each report should be removed. This procedure will not check if there is personal data in the body of a report, so, a doctor who will send it to this system should change any personal data in the body of the report.

4.8.3 Grouping files

This system support a main user or a teacher with the ability to create different groups such as a group of files or a group of users. This way is important to ease the control of a set of files. Each group has a unique name, ID, and collection of items. A group of items can be assigned to a group of users, which means that each user at the second group will have access to each item in the first group. The system asks for start and end dates for each assigning process to deny the access automatically to the items of the groups after the end dates are finished.

4.8.4 Security

The security issues are very crucial, since some DICOM files will be accessed by one or more users and might be transferred through untrusted networks, a clever attacker can sniff

all data being sent between the terminals; also, the passwords or other data might be cracked or used special if they are in clear texts. It is important to protect data from attackers and from users of this system.

First, there are many way that might be applied to protect the system against attackers in this system such as use MySQL server with SSL (secure socket layer) or with SSH (secure shell) tool, use dcm4chee TLS (transport layer secure) with passcode and user name, use virtual private network (VPN) OpenVPN or IPsec with MySQL server. Since this system works inside the University of Algarve, any user want to connect to the university from outside should install VPN (IPsec). Therefore, this system uses VPN to encrypt all data being sent and also a user name and password are mandatory to log in to deal with the system interfaces. Since all information between clients and dcm4chee server will be anonymized, dcm4chee might be used without TLS or VPN mechanisms, especially if the speed of the system is a big issue. The security in MySQL requires a user name and a password to connect to a MySQL server database, but only the password is encrypted in a client system using 4.1 hashing method (Zoratti, 2006). The hashing method is used to create hashing password to provide better security and the risk of disclosure will be reduced. By default, a client connect to MySQL server without encrypted connection, so the data will be vulnerable to attack. Therefore, a user can use Secure Sockets Layer (SSL) protocol to create a secure connection and to achieve more security fashion. In this system, MySQL server 5.5 is implemented with VPN (IPsec) protocol.

Second, to secure data inside tables of a database from being modified or deleted by users of the system, grant and revoke commands are applied. These commands support a developer with four options: select, insert, update, and delete to deal with a table in a database. As according to this work, different users will be created with limited access but only the admin user will have full permissions to all databases in MySQL server. A main user have full access to one database called dicom_db, which is the system database. While the rest users do not only have limited access to some tables of the dicom_db database, but also in some cases have limited rows inside some tables. Unfortunately, the grant command of MySQL does not support giving a user limited access to some rows inside a table; therefore, views are created to do that. The system contains many views to allow some rows in tables to be accessed by desired users. To be clear, let's imagine we have two tables: "Users" and "Files" as depicted in Figure 4.3. Both of them are main tables and created in the dicom_db database. We need to create a secure view that guarantee a user in "Users" table has limited access to some rows in "Files" table.

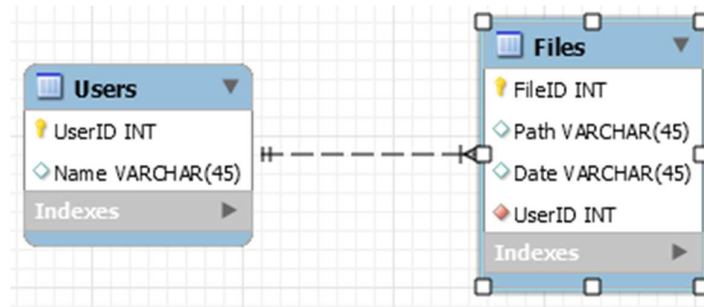


Figure 4.3: An example of the proposed views to give limited access to some rows in a table

To do that, we will create a view to show rows that belong to a user inside “Users” table by using the following command in MySQL workbench 6.2 CE:

```
CREATE VIEW dicom_db.user_view as
(SELECT u.* from dicom_db.Users u where u.name=SUBSTRING_INDEX(user(),'@',1) );
```

This command will create a new view with ‘user_view’ name. The created view allows to select all rows that have the same user name inside the “Users” table rows. “User()” is a secure function in MySQL that can be used by any user, which is responsible for getting the IP with user name in “userName@ip” form for a user who already logged to the system. Since the user name is required without IP, substring_index function is used to extract only the user name. It is possible to execute “select user from mysql.user where user=‘user_name’” directly to get the user name without using “substring_index” function, but the user must have select permission to “mysql.user” table in “mysql” database (the default database in MySQL server to store users’ information) and this is not secure to give a normal user access to this table (Oracle Corporation, 2014). Now, by using the previous view “user_view” with the following view, the user will have a limited access to the “File” table rows:

```
CREATE VIEW dicom_db.file_user_view as
(SELECT f.* from dicom_db.Files u where f.userID=select UserID from dicom_db. user_view);
```

Then, these views should be granted to the use by running the following commands:

```
GRANT SELECT ON dicom_db. user_view TO 'User_Name'@'IP';
GRANT SELECT, INSERT, UPDATE, DELETE ON dicom_db. file_user_view TO 'User_Name'@'IP';
```

Now, the user has access to all rows that he inserted by himself in “Files” table. This way grants that any user will only have access to rows that belong to her/him, even if any application of this system or other tools MySQL Workbench are used. Therefore, all views inside the database of the system are built according to this way to restrict access for some users. Besides to these views with assigned groups of files, the user only will have a limited access to the files of the system.

5 Discussion, conclusion, and future work

5.1 Discussion

In a final phase, a test and evaluation task was done. During the test phase, the PACS server was loaded with more than one hundred files and the work flow described in the user manuals was checked. Some minor errors were detected and corrected. During the evaluation phase, the functionalities of the present system was compared with others of similar goals, taking into account the state-of-the-art review done before the development of this system was started. This section will discuss and compare some functions of this system with other systems.

The anonymization procedure works well and better than some tools such as DICOM cleaner tool (PixelMed, 2015), DICOM Anonymizer (Slashdot Media, 2014), TUDOR DICOM (Hermen, 2009), DICOM browser (Archie, 2012), or Anonymize IJ DICOM (Cooper, 2009), either functioning automatically or manually. The system will collect all DICOM files that belong to the same patient, then replace the patient name and the patient ID in all DICOM files with a random patient name and a sequential ID. The blackout of DICOM cleaner tool is used to anonymize some parts of a DICOM image, but it can be applied to only one image at a time while the blackout of the proposed system not only does the same thing by using FIJI tool, but it can also be used to blackout a group of DICOM images simultaneously.

To annotate a DICOM image, DICOM exporter class of Tudor DICOM application can be used with the FIJI tool box, but this application does not use this class by any interface to annotate DICOM images (Hermena, 2008). Therefore, the library that contains DICOM exporter class is used. Unfortunately, it works locally, but when a derived DICOM file is sent to dcm4chee, an error happened about some tags that are not supported by dcm4chee. So the source code of DICOM exporter is used after some modifications such as changing some tag types into the supported types by dcm4chee. DICOM exporter has another problem, when it writes data to a DICOM file; it replaces the header of the DICOM file with a default header; Therefore, a new function is created to copy the missing header elements after writing data to a derived DICOM file from the original DICOM file, after these modifications, this system is able to annotate DICOM files and it works better than Tudor DICOM application regarding this issue.

There are some tools which can send, retrieve, or collect DICOM files such as Tudor, MRIdb, DCMTK, ClearCanvas, while others are built only to anonymize or view DICOM files. To our knowledge, there are no available tools that can perform all the needed functionalities

to implement a client-server system that handles medical image files as the developed system does.

This system and Tudor have similar functions such as sending or retrieving DICOM files from/to dcm4chee. They also works as plugins in ImageJ or FIJI. Tudor anonymization is manual (a user has to enter the anonymized data manually), while in this system implements an automatic anonymization that includes the possibility to change the anonymized data if a user so chooses. To send/retrieve files from dcm4chee using TUDOR system, a user needs to enter the names of the patients. In contrast, this system is simpler, not only offers support to a user choices with a combo box with the names of the patients, but also a mechanism for the auto-completion of names is supported. Management of different users in Tudor system is not supported, whereas it can be easily managed in this system using the interfaces of the admin and main user applications. Further details can be found in the corresponding user manuals.

There is a good library called “DICOM Image I / O Plugin” (Apteryx, 2014), which is a collection of Java libraries that helps a developer to collect, send, retrieve, annotate DICOM files, and also supports the developer with the server and the client libraries. Some experiments were applied to check its functionalities. This library works well, but unfortunately, it is not free, the cost of buying the license ranges from 50 to 8000 euros for client and server (Apteryx, 2014). Furthermore, this library is created to help only developers in the problems of reading, writing, sending, retrieving, and collecting DICOM files. Classification and access management are not supported.

Since there are several free PACS servers that collect DICOM files, several factors were considered when choosing a PACS server such as open source, free of charge, comprehensive documentation, multi-platform support, ease of use, works with a big organization, and allows to manage a lot of users (Schöch, 2014). According to the mentioned factors and also to the requirement of this thesis, dcm4chee was chosen as PACS server, while FIJI is chosen as DICOM viewer. Please review chapter 2 for more details.

5.2 Conclusion

Breast cancer is a malignant disease that starts in the cells of the breast, usually in the ducts and lobules. Mammography, Ultrasound, or Elastography modalities can be used to detect the breast lesions according to the age of a patient and a lesion’s type. Currently, most imaging equipment’s conform to one version or another of the DICOM standard. A DICOM file contains of a header and pixels of the image. Each header contains many elements and each element consist from a tag, type and value. PACS system is used to help in collecting DICOM files from

these modalities to be accessed through the Internet or through a local network. The developed system unites and integrates a group of tools and applications: dcmsnd, dcmrcv, dcmqr, dcm4chee, MySQL server, FIJI, DICOM exporter of TUDOR application, six new solutions, and the IPsec protocol. Dcm4chee contributes with different components such as Weasis, dcmsnd, dcmrcv, and dcmqr tools and is used to collect DICOM files. The source code of dcmsnd tool is used for sending DICOM files, while for retrieving DICOM files from dcm4chee, the source code of dcmqr is used. In order to annotate a DICOM image, FIJI, modified TUDOR DICOM exporter class, and a newly developed function to copy the missing elements of the DICOM headers are used. Six solutions (doctor, main user, researcher, teacher, and student plugins and the admin application) were developed targeting each of the user roles, interfacing the server with different privileges but also determining a different work flow on a limited group of DICOM files and reports to be accessed for classification or learning purposes. This system rests on three levels of security. First level, since this system is working inside the firewall of the University of Algarve, each user must use the intranet or the VPN (IPsec) to have access. Second level is enforced by using a user name and password at MySQL server to access to the system. Last level consists on the application of the anonymization procedure to all DICOM files and reports before sending them to dcm4chee. Health organizations use deletion of 18 identifiers to determine the important personal data in a DICOM file to be removed. This system not only anonymizes a DICOM file header elements, but also removes the personal information conveyed in the DICOM image by using the blackout procedure, which draws in black colour the pixels of the upper part of the DICOM image to remove the personal data. Some plugins are implemented to help a developer to run a group of methods (plugins) to a group of DICOM files.

5.3 Future work

Performance evaluation of the system under heavy load will open interesting lines of work. The extension of the researcher interface with the ability to retrieve a group of DICOM files and apply Matlab or OpenCV methods being studied and, eventually, to execute a group of commands in remote machine or in remote cluster would also be challenging.

In what concerns medical images, it would be beneficial if the annotation of DICOM images would support compressed DICOM files and video DICOM files.

6 User Manuals

6.1 MedDoctor plugin User Manual

This tutorial is arranged in two parts: installation, configuration, and usage instructions of the server side for an admin user; and also installation and usage instructions of the client side for a main user, a researcher, a doctor, a student, and a teacher.

6.2 Server side

6.2.1 The installation steps

The following steps are to install dcm4chee in the Debian operating system:

- 1- Download MySQL server from this link: <http://dev.mysql.com/downloads/mysql/>, for more details, go to : <http://dev.mysql.com/doc/refman/5.1/en/binary-installation.html>
- 2- Download the last version of dcm4chee, in this work dcm4chee-2.18.0 is used, it can be downloaded from this link: <http://sourceforge.net/projects/dcm4che/files/dcm4chee/2.18.0/dcm4chee-2.18.0-mysql.zip/download>
- 3- Download the binary distribution package of JBoss Application Server 4.2.3.GA: <http://sourceforge.net/projects/jboss/files/JBoss/JBoss-4.2.3.GA/> , download namely “jboss-4.2.3.GA.zip” package.
- 4- Create a new folder with a name such as “dcm4chee”
- 5- Copy these zip folders (dcm4chee-2.18.0-mysql.zip, jboss-4.2.3.GA.zip) into “dcm4chee” folder
- 6- Unzip these folders and make sure that these folders after the zip operation do not have the same folder name inside. For example, when you unzip jboss-4.2.3.GA.zip folder you will get a new folder with name “jboss-4.2.3.GA”, open this folder and check if there is another folder inside it with the same folder’s name “jboss-4.2.3.GA”, that means the unzipped folder will be like “path/jboss-4.2.3.GA/jboss-4.2.3.GA/all_files_and_folders”, if so, cut all files and folders into the first one and delete the second occurred of “jboss-4.2.3.GA” folder to make this folder as “path/ jboss-4.2.3.GA/all_files_and_folders”. If you do not do that, some errors will occur during execution time, also do the same for dcm4chee-2.18.0-mysql folder.
- 7- Execute the commands between quotation marks in the terminal of Debian to copy files from jboss-4.2.3.GA folder to dcm4chee-2.18.0-mysql folder, just write “*cd*”

path/dcm4chee/dcm4chee-2.18.0/bin” then write *“./install_jboss.sh path/dcm4chee/jboss-4.2.3.GA”*. If you get a permission denied, you must change the permission of *install_jboss.sh* file with this command *“chmod u+x install_jboss.sh”*

- 8- Create and initialize the dcm4chee database and the database of this system at MySQL server. First, write *“mysql –uroot -p”* in terminal, you should enter the password of the root user. And, write the following commands:

```
“create database pacfdb; /*dcm4chee database*/
```

```
grant all on pacfdb.* to 'pacs'@'localhost' identified by 'pacs';
```

```
create database dicom_db; /*the system database*/
```

```
mysql> \q
```

```
mysql -upacs -ppacs pacfdb < path/dcm4chee/dcm4chee-2.18.0/sql/create.mysql
```

```
mysql -uroot –proot_passwod dicom_db < path/Database_Script_v13.sql
```

```
mysql -uroot –proot_passwod dicom_db < path/inputs_v3.sql“. Please notice that, the last two lines contain “root_password” parameter, which it is the password of a root user at MySQL server.
```

- 9- You should set the environment variable *“JAVA_HOME”* to JDK location. Please install *jdk1.6.0_45* version to make this server works without mistakes, if you have an older version or newer, you will have errors during the run time. Write *“export JAVA_HOME=path/jdk1.6.0_45/”* and *“export PATH=\${JAVA_HOME}/bin:\${PATH}”* in terminal. You can download this version of *jdk* from this web site: <http://www.oracle.com/technetwork/java/javase/downloads/java-archive-downloads-javase6-419409.html>. After that, use *“cd”* command to enter to this folder and write *“chmod +x jdk-6u45-linux-i586.bin”* and *“./jdk-6u45-linux-i586.bin”*. Do not forget to write *“export JAVA_HOME=path/jdk1.6.0_45/”* and *“export PATH=\${JAVA_HOME}/bin:\${PATH}”* to set java environment for this server. You can write *“java – version”* to see your installed java version and *“which java”* to see the location of your installed java.

- 10- Write *“cd path/dcm4chee/dcm4chee-2.18.0/bin/”* command and write *“./run.sh”* to run the server.

- 11- Write on web browser this link: *“Server_IP:8080/dcm4chee-web3/”* or *localhost:8080/dcm4chee-web3/”* if you want to run the server locally. The default user name is *“admin”* and the password is also *“admin”*.

12- For more details on installation steps of dcm4chee, go to the following link :

<http://www.dcm4che.org/confluence/display/ee2/Installation>

13- In order to allow a user to connect remotely to MySQL server, you need to write “sudo nano /etc/mysql/my.cnf”, go under “bind-address = 127.0.0.1” line, add this “bind-address = 0.0.0.0” line, and save and exit by clicking ctrl&o and ctrl&x on the keyboard. Then, restart MySQL server by writing “sudo service mysql restart” on terminal. This means, you allow any user from any address to connect to MySQL server.

6.2.2 The configuration steps for dcm4chee

After you have already installed the server side successfully, you need to configure the dcm4chee to enforce users to enter a user name and a password for sending or retrieving DICOM files by applying the following:

1- Open this file “path/dcm4chee/dcm4chee-2.18.0-mysql/server/default/conf/login-config.xml” and comment out “<login-module> ... </login-module>”, so this part should be like:

```
<application-policy name = "dcm4chee-dicom">
  <authentication>
    <!-- Comment out following login-module to require/check
    passcode in
    DICOM User Identity negotiation -->
    <!-- ←
    <login-module
    code="org.dcm4chex.archive.security.TrustLoginModule"
      flag = "required" >
        <module-option name = "password-
    stacking">useFirstPass</module-option>
    </login-module>
    --> ←
    ...
  </application-policy>
```

Please, just add “<!-- “ before “<login-module” line and “-->” after “</login-module>” line

2- Open this file “path/dcm4chee/dcm4chee-2.18.0-mysql/server/default/conf/xmdesc/dcm4chee-dcmsec-xmbean.xml” and replace false value with true value as following:

```
<attribute access="read-write"
getMethod="isRejectIfNoUserIdentity"
  setMethod="setRejectIfNoUserIdentity">
  <description><![CDATA[
    Defines if Association Request w/o User Identity
    information from
    remote AEs with no default User Identity
    specified by the AE
    configuration shall be rejected!]]>
  </description>
  <name>RejectIfNoUserIdentity</name>
```

```

<type>boolean</type>
<descriptors>
  <value value="true" />
</descriptors>
</attribute>

```

Modify false
value into true

3- Use “cd” command to go to “/path/dcm4chee/dcm4chee-2.18.0/bin” in terminal and write “./shutdown.sh -S -u admin -p admin_password” to stop dcm4chee. And then, write “./run.sh” to restart dcm4chee again.

6.2.3 The usage steps for server side

6.2.3.1 Admin user Application

An admin user connects to the system locally or remotely. The system will ask for a user name, a password, and an IP. The IP can be “localhost” or the server’s IP as shown in the Figure 6.1. This user should be a root user in MySQL server because this user is responsible for creating or deleting users and databases.

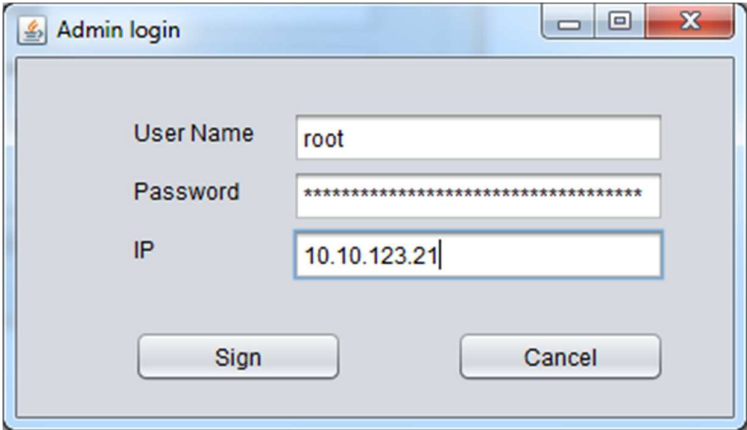


Figure 6.1: An admin login interface

After the admin user logged in to the system correctly, the new window is appeared with two buttons, “Add User” and “Connection” as shown in Figure 6.2. The admin user is able to add a new user by clicking on “Add User” button. The window of creating users and assign some roles to them is depicted in Figure 6.3.

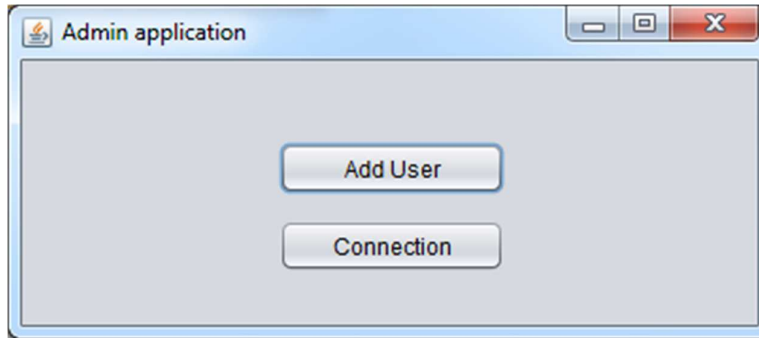


Figure 6.2: Add users and configure the connection interface

A new user can be inserted by entering a user name, a password, IP in “User Name” combo box, “Password” text fields, and IP text field. Choose the desired privilege for the current user such as teacher, student, main user, doctor, or researcher check boxes. “PACS User” check box allows the current user to send or retrieve DICOM files to/from dcm4chee. It requires to select a role to be assigned to the user and new password to be used automatically by the system in retrieving or sending DICOM files. Finally, you need to click on “Create” button to save these information to the system. For the security issues, set “IP” field to a specific IP value such as “10.10.234.192”. To allow the user to connect to the system from any device, set “IP” field to “%”. It is possible to enter one or more IPs in “IP” field to allow a user to connect from any one of these IPs, *i.e.*, set the “IP” field to “IP1;IP2;IP3;...”. Where each IP should be followed by semicolon as shown in Figure 6.3. The Admin is able to assign any role to the user, but it is better to create a new role in dcm4chee to create limited access to the user. You will know how to create new roles in dcm4chee in the next page. If you want to modify the user privilege from one to another or if you want to disable this user from sending or retrieving files to/from dcm4chee, just choose the user name from combo box to load his information automatically, and then click on “Modify” button, choose the destination privilege or unselect “PACS User” check box and click “Save Modifying” button.

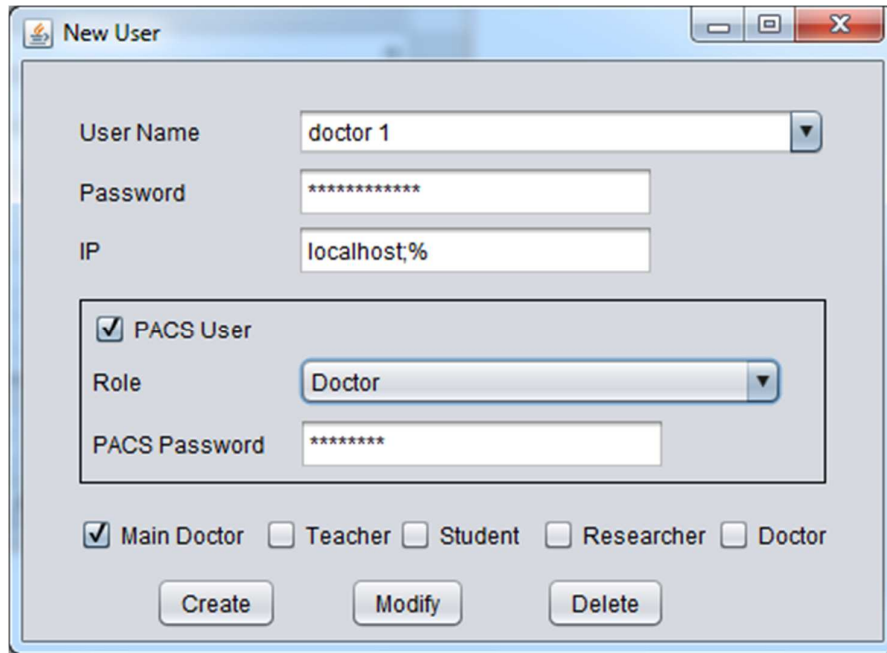


Figure 6.3: The interface of inserting a user to the system.

The connection configuration can be changed by clicking on “Connection” button in Figure 6.2. Since the default configuration of dcm4chee is used, you can set “Port” field to “11112”, “Address” field to “The IP of your Server”, “AET” field to “DCM4CHEE”, “AET” Lister to “DCMRCV”, and then click on “Save” button as shown in Figure 6.4.

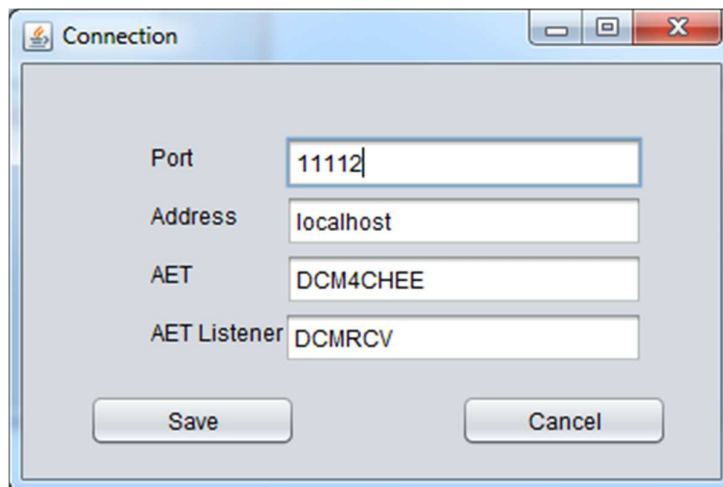


Figure 6.4: Using the default configuration of dcm4chee

6.2.3.2 Add a new role in dcm4chee

A new role should be added to let users only send or retrieve DICOM files without connecting to dcm4chee by a web browser. An admin will use dcm4chee-web3 tool to connect

to dcm4chee from the web browser. Only copy this link "*Server_IP:8080/dcm4chee-web3/*" into your web browser and change "Server_IP" to the IP of your server. You need to enter a user name and password. After you log in to dcm4chee successfully, go to "Roles" menu and click on "New role" button, write a name and a description as you want, for example, enter "Student" as the role name and "limited access for students" as the description, choose "DICOM" and "AET" check boxes, and click "save" button as shown in Figure 6.5.

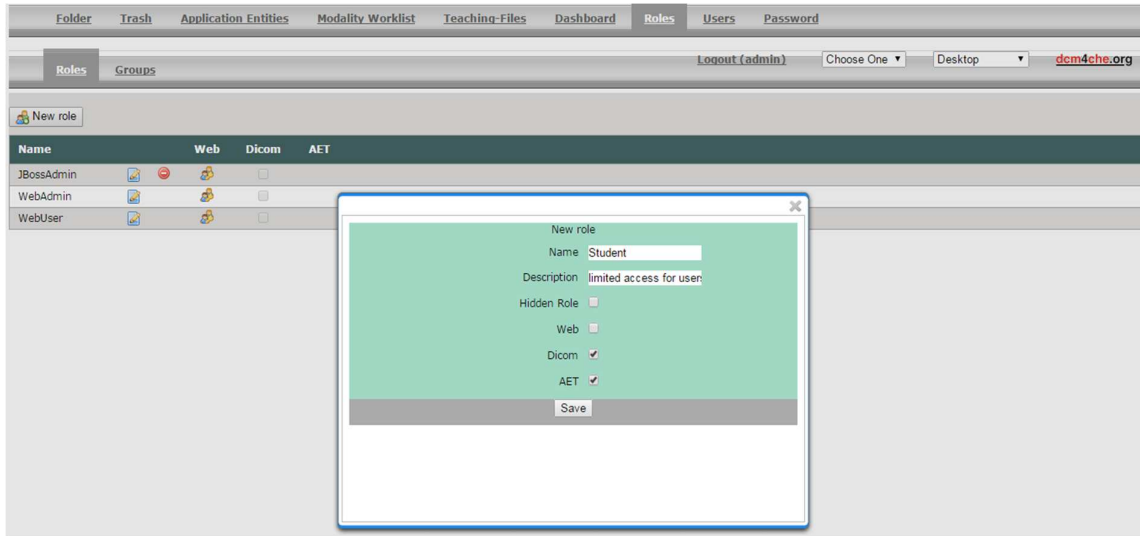


Figure 6.5: Add new role in dcm4chee

Then, if you want the “Student” role that you have already inserted to be displayed in the “Role” combo box in the Figure 6.3, you should assign this role to the admin user or any user at dcm4chee server. Therefore, you will click on “Users“ menu and click on “manage roles” shape. Choose “Student” role on the appeared window and click on close button as shown in Figure 6.6.

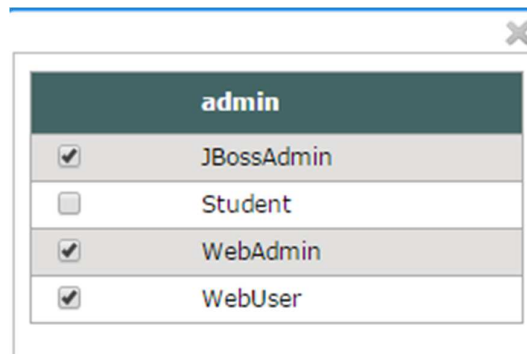


Figure 6.6: Assign a role to the admin user

6.2.3.3 Macro Listening Application

The Admin can run macro listening application to listen on special port to receive commands from researcher application as shown in Figure 6.7. Private and public keys must be created for the server and the client sides to protect the listening socket to be connected by any application. Key tool of Java is used to create these keys to encrypt sending and receiving packets and to allows a client that has the public key to communicate with the server.

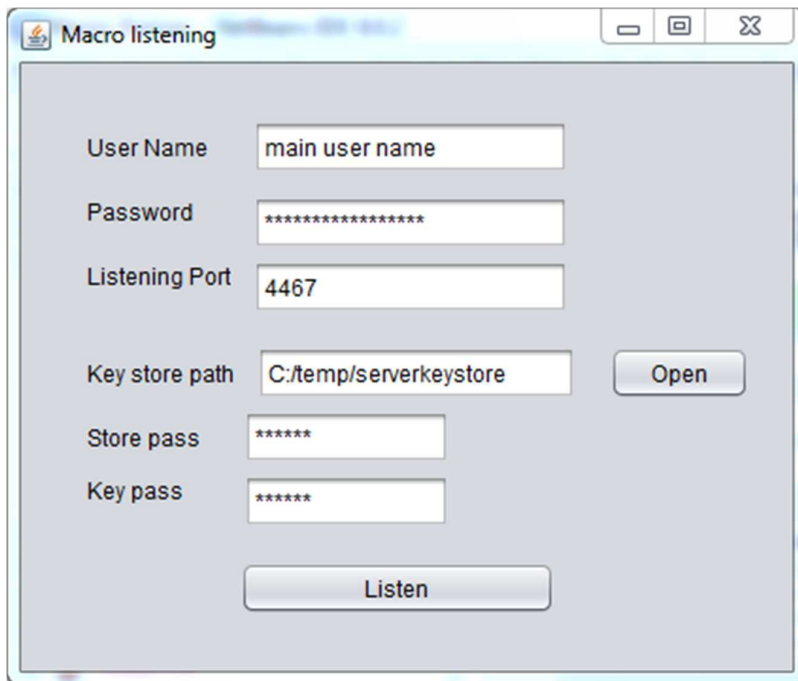


Figure 6.7: Receive and run commands of the researcher application at the server.

To create public and private keys using Key tool, please do the following steps:

- 1- Open the terminal and write the command between brackets: `[keytool -genkeypair -keyalg RSA -keysize 3072 -dname "cn=assylia, o=assylia.Inc, c=FR" -alias alias-name-for-server-key -keypass yourkeypassword -storepass yourstorepassword -validity 36500 -keystore path/keystore]`
- 2- After that, export a self-signed certificate by writing: `[keytool -export -alias alias-name-for-server-key -keystore path/keystore -rfc -file path/certificate.cer]`
- 3- Finally, create a trust key for client as following: `[keytool -import -alias new_alais_name-client-key -file path/certificate.cer -keystore path/truststore]`

The last command will ask to insert a password. It is possible to insert the same password of keystore but it is better to create a new one. “*alias-name-for-server-key*” and “*new_alais_name-client-key*” values are inner names for each key and they must be different, while “*path/keystore*” is the path to the server key with “keystore” name, “*path/truststore*” is the path to the client key with “truststore” name. When you run the macro listening application, choose the path of “keystore” file and enter the store and key passwords that you have already entered in first step of creating keys. In order to allow the researcher application to use the “truststore” file automatically, please copy this file into the “Medresearcher” folder inside FIJI plugins folder. Where “Medresearcher” is a folder that contains all files of the researcher application

to be run as plugin in FIJI tool. You will know more information about the researcher applications later.

6.3 Client side

6.3.1 The installation steps for a doctor, a student, a teacher, a hospital doctor, or a researcher

- 1- Download FIJ tool from this link: <http://fiji.sc/Downloads>, as according to your operating system, choose FIJI version that works in your operating system. In this version, FIJI for windows 7 32bit is used as shown in Figure 6.8.

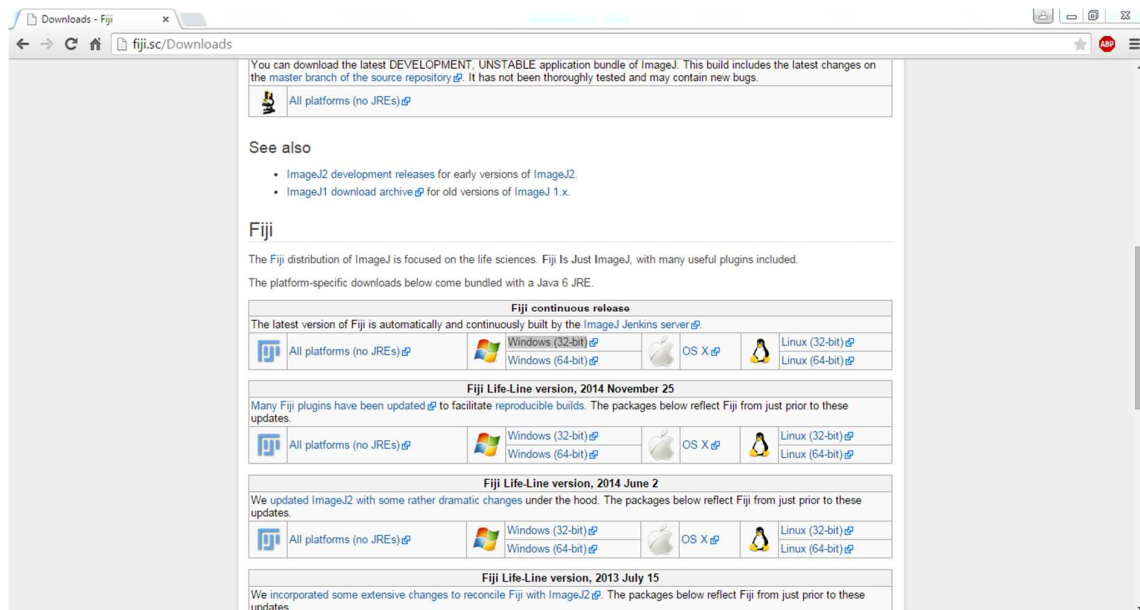


Figure 6.8: A web site for download FIJI for windows 7 (32bit).

- 2- Put the downloaded folder at any place you want. In this installation, FIJI win32 zip file is cut into a new folder called “FIJI application” at local disk (c) as shown in Figure 6.9.

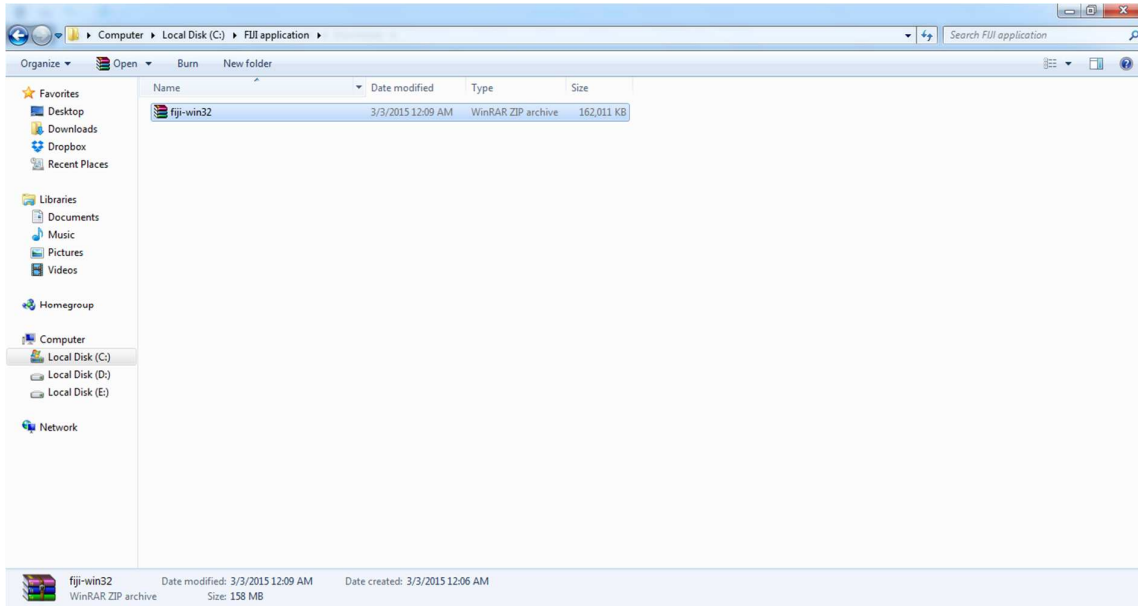


Figure 6.9: Put a Zipped FIJI file into a folder

- 3- Then, unzip Fiji-win32 file by mouse right-click on this file and choose “Extract to fiji-win32\” command as shown in Figure 6.10. Your operating system should have a zip application to such as WinRAR, you can download it from <http://www.winrar.com/affdownload/download.php>.

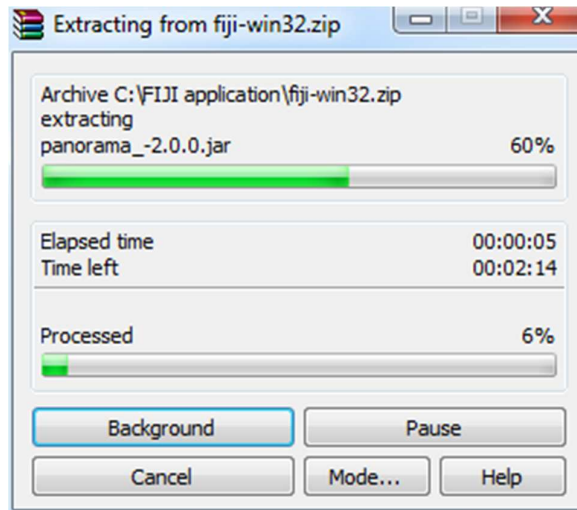


Figure 6.10: FIJI extracting

- 4- To run Fiji application, open “FIJI application/fiji-win32” folders, and double click on “ImageJ-win32.exe” as shown in Figure 6.11.

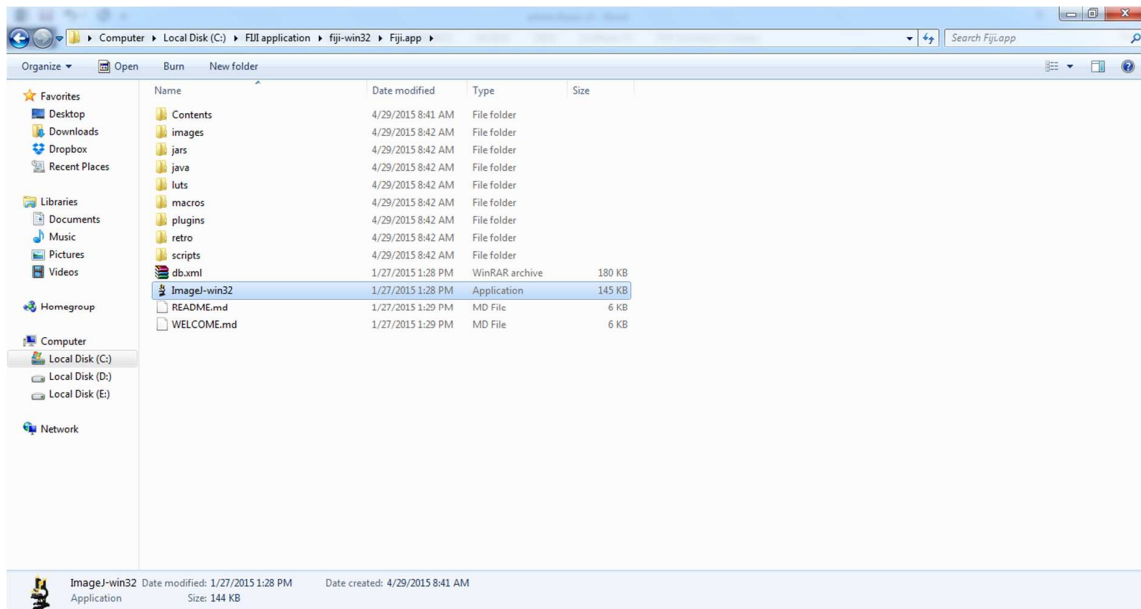


Figure 6.11: FIJI application path

5- Now, the Fiji interface will appear as shown in Figure 6.12.

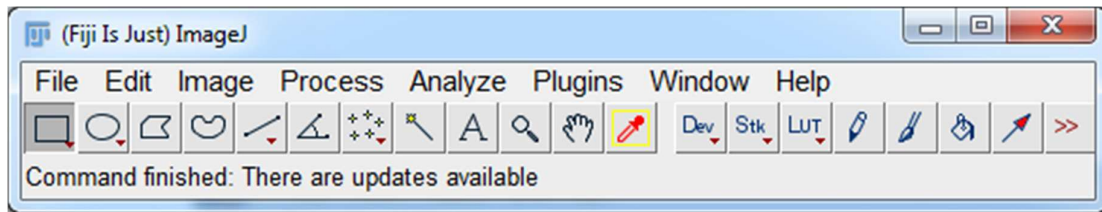


Figure 6.12: FIJI interface

6- The doctor, student, teacher, or researcher applications work as plugins inside FIJI, and each one has a special folder, therefore, copy your folder into “FIJI_Path/plugins/” folder. The admin of this system will give you a folder with all packages inside. After that, close FIJI application and run it again. These application will appear as commands (plugins) at FIJI’s plugins menu with different names such as MedDoctor, MedStudent, or MedTeacher.

6.3.2 The usage steps for a doctor, a student, a teacher, a hospital doctor, or a researcher

6.3.2.1 Main doctor

6.3.2.1.1 Doctor login

A user connects to the system locally or remotely according to the privilege that assigned by an admin user to this user. To login to the system, run FIJI application, go to FIJI plugins menu, click on “MedDoctor” plugin. Normally, the system will ask for a user name, a password, and IP as shown in Figure 6.13.

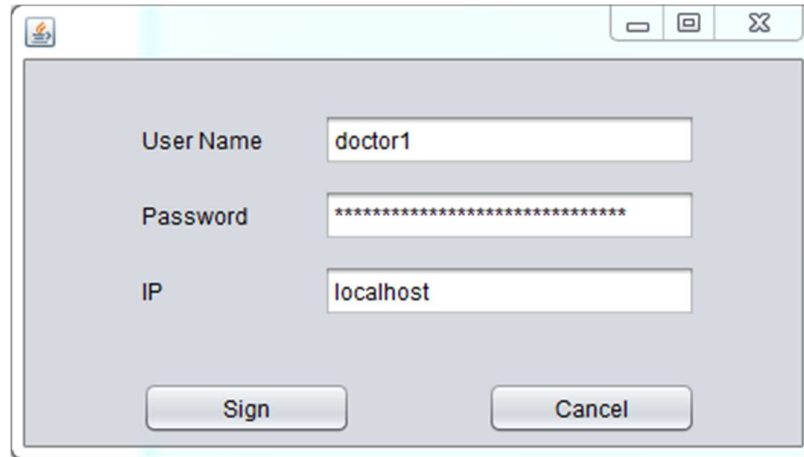


Figure 6.13: Login interface of a doctor

6.3.2.1.2 Anonymize and Send files to PACS

To run anonymization procedure, click on “Operation” tab, “File Upload” tab, “open” button, choose a desired directory that contains a group of DICOM files and reports, and click on “Open” button. Personal information in these files will be anonymized automatically and appeared in the table of the current opened window as shown in Figure 6.14. These changes will occur only after clicking on “send” button. That’s means, if the automatic changes are satisfied, the user can click on “send” button to keep these changes into files and send all the files and reports to dcm4chee. Otherwise, some changes can be occurred by selecting a destination cell to be changed to something else in the table of Figure 6.14, enter a new value in “Change selected value(s)” field, and press “enter” on keyboard. It is possible to change one cell value or more one time by clicking on the first cell, press and hold “ctrl” on the keyboard, click on the destination cells one-by-one, and click on “change selected value(s)” field and enter your value in and press “enter” on the keyboard.

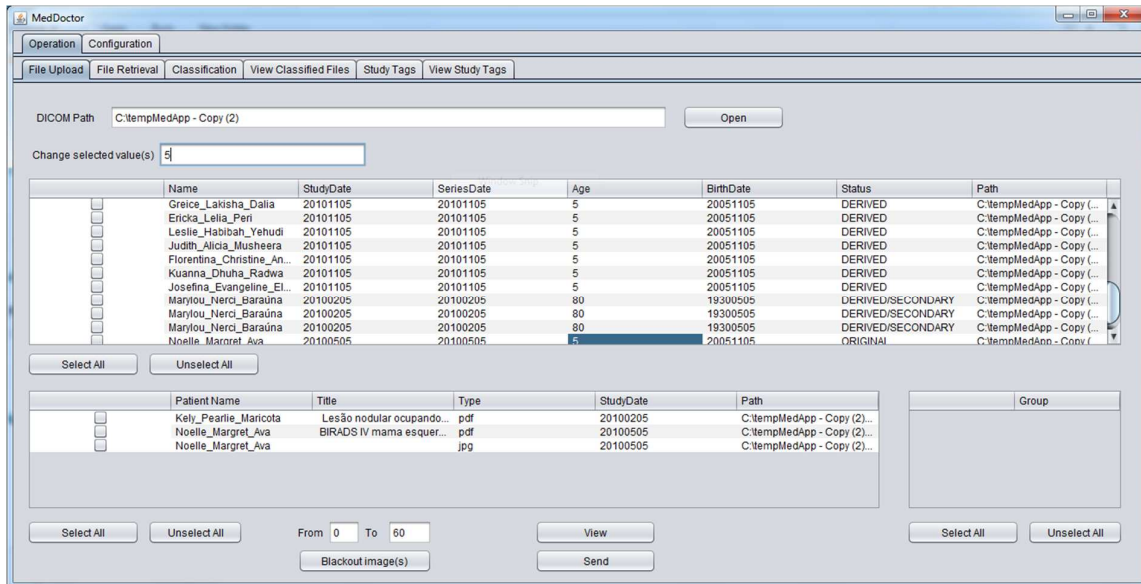


Figure 6.14: Anonymize and send DICOM files and reports to dcm4chee

6.3.2.1.3 Retrieve files from PACS

Click on “Operation” tab, “File Retrieve” tab, and choose a criteria of search from “DICOM Type”, “Status”, “Patient’s Name” combo boxes as shown in Figure 6.15. The system will load the criteria of searching when any patient name is chosen. Finally, select one or more rows from the table in the depicted window in Figure 6.15 and click on “Retrieve from the

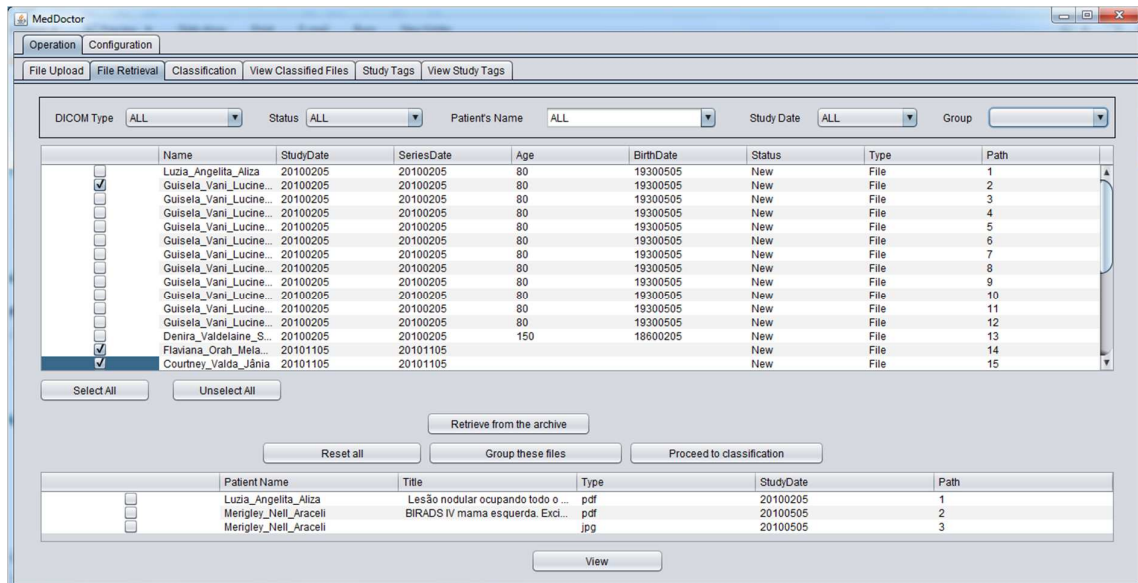


Figure 6.15: The Criteria of retrieve DICOM files from dcm4chee

archive” button. In the “status” combo box, different values can be chosen such as ‘all’, ‘new’, ‘Classified W marks’, or ‘Classified W/O marks’, where ‘new’ means a file is not classified

yet, while ‘Classified W marks’ means the file is classified with some annotations (shapes or texts), and ‘Classified W/O marks’ means that the file is classified without marks or annotations.

6.3.2.1.4 Classify DICOM files

After you have already retrieved DICOM files from dcm4chee to your local machine, you are able to classify them by clicking on “Proceed to Classification” button in the Figure 6.15. The system will active “Classification” tab as depicted in Figure 6.16.



Figure 6.16: Classifying DICOM files

6.3.2.1.4.1 Insert a new tag with its value to the system

Since each classification can have one or more tags according to the status of a current DICOM file, a user is able to choose different values from “Tag” and “Value” combo boxes as shown in Figure 6.16. If the default tags with its values are not enough for a classification, the user can add a new tag and its values by inserting a value in “Tag” combo box and a value in “Value” combo box and press “enter” on the keyboard before leaving “Value” combo box.

6.3.2.1.4.2 Annotate a DICOM file

In order to draw shapes or texts (annotation) on a DICOM image, click on “Import to FIJI” button in the depicted window in Figure 6.16. The system will show a new interface as shown in Figure 6.17. In order to draw a meandering line on an image, click on “Freeline” radio box, move the mouse cursor to the image and click and hold the left mouse button with keep

moving to draw your shape, free the button of the mouse and click “ctrl&d” on the keyboard to keep your drawing on the image, and click “Save” button to keep the modification to the DICOM file. If the user need more shapes that are not exist in Figure 6.17, FIJI toolbox can be used by clicking on any wanted shape and go back to this image and starts your drawing. Also, other FIJI’s commands or plugins can be applied to the current image. Now, go back to the

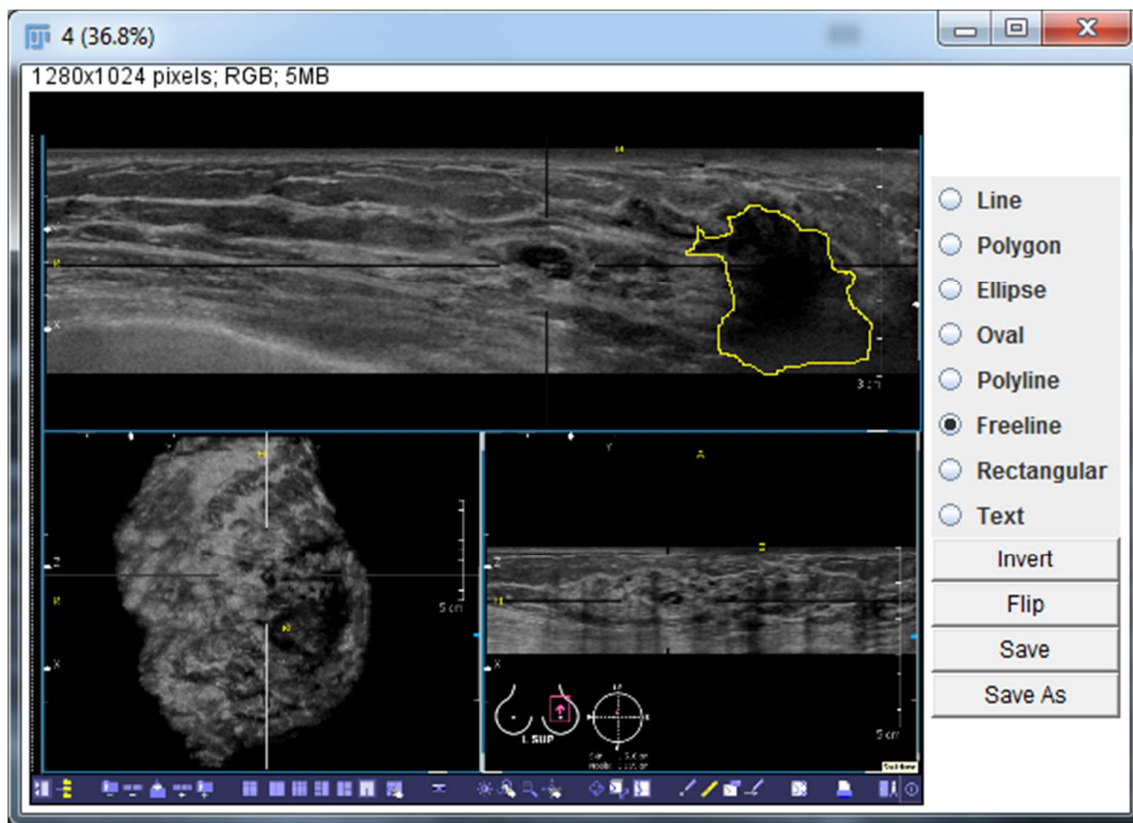


Figure 6.17: Drawing shapes on DICOM images

window in Figure 6.16 and click on “Reload modified image” button to load the modified image to this application.

6.3.2.1.4.3 Save and send a DICOM file to dcm4chee

You can click on “Save also modified image(s)” check box to enforce the system to send the derived image, and then click on “Save this classification” button to send the classification tags with the derived image to PACS. If “Save also modified image(s)” is not selected, the classification tags only will be stored at the system.

6.3.2.1.5 Create a global Classification for each study date of patient

To create a global classification for each study date of patient, click on “Operation” tab, “Study Tags” tab, choose patient’s name and study date from combo boxes, and click on “Update Table” button. Enter the desired classification tags from combo boxes and click on “Save this global Classification” button as shown in Figure 6.18.

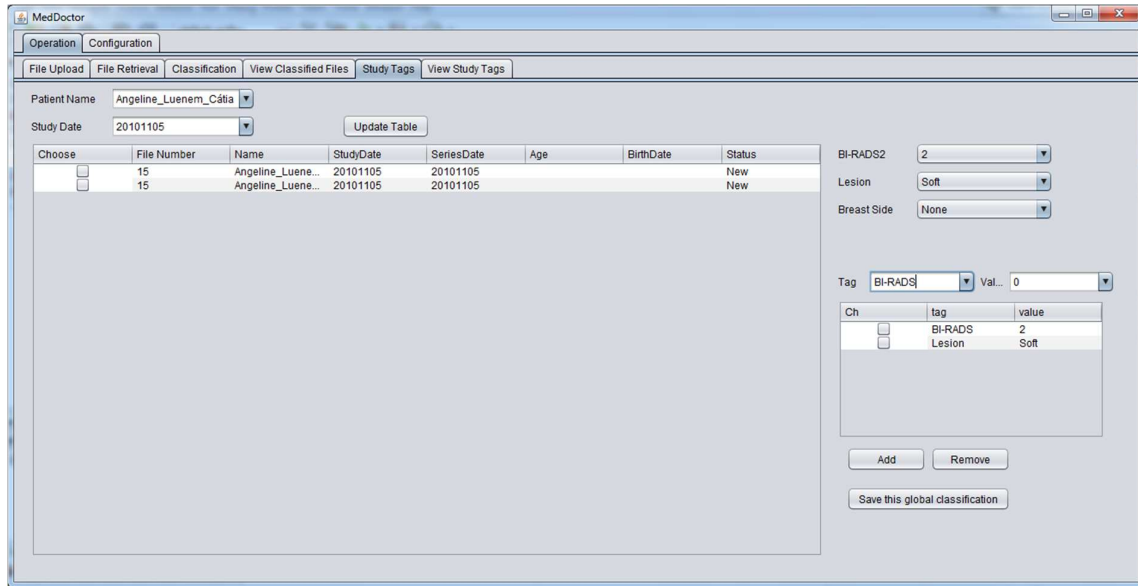


Figure 6.18: Classify a study date for a patient

This global classification can be viewed by clicking on “Show study Tags” tab, choose the criteria of search, and click on “Update Table” button.

6.3.2.1.6 Create a group of users, DICOM files, or reports

A main user can manage a group of users easily by creating a group and assign users to this group. To create the group of users, click on “Configuration” tab, “New Group” tab, “New Group” radio box, enter a new group name in “Group Name” field, choose “Users” radio box as a group type, and click on “Create” button as shown in Figure 6.19. With the same way, a group of files or reports can be created.

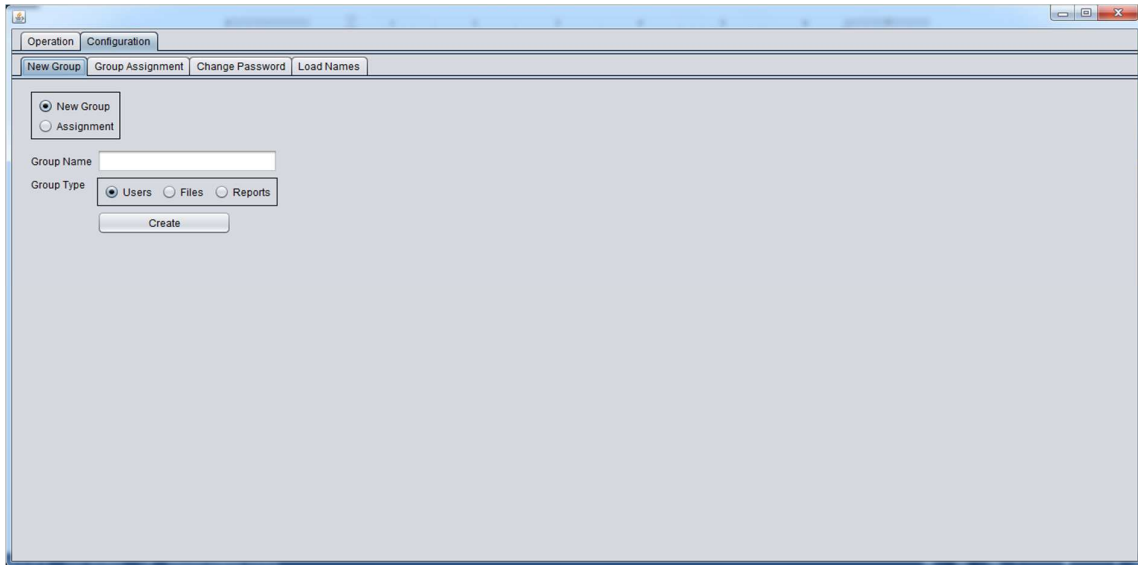


Figure 6.19: Create a group

After creating different groups, the main user is able to add a collection of users, files, or reports to the created groups by clicking on “Assignment” radio box, select the desired group type, click on “Load Type”, choose a collection of users, files, or reports, and select the destination group name and click “Add” as shown in Figure 6.20.

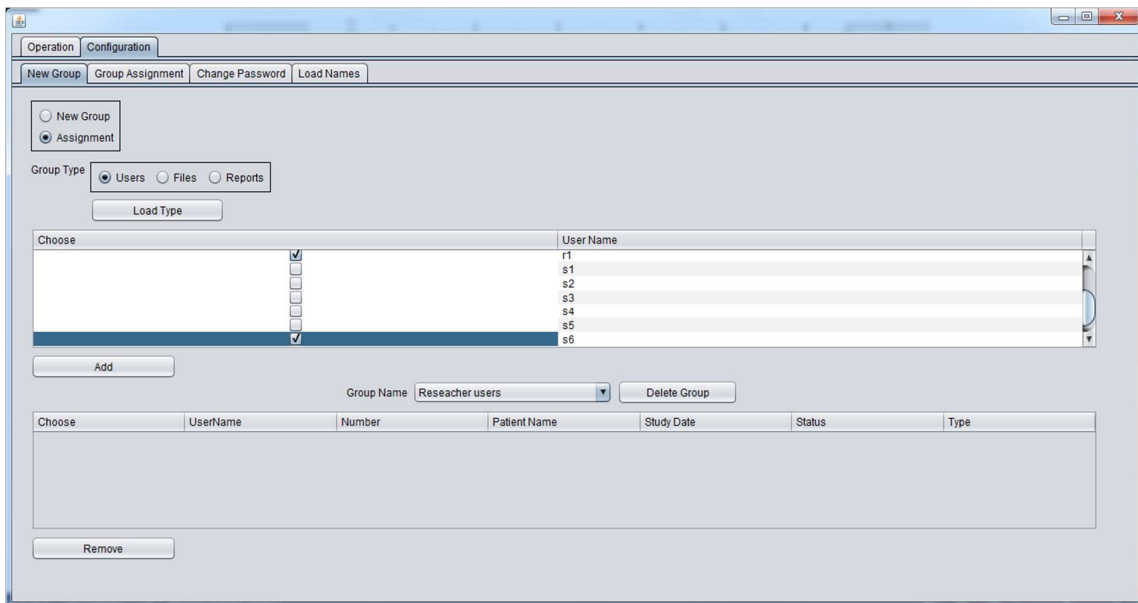


Figure 6.20: Add elements to a group

Finally, you need to assign groups to allow the users of a group to have access to other groups’ elements. To assign groups, go to “Configuration” tab, “Group Assignment”, and enter source and destination groups’ names, enter correct dates, and click on “Add” button as shown in figure 6.21.

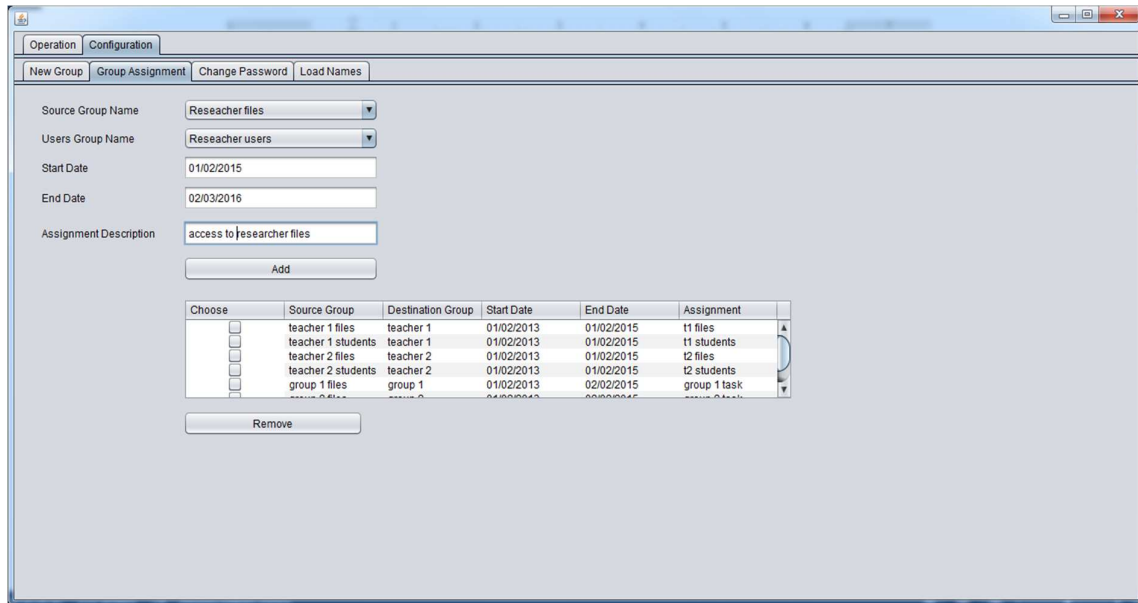


Figure 6.21: Assign a group to another group

6.3.2.1.7 Load one-word names to the system

A group of one-word names can be uploaded from a txt file to be used in creating patients' names in the anonymization procedure. To do that, go "Configuration" tab, "Load Names" tab, select a file and click on "load Names" as shown in figure 6.22.

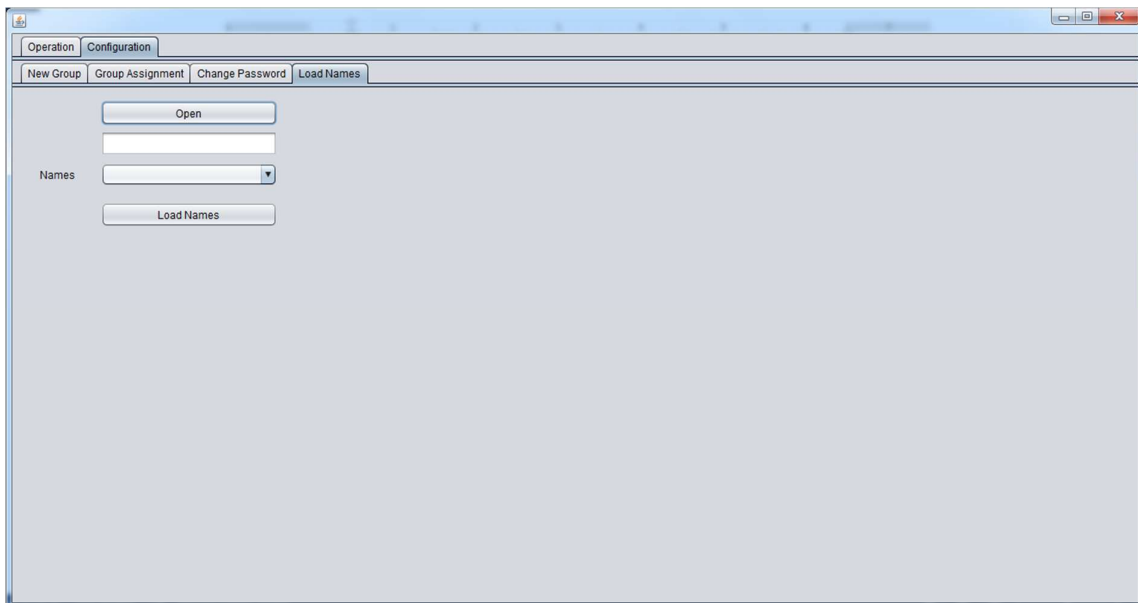


Figure 6.22: Load single names to the system database.

6.3.2.2 Teacher

Basically, to connect to the system, a user name, password, and ip are required. Different procedures will be available at least retrieve DICOM files, send derived files, and create groups of users, views reports, grade students' classifications.

6.3.2.2.1 Retrieve Files and classify DICOM files

Please review the previous parts of the main user to see how to retrieve and classify DICOM files.

6.3.2.2.2 Put grades for a student

To put a grade, go to "Operation" tab, "View Classified Files" tab, select the criteria of search from the combo boxes, the system will load the selected criteria after selecting any patient's name into the shown table in Figure 6.23. Select the desired row(s) from this table, click on "retrieve image" check box if you want to view the derived images that were created by students, and click on "Retrieve" button. To put a grade for a student, just enter a grade in "Grade" field and click "Save" button.

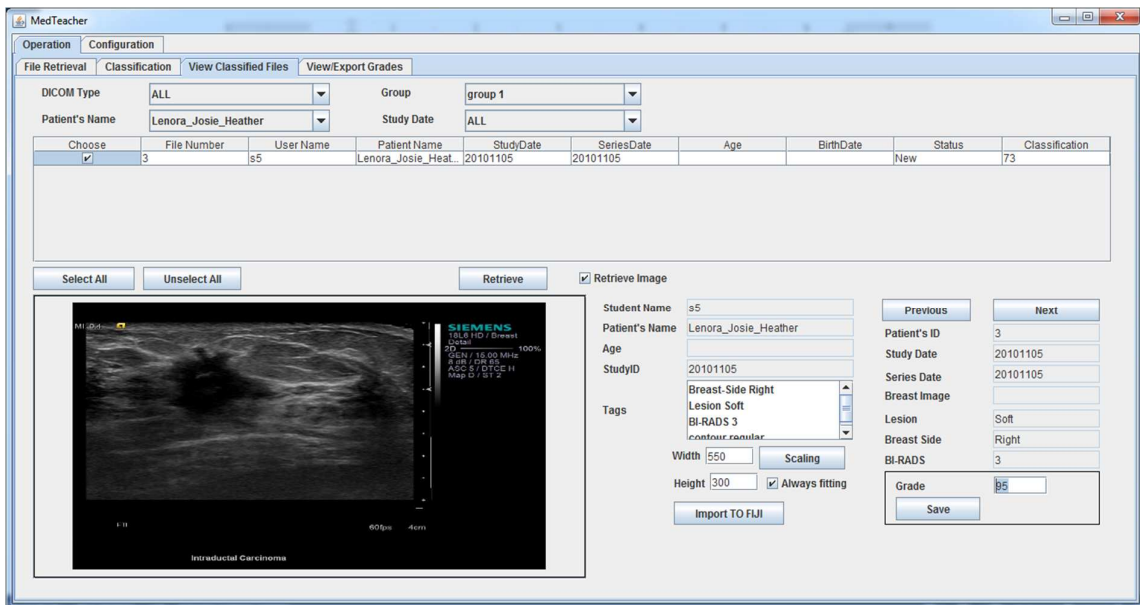


Figure 6.23 : Review the classification of students and put grades

You are able to export the grades of the students to csv files. To do that, click on "View/Export Grades" tab, choose your criteria from the appeared combo boxes, and click on "Export to CSV file" to export the results as shown in Figure 6.24. The system will calculate automatically the total grades for each student.

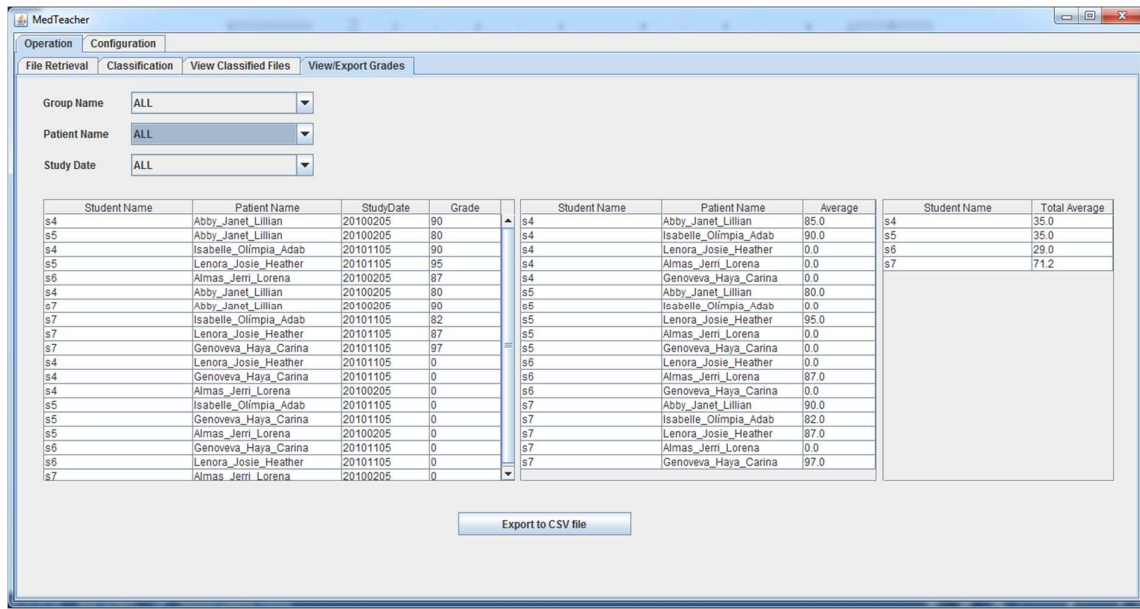


Figure 6.24 : Export students' grades to csv files

6.3.2.2.3 Create a group of students, DICOM files, or Reports

Please review the previous parts of the main user to see how to create a group of students, DICOM files, reports.

6.3.2.3 Student

The student's application has the same interfaces as the teacher's application except in the "configuration" tab, and the interfaces for creating groups and assigning groups are not exist. Besides, this application only allows a student to view grades of the classifications.

6.3.2.4 Doctor at hospital

A doctor at the hospital has special application with the same interface as the main user application except in the "configuration" tab, and the interfaces for creating groups, assigning groups, and upload DICOM files are not exist. Also this application does not have interface for loading one-word names to the system. Please review the usage instructions of the main user application to avoid repetition in this tutorial.

6.3.2.5 Researcher

The researcher application supports a user with two mechanisms for annotating DICOM files: annotate DICOM files directly without log in to the system, or annotate and classify files with login to the system.

6.3.2.5.1 Annotate DICOM files without login

FIJI allows to annotate a DICOM image by clicking on “File” menu, “open” command, choose a destination DICOM file, draw any shape you want, click on ctrl&d on the keyboard after drawing, go to the “plugins” menu, select “DICOM Save” from “MedResearcher” plugin to keep modifications on the DICOM file. You can use “DICOM Save as” plugin to save the DICOM file with a different name. If you want to apply a group of commands or plugins to a group of DICOM files, “Batch Processor” plugin will be used. Click on “Batch Processor” from “MedResearcher” plugin, enters a source path of DICOM files in “input” field and a destination path for the output DICOM files in “output” field, choose “DCM” item from “Output Format” combo box, add any macro command from “Add Macro Code” combo box, or enter the name of a plugin or a command in the text area manually, and then click on “Process” button as shown in Figure 6.25.

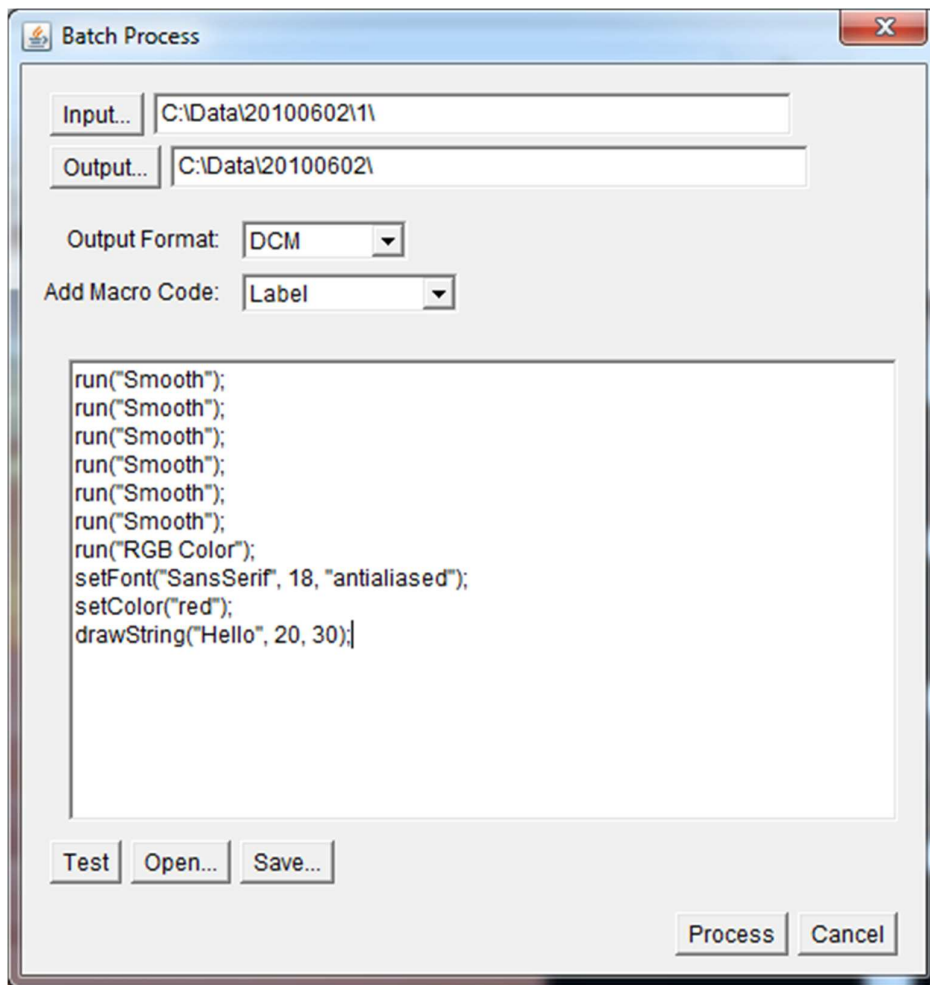


Figure 6.25 : Modified batch processor to apply a group of plugins on DICOM files.

6.3.2.5.2 Create a Macro file with one or more Plugins

A new macro file can be created in FIJI by going to “Plugins” menu, select “Macro” item from “New” sub-menu, write your commands or plugins, and save this file in “*.ijm” format. This file can be used later as a group of plugins to be applied to a group of DICOM files. You can record a group of plugins and save them into a macro file by clicking on “Plugins” menu, click on “Macros” sub-menu, and click on “Record” item to open new window to record all plugins in. To record one or more plugins, you need to open a DICOM file and click on the desired plugins to be applied to the current DICOM file such as “Smooth”, “Sharpen”, “Find Edges”, etc. When you click on any plugin, FIJI will write this plugin into the window of recording. After that, you need to keep all the recorded plugins to a macro file in the window of recording by clicking on “Create” button.

6.3.2.5.3 load a group of parameters of DICOM files from the system

Normally, a user should log in to the system to get a group of DICOM files by clicking on “Researcher login” plugin in FIJI plugins menu. The system asks for a user name, password, and IP. New interface will be appeared as shown in Figure 6.26. The user can select one or more criteria from “Tag”, “Value”, “DICOM Type” combo boxes to be used as the criteria of searching, click on “Add” button to add the criteria to the first table, enter the maximum number of files’ parameters that should be retrieved in “Number of Files” text field, and click on “Update Table” to get a set of files’ parameters in the second table. This criteria can be saved to be used later by entering a new value in “Criteria Name” and clicking on “Save Criteria”.

It is better to save the retrieved files parameters in a list if you want use them many times. To do that, enter a new name in “list name” field and click on “save list” button.

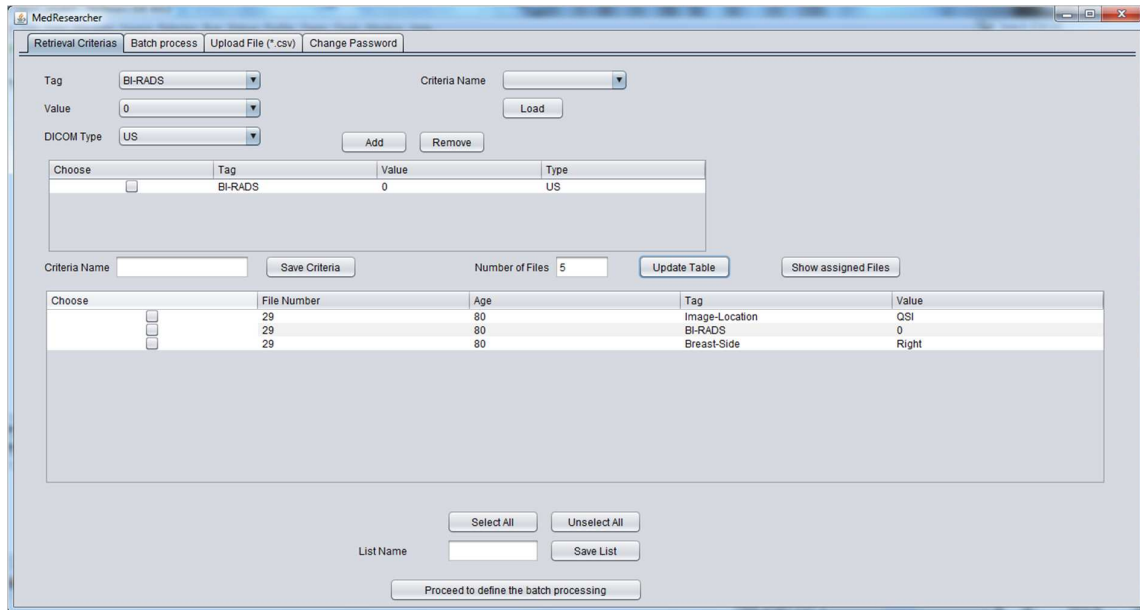


Figure 6.26: Retrieve a group of DICOM files parameters.

6.3.2.5.4 Apply plugins to DICOM files

A group of plugins can be applied to the loaded files’ parameters in the second table in Figure 6.26 by selecting one or more files’ parameters, click on “Proceed to define the batch processing” button to active “Batch Process” tab. And then, select “Selected Criteria Files” item from “Files list Name” combo box if you want to load the selected files’ parameters in the second table in “Retrieval criteria” tab, or you can select another item from “Files list Name” combo box to load an old list. After that, you have two options: apply plugins locally or remotely using FIJI tool, or apply external methods using other tools such as Matlab.

First, to retrieve the DICOM files to your local machine, select one or more file numbers from the first table in Figure 6.27 and click “Retrieve files locally” button. Then, click on “internal method” radio box, click on “Open macro File” to load plugins from a macro file into the method table or into the second table in Figure 6.27, click on ”Apply Methods locally” button to implement these plugins to the retrieved files in your machine. To send the modified DICOM files, click on “Send results to server” button.

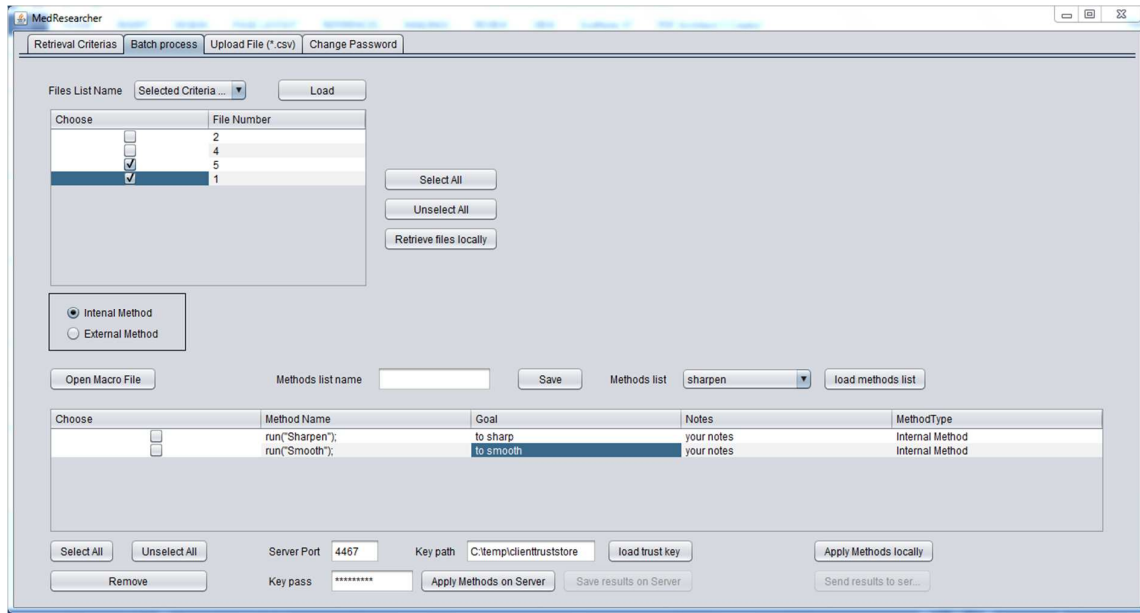


Figure 6.27: Apply a group of plugins to a group of DICOM files.

For retrieving and applying a group of plugins on the server, you will do the same as applying plugins locally, but in this time, you need to enter a port that server is listening on, load a trusted key from a directory, enter the password of the trusted key, click on “Apply Methods on Server”, and click on “Save results on Server” button.

The second option is to use other tools such as Matlab or OpenCV to annotate DICOM images if it is necessary and send the derived DICOM files back to dcm4chee. In this work, only Matlab is applied to deal with DICOM files. Since Matlab is used by different areas, new algorithms are added by different authors to detect breast lesions. In this case, a user will export a group of DICOM files from this system to be annotated by Matlab. The following is an example to show how to read a DICOM file, black out pixels from 0-100 rows of a DICOM image, and write the result to a new DICOM file:

```
img = dicomread('path/filename.dcm');
metadata = dicominfo('path/filename.dcm');
% You can put your algorithm in place of for loop.
[rows,columns,colors,frames]=size(img);
for i=1:frames
    img(1:100, :, :, i)=0;
end
dicomwrite(img, 'path/modifieddicom.dcm', metadata, 'CreateMode', 'copy');
```

Actually, the system will create a folder called “Med_Temp_Folder” to put the retrieved DICOM files from server in. This folder will be created in the temporary folder at your operating system. After DICOM files had been retrieved locally, you should copy these files into a folder and apply the Matlab code as mentioned above to generate derived DICOM files. You can modify this code as you want and apply any algorithm to detect breast lesion or any other purposes. The above Matlab code is an example to show a developer how to read and modify and write the data to DICOM files. A developer can replace ‘for loop’ in the previous code with any algorithm to do annotation, detect breast lesions, etc. To send the derived DICOM files, a csv file must be created, which contents paths of the original DICOM files and the new ones, and classification results for each DICOM file. For example, each line in CSV file should have a form like “filepath=path_to_original_DICOM, derivedfilepath=path_to_derived_DICOM, tag=tag_name, value=tag_value”. Where “path_to_original_DICOM” is a path to original DICOM file, “path_to_derived_DICOM” is a path to derived DICOM file, “tag_name” is a general name for procedure or method that was applied to get the derived DICOM file, or it can be a classification result such as lesion, contour, BI-RADS, etc. While, “tag_value” is a value of “tag_name” such as parameter for method or procedure, expected values for lesion, BI-RADS score, etc. After CSV file is generated, go to “Upload Files (*.CSV)” tab in the researcher application, click on “open” button, choose the path of the generated CSV file to load the file’s lines into the classification table in Figure 6.28, and click on “Send DICOM files with classification” to send the classification results.

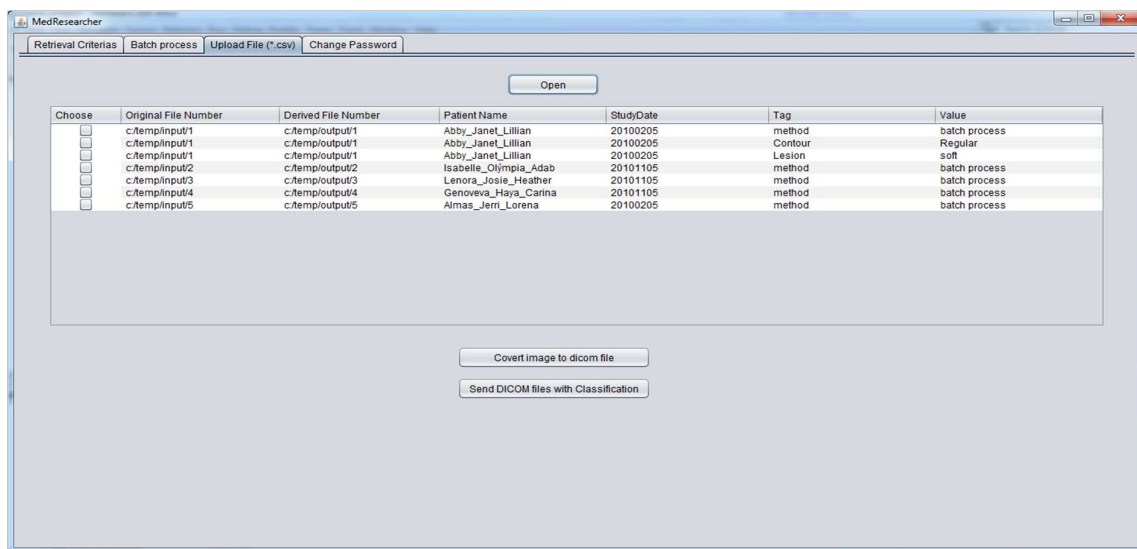


Figure 6.28: Read a csv file and send its lines to PACS server.

Bibliography

- American Cancer Society, Inc. (2013). *Do we know what causes breast cancer?* Retrieved April 25, 2015, from cancer.org website:
<http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-what-causes>
- American Cancer Society, Inc. (2013). *Types of breast cancers.* Retrieved April 25, 2015, from cancer.org website: <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-breast-cancer-types>
- American Cancer Society, Inc. (2013). *What are the risk factors for breast cancer?* Retrieved April 25, 2015, from cancer.org website:
<http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-risk-factors>
- Analogic. (2015). *Elastography Imaging for Ultrasound | Analogic Ultrasound.* (Analogic) Retrieved April 25, 2015, from BK Ultrasound website:
<http://www.analogicultrasound.com/ultrasonix/options/elastography>
- Apache Software Foundation. (2002-2014). *Welcome to Apache Maven.* Retrieved April 25, 2015, from Apache Maven website: <http://maven.apache.org/>
- Apteryx. (2014). *Dicom Image I/O Plugin.* Retrieved April 25, 2015, from Apteryx website:
<http://imaging.apteryx.fr/dicom/index>
- Archie, K. A. (2012). DicomBrowser: Software for Viewing and Modifying DICOM Metadata. *Journal of digital imaging, 25(5)*, 635-645. Retrieved April 25, 2015
- Bidgood, W. D., Horii, S. C., Prior, F. W., & Van Syckle, D. E. (1997). Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging. *Journal of the American Medical Informatics Association, 4(3)*, 199-212. Retrieved April 25, 2014, from Ncbi Web site:
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC61235/>
- Bradski, G. a. (2008). *Learning OpenCV: Computer vision with the OpenCV library.* US: O'Reilly Media, Inc. Retrieved April 25, 2015
- Bui, A. A. (2007). OpenSourcePACS: an extensible infrastructure for medical image management. *Information Technology in Biomedicine, IEEE Transactions on 11.1*, 94-109. Retrieved April 25, 2015
- Chetan Uberoy. (2010). *Dcmqr - dcm4che-2.x - Confluence.* Retrieved April 25, 2015, from dcm4che Web site: <http://www.dcm4che.org/confluence/display/d2/dcmqr>
- ClearCanvas Inc. (2013). *ClearCanvas Workstation, Personal - User's Guide.* Retrieved April 25, 2015, from Clearcanvas Web site:
http://www.clearcanvas.ca/Portals/0/ClearCanvasFiles/Documentation/UsersGuide/Personal/7_1/index.html?study_browser_tools.htm
- ClearCanvas Inc. (2009). *ClearCanvas ImageServer User's Guide.* Retrieved April 25, 2015, from ClearCanvas website:

- http://www.clearcanvas.ca/Portals/0/ClearCanvasFiles/Documentation/UsersGuide/ImageServer/1_5/index.html?users.htm
- Cooper, J. (2009). *Anonymize IJ DICOM*. Retrieved April 25, 2015, from imagej:
<http://imagej.nih.gov/ij/plugins/anonymize-ij-dicom/index.html>
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (Vol. 2). MIT press Cambridge. Retrieved from https://books.google.pt/books?id=NLngYyWFI_YC
- Costa, C. e. (2011). Dicoogle-an open source peer-to-peer PACS. *Journal of digital imaging* 24.5, 848-856. Retrieved April 25, 2015
- Enthought, Inc. (2013). *Downloads - Enthought Scientific Computing Solutions*. Retrieved April 25, 2015, from <https://www.enthought.com/downloads/>
- Evans, D. (2006). *Dcmrcv - dcm4che-2.x - Confluence*. Retrieved April 25, 2015, from dcm4che Web site: <http://www.dcm4che.org/confluence/display/d2/dcmrcv>
- Evans, D. (2008). *A Very Basic DICOM Introduction - dcm4che-2.x - Confluence*. Retrieved April 25, 2015, from dcm4che Web site:
<http://www.dcm4che.org/confluence/display/d2/A+Very+Basic+DICOM+Introduction>
- Evans, D. (2008). *Dcmsnd - dcm4che-2.x - Confluence*. Retrieved April 25, 2015, from dcm4che Web site: <http://www.dcm4che.org/confluence/display/d2/dcmsnd>
- Evans, D. (2008). *System Overview - dcm4chee-2.x - Confluence*. Retrieved April 25, 2015, from dcm4che Web site: <http://www.dcm4che.org/confluence/display/ee2/System+Overview>
- Evans, D. (2012). *Home - dcm4chee-2.x - Confluence*. Retrieved June 26, 2015, from dcm4che.org:
<http://www.dcm4che.org/confluence/display/ee2/Home>
- FIPS, P. (2002, April 8). SECURE HASH STANDARD. *National Institute of Standards and Technology*, 180-2. Retrieved April 25, 2015, from nist.gov: <http://csrc.nist.gov/publications/fips/fips180-2/fips180-2.pdf>
- Gray, H. (2014). *XI. Splanchnology. 3d. 6. The Mammae; Gray, Henry. 1918. Anatomy of the Human Body*. Retrieved April 25, 2015, from Bartleby Web site:
<http://www.bartleby.com/107/271.html>
- Hall, M. C. (2012). *Aesthetic Surgery*. lulu.com. Retrieved April 25, 2015
- Hermen, J. C. (2009). Solving annoying dicom problems using imagej and the tudor dicom tools. *Global Telemedicine and eHealth Updates: Knowledge Resources*. Retrieved April 25, 2015
- Hermena, J. C. (2008). Managing and viewing DICOM images with ImageJ. *Proceedings of the ImageJ user and developer conference*. Retrieved April 25, 2015
- Itseez. (2013). *OpenCV | OpenCV*. Retrieved April 25, 2015, from opencv Web site:
<http://opencv.org/>

- Jodogne, S. e. (2013). Orthanc-A lightweight, RESTful DICOM server for healthcare and medical research. *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. Retrieved April 25, 2015
- Leighton, T. G. (2007, August 15). What is ultrasound? *Progress in biophysics and molecular biology, 93(1)*, 3-83. Retrieved April 25, 2015
- M.D., P. S. (2007). *HOW TO DEPLOY A WEB PACS IN MINUTES*. Retrieved April 25, 2015, from Cdmedic Web site: <http://cdmedicpacsweb.sourceforge.net>
- Marcel van Herk, L. Z. (2013). *Conquest DICOM software*. Retrieved April 25, 2015, from <http://ingenium.home.xs4all.nl/dicom.html>
- Mason, D. (2012). *Getting Started with pydicom — pydicom 1.0a documentation*. Retrieved June 26, 2015, from readthedocs.org: http://pydicom.readthedocs.org/en/latest/getting_started.html
- MathWorks, Inc. (2015). *DICOM Support in the Image Processing Toolbox*. Retrieved April 25, 2015, from MathWorks Web site: <http://www.mathworks.com/help/images/dicom-support-in-the-image-processing-toolbox.html>
- MathWorld Team. (2014). *One-Way Function*. Retrieved April 25, 2015, from wolfram: <http://mathworld.wolfram.com/One-WayFunction.html>
- Medfloss.org. (2014). *Open Source Picture Archiving and Communication System (OSPACS)*. Retrieved April 25, 2015, from Medfloss website: <http://www.medfloss.org/node/115>
- MediaWIKI. (2013). *Fiji is Just ImageJ*. Retrieved April 25, 2015, from Fiji Web site: <http://fiji.sc/wiki/index.php/Fiji>
- MediaWiki. (2015, April 24). *Batch Processing - Fiji*. Retrieved from <http://fiji.sc/>: http://fiji.sc/Batch_Processing
- Medical Connections Ltd. (2014). *Photometric Interpretations*. Retrieved April 25, 2015, from Medicalconnections Web site: https://www.medicalconnections.co.uk/kb/Photometric_Interpretations
- Medical Connections Ltd. (2014). *Transfer Syntax*. Retrieved April 25, 2015, from Medicalconnections Web site: https://www.medicalconnections.co.uk/kb/Transfer_Syntax
- Morate, D. G. (2012, Feb 07). *Open Source Clinical Image and Object Management*. Retrieved April 25, 2015, from dcm4che Web site: <http://www.dcm4che.org/confluence/display/ee2/Home>
- Mwoodbri. (2014). *MRIdb Reference Guide*. Retrieved April 25, 2015, from GitHub Web site: <https://github.com/mwoodbri/MRIdb/wiki/MRIdb-Reference-Guide>
- National Institutes of Health. (2014). *Breast Cancer Prevention - National Cancer Institute*. Retrieved April 25, 2015, from National Cancer Institute website: http://www.cancer.gov/cancertopics/pdq/prevention/breast/Patient/page2#figure_143_e

- National Library of Medicine. (2013). *Mammography: MedlinePlus Medical Encyclopedia*. Retrieved April 25, 2015, from National Institutes of Health's Web site:
<http://www.nlm.nih.gov/medlineplus/ency/article/003380.htm>
- NEMA. (2013). *7 DICOM File Format*. Retrieved April 25, 2015, from Medical Nema Web site:
http://medical.nema.org/dicom/2013/output/chtml/part10/chapter_7.html
- NEMA. (2014). *Data Structures and Encoding*. NEMA Publications. Retrieved April 25, 2015, from medical.nema Web site:
<ftp://medical.nema.org/medical/dicom/2014a/output/pdf/part05.pdf>
- NEMA. (2015, April 24). *DICOM Homepage*. Retrieved from [nema.org](http://dicom.nema.org/): <http://dicom.nema.org/>
- O'Dell, W. (2003). *DICOM Rewriter*. Retrieved April 23, 2015, from ImageJ:
<http://rsb.info.nih.gov/ij/plugins/dicom-rewriter.html>
- OFFIS. (2014). *dicom.offis.de - DICOM Software made by OFFIS - DCMTK - DICOM Toolkit*. Retrieved April 25, 2015, from <http://dicom.offis.de/dcmtk.php.en>
- Oracle Corporation. (2014). *General Security Issues*. Retrieved April 25, 2015, from Mysql Web site:
<http://dev.mysql.com/doc/refman/5.6/en/general-security-issues.html>
- Oracle Corporation. (2015). *MySQL :: The world's most popular open source database*. Retrieved April 24, 2015, from [mysql.com](http://www.mysql.com/): <http://www.mysql.com/>
- PixelMed. (2015). *Software, tools and images*. Retrieved April 25, 2015, from PixelMed Publishing Web site: <http://www.pixelmed.com/index.html#PixelMedJavaDICOMToolkit>
- Python Software Foundation. (2013). *The Python Tutorial*. Retrieved April 25, 2015, from Python Web site: <http://docs.python.org/2/tutorial/>
- Radiological Society of North America, Inc. (2013). *Breast Cancer Screening*. Retrieved April 25, 2015, from Radiologyinfo Web site: <http://www.radiologyinfo.org/en/info.cfm?pg=screening-breast>
- Rasband, W. S. (2012). ImageJ. *US National Institutes of Health, Bethesda, MD, USA, 1997–2012*. Retrieved April 25, 2015
- Schöch, C. i. (2014). *Diploma thesis, List of DICOM Toolkits*. Retrieved April 25, 2015, from University of Lübeck Web site: <http://www.schoech.de/diploma/toolkits.html#cpp>
- Slashdot Media. (2014). *DICOM Anonymizer*. Retrieved April 25, 2015, from Sourceforge Web site:
<http://sourceforge.net/projects/dicom-anonymizer/>
- Stott, W. e. (2008). Source Code for Biology and Medicine. *Source code for biology and medicine*, 3, 11.
- Synaptive Medical. (2014). *ClearCanvas*. Retrieved April 25, 2015, from <http://sourceforge.net/projects/clearcanvas/>

- Synaptive Medical. (2014). *ClearCanvas*. Retrieved April 25, 2015, from Clearcanvas Web site:
<http://www.clearcanvas.ca/Home/tabid/421/Default.aspx>
- The Johns Hopkins University. (2014). *Breast Imaging, Reporting*. Retrieved April 25, 2015, from Hopkinsmedicine Web site:
http://www.hopkinsmedicine.org/avon_foundation_breast_center/treatments_services/breast_cancer_screening/digital_mammography/breast_imaging_reporting_data_system.html
- Tutorialspoint. (2013). *Java – Overview*. Retrieved April 25, 2015, from tutorialspoint Web site:
http://www.tutorialspoint.com/java/java_overview.htm
- Tutorialspoint. (2015). *MATLAB Overview*. Retrieved April 25, 2015, from Tutorialspoint Web site:
http://www.tutorialspoint.com/matlab/matlab_overview.htm
- Tutorialspoint. (2015). *MySQL Introduction*. Retrieved April 24, 2015, from tutorialspoint.com:
<http://www.tutorialspoint.com/mysql/mysql-introduction.htm>
- UC Davis Health System. (2014). *De-identification*. Retrieved April 23, 2015, from UC Davis Health System website: <http://www.ucdmc.ucdavis.edu/compliance/guidance/privacy/deident.html>
- University of Connection. (2014). *Breast Anatomy*. (University of connection health center) Retrieved April 25, 2015, from fitsweb Web site:
http://fitsweb.uhc.edu/student/selectives/Luzietti/Breast_anatomy.htm
- WillStott. (2015, June 27). *OSPACS - Home*. Retrieved from codeplex.com:
<http://ospacs.codeplex.com/documentation>
- Woodbridge, M. G. (2013). MRIdb: Medical Image Management for Biobank Research. *Journal of digital imaging*, 886-890. Retrieved April 25, 2015
- Zoratti, I. (2006). MYSQL security best practices. In Crime and Security. *The Institution of Engineering and Technology Conference on., IET*, 183-198. Retrieved April 25, 2015