**DOKUZ EYLÜL UNIVERSITY**
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# ONTOLOGY-BASED MEDICAL IMAGE ANNOTATION AND RETRIEVAL

**by**
**Hakan BULU**

**September, 2013**
**İZMİR**

# ONTOLOGY-BASED MEDICAL IMAGE ANNOTATION AND RETRIEVAL

**A Thesis Submitted to the**
**Graduate School of Natural and Applied Sciences of Dokuz Eylül University**
**In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Computer Engineering**

**by**
**Hakan BULU**

**September, 2013**
**İZMİR**

**Ph.D. THESIS EXAMINATION RESULT FORM**

We have read the thesis entitled **"ONTOLOGY-BASED MEDICAL IMAGE ANNOTATION AND RETRIEVAL"** completed by **HAKAN BULU** under supervision of **ASSOC. PROF. DR. ADİL ALPKOÇAK** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. Adil ALPKOÇAK

Supervisor

Asst. Prof. Dr. Şen ÇAKIR

Thesis Committee Member

Prof. Dr. Cüneyt GÜZELİŞ

Thesis Committee Member

Prof. Dr. Alp KUT

Examining Committee Member

Prof. Dr. Utku ŞENOL

Examining Committee Member

Prof.Dr. Ayşe OKUR

Director

Graduate School of Natural and Applied Sciences

## ACKNOWLEDGEMENTS

First, I want to thank my advisor Associate Professor Dr. Adil ALPKOÇAK, for his valuable suggestions, encouragement, patience and support during this study. His academic perspective has very strongly impacted my academic career in all aspects. It was a great privilege for me to work with him.

I would like to thank my thesis tracking committee members Professor Dr. Cüneyt GÜZELİŞ and Assistant Professor Dr. H. Şen ÇAKIR for their contributions to this study and sharing their ideas during the development and writing of this thesis.

I also thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting the development of this thesis under project number 107E217. I would also like to acknowledge project team members Associate Professor Dr. Adil ALPKOÇAK, Professor Dr. Pınar BALCI, Professor Dr. Cüneyt GÜZELİŞ, Professor Dr. Oğuz DİCLE, Research Assistant Tolga BERBER, Research Assistant Ozan AKÇAY for their collaborative work and contributions to this thesis.

Also, I would like to thank Dr. Daniel Rubin from Stanford University for giving me a chance to study at Stanford University for six months. I appreciate all his contributions of time, ideas and funding to make my researches.

I want to give special thanks to my spouse Semra Serbest BULU, my parents Neziha and Bircihan Mustafa BULU, Suada and Aziz SERBEST for their endless support, patience and encouragement during preparation of this thesis.

Finally, I want to thank my angel Eda BULU and my coming angel ... BULU. Their presence give me the strength to work harder through my Ph.D. education.

Hakan BULU

**ONTOLOGY-BASED MEDICAL IMAGE ANNOTATION AND RETRIEVAL**

## ABSTRACT

In this thesis, we proposed a new ontology-based medical image annotation and retrieval system for mammographic examinations. For that purpose, we have first developed a mammography annotation ontology (MAO) which is a domain ontology and it provides shared vocabulary for mammography interpretation. Then we have developed a new ontology-based mammography annotation and retrieval tool (MART) to create our mammography dataset. Then, we have developed a content based image retrieval system where a breast mass is described with three sets of features: low, mid and high-level feature. Mathematical model of similarity calculation between two breast lesions and implementation of the model with SQWRL and XQuery explained in detail. To test our CBIR system, we performed set of queries on the DEMS. Furthermore, we present an approach to model uncertainty in mammography, and perform SQWRL rules to infer BI-RADS scores for a given mass instance. Experimentation results showed that uncertainty exists in interpretation of BI-RADS scoring in mammography and average level of uncertainty for crisp logic is clearly greater than our approach. Additionally, we show that using low-level features together with high and mid-level features in the content based image retrieval of breast masses improves the overall system performance and it is found statistically significant ($p$ is lower than 0.001).

**Keywords:** Ontology, content-based image retrieval, low-level image features, breast mass, medical image retrieval, mammography, uncertainty.

# ONTOLOJİ TABANLI TIBBİ GÖRÜNTÜ BETİMLEME VE ERİŞİMİ

## ÖZ

Bu tez kapsamında, mamografi incelemelerinde kullanılmak üzere yeni bir ontoloji tabanlı tıbbi resim betimleme ve geri getirim sistemi önerilmiştir. Bu amaçla ilk olarak mamografi incelemelerde kullanılmak üzere ortak bir kelime haznesi sağlayan yeni bir mamografi betimleme ontolojisi geliştirdik. Daha sonrasında, veri setimizi oluşturmak amacıyla, ontolojimiz ile uyumlu, yeni bir ontoloji tabanlı mamografi betimleme ve geri getirim uygulaması geliştirdik. Sonrasında, her bir meme kitlesinin üç farklı seviyede öznitelik (yüksek, orta ve düşük) ile temsil edildiği içerik tabanlı resim geri getirim modelimizi geliştirdik. İlgili modelin matematiksel modeli SQWRL ve XQuery kullanılarak uygulamaya geçirilmesine ilişkin detaylar tez içersinde verilmiştir. İçerik tabanlı resim geri getirim sistemimizi test etmek amacıyla bir grup sorguyu veri setimiz üzerinde çalıştırdık. Ayrıca, mamografi incelemeleri sırasında ortaya çıkabilen belirsiz durumları modellemek üzere yeni bir yaklaşım önerdik ve verilen bir meme kitlesinin BI-RADS skorunu belirmek için SQWRL kuralları geliştirdik. Yapılan deneyler sonucunda, mamografi incelemeleri sırasında BI-RADS skorlarının belirlenmesi aşamasında bir belirsizlik durumunun olduğu ve formüle edilen belirsizlik seviyesinin kesin mantık için bizim yaklaşımımızdan açık bir şekilde daha yüksek olduğu görülmüştür. Ek olarak, içerik tabanlı resim geri getirim sistemlerinde, düşük seviyeli özniteliklerin, yüksek ve orta seviyeli öznitelikler ile birlikte kullanılması, sistem performansını iyileştirmiştir. Bu iyileştirme istatistiksel olarak anlamlı bulunmuştur ($p$ küçüktür 0.001).

**Anahtar Sözcükler:** Ontoloji, içerik tabanlı resim geri getirim, düşük seviyeli resim öznitelikleri, meme kitlesi, tıbbi resim geri getirim, mamografi, belirsizlik.

**CONTENTS**

**Page**

## LIST OF FIGURES

**Page**

## LIST OF TABLES

**Page**

# CHAPTER ONE
# INTRODUCTION

## 1.1 Overview

Breast cancer is the most common tumor for the women, in Western countries. Some statistics of breast cancer shows that nearly 1 in 8 women in the United States will develop invasive breast cancer over their lifetime (Breastcancer.org, 2013). But, breast cancer is most treatable when it is early detected. In this sense, a mammography examination, called a mammogram, is the gold standard for breast cancer screening, early detection and diagnosis. Mammography is a specific type of imaging that uses a low dose x-ray system to examine breasts. Mammograms can help to detect up to 90% of breast cancers, even before they are felt like a lump (Stephan, 2013). The American Cancer Society recommends that women 40 years old and older have an annual mammogram. Therefore, many researchers have been working on computer-aided diagnosis system (CADx) to detect and identify breast masses automatically in digital mammograms over several decades. All these researches aim to support radiologists in the difficult task of discriminating benign and malignant breast lesions. Hence, it is not surprising that typically only 15% to 30% of breast biopsies performed on calcifications will be positive for malignancy (Hall et al., 1988). To improve the level of CADx in mammography, there is a need to a system taking the background knowledge of radiologist into account in decision-making process with a more computable way. In this point, ontologies can be a solution to improve the performance of CADx systems in Mammography.

Ontology is the most common way to represent the knowledge for computers, and defined as a formal, explicit specification of a shared conceptualization and encodes a partial view of the world, with respect to a given domain. It is composed of a set of concepts, their definitions and their relations that can be used to describe and reason about a domain. Ontological modeling of knowledge is vital in many real world applications and in medicine. In intelligent systems, ontologies are the way to transform background knowledge of a domain to machine understandable form. For

example, the interpretation of radiological examinations includes years of experience, the knowledge on the respective domain. The medical image interpretation is not solely reached by pattern recognition and it also includes a deep knowledge in medical domain. Therefore, a successful implementation of radiological imaging system should be able to model and incorporate such knowledge into a more computable format. In this point, ontology is a tool to be able to solve this issue. Medical ontologies are developed to solve problems such as reusing and sharing of patient data, required of semantic-based queries/inference or the transmission of these data. The communication of complex and detailed medical concepts is a very important task in current medical information systems. In this way, more complex tools such as case-based retrieval or evidence based medicine can be possible in medicine.

Radiology department of an average hospital may produce hundreds of mammograms per day. Thus, annotation and retrieval of mammographic examinations in an acceptable time is important for right diagnosis. In this respect, Hung and Chen propose a Case based Retrieval (CBR) system for mammographic cases (Hung & Chen, 2006). On the other hand, in recent years, many researches aim to develop ontology-based medical image annotation and retrieval approach to reduce the occurrence of irrelevant resource retrieval in a medical imaging information system. The main goal is to answer the user queries based on semantic relations that can be inferred from meaningfully between the data items. Hu et al. built a semantically rich system by accommodating image annotation and retrieval services around a rigidly defined ontology for medical images used in breast cancer treatment, in 2003. The aim of the their Breast Cancer Imaging Ontology (BCIO) is to provide a commonly agreed vocabulary with formal definitions that can be used to represent breast X-ray and MRI images, abnormal findings and medical assessments (Hu et al., 2003). In 2006, Qi et al. developed a mammography ontology called as Pocket-Ontology. They use ontology-based comparison method for finding groups of diagnosis that radiologists detect using the same analysis process. Their comparison method is based on an edit distance, which is a similarity measurement between two concepts (Qi et al., 2006). Ren and Barnaghi created a framework for medical specialists to be able to annotate digital mammograms, and to retrieve relevant

resources based on semantic relations, in 2007 (Ren & Barnaghi, 2007). . In 2008, Rubin et al. develop an ontology-based annotation and retrieval framework, which is called Annotation and Imaging Markup (AIM) (Rubin et al., 2008). Levy et al. perform a SWRL rule on AIM to identify the malignancy of liver lesions, depend on its length (Levy et al., 2009). Shanbolt et al. developed an ontology-based knowledge management system which is called MIAKT (Medical Imaging with Advanced Knowledge Technologies) for the data that the screening process generates, as well as providing a means for medical staff to investigate, annotate and analyze the using web, in 2004 (Shadbolt et al, 2004).

## 1.2 Aim of This Thesis

Aim of this thesis is to develop ontology-based content based image retrieval system for breast masses. Hence, a successful implementation of radiological imaging system could be able to model and incorporate such knowledge into a more computable format. In this way, more complex tools such as case-based retrieval or evidence-based medicine can be possible in mammography. In order to achieve this goal, we propose several improvements; …iyileştirmelerin neler olduğunu yazmak lazım.

## 1.3 Thesis Organization

This thesis is organized as follows. In chapter 2, we present our Mammography Annotation Ontology (MAO) and Mammography Annotation Retrieval Tool (MART). In chapter 3, we propose a sample mammogram dataset (DEMS: Dokuz Eylul University Mammogram Set), which is fully annotated with the MART. Chapter 4 introduces mathematical model of our CBIR system for digital mammograms and figure out the performance effect of different level of features in the system. In chapter 5, we propose a new ontology-based mammography annotation system with a capability of uncertainty modeling in ontologies. Implementation of our ontology-based CBIR system with XQuery is given in chapter 6. Finally, chapter 7 concludes this thesis and provides future direction.

# CHAPTER TWO
# MAMMOGRAPHY ONTOLOGY WITH ANNOTATION AND RETRIEVAL TOOL

## 2.1 Overview

Ontology is the most common way to represent the knowledge for computers, and defined as a formal, explicit specification of a shared conceptualization and encodes a partial view of the world, with respect to a given domain. It is composed of a set of concepts, their definitions and their relations that can be used to describe and reason about a domain. Ontological modeling of knowledge is vital in many real world applications and in medicine. In intelligent systems, ontologies are way to transform background knowledge of a domain to machine understandable form. For example, the interpretation of radiological cases includes years of experience, the knowledge on the respective domain. The medical image interpretation is not solely reached by pattern recognition and it also includes a deep knowledge in medical domain. Therefore, a successful implementation of radiological imaging system should be able to model and incorporate such knowledge into a more computable format. In this point, ontology is a tool to be able to solve this issue. Medical ontologies are developed to solve problems such as reusing and sharing of patient data, required of semantic-based queries/inference or the transmission of these data. The communication of complex and detailed medical concepts is a very important task in current medical information systems. In this way, more complex tools such as case-based retrieval or evidence-based medicine can be possible in medicine.

In this chapter, we present mammography annotation ontology (MAO), Mammography Annotation Retrieval Tool (MART). MAO is a domain ontology for mammography and it was created based on the 3$^{th}$ edition of ACR (American College of Radiologists) BI-RADS (Breast Imaging Reporting and Data System) Mammography Atlas (The American College of Radiology, 2012). MART is a software tool to annotate and retrieve mammographic examinations based on MAO.

## 2.2 Mammography Annotation Ontology (MAO)

In terms of computer science, ontologies are state-of-the-art method to represent knowledge and become more important in image annotation. Ontology includes a set of concepts and the relationships between them. We divide ontologies into two main groups; *upper ontology* and *domain ontology*. Upper ontologies model the common objects, which are generally used in the domain ontologies while the domain ontologies model a specific domain or part of the world. Domain ontologies generally provide a shared vocabulary. Main role of these vocabularies is to help data integration by representing the knowledge and to aid decision-making processes. In that respect, ontologies are important for health care systems.

In this study, Mammography Annotation Ontology (MAO) is an essential part of the system. In development of MAO we used the 3$^{rd}$ edition of BI-RADS Mammography Atlas, and used the ontology to annotate any abnormality observed in mammograms. However, some mammograms may not contain any abnormalities. Principally, MAO provides a shared vocabulary and knowledge that makes annotations understandable and computable by the computer. Prominently, it makes reasoning of any other information possible.

In literature, some research suggests a framework for ontology-based medical image annotation and retrieval as an approach to reduce the occurrence of irrelevant resource retrieval in a medical imaging information system. Hu et al. built a semantically rich system by accommodating image annotation and retrieval services around a rigidly defined ontology for medical images used in breast cancer treatment, in 2003. They developed the Breast Cancer Imaging Ontology (BCIO) to provide a commonly agreed vocabulary with formal definitions that can be used to represent breast X-ray and MRI images, abnormal findings and medical assessments (Hu et al., 2003). In 2006, Qi et al. developed mammography ontology and used an ontology-based comparison method for finding groups of diagnosis that radiologists detect using the same analysis process based on edit distance, which is a similarity measurement between two concepts (Qi et al., 2006). Ren and Barnaghi suggested a framework for medical specialists to be able to annotate digital mammograms and to

retrieve relevant resources based on semantic relations, in 2007 (Ren & Barnaghi, 2007). In 2008, Rubin et al. developed a generalized ontology-based annotation and retrieval framework, which is called Annotation and Imaging Markup (AIM) (Rubin et al., 2008). Shanbolt et al. developed an ontology-based knowledge management system which is called MIAKT (Medical Imaging with Advanced Knowledge Technologies) for the data that the screening process generates, as well as providing a means for medical staff to investigate, annotate and analyze the using web, in 2004 (Shadbolt et al., 2004). And, in 2012, we have proposed a system for Ontology-based annotation and retrieval of breast masses (Bulu et al., 2012).

Ontology development is an iterative process and there is no one best way or methodology to develop ontologies. In development process of the MAO, we consider the domain covered with intended use of the ontology. We use middle-out strategy as ontology development methodology (Fernández-López, 1999). To achieve this, we choose the base concepts in mammography (i.e., Case, Breast, Image, Abnormality etc.) and some of their basic relationships. Then, we describe the other necessary concepts (i.e., ROI, 2D Point etc.). Furthermore, the MAO is also used to handle uncertainties and to infer the BI-RADS score for a particular breast mass (Bulu et al., 2013). Figure 2.1 shows the important concepts of MAO and the relationships between them, excluding details.

In the MAO, a mammography examination is represented by a `MammoCase` concept having `Breast` and `Image` concepts. Each abnormality in a case has a `BI-RADS` concept to show its BI-RADS score. Thus, a `MammoCase` may contain more than one `BI-RADS` concept. In this case, the highest BI-RADS score is assigned to the case as final score. In other words, we set the case's BI-RADS score automatically from the abnormalities found in the case.

The `Image` concept represents the digital images of examination such as MRI, CT, mammography etc. Screening mammography generally involves two views of the breast: one from above (Cranial-Caudal view, CC) and the other from oblique or angled views (Mediolateral-Oblique, MLO). Therefore, a typical mammography

examination contains four mammograms; two MLO and two CC views for two breasts.

The `ROI` concept describes any region of interest (ROI) on an image. The radiologists draw or select a predefined shape for ROI by using the annotation tool and we assume that each ROI represents an abnormality with its additional properties such as mean intensity value, area of the abnormality as pixel count, etc. Abnormality concept describes an abnormality in an image, such as mass, calcification, associated finding, special case and other. As a rule, each Abnormality concept must have at least one ROI and one BI-RADS concept associated with it. Mass concept is used for masses, and it is a subclass of Abnormality concept. Mass concept has additional MassDescriptor, which is the super class of MassShape, MassMargin and MassDensity classes, to describe any particular mass.

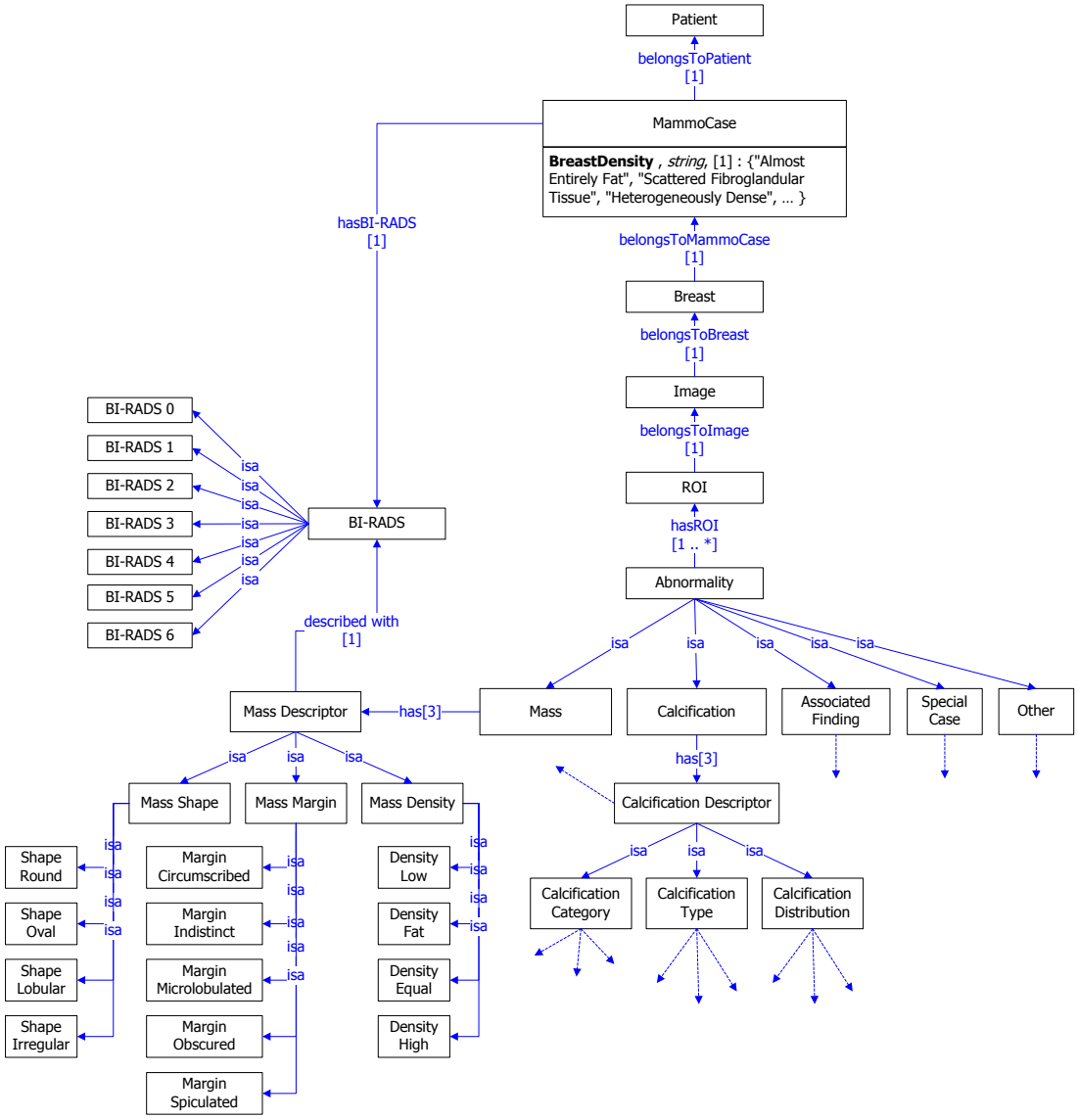



Figure 2.1 Simplified view of the Mammography Annotation Ontology (MAO).

In Figure 2.2 , we illustrate a sample mass annotation, which is in the left breast, and annotated by irregular shape, speculated margin, equal density and BI-RADS score 5. The mean intensity value of the mass is 35598.1 in 16 bits level and area of the mass is 81765 pixel$^2$.

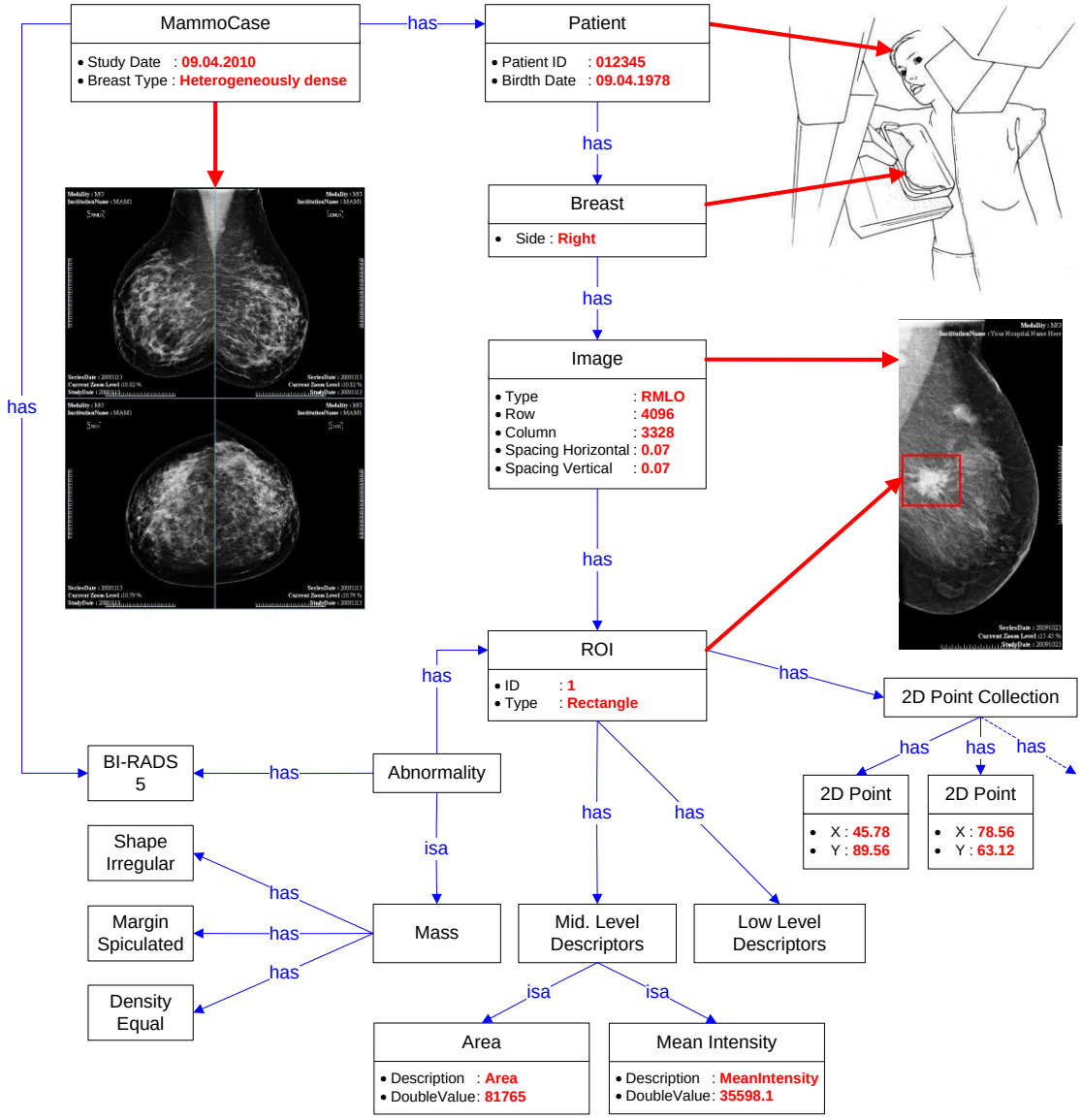



Figure 2.2 Annotation of mammograms with the MAO.

## 2.3 Mammography Annotation and Retrieval Tool (MART)

Mammography Annotation and Retrieval Tool (MART) allows radiologists to examine four images in total, CC and MLO projection of the right and left breasts, for a typical mammography case. In interpreting mammograms, radiologists mark and annotate the abnormalities on images by using a variety of tools, and specify the breast type. MART stores all annotations in XML format, which is then easily, converted into a variety formats, such as; OWL (Web Ontology Language) (W3C, 22.03.2013), radiology reports in natural language etc. We developed the MART

using C++ programming language with QT framework (QT Digia, n.d.) with a cross-platform support. Figure 2.3 depicts a sample screenshot of the MART.

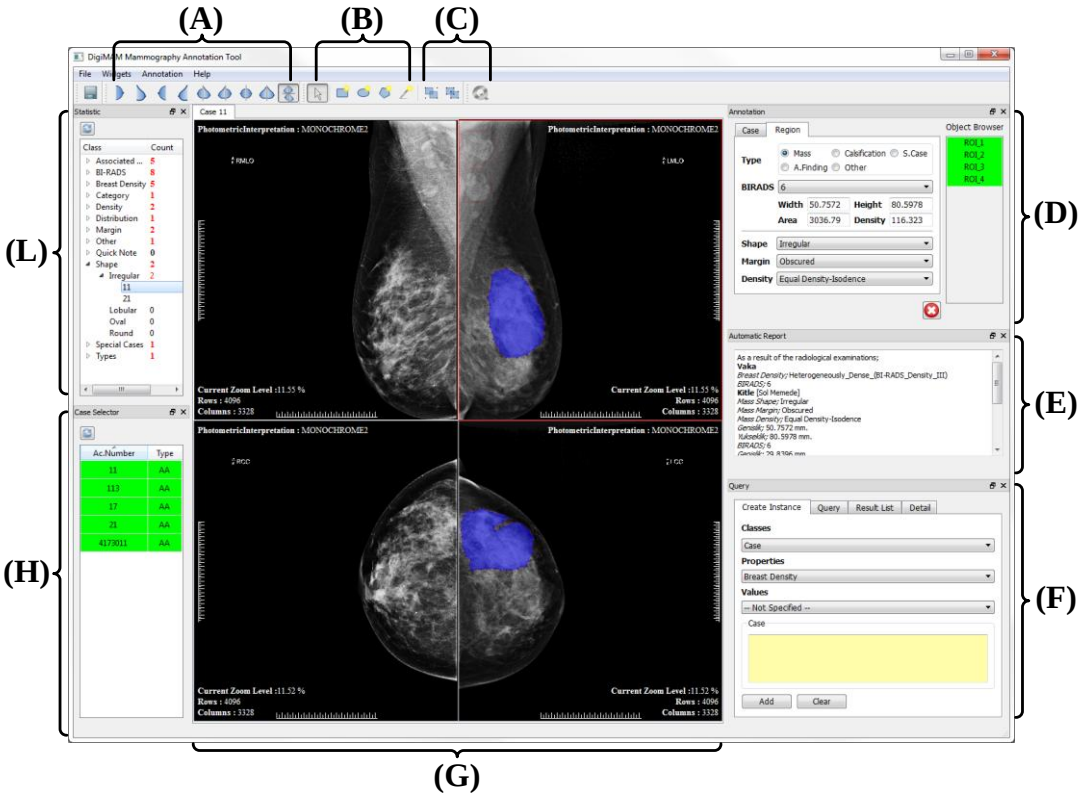



Figure 2.3 Mammography annotation tool.

### *2.3.1 General System Overview of the MART*

MART has two inputs, MAO and Mammograms in DICOM format. MAO represents the expert knowledge used in MART. The second input is mammogram in DICOM format. User must put four mammograms, which are CC and MLO views of the two breasts into a *Case Folder*. Before starting to annotation process, MART converts DICOM images into lossless PNG format and renames the PNG files with respect to their view and then produces an initial XML file (Annotation.xml) using from DICOM header information. As a result, *Case Folder* contains the following files; LCC.png, LMLO.png, RCC.png, RMLO.png and Annotation.xml. After all, user can start to annotate the mammograms by using predefined drawing tools and annotations controls. The annotations are stored in Annotation.xml file. Then, it is possible to convert the XML file to any other format by using predefined XSD

(XML Schema Definition) files, such as MAO class instances and radiological report etc. All these process are illustrated in Figure 2.4.

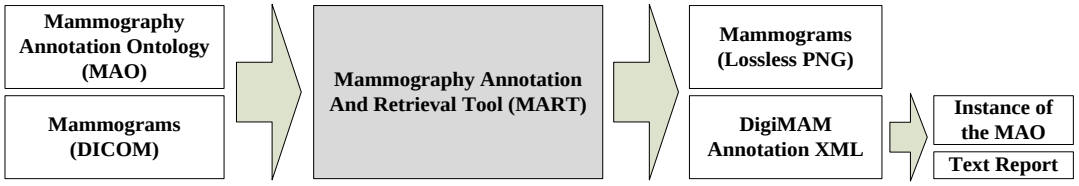



Figure 2.4 General system overview.

### *2.3.2 Main Components of the MART*

*Statistic Widget* (shown in Figure 2.3-L) is the widget gives quick statistics for the selected repository, where it shows count of the mammographic cases for each abnormality properties and their possible values. In this way, user can easily see the distribution of the MAO instances for the selected repository and can easily search and access mammographic cases for a particular abnormality. For example, user can easily list all mammographic cases having at least one mass with *lobular* shape. When double clicked on a case number in the list, then it loads the case selected.

*Case Selection Widget* (shown in Figure 2.3-H) is the widget to browse the repository and to load any mammographic case with double click. Green background indicates the case is already annotated and red background means that the case has not been annotated yet.

*Annotation Widget* (shown Figure 2.3-D) is used to annotate any selected lesion and breast density based on MAO. So, when the MAO is updated, annotation options in the widget are also updated automatically. In practice, first user chooses the ROI to annotate, and then selects the type of the abnormality (i.e. mass, calcification, spatial case etc.) from top part of the widget shown in Figure 2.5-F. Depends on the selected type, below section of the widget (shown in Figure 2.5-C) is changed. For example, if the type of the abnormality is selected as “mass”, then below section asks shape, margin and density of the mass. If “calcification” is selected, below section asks category, type and distribution of the calcification. All possible values in the drop-down-list controls come from MAO file in run-time. Additionally, the widget

also calculates width, height, area and mean-intensity (density) values of the selected ROI shown in Figure 2.5-D. In the right side of the widget, object browser (shown in Figure 2.5-A) lists the ROIs where green rows indicates the ROI is annotated while red background indicates the ROI is not annotated yet. The MART does not allow user to save un-annotated ROIs. User can clear the annotation of the selected ROI by clicking on the button shown in Figure 2.5-B.

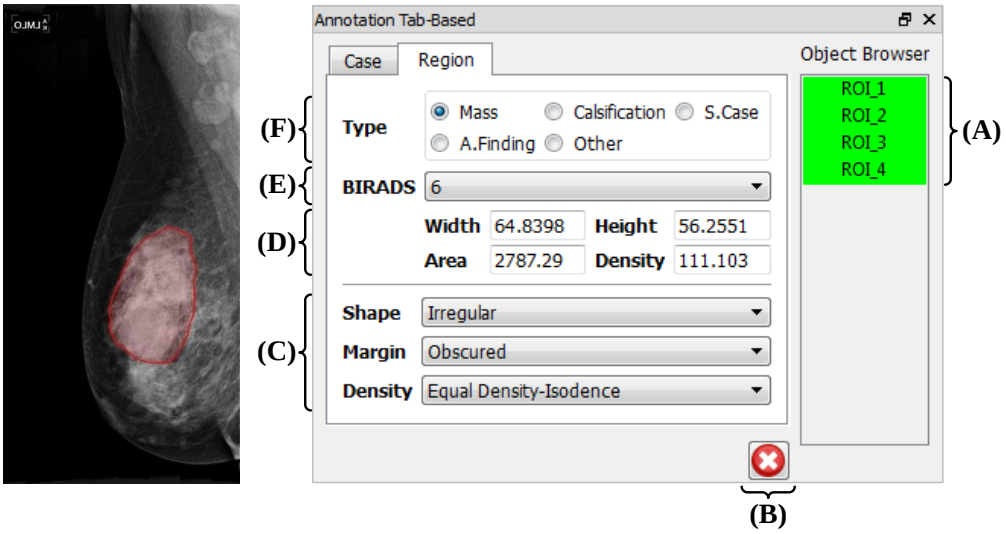



Figure 2.5 Sample mass annotation.

*Case*-based Retrieval Widget (shown in Figure 2.3-F) performs Case-based Retrieval (CBR) functionality, for a given query of abnormality or mammography case. In practice, user clicks "Q" button (Figure 2.3-C) and sends the selected abnormality to the widget as query. Then clicks "Execute" button (Figure 2.6-A) to perform CBR on the selected mammography repository. The result list is shown in the "Result List" tab (Figure 2.6-B), where the list is sorted from most similar to least similar. To see the detail of the similarity calculation between the query and results, user double clicks on any row in the Result List and "Detail" tab (Figure 2.6-C) is opened. During the similarity calculation, CBR algorithm (Bulu et al., 2012) uses both high-level (semantic) and mid-level features (e.g., mean intensity (density) and size of area). In this way, it is possible to sort the abnormalities having same high-level feature values by most relevant to least similar. This ranking improves the accuracy of the CBR result. Additionally, user can create his own query by using "Create Instance" tab (Figure 2.6-D).

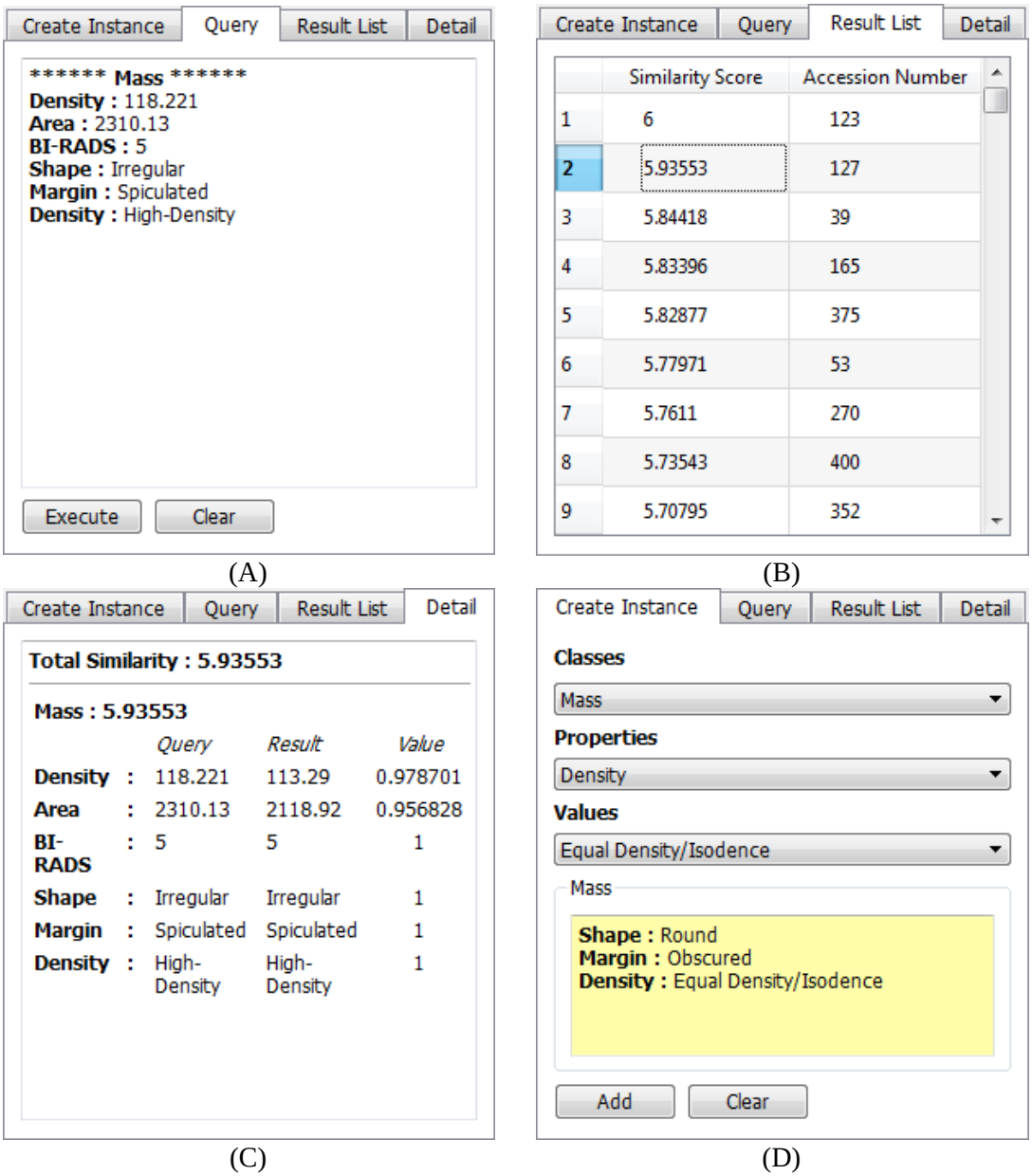



Figure 2.6 Tab views of the case based retrieval widget.

*Mammograms Display Tool* (shown in Figure 2.3-A) enables user to display the mammograms in various options. A standard mammography examination contains four mammograms; two for left breast and two for right breast. This provides user to compare breasts easily or focus on one view to examine the abnormalities in detail.

*Lesion Selection Tool* (shown in Figure 2.3-B) presents four different drawing options to user for marking any abnormality seen in the mammograms. These are rectangle, oval, polygon and free-hand. Generally, an abnormality is seen in both

view (CC and MLO) and in practice user marks them individually and connects them to express ROIs belonging to same abnormality. Then, user annotates the ROIs with single annotation. When the user clicks any of them, both of them are selected and user can easily see the abnormality in both views.

# CHAPTER THREE
# DEMS: DOKUZ EYLUL UNIVERSITY MAMMOGRAM SET

## 3.1 Overview

In this chapter, we present a sample mammogram dataset (DEMS: Dokuz Eylul University Mammogram Set) which is fully annotated with the MART. DEMS contains fully-annotated digital mammograms for computer-aided diagnosis (CAD) studies. It is also compliant with the state-of-the-art semantic-web knowledge representation technologies. During the preparation process of the DEMS, case selection performed in two stages. In first step, candidate cases are selected retrospectively from PACS of Radiology Department of Dokuz Eylul University Medical Faculty Hospital, among more than 50K mammography examination diagnosed between 2004 January and 2008 November. Each candidate case includes four images in DICOM format, which are CC MLO views of both breasts. All of the patients and physicians identifications are manually removed and the whole dataset were anonymized. To select initial candidate cases, we developed a textual Boolean information retrieval system to speed up selection process for each concept in the ontology. In final form, DEMS contains 485 mammographic cases where 255 of them contain one or more lesion. Radiologist expert in mammography annotated each case in three phases using MART.

## 3.2 Existing Mammogram Dataset in Literature and DEMS

There are several mammogram datasets available to researchers who want to measure performance of their lesion detection and classification approaches. But most of them loses their majority or are no longer available. Major mammography datasets are described in following sections.

*Nijmegen Digital Mammogram Dataset*; This dataset includes 40 digitized mammograms of 20 patients. Dataset created by Department of Radiology, University of Nijmegen in the Netherlands and The Dutch National Expertise and Training Center for Breast Cancer Screening. Images are obtained by using

combination of Kodak MIN-R/SO177 and a variety of hardware. Then images are digitized by using Eikonix 1412 12-bit CCD camera with 50 μm sampling aperture and 100 μm sampling distance settings (effective pixel resolution 100 μm). Each image size is 2048 × 2048 pixels. Subsequently, regional light inequality in the images is corrected. All images include at least one cluster of microcalcifications, and dataset consists of 7 malignant, 13 benign lesions. This dataset is not available now (University of South Florida, n.d.).

*Washington University Digital Mammogram Dataset*; This dataset consists of 80 cases acquired by LoRad CCD-based stereotactic core biopsy system to locate the lesion in the breast with single point of view-angle images of digital mammography. The number of benign and malign lesions is equal like the number of microcalcifications and masses. Each image size has 512 × 512 pixels, 100 μm pixel resolution and 12 bits intensity depth Although this dataset is no longer available, this is the first example of digitally captured dataset and could have been accessible by anyone via FTP (Nishikawa, 1997).

*OWH (Office of Women's Health) Dataset*; According to the Nishikawa's article (Nishikawa, 1997), this dataset which is not freely available to everyone developed by Office of Women Health under U.S. Ministry of Health. It contains totally 900 diagnoses from 5 different regions (University of Pennsylvania, University of Virginia, UCLA, UCSF and the American National Naval Medical Center) to provide a national training dataset for CAD developers. Each case include CC and MLO view of both right and left breast acquired using Lumiscan 85 film scanner at 50 μm pixel resolution and 12-bit color depth. Dataset contains 540 normal subjects (proved by biopsy or diagnosed after two years of examination), and 180 benign and 180 malignant lesions. Additionally, the dataset includes the location and properties of the lesion, and pathological features.

*(Mini-)MIAS (Mammographic Image Analysis Society) Dataset*; This dataset is developed by Mammographic Image Analysis Society, formed by more than twenty research institutes in the UK (Davies, 1993). Dataset includes 161 cases selected from British National Mammography Screening Program. Each case includes MLO view of left and right breast (total number of images 322). The original dataset

images have 50 μm pixel resolution with 8-bit color depth, but this set of data is not available now (Nishikawa, 1997). Moreover, a new dataset named mini-MIAS containing cropped versions of original images at 1024×1024 image size and 200 μm pixel resolutions was created according to intensive demand.

*LLNL/UCSF Dataset*; This dataset prepared jointly by the U.S. Lawrence Livermore National Laboratory (LLNL) and The Department of Radiology of University of California at San Francisco (UCSF) to help researchers working on microcalcifications. Dataset contains 197 digitized mammograms of 50 patients (CC and ML views of both left and right breast for each patient, 2 images instead of 4 for one patient who had mastectomy, and 1 corrupted image) (Ashby et al., 1995). Images are digitized by using Du Pont Industrial NDT film digitizer with 35 μm pixel resolution and 12-bit intensity depth and stored using ICS (Image Cytometry Standard) format. Moreover dataset contains two binary truth files describing calcification clusters and major calcification boundaries. Additionally, dataset contains a text file including case history and expert radiologist comments (Nishikawa, 1997).

*DDSM (Digital Database for Screening Mammography)*; This dataset is developed by co-operation of Massachusetts General Hospital, University of South Florida (USF), American Sandia National Laboratories and the U.S. Army Medical Research and Material Unit Breast Cancer Research Program's fund. Each case in the dataset contains two standard views (CC and MLO) of two breasts and is selected from patients diagnosed between October 1988 and February 1999 at Massachusetts General Hospital, Wake Forest University School of Medicine, St. Sacred Heart Hospital and Washington St. Louis University School of Medicine (Heath et al., 2001). The dataset has a total number of 2620 studies. Besides, dataset also contains demographic data for each case like, age of the patient, the mammogram acquisition date, mammogram digitization date and ACR breast density determined by an expert, as well as abnormality verification file containing lesion markings, BI-RADS assessment made by a radiology expert, with the degree of difficulty.

*GPCALMA (Grid Platform for a Computer-Aided Library in Mammography) Dataset*; This dataset was started to be developed by a group of physician working in

Italian National Institute for Nuclear Physics (INFN) with radiologists at 1999. Dataset contains totally 3369 digitized mammography images of 967 cases (each case has varying number of images from 1 to 6) (Lauria, 2009). Mammograms from participating Italian Hospitals are digitized by using single CCD film scanner at 2067x2657 size with 85 μm effective resolution and 12-bit intensity depth and stored using CALMA format (Lauria, 2006). No normalization is applied to the images during the digitization phase due to unavailability of acquisition parameters of films. Dataset contains some assessments made by expert radiologists like breast tissue, lesion presence, lesion location and lesion type. Moreover, dataset includes some demographical information and follow-up studies.

*INbreast Dataset*; This dataset is developed in Breast Centre in CHSJ, Porto. Cases in dataset belong to patients who diagnosed between April 2008 and July 2010. All images acquired by MammoNovation Siemens FFDM at 70 μm effective resolution and 14-bit intensity depth. Acquired images are stored in DICOM files. Dataset includes a total number of 115 cases and 56 of them have biopsy data (Moreira et al., 2012). General properties of all dataset are summarized in Table 3.1 for easy comparison.

*DEMS*: This dataset contains 485 cases, where each case contains four mammograms, MLO and CC views for two breasts, and one XML file called as "DEMS Annotation XML". Each image converted from DICOM images into lossless PNG and name of the each images is set according to its view, e.g. LCC.png, LMLO.png. Resulting PNG images have 16-bit intensity depth, 70 μm effective resolution and, 2560×3328 or 3328×4096 size. Figure 3.1 shows sample mammography case in DEMS which have more than one abnormality. The case contains one mass and two associated findings in the left breast. The mass is indicated with red contour and it has *irregular* shape, *spiculated* margin and *equal* density. Additionally, there are *skin retraction* and *skin thickening* as the associated findings. The breast density of the case is *Almost Entirely Fat* and final BI-RADS score of the case is *6*. This means that the mass is pathologically proven malignancy.

Table 3.1 General overview of datasets (FT-DM: Digitized Mammography; SR-FFDM: Full Field Digital Mammography; in formula of Image Count column a×b=c where a: Images in each Case, b: Number of Cases, c: reported number of images in the dataset, N/A: unknown).

| Dataset | Image Acquisition Method | Resolution (μm) | Intensity Depth (bits) | Image Size (pixel) | Case count w.r.t. pathology | | | | Class Count | Image Count | Access |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Normal | Benign | Malign | Total | | | |
| Nijmegen | DM | 100 | 12 | 2048×2048 | 0 | 7 | 13 | 20 | N/A | 2×20=40 | No |
| Washington Univ. | FFDM | 100 | 12 | 512×512 | 0 | 40 | 40 | 80 | N/A | 1×80=80 | No |
| OWH | DM | 50 | 12 | N/A | 540 | 180 | 180 | 900 | N/A | 4×900 | N/A |
| Mini-MIAS | DM | 200 | 8 | 1024×1024 | ? | ? | ? | 161 | 12 | 2×161=322 | Web |
| LLNL /UCSF | DM | 35 | 12 | N/A | 10 | 32 | 8 | 50 | 5 | 4×50≈397 | Mail |
| DDSM | DM | 42-50 | 12-16 | N/A | 695 | 1011 | 914 | 2620 | 4 | 4×2620 | Web |
| GPCALMA | DM | 85 | 12 | 2067×2657 | 306 | 661 | | 967 | 19 | 3369 | N/A |
| INBreast | FFDM | 70 | 14 | 3328×4084 | | 11 | 45 | 115 | | 410 | Web |
| DEMS | FFDM | 70 | 16 | 2560×3328 or 3328×4096 | 230 | 225 | | 485 | 5 | 4×485 | Web |

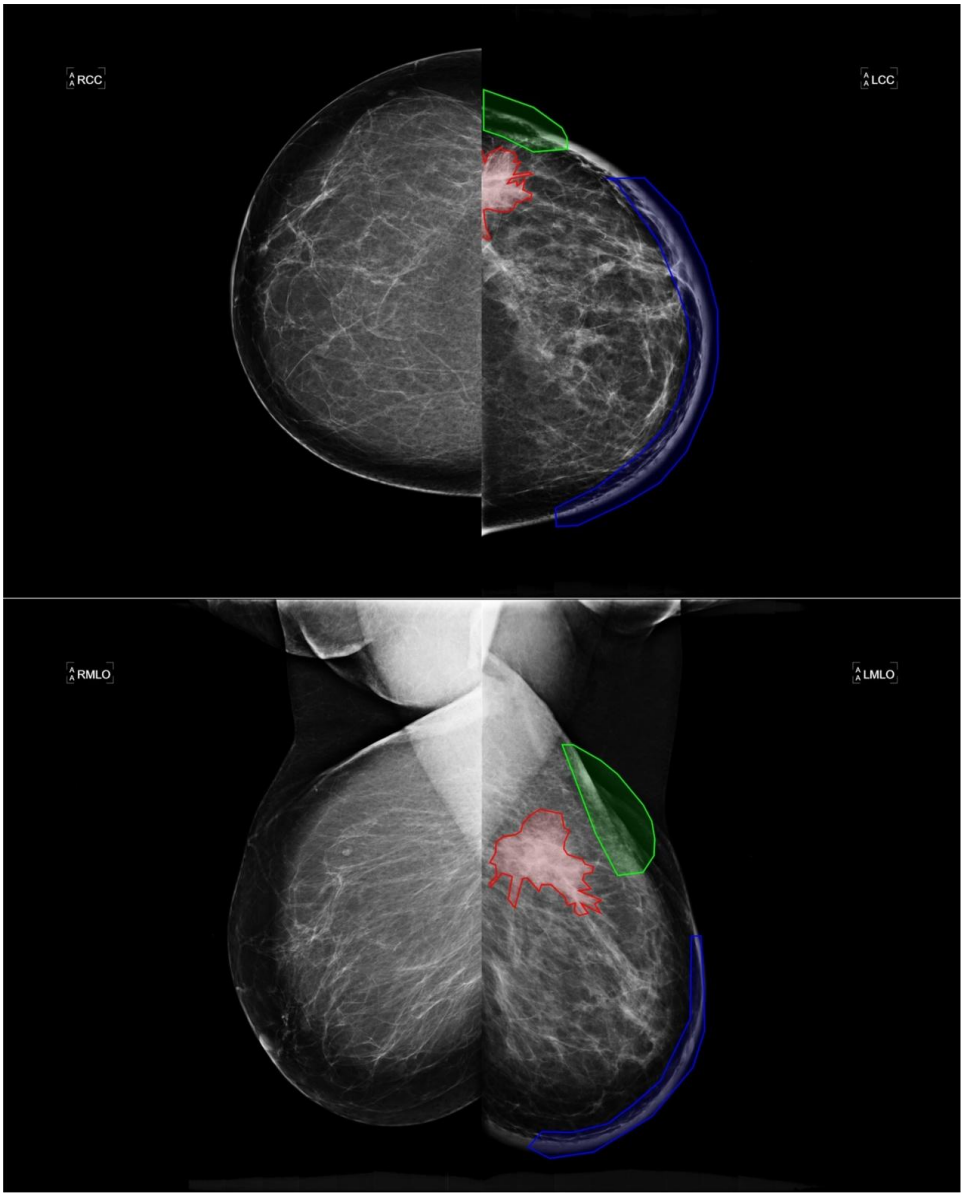



Figure 3.1 Sample mammography case with its ROI's in DEMS, where RCC view is on the right-top corner, LCC view is on the left-top corner, RMLO view is on the right-bottom corner and LMLO view is on the left-bottom corner.

## 3.3 DEMS Annotation XML

DEMS Annotation XML file contains *Patient* and *Case* tags. For privacy reasons we just store *birth date* of the patient. On the other hand *Case* tag includes all image and annotation data with *date of study* which is important to calculate age of patient during examination date. Images are described by *Image* tag, which contains important DICOM headers and lesion annotations denoted by *GraphicItem* tag. A sample *GraphicItem* tag is shown in Figure 3.2.

```
<GraphicItem id="1" type="4" groupId="1" uniqueId="123">
  <PointCollection>
    <Point x="2658.057469371625" y="1941.542446510455" />
    ...
  </PointCollection>
  <Annotation>
    <Instance classId="03">
      <Property valueId="11" id="04" />
      <Property valueId="03" id="05" />
      <Property valueId="03" id="06" />
      <Property valueId="02" id="07" />
    </Instance>
    <MiddleLevelFeatures>
      <Property valueText="94.6206" id="13" />
      <Property valueText="5215.78" id="14" />
      <Property valueText="67.5157" id="15" />
      <Property valueText="101.529" id="16" />
    </MiddleLevelFeatures>
  </Annotation>
</GraphicItem>
```

Figure 3.2 Sample GraphicItem tag in DEMS annotation XML.

Each *GraphicItem* tag includes lesion boundary in *PointCollection* tag and annotation data in *Annotation* tag. The *Annotation* tag describes set of MAO instances in two child tags, namely, *Instance* and *MiddleLevelFeatures*. Value of each *id* attribute in *Annotation* tag is coming from a mapping XML file which derived from MAO to simplify representation of OWL.

### 3.4 Statistics of DEMS

All cases in DEMS annotated for BI-RADS breast type; *Almost Entirely Fat, Scattered Fibroglandular Tissue, Heterogeneously Dense* or *Extremely Dense.* Figure 3.3-A shows breast type distribution, where *Extremely Dense* has the lowest percentage. Secondly, lesions in DEMS are belong to one of the category; *mass*, *calcification*, *special case* and *associated finding*. Additionally, in some of the mammograms metallic clips appear. To be able to distinguish them from any other lesions, we create one more lesion category as *other* and we consider them in this group. Distribution of the abnormalities is shown in Figure 3.3-B.

*Mass is* the one of the major lesion type in DEMS. According to BI-RADS mammography atlas each mass has three attributes; *shape*, *margin* and *density*. Furthermore, each attribute has a set of allowed values (e.g., mass shape can be *round*, *lobular*, *oval* or *irregular*). Table 3.3 shows the distribution of masses in

DEMS according to their attributes in detail. In the table count of lesions in DEMS are given, where *Case* column shows the number of unique mammographic case containing related value, *Lesion* column shows the number of unique lesion marked with related value. For instance, total number of cases containing lobular shaped mass is 28. On the other hand, the total number of lobular shaped mass is 29 since a case contains more than one mass with lobular shape. The masses annotated as BI-RADS category 6 are pathologically proven malignant lesion, while all other masses require pathologic examination to determine if they are benign or malign.

Table 3.2 Features of masses with their count.

| Feature | | Case | Lesion |
|---|---|---|---|
| **BI-RADS** | 2 | 23 | 27 |
| | 3 | 26 | 29 |
| | 4A | 9 | 9 |
| | 4B | 6 | 6 |
| | 4C | 10 | 10 |
| | 5 | 37 | 39 |
| | 6 | 14 | 16 |
| **Shape** | Round | 21 | 27 |
| | Lobular | 28 | 29 |
| | Irregular | 56 | 59 |
| | Oval | 21 | 21 |
| **Margin** | Circumscribed | 44 | 52 |
| | Microlobular | 5 | 5 |
| | Obscured | 16 | 16 |
| | Illdistinct / Illdefined | 22 | 26 |
| | Spiculated | 37 | 37 |
| **Density** | High | 51 | 62 |
| | Equal / Isodence | 55 | 59 |
| | Low / Not Fat Containing | 3 | 3 |
| | Fat Containing Radiolucent | 11 | 12 |

*Calcification* is the second major abnormality type in DEMS. Like masses, annotation of calcifications is determined according to BI-RADS mammography atlas. So, each calcification has *category*, *type* and *distribution* attributes with their allowed values. Figure 3.3-C shows distribution of the *calcification* according to *category* attribute, where *typically-benign* calcification has the highest percentage. Furthermore, distribution of calcification *categories* is shown in Figure 3.3-D.

*Special-cases* and *associated-findings* are the other abnormality types in DEMS. There are six allowed values for *special-cases*, and seven allowed for

*associated-findings*. Different from the other abnormalities in DEMS, some associated findings may not have BI-RADS scores. For these types of lesions we have added one more BI-RADS score as *N/A*. Distributions of the *special-cases* and *associated-findings* in DEMS are shown in Figure 3.3-E and Figure 3.3-F, respectively.

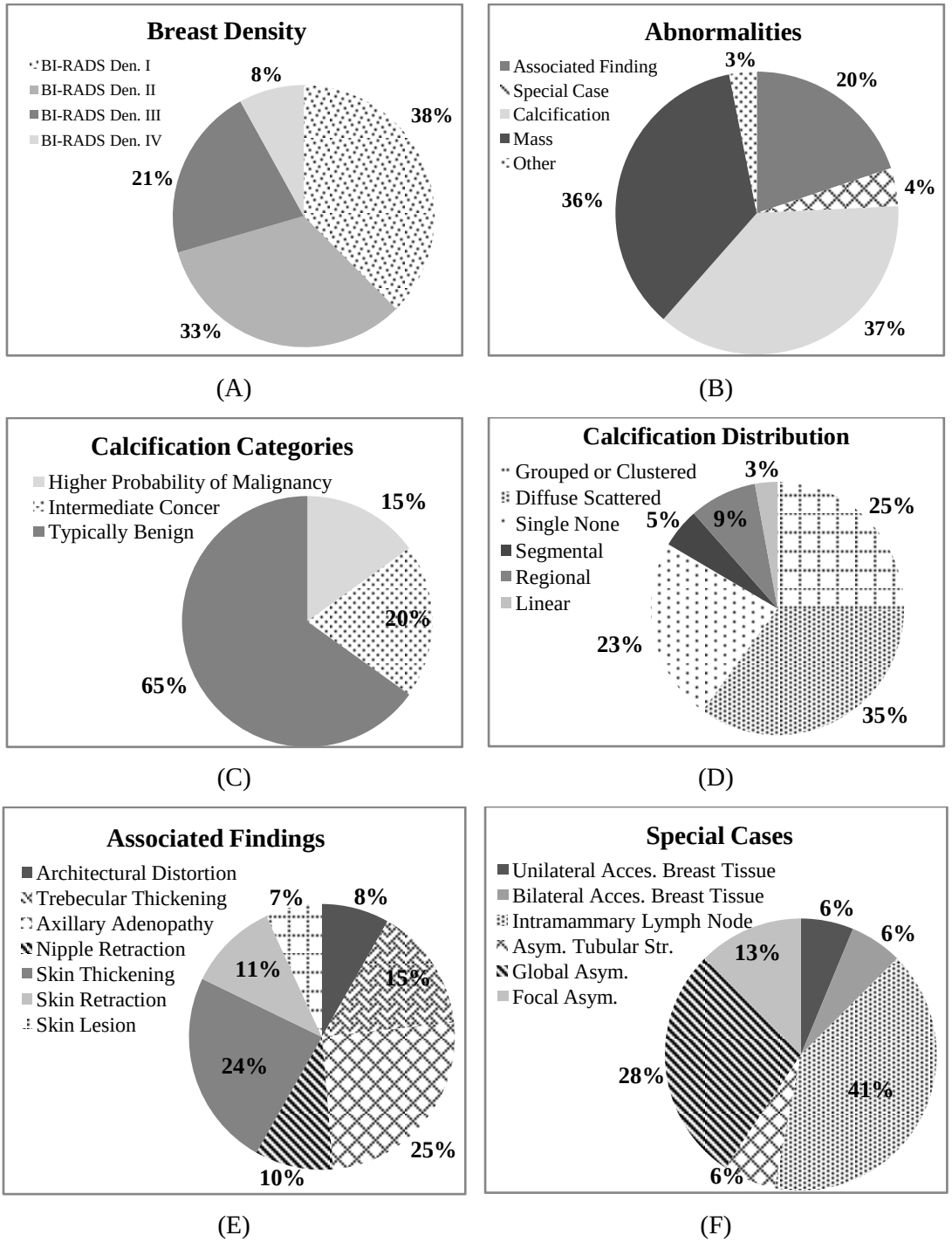



Figure 3.3 Distribution of breast types and abnormalities.

## 3.5 DEMS Web Browser

We have developed a web application (shown in Figure 3.4) to make browsing DEMS possible and easier. The web application has three main parts; *filter*, *display* and *annotations*. In the *filter* part, user can easily filter the cases in DEMS by using combo-boxes. Then, the filtered cases are listed in 'Case List' list-box. When user selects one case from the list, display options are listed above list-box. The list-box is filled dynamically according to lesion types of the selected case. When user clicks on 'Load The Case' button, selected case is displayed with selected display option. In the *display* part, mammogram(s) are displayed. Finally, *annotation* part shows annotations of all lesions in selected case. We use color-coding to connect the ROIs and annotations. In other words, same color is used for both ROI and its annotation.

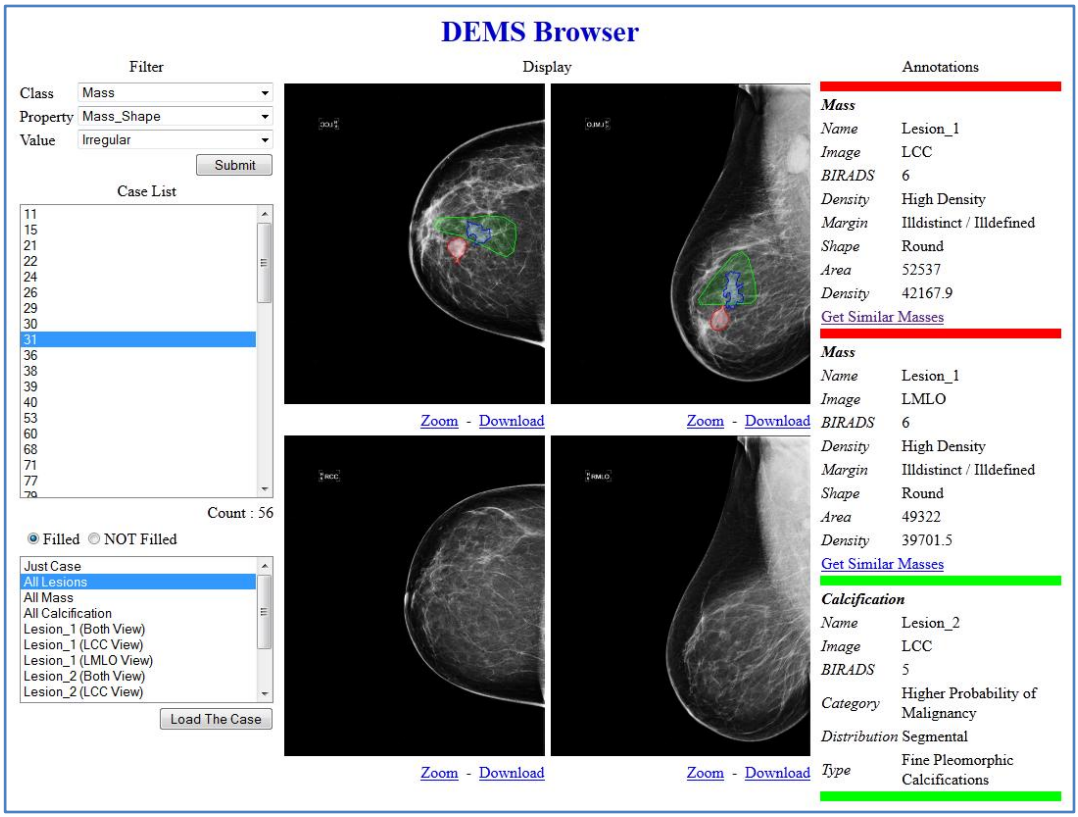



Figure 3.4 Screen shot of DEMS browser

### 3.6 DEMS Low-level Features

Contrary to any existing mammogram datasets, we provide a set of low-level features of mammogram cases in DEMS, where these features can be used to improve CBR results. Moreover, low-level features are mandatory components of a Content-based Image Retrieval (CBIR) system. Without them, the system becomes a metadata-based retrieval system. Each mass in DEMS has 29 different low-level features describing the content and the characteristics of the masses for their shape, texture, margin, mass intensity and size. Low-level features are represented with a vector of floating point numbers, whose the total length of the all feature vector is 578. These features are typically used for classification and clustering of breast masses (Berber, 2013), and to improve CBR results of breast masses.

# CHAPTER FOUR
# CONTENT-BASED IMAGE RETRIEVAL OF BREAST MASSES WITH HIGH-, MID- AND LOW-LEVEL IMAGE FEATURES BY USING SEMANTIC WEB TECHNOLOGIES AND PERFORMANCE COMPARISION OF THE FEATURES

## 4.1 Overview

Computer aided diagnosis (CAD) of breast cancer becomes significant topic for mammography (Mousa et al., 2005), (Verma et al., 2010) and (Keles & Yavuz, 2011). Hence, there is an urgent need to browse medical image databases by their visual content to find cases, and to compare visually similar images and their diagnoses (Müller at al., 2004). Case based reasoning (CBR) is one of the most common problem solving methods for both human and computer, which is based on the solutions of similar past problems and, consequently, it is a popular method for CAD systems. CBR has been formalized for purposes of computer reasoning as a five-step process, namely, retrieve, reuse, revise, review and retain (Domeshek & Kolodner, 1993), (Watson, 1999). Retrieve step is the first and the most important steps of case-based reasoning. Likewise, content-based image retrieval (CBIR) becomes integral part of the case-based reasoning scenario when medical images are considered.

CBIR systems allow searching large image archive for a given query based on visual similarity. For example, radiology department of an average hospital may produce thousands of medical images per day. Currently, retrieval of medical images stored in archives is generally provided by external attributes (e.g. patient ID, patient name, reports or annotations etc.) associated to each case. Search by textual keyword from the radiology report or the electronic patient record is also possible. Besides, CBIR systems allow to browse and search in large image collections based on visual features that are automatically extracted from the images, as well as external attributes.

To date, many CBIR systems were developed for mammographic examinations (Alto & Rangayyan, 2005), (Kinoshita et al., 2007), (Wei et al., 2005) and (El et al., 2002). However, it is clearly known that most CIBR systems have semantic gap between low-level image features and high-level semantic descriptors. Therefore, closing the semantic gap is an important issue in CBIR area. From this point of view, some researchers aim to reduce this gap with combination of high- and low-level features in medical domain (Nair, 2011), (Selvarani & Annadurai, 2007). Since high-level semantic descriptions of medical images are subjective, description may be change from one expert to another. For instance, for a given breast mass, one expert can describe it as round shape while other one can interpret as oval shape. This is the nature of medical image interpretation. However, mid and low-level features are objective since computers calculate these features automatically without human intervention. So, proper combination of different features and similarity score calculations will result more similar cases. Hence, there is an urgent need to a CBIR system to find similar medical cases for case based reasoning and evidence-based medicine.

In this chapter, we aim to figure out performance effect of different level of features in CBIR system for digital mammograms. In this respect, we develop a CBIR system where a breast mass is described with three sets of features: low, mid and high-level feature. High level (HL) features are expert interpretation of a mass for shape, margin and density characteristics. Mid level (ML) features are computer-calculated values for mass intensity and mass size. Both high- and mid level features are human readable. For low-level (LL) features, we have first examined 25 different features and then choose the most three successful of them: Zernike Moments, Texture Browsing and Mean Margin Difference. Then, we compared the performance of individual feature set as well as different combination of them. The experimentations show that using low-level together with high and mid level features improves the system precision and our CBIR system also helps to close the semantic gap between high and low-level features.

## 4.2 Features for Content-based Image Retrieval (CBIR)

We divide the features into three main groups, namely, high-level, mid-level and low-level. We use all of them to describe any existing breast mass observed in the mammograms. High level features are semantically meaningful labels describing an entity (e.g., round, oval, lobular or irregular to describe shape of a mass). Intuitively, high level features are expected to set by human experts. Instead, mid level features are generally computed automatically, and someway can be interpreted by both computers and humans. Typical mid level features are size, length or average intensity of a mass. The third, low level features are extracted by computers and generally represented as a vector numbers. Thus, low-level features are not semantically meaningful by human. It requires an intensive processing to make them interpretable by human. Table 4.1 lists the features with their significant properties.

Table 4.1 CBIR Features

| Feature Level | Readability | Data Type | Acquisition | Match Type | Similarity |
|---|---|---|---|---|---|
| High | Human | Scalar | Human | Exact | Equality |
| Mid | Human/computer | Scalar | Computer | Exact | Equality |
| Low | Computer | Vector | Computer | Similarity | Euclidean Distance |

We propose to use combination of all features to improve the CBIR performance, instead of using high, mid and low-level features individually. In many cases, although the masses have exactly same high-level features, mid and low-level features can be significantly different and ranking the results from most relevant to less is an important task. So, if we use high-level features only, it is impossible to rank the masses from the most to the least relevant cases. Hence, to be able to retrieve more accurate query results, we need to take all level of features into consideration. Details of the features are given in following sections.

### *4.2.1 High-Level Features*

According to ACR BI-RADS (Breast Imaging Reporting and Date System) mammography atlas (The American College of Radiology,2012) any breast mass is

annotated with three high-level features; *shape*, *margin* and *density*. All possible feature values are shown in Table 4.2, which make mammographic examinations more meaningful to process for the human as well as computers.

Table 4.2 High-Level features with allowed values where values in parenthesis are acronym.

| Feature | Allowed Values |
|---|---|
| *Shape* | Round (*Ro*), Oval (*Ov*), Lobular (*Lb*), Irregular (*Ir*) |
| *Margin* | Circumscribed (*Ci*), Microlobulated (*Mi*), Obscured (*Ob*), Indistinct (*In*), Spiculated (*Sp*) |
| *Density* | High (*Hi*), Equal (*Eq*), Low (*Lo*), Fat (*Fa*) |

Today, PACS systems use text-based image retrieval techniques to annotate and retrieve of medical images. However, using high-level features only makes a system as Boolean retrieval systems, where no ranking is possible for cases annotated with same feature values. This is the most important problem of text-based image retrieval. Another major problem is that the task of describing image content with keywords is very subjective. An image can mean different things to different people. Moreover, even with the same view, the words used to describe the content could vary from one person to another. In other words, there could be a variety of inconsistencies between user textual queries and image annotations or descriptions (Hung & Chen, 2006).

### *4.2.2 Mid-Level Features*

The features in this category are calculated automatically by computer for each individual breast mass and they can be scalar or vector. One of the advantages of using calculated features is being objective. Mid-level features can be read and understand by human as well as computers. In this work we proposed to use two simple features, *Area* and *Mean Intensity*, of the masses.

### *4.2.3 Low-level Features*

Low-level features are set of real numbers, so that they are meaningful only for computer and heavily used in CBIR systems. Unlike other feature types, low-level features lack of semantic information. On the other hand, low-level features are more

objective than high-level feature type since they are calculated from image pixel data without human intervention. Therefore, we believe that we will be able to get better query results by using objective features in similarity calculation.

In literature, several low-level features and their combinations are used in Mammography CADx systems. In this study, we used three low-level features. One of them is used in several mammography related works. The others are proposed in CBIR literature but not used for mammography images before. Three low-level features used in this work are described in following shortly.

*4.2.3.1 Zernike Moments*

Zernike Moments (Khotanzad & Hong, 1990) are orthogonal moments, which use unit vector representation of an image. They are rotation and scale invariant and denoted as in the following formula.

$$A_{nm} = \frac{n+1}{\pi} \sum_{x} \sum_{y} I(x,y) V_{nm}^{*}(\rho,\theta), n > 0, n - |m| \text{ is even, } |m| < n$$

where |·| denotes absolute value of a real number, ρ is the length of the vector from origin to point (x,y), angle between x axis to the vector and $A_{nm}^{*} = A_{n,-m}$. Here, $V_{mn}^{*}(\rho,\theta)$ are the Zernike polynomials and denoted as in the following formula.

$$V_{nm}(\rho,\theta) = R_{nm}(\rho) e^{jm\theta}$$

where

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^{s} \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} p^{n-2s}$$

Rosa et al. (Rosa et al., 2008) used Zernike Moments in a mammography CBIR system, reported that experimentations on DDSM dataset achieves 90% of precision with respect to the recall.

#### *4.2.3.2 Texture Browsing*

Texture Browsing feature aims to describe texture of a region similar to human perception in terms of regularity, coarseness and directionality (Wu et al., 2000) and it is an element of MPEG-7 standard. To the best of our knowledge, this texture descriptor, like Homogeneous texture descriptor, is not used in medical domain. Representation of this descriptor is defined as the following feature vector.

$$TBD = \langle v_1, v_2, v_3, v_4, v_5 \rangle$$

Elements of the feature vector represent regularity ($v_1$) of texture, dominant orientations ($v_2$, $v_3$) of texture and dominant scales ($v_4$, $v_5$) of texture. Extraction of this descriptor uses Gabor filter functions with 6 orientations and 4 scales.

Gabor functions are Gaussian functions that are modulated by complex sinusoids. In image processing, these functions are used for edge and bar detections. In medical image domain, this feature set is used in some works to represent texture information of the image. Müller et al. (Müller et al., 2004) uses a generic CBIR system to create a reference medical dataset. Their system uses Gabor filters to describe image textual content. Zheng (Zheng, 2009) refers these features as "commonly used visual descriptors" in mammographic CAD systems. Yu and Huang (Yu & Huang, 2010) show that using Gabor filters in conjunction with windowed Fourier transform shows similar performance with high order statistical methods in microcalcification detection.

#### *4.2.3.3 Mean Margin Difference*

Margin of a mass includes very important clues for determining malignancy of a mass. Therefore, a low-level feature modeling the mass margin formally is needed to assign margin property to a mass. There are several works attempting to model margin of a mass using shape features (Rangayyan et al., 2000), (Deloguet al., 2007). Although shape descriptors are useful for margin characterization, intensity difference between inner and outer object areas is another important feature.

Mean Margin Difference Feature aims to model marginal intensity characteristics of a mass. Polar representation of mass's bounding box that is centered on mass center is used to extract angular properties of a mass. Additionally, manually or automatically segmented binary mass region is used to determine inner and outer regions of a mass in polar representation. Moreover, a dilation and erosion mask is used to find inner and outer margin areas. Generating a polar representation of an image is given with following formula.

$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \theta = \tan^{-1}(\frac{y}{x})$$

where (x,y) is the coordinates of original image, (r,θ) are length and angle axis of the polar coordinate system. Figure 4.1 contains both original and segmented regions and their polar representations.

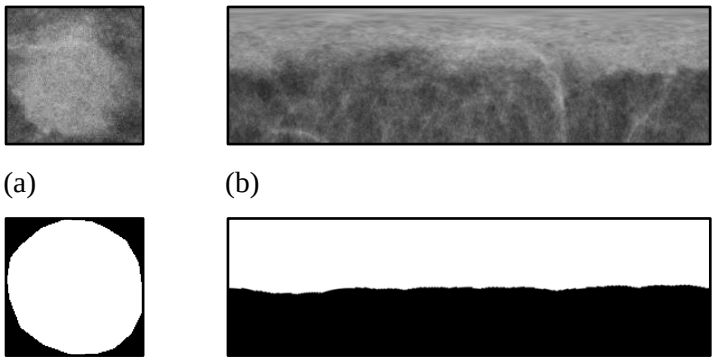



Figure 4.1 (a) Original ROI (b) Polar representation of original ROI (c) Binary segmentation of the mass (d) Polar representation of the segmented ROI (c).

Herein, using polar representation of region mask we determine inner (IR) and outer (OR) regions of the mass in polar coordinate representation. Furthermore, approximate margin area is determined by subtracting original image from eroded (inner margin area, IMA) and dilated (outer margin area, OMA) mask region. After obtaining all required regions, we calculate mean margin difference using following formula.

$$\mu_D(i) = \mu_{IR}(i) - \mu_{OR}(i)$$

where i denotes i-th column of polar image of region, $\mu_{IR}$ and $\mu_{OR}$ are average intensity values of inner and outer margin area of polar mass image.

### 4.3 Similarity Calculation

In CBIR systems, the similarity calculation between two images is inevitable, thus, we explain the similarity calculation of masses in detail. However, system can be easily extended to any other type of abnormalities in mammography (e.g. calcification, associated finding etc.).

Let us assume that we have a database of masses, each described with a set of features. More formally, a mass database, M, is defined as follows:

$$M = \{m_1, m_2, ..., m_n\} \quad (1)$$

Furthermore, each mass is defined with a set of features: high, mid and low-level. For a given arbitrary mass, $m_x$, is defined as follows:

$$m_x = \{F_x^{\mathrm{H}}, F_x^{\mathrm{M}}, F_x^{\mathrm{L}}\} \quad (2)$$

where the terms $F_x^{\mathrm{H}}, F_x^{\mathrm{M}}$ and $F_x^{\mathrm{L}}$ denotes high, mid and low-level features of the mass $m_x$, respectively. Formally, each type of feature may be defined with a different set of sub attributes.

$$F_x^{\mathrm{H}} = \{f_x^{\mathrm{s}}, f_x^{\mathrm{m}}, f_x^{\mathrm{d}}\}\ ,\ F_x^{\mathrm{M}} = \{f_x^{\mathrm{a}}, f_x^{\mathrm{mi}}\}\ ,\ F_x^{\mathrm{L}} = \{f_x^{\mathrm{mmd}}, f_x^{\mathrm{tb}}, f_x^{\mathrm{zm}}\} \quad (3), (4), (5)$$

where superscript represents features names: such as s for shape, m for margin, d for density, a for area, mi for mean intensity, *mmd* for margin mean differences, *tb* for texture browsing and *zm* for Zernike moments. Using Equation 3, 4 and 5, we can represent the $m_x$ as follows,

$$m_x = \left\{\left\{f_x^{\mathrm{s}}, f_x^{\mathrm{m}}, f_x^{\mathrm{d}}\right\}, \left\{f_x^{\mathrm{a}}, f_x^{\mathrm{mi}}\right\}, \left\{f_x^{\mathrm{mmd}}, f_x^{\mathrm{tb}}, f_x^{\mathrm{zm}}\right\}\right\} \quad (6)$$

where, $f_x^{\mathrm{a}}$ and $f_x^{\mathrm{mi}}$ are real numbers, $f_x^{\mathrm{mmd}}$, $f_x^{\mathrm{tb}}$ and $f_x^{\mathrm{zm}}$ are real number vectors, and they are calculated by using pixel values of the mass and,

$f_x^{\mathrm{s}} \in \{\text{Round, Oval, Lobular, Irregular}\}$

$f_x^{\mathrm{m}} \in \{\text{Circumscri bed, Microlobular, Obscured, Illdistict , Spiculated}\}$

$f_x^{\mathrm{d}} \in \{\text{Low, Fat, Egual, High}\}$

The similarity function, S, between masses, $m_a$ and $m_b$ is given in Equation 7.

$$S(m_a, m_b) = \alpha \times S^H(m_a, m_b) + \beta \times S^M(m_a, m_b) + \gamma \times S^L(m_a, m_b) \quad (7)$$

where superscript of H, M, and L shows high, mid and low-level similarity score of the masses, respectively. And $\alpha$, $\beta$ and $\gamma$ represent weight value of each similarity score. In this study, we set each weight value as 1.0. In detail,

$$S^H(m_a, m_b) = \alpha \cdot sim^S(f_a^s, f_b^s) + \beta \cdot sim^M(f_a^m, f_b^m) + \omega \cdot sim^D(f_a^d, f_b^d) \quad (8)$$

where the functions $sim^S$, $sim^M$ and $sim^D$ use Table 4.3. The arguments of these functions denote row and column of the table. The table is similarity matrix for each high-level property (i.e., shape, margin and density); its values are empirically set within a range of zero and 1. Here, the value of 1 means that the features are identically similar while zero indicates no similarity. Similarity calculation of mid-level features is given as follows:

$$S^M(m_a, m_b) = sim^{\mathrm{A}}(m_a, m_b) + sim^{\mathrm{Mi}}(m_a, m_b) \quad (9)$$

where $sim^{\mathrm{A}}(m_a, m_b)$ and $sim^{\mathrm{Mi}}(m_a, m_b)$ are functions to calculate similarity score between the masses $m_a$ and $m_b$, depends on their area and mean intensity values, respectively. In detail,

$$sim^{A}(m_a, m_b) = 1 - \frac{\left| f_a^{a} - f_a^{a} \right|}{\text{maxArea}} \tag{10}$$

$$sim^{Mi}(m_a, m_b) = 1 - \frac{\left| f_a^{mi} - f_b^{mi} \right|}{\text{maxMeanIntensity}} \tag{11}$$

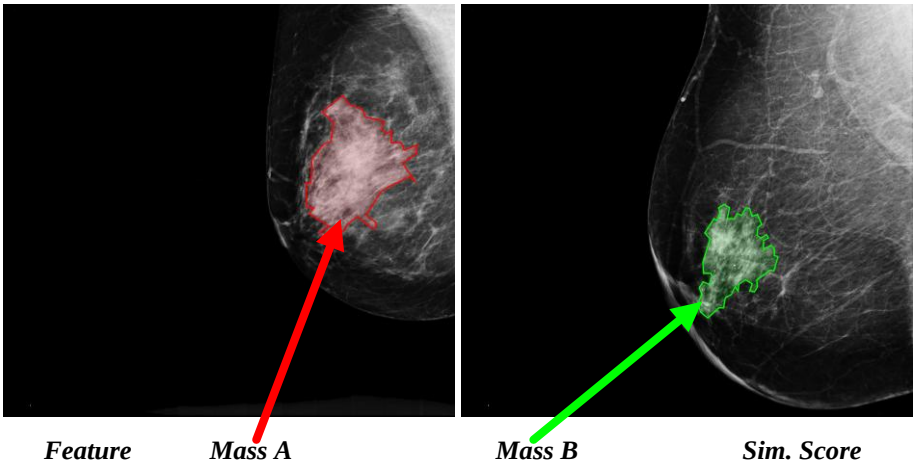
where $f_a^{a}$ and $f_b^{a}$ are area, $f_a^{mi}$ and $f_b^{mi}$ are mean intensity values of masses $m_a$ and $m_b$, respectively. maxArea and maxMeanIntensity denotes the maximum area and mean intensity values in all over masses. Finally, to calculate similarity score for the low-level features, we use the distance function shown in Equation 12.

$$S^{L}(m_a, m_b) = \sum_{i=1}^{3} 1 - \frac{\text{Euclidean}(f_a^{i}, f_b^{i})}{\text{maxDistanc e(i)}} \quad \text{where,} \quad i \in \{mmd, tb, zm\} \tag{12}$$

Figure 4.2 illustrates a sample similarity calculation.

Table 4.3 Similarity matrixes for high-level features of the masses. Meaning of the row and column headers is given in Table 4. 2.

| **Shape (simS)** | | | | | **Density (simD)** | | | | | **Margin (simM)** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Ro* | *Ov* | *Lb* | *Ir* | | *Hi* | *Eq* | *Lw* | *Fa* | | *Ci* | *Mi* | *Ob* | *In* | *Sp* |
| *Ro* | 1 | 0.8 | 0.8 | 0.4 | *Hi* | 1 | 0.5 | 0.2 | 0 | *Ci* | 1 | 0.6 | 0.5 | 0.4 | 0.2 |
| *Ov* | 0.8 | 1 | 0.6 | 0 | *Eq* | 0.5 | 1 | 0.7 | 0.6 | *Mi* | 1 | 1 | 0.5 | 0.6 | 0.3 |
| *Lb* | 0.8 | 0.6 | 1 | 0.6 | *Lw* | 0.2 | 0.7 | 1 | 0.9 | *Ob* | 1 | 0.5 | 1 | 0.8 | 0.7 |
| *Ir* | 0.4 | 0 | 0.6 | 1 | *Fa* | 0 | 0.6 | 0.9 | 1 | *In* | 0 | 0.6 | 0.8 | 1 | 0.8 |
| | | | | | | | | | | *Sp* | 0 | 0.3 | 0.7 | 0.8 | 1 |



| *Feature* | *Mass A* | *Mass B* | *Sim. Score* |
|---|---|---|---|
| ***Shape*** | Irregular | Irregular | 1.0 |
| ***Margin*** | Spiculated | Indistinct | 0.8 |
| ***Density*** | High Density | Equal Density | 0.7 |
| ***Area*** | 472920 | 273012 | 0.81 |
| ***Mean Int.*** | 37753.3 | 27985.8 | 0.84 |
| ***TB*** | {1.0, 1.0, 2.0, 2.0, 1.0} | {1.0, 1.0, 1.0, 3.0, 2.0} | 0.72 |
| ***ZM*** | {0.44, -0.10, ... , -0.05} | {0.28, -0.04, ... , -0.002} | 0.7 |
| ***MMD*** | {0.05, 0.045, ... , 0.021} | {0.04, 0.06, ... , 0.04} | 0.9 |
| | | ***Total Similariy Score*** | **6.47** |

Figure 4.2 Sample similarity calculation.

## 4.4 Semantic Query-enhanced Web Rule Language (SQWRL)

SQWRL is built on the Semantic Web Rule Language (SWRL), which is an expressive OWL-based rule language. SWRL makes possible to write inference rules and this provides more powerful deductive reasoning capabilities than OWL alone. Semantically, SWRL is built on the same description logic foundation as OWL and provides similar strong formal guarantees when performing inference (Protege, 02.09.2013). For example, as we mentioned before if a mammographic case contains one or more abnormality then the highest BI-RADS score of the masses in that case is assigned as the BI-RADS score of the case. So, following SWRL rule, Figure 4.3, infers and sets BI-RADS score of a particular mammographic case (MammoCase) according to the BI-RADS score of its abnormalities;

```
MammoCase(?case) ∧ hasBreast(?case, ?breast) ∧
hasImage(?breast, ?image) ∧ hasROI(?image, ?roi) ∧
hasAbnormality(?roi, ?mass) ∧ hasBirads(?mass, ?birads) ∧
sortIndex(?birads, ?sortIndx) ˚ sqwrl:makeSet(?setBirads,
?sortIndx) ∧ sqwrl:groupBy(?setBirads, ?case) ˚
sqwrl:max(?maxIndex, ?setBirads) ∧ swrlb:equal(?maxIndex,
?sortIndx)
→
hasBirads(?case, ?birads)
```

Figure 4.3 SWRL Rule to Infer and Set BI-RADS Score of a MammoCase

On the other hand, SQWRL takes a standard SWRL rule. Both of them have an antecedent part, which is referred to as the body, and a consequent part, which is referred to as the head. Differently, SQWRL replaces the rule consequent with a retrieval specification for retrieving knowledge from OWL by providing SQL-like operations (O'Connor & Das, 2009). For example, the following SQWRL rule, Figure 4.4, retrieves maximum mean intensity value in all mass instances.

```
Mass(?mass) ∧ hasROI(?mass, ?roi) ∧
hasMidLevelDescriptor(?roi, ?meanIntensity) ∧
description(?meanIntensity, ?desMeanIntensity) ∧
swrlb:equal(?desMeanIntensity, "MeanIntensity") ∧
doubleValue(?meanIntensity, ?meanIntensityValue) ˚
sqwrl:makeSet(?setMeanIntensity, ?meanIntensityValue) ˚
sqwrl:max(?maxMeanIntensityValue, ?setMeanIntensity)
→
sqwrl:select(?maxMeanIntensityValue)
```

Figure 4.4 SQWRL Rule to Retrieve Maximum Mean Intensity Value of the Masses

### 4.5 Similarity Calculation with SQWRL

We use OWL and SQWRL to retrieve similar masses for a given mass query. We process SQWRL rules with Jess rule engine (Friedman-Hill, n.d.) in SWRLTab of Protégé (Stanford Center for Biomedical Informatics Research, n.d.). To be able to perform SQWRL rule for retrieval, we also store high-level similarity matrixes in the MAO. To do this, we defined object type `owl:AnnotationProperty` between each `Mass_Descriptor` class and use *`rdfs:range`* and *`rdfs:domain`* tags

of owl:AnnotationProperty to store row and column of our similarity matrix, shown in Table 4.3. rdfs:label tag of owl:AnnotationProperty stores the similarity score in string format, and converted into float while calculation. Figure 4.5 shows OWL syntax of a sample owl:AnnotationProperty between DensityHigh and DensityEqual classes, where similarity score is equal to 0.5 as shown in Table 4.3.

```
<owl:AnnotationProperty rdf:ID="AnnoMassDensityHighEqual">
<rdfs:domain rdf:resource="#DensityHigh"/>
<rdfs:range rdf:resource="#DensityEqual"/>
<rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
0.5
</rdfs:label>
<rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</owl:AnnotationProperty>
```

Figure 4.5 Sample annotation property for classes DensityHigh and DensityEqual.

To retrieve similar masses for a given query, we run the SQWRL rule given in Figure 4.6. Where variable ?m-q denotes querying mass instance and ?m-r denotes all other mass instances different from the query mass in the ontology (tbox:notEqualTo(?m-q, ?m-r)). As we mentioned in Equation 6, total similarity score between two masses is sum of their high-level and mid-level similarity scores. So the antecedent part of the rule consist of two main parts, one for calculation high-level similarity score (?sumHigh) and the other part for mid-level similarity score (?sumMid) calculation.

To obtain the high-level similarity score we process owl:AnnotationProperty's for each high-level feature of the masses (i.e. tbox:isAnnotationProperty(?ap-s) where ?ap-s denotes the owl:AnnotationProperty for high-level feature *Shape*). As a result of calculation high-level similarity score (?sumHigh), we sum similarity values of each high-level feature of mass (swrlb:add(?temp1, ?sv-s, ?sv-m) ∧ swrlb:add(?temp2, ?sv-d, ?sv-b) ∧ swrlb:add(?sumHigh,

?temp1, ?temp2), where; ?sv-s, ?sv-m, ?sv-d and ?sv-b denotes similarity values for mass high-level features *Shape*, *Margin*, *Density* and *BI-RADS*, respectively). On the other hand, to calculate mid-level similarity score (?sumMid) between the masses, first we calculate maximum *Area* and *Mean Intensity* values (?maxArea and ?maxMeanIntensity) depend on the all mass instances in the MAO. Then, we perform Equation 9 and 10 for the normalization process (i.e. calculation of $sim^{A}$ in the Equation 9; swrlb:subtract(?subArea, ?valArea-q, ?valArea-r) ∧ swrlb:abs(?absArea, ?subArea) ∧ swrlb:divide(?divArea, ?absArea, ?maxArea) ∧ swrlb:subtract(?simResArea, 1, ?divArea), where, ?valArea-q and ?valArea-r denotes *Area* value of the query and retrieved masses, respectively and ?simResArea denotes result value of the function $sim^{A}$). Finally, by summing ?sumHigh and ?sumMid, we obtain result similarity score (?sumSim) between the mass (swrlb:add(?sumSim, ?sumHigh, ?sumMid)). We sort descending the result set according to ?sumSim value to list most similar masses on the top of the result list.

```
Mass(?m-q) ∧ ID(?m-q, ?id-q) ∧ swrlb:equal(?id-q, "A") ∧
Mass(?m-r) ∧ tbox:notEqualTo(?m-q, ?m-r) ∧ hasROI(?m-q, ?roi-
q) ∧ hasMidLevelDescriptor(?roi-q, ?midArea-q) ∧
description(?midArea-q, ?desArea-q) ∧ swrlb:equal(?desArea-q,
"Area") ∧ doubleValue(?midArea-q, ?valArea-q) ∧
hasMidLevelDescriptor(?roi-q, ?midMeanIntensity-q) ∧
description(?midMeanIntensity-q, ?desMeanIntensity-q) ∧
swrlb:equal(?desMeanIntensity-q, "MeanIntensity") ∧
doubleValue(?midMeanIntensity-q, ?valMeanIntensity-q) ∧
hasROI(?m-r, ?roi-r) ∧ hasMidLevelDescriptor(?roi-r, ?midArea-
r) ∧ description(?midArea-r, ?desArea-r) ∧
swrlb:equal(?desArea-r, "Area") ∧ doubleValue(?midArea-r,
?valArea-r) ∧ hasMidLevelDescriptor(?roi-r, ?midMeanIntensity-
r) ∧ description(?midMeanIntensity-r, ?desMeanIntensity-r) ∧
swrlb:equal(?desMeanIntensity-r, "MeanIntensity") ∧
doubleValue(?midMeanIntensity-r, ?valMeanIntensity-r) ∧
hasMassShape(?m-q, ?ds-q) ∧ hasMassMargin(?m-q, ?dm-q) ∧
hasMassDensity(?m-q, ?dd-q)∧ hasBirads(?m-q, ?db-q) ∧
abox:hasClass(?ds-q, ?cs-q) ∧ abox:hasClass(?dm-q, ?cm-q) ∧
abox:hasClass(?dd-q, ?cd-q)∧ abox:hasClass(?db-q, ?cb-q) ∧
hasMassShape(?m-r, ?ds-r) ∧ hasMassMargin(?m-r, ?dm-r) ∧
hasMassDensity(?m-r, ?dd-r)∧ hasBirads(?m-r, ?db-r) ∧
abox:hasClass(?ds-r, ?cs-r) ∧ abox:hasClass(?dm-r, ?cm-r) ∧
abox:hasClass(?dd-r, ?cd-r)∧ abox:hasClass(?db-r, ?cb-r) ∧
tbox:isProperty(?ap-s) ∧ tbox:isAnnotationProperty(?ap-s) ∧
tbox:isInDomainOf(?cs-q, ?ap-s) ∧ tbox:isInRangeOf(?cs-r, ?ap-
s) ∧ rdfb:hasLabel(?ap-s, ?lb-s) ∧ StringToDoble(?s2d-s) ∧
stringValue(?s2d-s, ?str-s) ∧ swrlb:equal(?str-s, ?lb-s) ∧
doubleValue(?s2d-s, ?sv-s) ∧ tbox:isProperty(?ap-m) ∧
tbox:isAnnotationProperty(?ap-m) ∧ tbox:isInDomainOf(?cm-q,
?ap-m) ∧ tbox:isInRangeOf(?cm-r, ?ap-m) ∧ rdfb:hasLabel(?ap-m,
?lb-m) ∧ StringToDoble(?s2d-m) ∧  stringValue(?s2d-m, ?str-m)
∧ swrlb:equal(?str-m, ?lb-m) ∧ doubleValue(?s2d-m, ?sv-m) ∧
tbox:isProperty(?ap-d) ∧ tbox:isAnnotationProperty(?ap-d) ∧
tbox:isInDomainOf(?cd-q, ?ap-d) ∧ tbox:isInRangeOf(?cd-r, ?ap-
d) ∧ rdfb:hasLabel(?ap-d, ?lb-d) ∧ StringToDoble(?s2d-d) ∧
stringValue(?s2d-d, ?str-d) ∧ swrlb:equal(?str-d, ?lb-d) ∧
doubleValue(?s2d-d, ?sv-d) ∧ tbox:isProperty(?ap-b) ∧
tbox:isAnnotationProperty(?ap-b) ∧ tbox:isInDomainOf(?cb-q,
?ap-b) ∧ tbox:isInRangeOf(?cb-r, ?ap-b) ∧ rdfb:hasLabel(?ap-b,
?lb-b) ∧ StringToDoble(?s2d-b) ∧ stringValue(?s2d-b, ?str-b) ∧
swrlb:equal(?str-b, ?lb-b) ∧ doubleValue(?s2d-b, ?sv-b) ∧
swrlb:add(?temp1, ?sv-s, ?sv-m)∧ swrlb:add(?temp2, ?sv-d, ?sv-
b) ∧ swrlb:add(?sumHigh, ?temp1, ?temp2) ˚
sqwrl:makeSet(?setmidArea-q, ?valArea-q) ∧
sqwrl:makeSet(?setmidArea-r, ?valArea-r) ∧
```

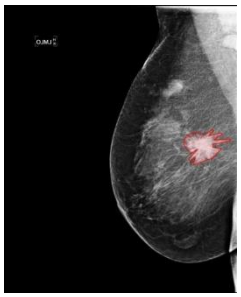
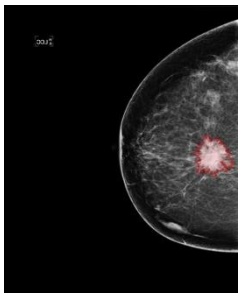
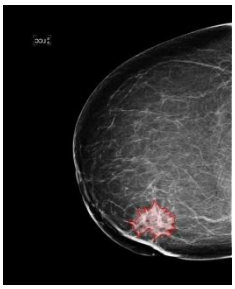
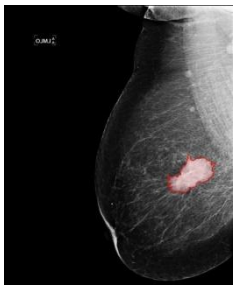
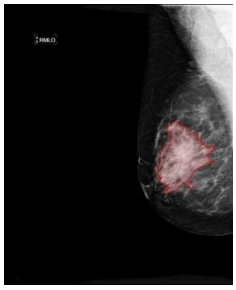
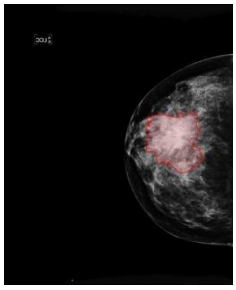
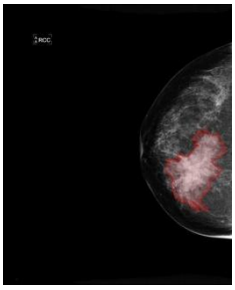
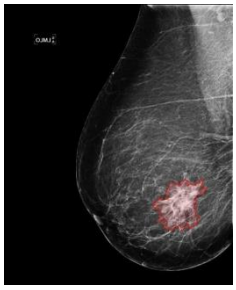
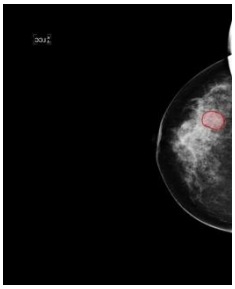
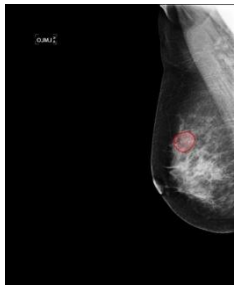
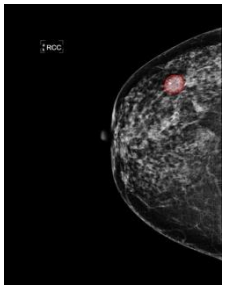
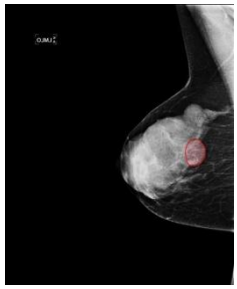
Figure 4.6 The SQWRL rule to retrieve similar masses for a given mass.

```
 sqwrl:makeSet(?setmidMeanIntensity-q, ?valMeanIntensity-q) ∧
 sqwrl:makeSet(?setmidMeanIntensity-r, ?valMeanIntensity-r) ∘
  sqwrl:union(?setUnionArea, ?setmidArea-q, ?setmidArea-r) ∧
 sqwrl:union(?setUnionMeanIntensity, ?setmidMeanIntensity-q,
?setmidMeanIntensity-r) ∧ sqwrl:max(?maxArea, ?setUnionArea) ∧
    sqwrl:max(?maxMeanIntensity, ?setUnionMeanIntensity) ∧
      swrlb:subtract(?subArea, ?valArea-q, ?valArea-r) ∧
    swrlb:abs(?absArea, ?subArea) ∧ swrlb:divide(?divArea,
?absArea, ?maxArea) ∧ swrlb:subtract(?simResArea, 1, ?divArea)
   ∧ swrlb:subtract(?subMeanIntensity, ?valMeanIntensity-q,
     ?valMeanIntensity-r) ∧ swrlb:abs(?absMeanIntensity,
     ?subMeanIntensity) ∧ swrlb:divide(?divMeanIntensity,
           ?absMeanIntensity, ?maxMeanIntensity) ∧
 swrlb:subtract(?simResMeanIntensity, 1, ?divMeanIntensity) ∧
   swrlb:add(?sumMid, ?simResArea, ?simResMeanIntensity) ∧
            swrlb:add(?sumSim, ?sumHigh, ?sumMid)
                              →
 sqwrl:select(?m-q, ?m-r, ?cs-q, ?cs-r, ?cm-q, ?cm-r, ?cd-q,
?cd-r, ?sumHigh, ?valArea-q, ?valArea-r, ?valMeanIntensity-q,
       ?valMeanIntensity-r, ?sumMid, ?sumSim) ∧ sqwrl:
                 orderByDescending(?sumSim)
```

Figure 4.6 The SQWRL rule to retrieve similar masses for a given mass. (Cont.)

To test our approach, we processed a list of queries. Table 4.4 shows the results of three sample queries. In Query-A and Query-B, all high-level features of the query masses are equal. Both of the queries mass has irregular shape, spiculated margin, high density and BI-RADS 5. But there is a significant difference between their mid-level features, where area of the mass in the Query-B is larger than Query-A. Therefore, when we look at the query results, we can see that, the size of the masses in result list of the Query-A is smaller than the masses in result list of Query-B. In other words, mid-level feature helps to rank result set, effectively. And we query a mass with oval shape, circumscribed margin, equal density and BI-RADS 2 in Query-C.

Table 4.4 Sample mass queries with their results

| Query-A | Result-A.1 | Result-A.2 | Result-A.3 |
|---|---|---|---|
|  |  |  |  |
| **Query-B** | **Result-B.1** | **Result-B.2** | **Result-B.3** |
|  |  |  |  |
| **Query-C** | **Result-C.1** | **Result-C.2** | **Result-C.3** |
|  |  |  |  |

## 4.6 Performance Effect of Low Level Image Features to Content based Image Retrieval of Breast Masses

To evaluate performance effect of low-level image features to CBIR of breast masses, we developed an experimental system and use DEMS (Dokuz Eylul Mammography Set) which is a mammogram dataset and contains fully-annotated digital mammograms (Dokuz Eylul University, 2012). DEMS is compliant with the state-of-the-art semantic-web knowledge representation technologies, and case selection has been performed in two stages. In first step, candidate cases were selected retrospectively from PACS system of Radiology Department of Dokuz

Eylul University Medical Faculty Hospital, among more than 50K mammography examination diagnosed between 2004 January and 2008 November. Each candidate case includes four images in DICOM format, which are CC and MLO views of both breasts. The dataset were manually anonymized by removing all patients and physicians IDs. DEMS contains 485 mammographic cases where 255 of them contain one or more abnormality and 260 mass annotations in total. DEMS case images and the features (i.e., high, mid, low-level) are all downloadable at http://demir.cs.deu.edu.tr/index.php/downloads.

In the dataset, all cases are examined by two radiologists and the lesions are marked by a radiologist who has more than 25 years experience on mammography interpretation. We call these marks as ROI (Region of interest) of the mass. In Figure 4.7, first column shows sample masses, and second column shows their ROI's. All calculated features (i.e., mid and low-level features) are extracted directly from the ROI of the masses. The last column indicates the region of the masses in black and white.

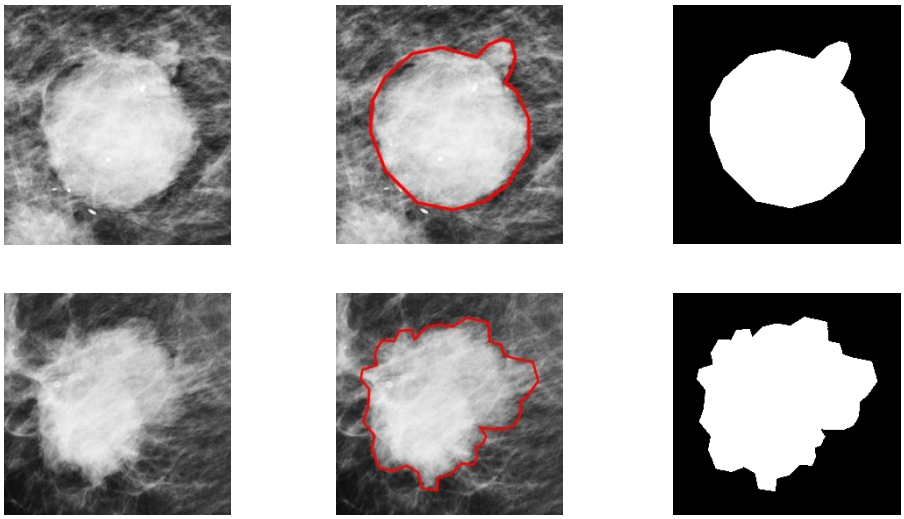


Figure 4.7 Sample masses in DEMS.

According to ACR Mammography Atlas, each mass should have a single BI-RADS score which can be 2, 3, 4A, 4B, 4C, 5 and 6. As the score increases, probability of being malignancy increases as well. The test set contains 4 query

masses from each BI-RADS score, 28 query masses in total. In Equation 13, query set is represented as Q where an arbitrary qi represents a query mass.

$$Q = \{q_1, q_2, q_3, ..., q_{28}\} \tag{13}$$

Then, we run our CBIR system for each query mass, qi, and it produces initial result list, riin, where, each riin, contains k result masses, mi, shown in Equation 15.

$$r_i^{in} = \{m_1, m_2, m_3, ..., m_k\} \tag{14}$$

After that, we ask two expert radiologists, a and b, to judge if masses are relevant to given query, qi, in riin. And, we obtain ria and rib , shown in equation 17 and 19, where length of the lists, x and y, are changed according to expert's judgment.

$$r_i^a = \{m_1, m_2, m_3, ..., m_x\} \tag{15}$$

$$r_i^b = \{m_1, m_2, m_3, ..., m_y\} \tag{16}$$

Finally, we consolidated the two sets ria and rib to obtain final relevant list ri by taking intersection of them. Equations 17 shows the final relevant set for query qi.

$$r_i = \{m_1, m_2, m_3, ..., m_n\} \tag{17}$$

We evaluate our system using these relevance lists in terms of P@10, Precision and Recall metrics. To distinguish the power of each feature type (i.e., low, mid and high), we set up a series of experimentations for different combinations of them. These are HL, LL, ML+LL, HL+ML and HL+ML+LL, where '+' represents combination of individual feature type (e.g. HL+LL means that using high and low-level features together).

Figure 4.8 shows the results of P@10 values obtained from experimentations. Average P@10 values for HL, LL, ML+LL, HL+ML and HL+ML+LL are 0.29, 0.13, 0.21, 0.43 and 0.71, respectively. It shows that, using high, mid and low-level features all together produces the best P@10 values.

Figure 4.9 shows Precision versus Recall graph, where it is clearly seen that, when we use computer-calculated features alone or together (LL and ML+LL), their performance is worse than the performance of using HL features alone. But, when we add some computer-calculated features to CBIR system, which use only HL features, performance is increased (HL+ML). Finally, using HL features with LL and ML features have the best performance on CBIR of breast masses.

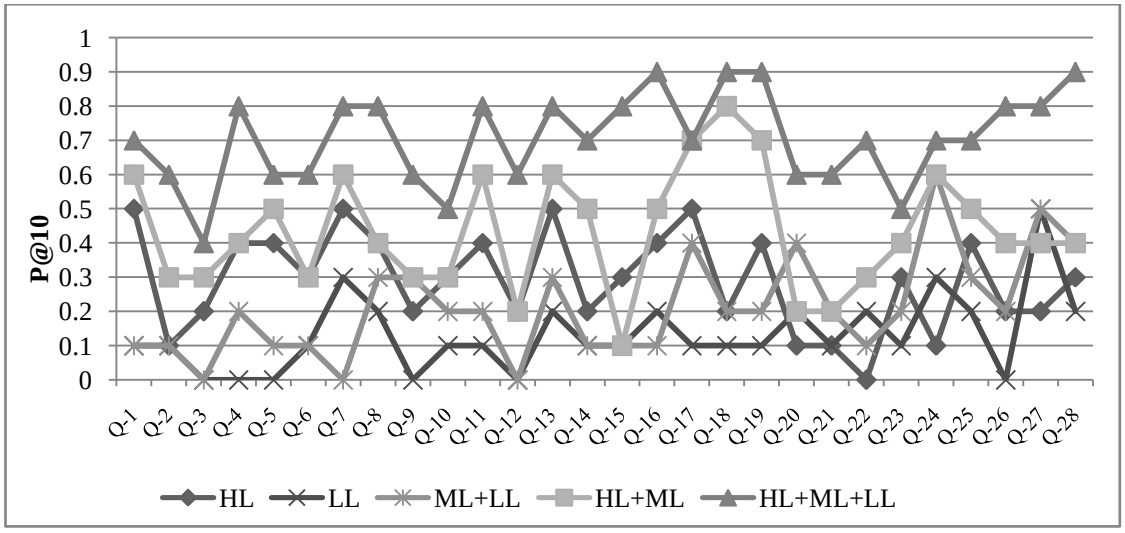



Figure 4.8 P@10 values for individual query IDs.

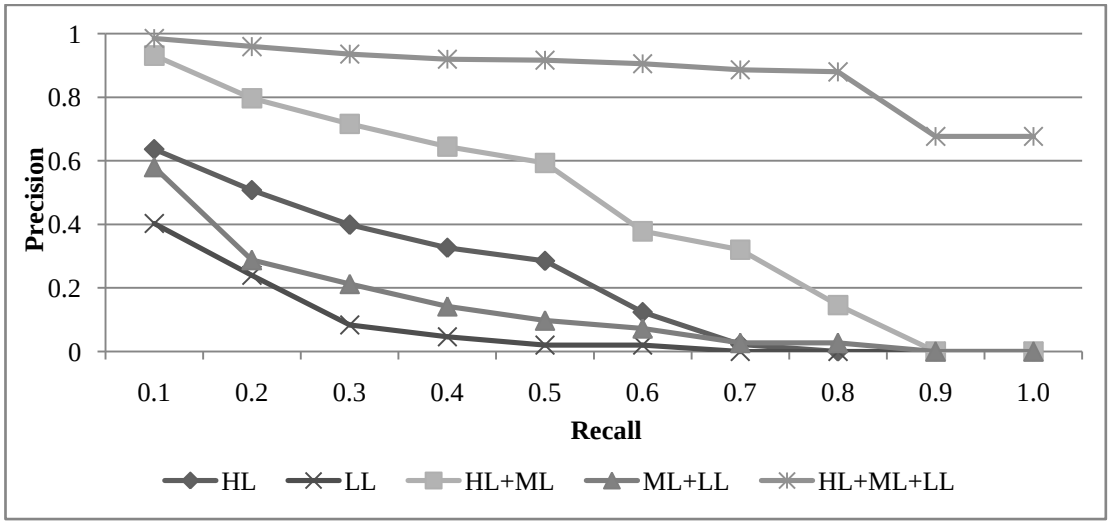



Figure 4.9 Precision vs Recall graph.

# CHAPTER FIVE
# UNCERTAINTY MODELING FOR ONTOLOGY-BASED MAMMOGRAPHY ANNOTATION WITH INTELLIGENT BI-RADS SCORING

## 5.1 Overview

In 1904, Sir William Osler mused that "Medicine is a science of uncertainty and an art of probability". As time passed, the emergence of science in medicine has done much in the last century to reduce the uncertainty surrounding of medicine. Apparently, even a simple search in PubMed, using the terms medical and uncertainty, results more than a thousand of recent articles. It evidently shows that it is still an active research area and many researchers still work on to reduce uncertainty. On the other hand, evidence-based medicine aims to provide ways to quantify and communicate uncertainty from a probabilistic way. Nevertheless, uncertainty remains in the nature of medicine as in the very famous quote of Sir William Osler.

A major weakness of usual ontological technologies is their inability to represent and to reason with uncertainty and imprecision (Hudelot et al., 2008). However, medicine, being a science whose subject is people, is inherently a science of certainty, and mostly deals with uncertain knowledge and imprecise and vague information. In order to achieve maximum advantage from ontologies, we need an extension of ontologies, which has the capability of capturing uncertainty knowledge about concepts, properties and relations in domains and of supporting reasoning with inaccurate information. Along this direction, researchers have attempted in the past to use different approaches on modeling uncertainty in ontologies.

In this chapter, we propose a new ontology-based mammography annotation system with a capability of uncertainty modeling in ontologies. To achieve this, we use Bayesian probability-based approach, without extending description logic and ontology languages. In addition, we also propose a rule based BI-RADS score reasoning approach using Semantic Web technology of SQWRL (Semantic Query-enhanced Web Rule Language).

## 5.2 Background and Literature Survey

The current state–of–the–art in image retrieval has two major approaches: Content–Based Image Retrieval (CBIR) and Annotation–Based Image Retrieval (ABIR). CBIR methods only operate on images by the extraction of visual primitives such as color, texture or shape. However, there is an important shortcoming of this approach: it is not possible to extract all semantic information from images alone, which is known as the 'semantic gap'. For mammography some CBIR systems were developed for retrieving and classification (Oliveira et al., 2010; Kinoshita et al., 2007; Oliver et al., 2008; Castella et al., 2007). The second approach, ABIR, mostly deals with high-level semantic annotations, which are generally performed by humans. Such, Ontology-based Annotation and Retrieval is a kind of ABIR approach.

### *5.2.1 Ontology-based Annotation and Retrieval of Mammograms*

In terms of computer science, ontologies are state-of-the-art method to represent knowledge and become more important when annotating an image. Ontology includes a set of concepts and the relationships between the concepts. We can divide the ontologies into two main types; *upper ontology* and *domain ontology*. Upper ontologies model the common objects, which are generally used in the domain ontologies while the domain ontologies model a specific domain or part of the world. Domain ontologies generally provide a shared vocabulary. Main role of these vocabularies is to help data integration by representing the knowledge and to aid decision-making processes. In that respect, ontologies are important for health care systems.

In literature, there is some research that suggests a framework for ontology-based medical image annotation and retrieval as an approach to reduce the occurrence of irrelevant resource retrieval in a medical imaging information system. Hu et al. built a semantically rich system by accommodating image annotation and retrieval services around a rigidly defined ontology for medical images used in breast cancer treatment, in 2003. They developed the Breast Cancer Imaging Ontology (BCIO) to

provide a commonly agreed vocabulary with formal definitions that can be used to represent breast X-ray and MRI images, abnormal findings and medical assessments (Hu et al., 2003). In 2006, Qi et al. developed mammography ontology and used an ontology-based comparison method for finding groups of diagnosis that radiologists detect using the same analysis process based on edit distance, which is a similarity measurement between two concepts (Qi et al., 2006). Ren and Barnaghi suggested a framework for medical specialists to be able to annotate digital mammograms and to retrieve relevant resources based on semantic relations, in 2007 (Ren & Barnaghi, 2007). In 2008, Rubin et al. developed a generalized ontology-based annotation and retrieval framework, which is called Annotation and Imaging Markup (AIM) (Rubin et al., 2008). Shanbolt et al. developed an ontology-based knowledge management system which is called MIAKT (Medical Imaging with Advanced Knowledge Technologies) for the data that the screening process generates, as well as providing a means for medical staff to investigate, annotate and analyze the using web, in 2004 (Shadbolt et al., 2007). And, in 2012, we have proposed a system for Ontology-based annotation and retrieval of breast masses (Bulu et al., 2012).

### *5.2.2 BI-RADS Scoring and Mass Descriptors in Mammography*

In mammography, Breast Imaging Reporting and Data System (BI-RADS) is a standard for rating mammograms and breast ultrasound images, so it represents the radiologist's final opinion of the absence or likelihood of breast cancer. The scores are assigned to abnormalities to give quick information about malignancy percentage of the abnormalities. It was developed by the American College of Radiologists (ACR)and the scores change from 0 to 6. BI-RADS scores are highly related to the probability of malignancy. Thus, BI-RADS score of a mass increases as probability of the malignancy increases. In other words, BI-RADS scoring is a decision making process for malignancy probability of the mass. For example, ACR BI-RADS atlas states that BI-RADS score is 4A, if the probability of malignancy is between 3% and 29%. Table 5.1 shows the meaning of each score (The American College of Radiology, 2012).

Table 5.1 Breast Imaging Reporting and Database System (BI-RADS)

| Category | Diagnosis | Number of Criteria |
|---|---|---|
| 0 | Incomplete | Mammogram or ultrasound didn't give the radiologist enough information to make a clear diagnosis; follow-up imaging is necessary. |
| 1 | Negative | There is nothing to comment on; routine screening recommended. |
| 2 | Benign | A definite benign finding; routine screening recommended. |
| 3 | Probably Benign | Findings that have a high probability of being benign (>98%) |
| 4 | Suspicious Abnormality | Not characteristic of breast cancer, but reasonable probability of being malignant;<br>4A: Finding needing intervention with a low suspicion for malignancy. Probability of being malignant (3 to 29%)<br>4B: Lesions with an intermediate suspicion of malignancy. Probability of being malignant (30 to 59%)<br>4C: Findings of moderate concern, but not classic for malignancy. Probability of being malignant (60 to 94%)<br>Biopsy should be considered. |
| 5 | Highly Suspicious of Malignancy | Lesion that has a high probability of being malignant (>= 95%); take appropriate action. |
| 6 | Known Biopsy Proven Malignancy | Lesions known to be malignant that are being imaged prior to definitive treatment; assure that treatment is completed. |

On the other hand, relation between mass descriptors and morphology is highly correlated. Most descriptors of breast masses that define malignancy are related to mass morphology. For instance, malign masses tend to spread to other areas, while benign masses remain stable. As a result, malign masses commonly form irregular shapes; conversely benign masses commonly form regular shapes. Similarly, contour of the masses becomes uncertain and shows spicules while malignancy of mass increases. Figure 5.1 shows relationships between mass descriptors and morphology. If BI-RADS scores are related with probability of malignancy of the mass like mass descriptors we deduce that BI-RADS scores are related with mass descriptors.

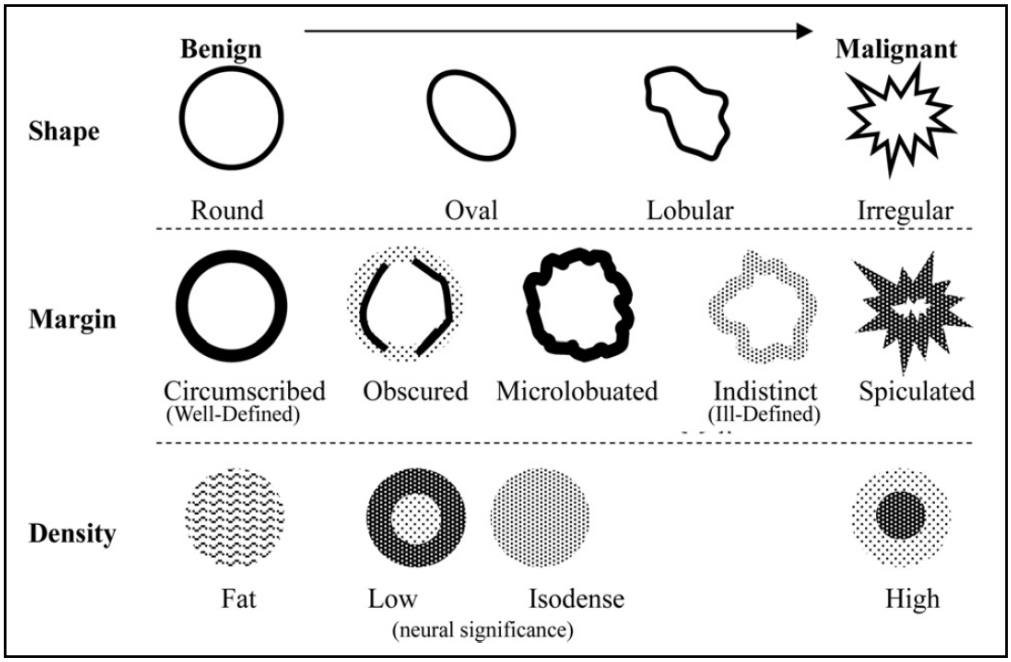



Figure 5.1 Relations between mass descriptors and morphology (Wei et al., 2012).

### *5.2.3 Uncertainty in Mammography Interpretation*

Mammography interpretation is a subjective process and even same expert may not have certain interpretation about an abnormality, in time. There is always a trade-off between objectivity and relevance in annotation process. For instance, an expert may have difficulty when describing an abnormality if its shape is oval or round. Naturally, when we face an uncertain situation we tend to use more than one attribute together with different percentages. (e.g., 60% round and 40% oval).

Besides, there is uncertainty between mass descriptors and BI-RADS scores. In other words, possibility of the BI-RADS scores are not same for each mass descriptors, while some of the descriptors tend to get higher BI-RADS scores, others tend to get lower. However, current Semantic Web technologies are based on crisp logic where the relations are binary and they are unable to represent these uncertainties in mammography. Thus, if we want to represent expert knowledge with respect to mass descriptors and BI-RADS scores, we develop the ontology by using current Semantic Web technologies, shown in Figure 5.3-a. Here, it is important that there is no relation between `Oval Shape` and `BI-RADS 5` and `6`, as well as

there is no relation between `BI-RADS 2` and `3` with `Indistinct Margin`. In this study, we propose an approach to deal with these kinds of uncertainty problems in ontologies.

### *5.2.4 Uncertainty Modeling in Ontologies*

The inherent way to model uncertainty is to use the probabilities. In literature, probability has been already used to model uncertainty in classification (Kahn et al., 1997; Fischer et al., 2004; Burnside et al., 2000). Today, OWL (Web Ontology Language) (W3C, 22.03.2013) is the standard web ontology language to model the knowledge. However, OWL is based on crisp logic and it cannot manage uncertainty without complete knowledge about an application domain. Moreover, reasoning in the Semantic Web is a deterministic process of verifying if statements are true or false. However, in real world this is not always possible to represent the whole knowledge with crisp logic. The studies in literature, aiming to model uncertainty in ontologies, can be grouped in to two categories: (1) works proposing an extension to description logic and (2) not proposing an extension.

First group, some works extend the syntax of description logic with varying degrees. They propose solutions in terminological knowledge and/or assertional knowledge by adding probability. In description logic, knowledge is represented in two ways: TBox and ABox. TBox contains terminological (intensional) knowledge in the form of a terminology and is built through declarations that describe general properties of concepts. ABox contains assertional (extensional) knowledge, which is specific to the individuals of the discourse domain. In other words, TBox contains definitions of concepts and roles, while ABox contains definitions of individuals (instances). Some of the works suggest an extension in ABox (Dürig & Studer, 2005; Dürig & Studer, 2008) while some suggest TBox only (Heinsohn, 1994; Kolleret al., 1997) and some researches propose both in ABox and TBox (Lukasiewiczal, 2008; Jaeger, 1994).

The second group includes studies not proposing an extension on description logic. Some of the works focus on embedding the probabilities into the OWL (Yang

& Calmet, 2005; Nottelmann & Fuhr, 2006; Costa & Laskey, 2006; Ding et al., 2006), while some focus on the RDF (Resource Description Framework) (W3C, 22.03.2013; Fukushige, 2004; Udrea et al., 2006). Our proposal falls into second group, where it does not require any extension neither in description logic nor in OWL syntax. We use only the state-of-the-art Semantic Web technologies to model uncertainty.

## 5.3 Uncertainty Modeling with Bayesian Probability in Ontologies

In probability and statistics, an outcome is a particular result of the experiment, which is also called Random Variables. Random variables can be discrete or continuous. If a random variable can take only a finite number of distinct values, then it must be discrete, such as the "number of student in a class" or "one of six outcomes from rolling a dice". Discrete random variables are usually (but not necessarily) counts. Boolean-valued random variable if it denotes an event (a hypothesis) and there is some degree of uncertainty as to whether *A* occurs. For example, "Tomorrow will be a rainy day" or "You will enjoy this film". When the variable can assume uncountable infinite values in a given range of values is called Continuous Random Variables. Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile (Bernardo & Smith, 2001).

For given two events *A* and *B*, P(*A*) or P(*B*), i.e. probabilities of each of the events occurring without the knowledge of the other events occurrence is called the prior probability. On the other hand, given two events *A* and *B*, the probability of event *A* occurring given that event *B* has occurred is called the conditional probability.

In our case, each abnormality observed in a mammography has a BI-RADS score. Radiologists assign a BI-RADS score to an abnormality based on its descriptors (i.e. shape, margin, density etc.). In this point of view, description and BI-RADS score assignment are two related but distinct actions. However these distinct actions need to be connected in some way. So, Bayes theorem suits well to this issue, because the theorem makes it possible to calculate probability of conditional events. The

Bayesian interpretation of probability can be seen as an extension of logic that enables reasoning with uncertain statements. The notation of conditional probability is P($A$|$B$), where $A$ and $B$ are random variables, and the notation can read as "the probability of A for B". Let us assume that A represents being of oval mass shape, B represents being BI-RADS score of 3. Then, the probability of oval mass shape for given BI-RADS score of 3, P($A$|$B$), is defined as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{18}$$

where $P(B) > 0$.

In our dataset, we have totally 260 mass annotations, and 40 of them have oval shape. With respect to their BI-RADS score, the numbers of oval masses are 21, 15 and 4 for BI-RADS 2, BI-RADS 3 and BI-RADS 4, respectively. Here, the notation P(oval|birads-3)=0.27 shows the probability of being oval shape under the condition for BI-RADS 3. According to Table 5.2, detail of this calculation as follows;

$$P(\text{oval}|\text{birads-3}) = \frac{P(\text{oval} \cap \text{birads-3})}{P(\text{birads-3})} = \frac{15/260}{55/260} = \frac{15}{55} \cong 0.27$$

As a result, by using Table 5.2 and Equation (18), we calculate probability of each mass feature for related BI-RADS score, which are shown in Table 5.3. So, we augment the relations between BI-RADS scores and mass descriptors by adding probability, as it shown in Figure 5.2-b. Now, representation of our ontology is a weighted graph, and it makes probabilistic inference possible for BI-RADS scores.

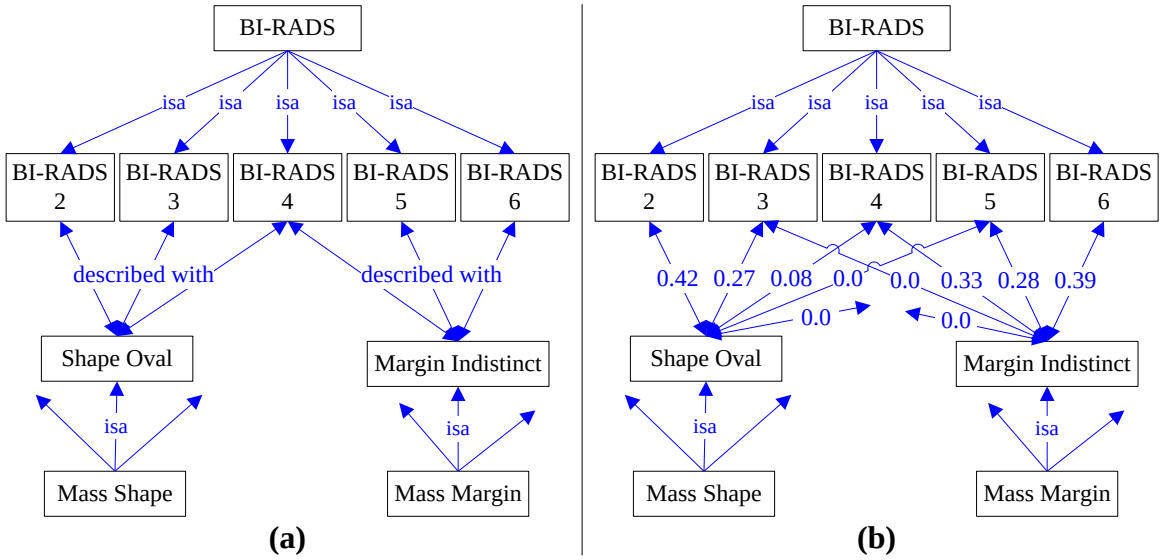



Figure 5.2 Relationships between BI-RADS scores and mass descriptors based on;(a) crisp logic (b) non-crisp logic.

For a given mass with already annotated with the descriptors, we calculate probability of the BI-RADS score as follows;

$$P(B) = P(S|\mathrm{B}) \times P(\mathrm{M}|\mathrm{B}) \times P(\mathrm{D}|\mathrm{B})$$

where B, S, M and D represents BI-RADS, shape, margin and density, respectively.

Table 5.3 shows conditional probability values of mass descriptors for BI-RADS scores. For a better understanding, a mass shown in Figure 5.3, with irregular shape, spiculated margin and high density, probability of being BI-RADS 5 is calculated as follows:

$$P(\mathrm{birads}-5) = P(\mathrm{irregular} \mid \mathrm{birads}-5) \times P(\mathrm{spiculated} \mid \mathrm{birads}-5) \times P(\mathrm{high} \mid \mathrm{birads}-5) =$$

$$\frac{P(\mathrm{irregular} \cap \mathrm{birads}-5)}{P(\mathrm{birads}-5)} \times \frac{P(\mathrm{spiculated} \cap \mathrm{birads}-5)}{P(\mathrm{birads}-5)} \times \frac{P(\mathrm{high} \cap \mathrm{birads}-5)}{P(\mathrm{birads}-5)} =$$

$$\frac{67/260}{75/260} \times \frac{54/260}{75/260} \times \frac{52/260}{75/260} = \frac{67}{75} \times \frac{54}{75} \times \frac{52}{75} = 0.89 \times 0.72 \times 0.69 \cong 0.45$$

As a result of above calculation, we can infer that probability of BI-RADS 5 is nearly 0.45, while probability of BI-RADS 2 nearly 0 for the annotated mass, shown in Figure 5.3. Accordingly, the calculation is equal to Naive Bayes classifier, which is a simple probabilistic classifier, based on applying Bayes' theorem, Equation 19.

$$classify(f_1,..., f_n) = \arg\max_c P(C = c)\prod_{i=1}^{n} P(F_i = f_i \mid C = c) \quad (19)$$

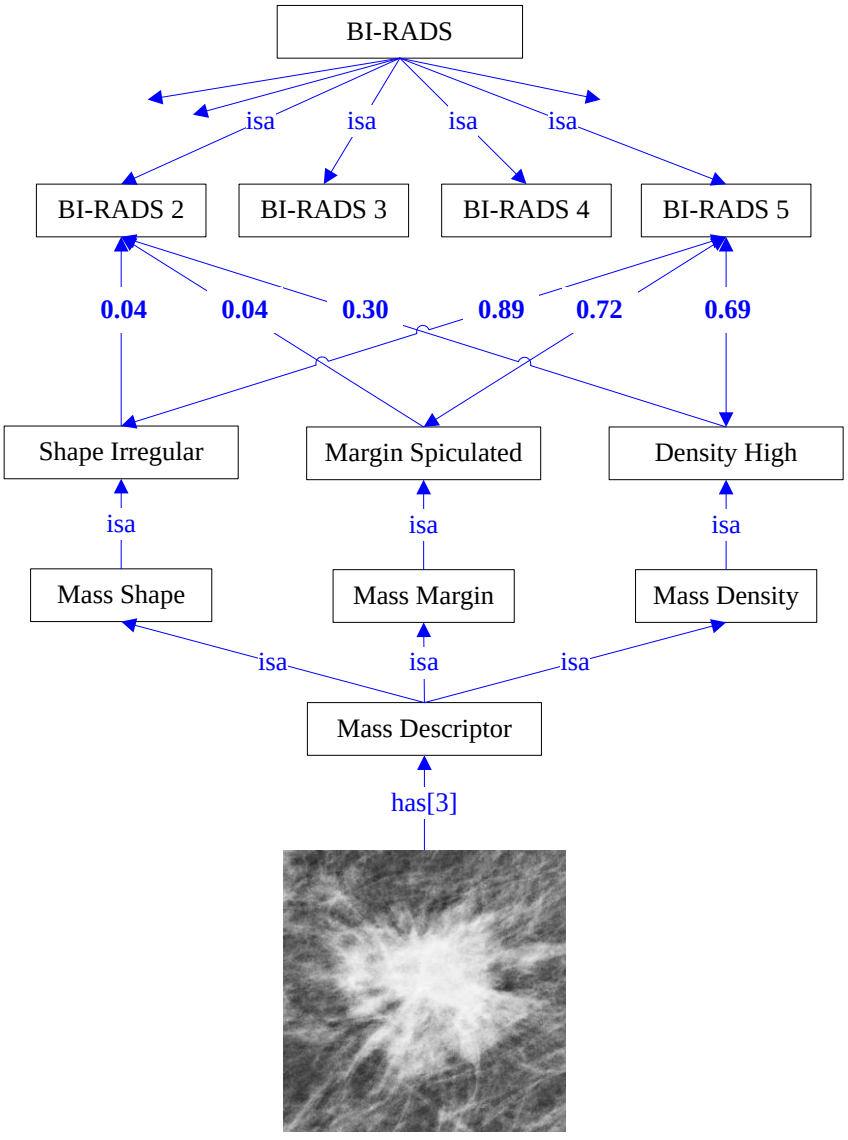



Figure 5.3 Probability of BI-RADS scores 2 and 5 for a mass with irregular shape, spiculated margin and high density.

## 5.4 Intelligent BI-RADS Scoring with SQWRL

BI-RADS scoring is a decision making process to assign predefined grades for a given abnormality with different characteristics and based on previous experiences of radiologist. In our approach, for a given abnormality, the annotation tool asks to define or choose a set of predefined descriptors based on MAO. Then, the tool automatically suggests a BI-RADS score for the given abnormality using Bayesian probability of selected mass descriptors based on previous cases. We call this process as intelligent BI-RADS scoring.

We propose two approaches to handle uncertainty in MAO. First approach keeps the probability values between mass descriptors and BI-RADS instances in OWL annotation properties. The second approach does not require any additional class or property to store. Instead, the probability scores calculated dynamically from knowledge base of previous cases, during BI-RADS scoring process. We use OWL and SQWRL to perform this kind of reasoning about probability of BI-RADS scores for a particular abnormality.

SWRL is an expressive OWL-based rule language. SWRL makes it possible to write inference rules and this provides more powerful reasoning capabilities than OWL alone. Semantically, SWRL is built on the same description logic foundation like OWL and it provides similar formal guarantees when performing inference (Protege, 02.09.2013). For example, in mammography, if a mammographic case (examination) contains one or more abnormality, its BI-RADS score is set to the maximum BI-RADS score of its abnormalities. So, we use the following SWRL rule to infer and set BI-RADS score of the whole case (`MammoCase`) depending on the BI-RADS score of the abnormalities it contains;

```
MammoCase(?mc) ∧ hasBreast(?mc, ?b) ∧ hasImage(?b, ?i)
∧ hasRoi(?i, ?r) ∧ containsAbnormality(?r, ?a) ∧
hasMassDescriptor(?a, ?md) ∧
hasBiradsMassDescriptor(?md, ?br) ∧ BI-RADS(?br) ∧
biradsScore(?br, ?bs) ˚ sqwrl:makeBag(?bag, ?bs) ˚
```

```
sqwrl:max(?maxbs, ?bag) ∧ swrlb:equal(?bs, ?maxbs) →
hasBiradsMammoCase(?mc, ?br)
```

On the other hand, SQWRL is built on SWRL and it takes a standard SWRL rule. Both of them have an antecedent part, which is referred to as the body, and a consequent part, which is referred to as the head. Differently, SQWRL replaces the rule consequent with a retrieval specification, so it provides SQL-like operations to retrieve knowledge from OWL (O'Connor & Das, 2009). For example, the following query retrieves all masses with mean intensity value is higher than 180;

```
Mass(?mass) ∧ containsAnAbnormality(?roi, ?mass) ∧
hasMidLevelDescriptor(?roi, ?mid) ∧ roiIntensity(?mid,
?inten) ∧ swrlb:greaterThan(?inten, 180) →
sqwrl:select(?mass, ?inten)
```

In the first approach, to model uncertainty we store conditional probability values in OWL annotation properties that represent relationships between concepts. There are two main types of properties, *object properties* and *datatype properties*. Object properties are relationships between individuals. *Datatype* properties link an individual to an XML Schema Datatype value or an RDF literal. In other words, they describe relationships between an individual and data values. OWL also has another type of property; *annotation properties* which is used to add extra information (metadata or data about data) to classes, individuals and object/datatype properties. OWL has five pre-defined annotation properties that can be used to annotate classes, properties and individuals: owl:versionInfo, rdfs:label, rdfs:comment, rdfs:seeAlso, rdfs:isDefinedBy.

In this approach we use object type annotation properties between BI-RADS and mass descriptors. And we use *rdfs:label* to store conditional probability value between them. We create 65 OWL annotation properties between 5 `BI-RADS` and 13 `MassDescriptor` classes. Figure 5.4 shows one of them, between mass descriptor oval and BI-RADS 3.

```
<owl:AnnotationPropertyrdf:ID="CondPropOvalBI-RADS3”>
<rdfs:domainrdf:resource="#ShapeOval"/>
<rdfs:rangerdf:resource="#BI-RADS3"/>
<rdfs:labelrdf:datatype="http://www.w3.org/2001/XMLSchema#string">
0.27
</rdfs:label>
<rdf:typerdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"
/>
</owl:AnnotationProperty>
```

Figure 5.4 OWL Syntax of sample annotation property.

It is possible to access label of annotation properties by using SQWRL with *rdfb:hasLabel* function. But, type of the label is string and it cannot be used in mathematical calculations, so we convert these values to float before use in calculations. Then, according to given mass, we get the labels and we infer probability values for each BI-RADS score. We use SWRL Tab of Protégé 3.4.5 (Stanford Center for Biomedical Informatics Research, n.d.) to execute SWRL and SQWRL rules, with Jess Rule Engine (Friedman-Hill, n.d.). We infer the probability value of each BI-RADS score by executing the SQWRL rule in Figure 5.5. The rule uses the descriptor of mass *?q* with irregular shape, spiculated margin and high density. Furthermore, Figure 5.6 shows screen-shot of Protégé SWRL tab where the results are shown in bottom half of the figure and the probability of BI-RADS 5 has the highest value, 0.45.

```
                    Mass(?mass) ∧
                abox:hasURI(?q, "http://www.owl-
ontologies.com/Ontology1299746090.owl#MassTest") ∧ hasMassDescriptor(?q,
  ?md) ∧ abox:hasClass(?md, ?mdc) ∧ tbox:isProperty(?annotationProp) ∧
  tbox:isAnnotationProperty(?annotationProp) ∧ tbox:isInDomainOf(?mdc,
     ?annotationProp) ∧ tbox:isInRangeOf(?range, ?annotationProp) ∧
   rdfb:hasLabel(?annotationProp, ?label) ∧ stringValue(?s2d, ?str) ∧
     floatValue(?s2d, ?doubleLabel) ∧ swrlb:matches(?str, ?label) ˚
   sqwrl:makeBag(?bag, ?doubleLabel) ∧ sqwrl:groupBy(?bag, ?range) ˚
      sqwrl:nth(?first, ?bag, 1) ∧ sqwrl:nth(?second, ?bag, 2) ∧
    sqwrl:nth(?third, ?bag, 3) ∧swrlb:multiply(?probability, ?first,
      ?second, ?third) ∧ swrlb:multiply(?h, ?probability, 100) ∧
  swrlb:round(?r, ?h) ∧ swrlb:divide(?roundedProbability, ?r, 100) →
      sqwrl:select(?mass, ?mdc, ?annotationProp, ?range, ?label,
 ?roundedProbability) ∧ sqwrl:orderBy(?range, ?mdc, ?annotationProp)
```

Figure 5.5 SQWRL rule for inference of BI-RADS probability for a mass.

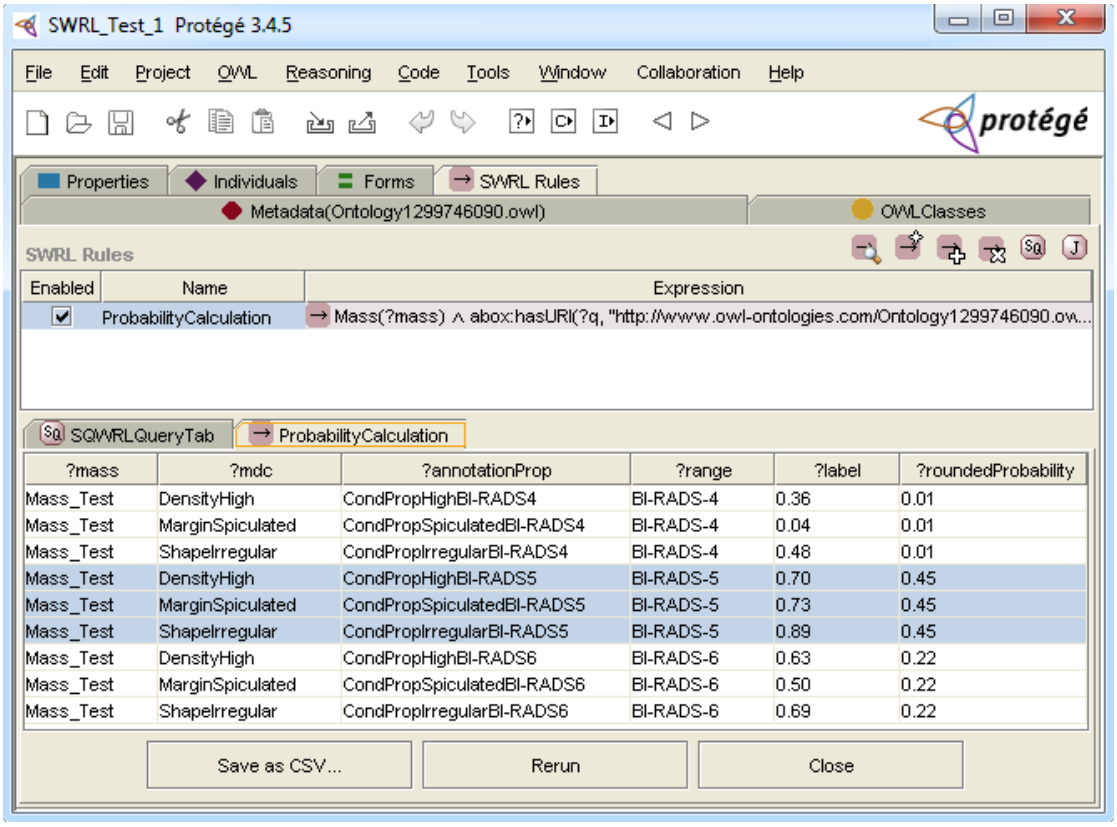



Figure 5.6 SWRL tab of Protégé.

Alternatively, it is also possible to calculate conditional probability value between BI-RADS and mass descriptor without using any additional classes or properties in MAO. In this case, the probability values are calculated dynamically by using SQWRL. Consequently, probability values change when the new instance is added into instance knowledge base. As a result, the accuracy of the whole system will be improved as time changes and as more instances are populated into the knowledge base. However, scanning the whole knowledge base is very time consuming and it is far more efficient with current Semantic Web technologies.

## 5.5 Experimentations

To evaluate our proposal on uncertainty modeling in ontologies, we set up experimentations on two mammography datasets: DEMS (Dokuz Eylul University Mammography Set) (Dokuz Eylul University, 2012) and DDSM (Digital Database for Screening Mammography) (Heath et al., 2001).

DEMS is a fully-annotated digital mammogram dataset, where cases were retrospectively selected from the PACS system of Radio-diagnostics Department of Dokuz Eylul University Hospital among 50K mammography cases examined during from 2004 to 2008. Then, all cases in the dataset were annotated in three phases using MAO. Hence, all masses have shape, margin, density and BI-RADS values individually. The resulting dataset has a total number of 485 mammography cases where 255 of them contain one or more abnormality and 260 mass annotations in total.

DDSM is developed by co-operation of Massachusetts General Hospital, University of South Florida (USF), American Sandia National Laboratories and the U.S. Army Medical Research and Material Unit Breast Cancer Research Program's fund. Each case in the dataset contains two standard views (CC and MLO) of two breasts and is selected from patients diagnosed between October 1988 and February 1999 at Massachusetts General Hospital, Wake Forest University School of Medicine, St. Sacred Heart Hospital and Washington St. Louis University School of Medicine. The dataset has a total number of 2620 studies. However, DDSM does not include density description for masses. In our study, we used 2235 masses in DDSM annotated with shape, margin and BI-RADS values.

Table 5.2 shows the distribution of masses in DEMS based on their BI-RADS scores. For example, we have 55 masses with BI-RADS 3, where 25 of them have round shape, 15 of them have oval shape, 15 of them have lobular shape and none of them has irregular shape.

Table 5.2 Distribution of masses in DEMS based on their BI-RADS scores.

| | Descriptors | BI-RADS Scores | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | **2** | **3** | **4** | **5** | **6** | |
| **Shape** | *Round* | 15 | 25 | 4 | 0 | 6 | **50** |
| | *Oval* | 21 | 15 | 4 | 0 | 0 | **40** |
| | *Lobular* | 12 | 15 | 18 | 8 | 4 | **57** |
| | *Irregular* | 2 | 0 | 23 | 67 | 21 | **113** |
| **Margin** | *Circumscribed* | 44 | 50 | 4 | 0 | 0 | **98** |
| | *Obscured* | 4 | 5 | 18 | 0 | 4 | **31** |
| | *Microlobulated* | 0 | 0 | 9 | 0 | 0 | **9** |
| | *Indistinct/Ill-Defined* | 0 | 0 | 16 | 21 | 12 | **49** |
| | *Spiculated* | 2 | 0 | 2 | 54 | 15 | **73** |
| **Density** | *Fat-Containing Radiolucent* | 24 | 0 | 0 | 0 | 0 | **24** |
| | *Low Density* | 2 | 4 | 0 | 0 | 0 | **6** |
| | *Equal Density/Isodence* | 9 | 36 | 32 | 23 | 11 | **111** |
| | *High-Density* | 15 | 15 | 17 | 52 | 20 | **119** |
| **Total** | | **50** | **55** | **49** | **75** | **31** | **260** |

As we already mentioned above, we calculate all conditional probability values of mass descriptors for BI-RADS scores by using Equation (18) with frequencies in Table 5.2. As a result we obtain Table 5.3 showing the conditional probabilities for DEMS.

Table 5.3 Conditional probability values of mass descriptors for BI-RADS scores in DEMS.

| | Descriptors | BI-RADS Scores | | | | |
|---|---|---|---|---|---|---|
| | | **2** | **3** | **4** | **5** | **6** |
| **Shape** | *Round* | 0.30 | 0.45 | 0.08 | 0.00 | 0.19 |
| | *Oval* | 0.42 | 0.27 | 0.08 | 0.00 | 0.00 |
| | *Lobular* | 0.24 | 0.27 | 0.37 | 0.11 | 0.13 |
| | *Irregular* | 0.04 | 0.00 | 0.47 | 0.89 | 0.68 |
| **Margin** | *Circumscribed* | 0.88 | 0.91 | 0.08 | 0.00 | 0.00 |
| | *Obscured* | 0.08 | 0.09 | 0.37 | 0.00 | 0.13 |
| | *Microlobulated* | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 |
| | *Indistinct/Ill-Defined* | 0.00 | 0.00 | 0.33 | 0.28 | 0.39 |
| | *Spiculated* | 0.04 | 0.00 | 0.04 | 0.72 | 0.48 |
| **Density** | *Fat-Containing Radiolucent* | 0.48 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Low Density* | 0.04 | 0.07 | 0.00 | 0.00 | 0.00 |
| | *Equal Density/Isodence* | 0.18 | 0.65 | 0.65 | 0.31 | 0.35 |
| | *High-Density* | 0.30 | 0.27 | 0.35 | 0.69 | 0.65 |

We assume that the system learning abilities is equivalent to Bayesian Classifier, and test our approach with Naïve Bayesian classifier using Rapid Miner tool (Rapid-I, n.d.) and we obtained the confusion matrixes for DEMS and DDSM shown in Table 5.4 and Table 5.5,which is evaluated with10-fold cross-validation method. Average accuracy and sensitivity values for DEMS are 0.86, and 0.60, respectively. For DDSM, the values are 0.81 and 0.5.

Table 5.4 Confusion matrix of DEMS.

| | **BI-RADS** | **Expert Judgment** | | | | | **Precision** | **Accuracy** |
|---|---|---|---|---|---|---|---|---|
| | | **2** | **3** | **4** | **5** | **6** | | |
| **Class Prediction** | **2** | 26 | 7 | 0 | 0 | 0 | 0.78 | 0.88 |
| | **3** | 19 | 46 | 8 | 0 | 0 | 0.63 | 0.86 |
| | **4** | 3 | 2 | 28 | 7 | 4 | 0.63 | 0.85 |
| | **5** | 2 | 0 | 11 | 68 | 21 | 0.66 | 0.84 |
| | **6** | 0 | 0 | 2 | 0 | 6 | 0.75 | 0.89 |
| | **Sensitivity** | 0.52 | 0.84 | 0.57 | 0.91 | 0.19 | | |

Table 5.5 Confusion matrix of DDSM.

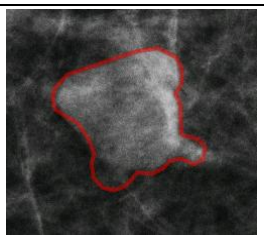
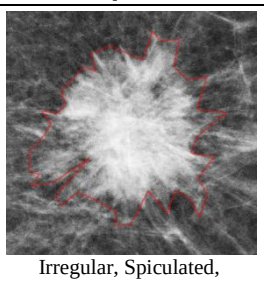
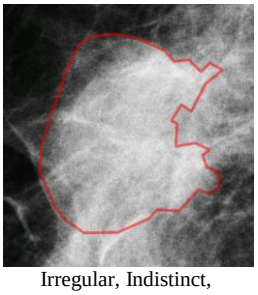
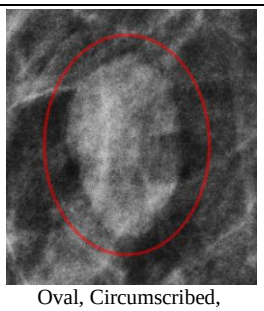
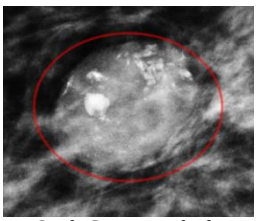
| | **BI-RADS** | **Expert Judgment** | | | | **Precision** | **Accuracy** |
|---|---|---|---|---|---|---|---|
| | | **2** | **3** | **4** | **5** | | |
| **Class Prediction** | **2** | 0 | 0 | 0 | 0 | 0 | 0.97 |
| | **3** | 51 | 337 | 298 | 27 | 0.47 | 0.76 |
| | **4** | 5 | 131 | 635 | 138 | 0.69 | 0.68 |
| | **5** | 0 | 28 | 127 | 458 | 0.74 | 0.85 |
| | **Sensitivity** | 0 | 0.67 | 0.59 | 0.73 | | |

Table 5.6 shows sample BI-RADS score reasoning results using both our approach and crisp logic where BI-RADS columns indicate the prediction value for related approach. We define a new measure showing the level of uncertainty, *U*. For this measure, 0 means there is no uncertainty, while 1 represents full uncertainty. Formally, the level of uncertainty is defined as follows;

$$U = 1 - \frac{1}{n}$$

where *n* represents the number of BI-RADS scores with the same maximum value. All these numbers are underlined in the Table 5.6. For example, the example mass in second row has irregular shape, spiculated margin and high density. Our approach produces the prediction values as, <0.0, 0.0, 0.0, 0.44, 0.21> for BI-RADS scores from 2 to 6, respectively. Here, the maximum probability value is 0.44 and it repeats once, so the value of *n* is 1 and *U* is zero. On the other hand, result of crisp logic are <0.0, 0.0, 1.0, 1.0, 1.0>, where *n* is equal to 3, and *U* is equal to 0.67. Average level of uncertainty for our approach and crisp logic are 0.04 and 0.53 for DEMS, respectively. For DDSM dataset, the average level of uncertainty values is 0.02 and 0.56, respectively.

Table 5.6 Example results in DEMS dataset.

| # | Image (Shape, Margin, Density, BI-RADS) | Our Approach | | | | | | Crisp Logic | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Prediction value for BI-RADS scores | | | | | *U* | Prediction value for BI-RADS scores | | | | | *U* |
| | | 2 | 3 | 4 | 5 | 6 | | 2 | 3 | 4 | 5 | 6 | |
| 1 |  Lobular, Circumscribed, Equal, 3 | 0.04 | 0.16 | 0.02 | 0.00 | 0.00 | **0.00** | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | **0.67** |
| 2 |  Irregular, Spiculated, High, 5 | 0.00 | 0.00 | 0.00 | 0.44 | 0.21 | **0.00** | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | **0.67** |
| 3 |  Irregular, Indistinct, High, 5 | 0.00 | 0.00 | 0.05 | 0.17 | 0.17 | **0.50** | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | **0.67** |
| 4 |  Oval, Circumscribed, Fat, 2 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| 5 |  Oval, Circumscribed, Equal, 2 | 0.07 | 0.16 | 0.00 | 0.00 | 0.00 | **0.00** | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | **0.50** |

# CHAPTER SIX
# ONTOLOGY-BASED CONTENT BASED IMAGE RETRIEVAL SYSTEM FOR BREAST MASSES BY USING XQUERY

## 6.1 Overview

As we mentioned in above sections our low-level features are computer calculated decimal vectors and we use Euclidean distance to compute similarity score between them. So, to be able to use them in our ontology-based CBIR system, we need to define our similarity functions. However, current Semantic Web Technologies does not support extendible similarity/comparison functions. The technologies can just provide basic comparison functions, such as; equal, smaller, higher etc. All these functions work with scalar values and return Boolean results. For example, result of the equality functions can be equal or not. But, there is no value/measurement to indicate how much equal one to other.

Let assume that we have two ontology-based systems, where both of them need to compare *banana* with *bandana* and *apple*. In the system one, we consider nature of the objects, so we can say that *apple* is more similar to *banana* than *bandana*. And, in the system two, we consider spelling of the objects, so we can say that *bandana* is more similar to *banana* than *apple*. It is clearly understand that, the comparison functions should be extendible, their structures and/or return values should have been updatable depends on the research fields. In other words, the technologies should support user defined functions. Nowadays, two languages are widely used on ontologies, SPARQL and SWRL/SQWRL. But both of them have only Boolean comparison functions. And there is no way to extend existing functions and create user defined functions.

OWL and RDF is based on XML technologies, so it is possible to process them by using XQuery which support user defined functions. Because of that we decided to use XQuery in our ontology-based CBIR system. We create our similarity functions and we could perform our CBIR algorithms.

## 6.2 General System Overview

To perform our XQuery functions on our mammography ontology, we have set of process to convert the OWL file to XML file. It is also possible to perform XQuery on the OWL file, but to increase the speed of our CBIR system, we prefer to use a XML file. To obtain the XML file, first we perform a SPARQL query to get all breast mass instances as RDF triples and we saved them as a XML file. Then, we convert this XML file to our target XML file with a XSLT file. As a result, we perform our XQuery functions on the final XML file. Figure 6.1 shows the all conversation steps.

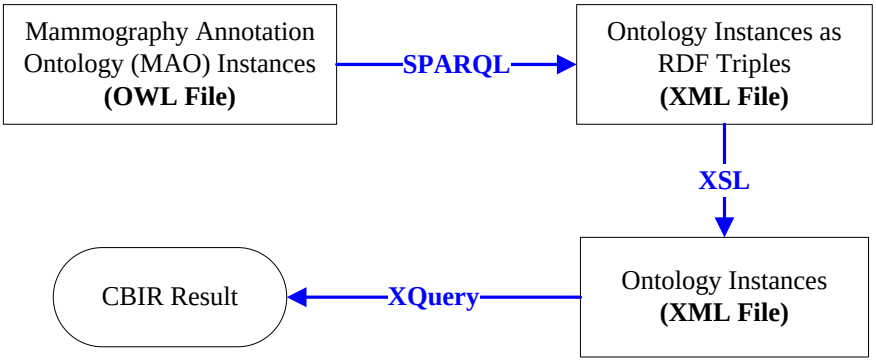



Figure 6.1 General system overview of XQuery calculation.

## 6.3 XQuery

XQuery is a language to extract and find attributes and elements from XML files. It is compatible with several W3C standards, such as Namespaces, XML, XSLT, XPath, and XML Schema. The main advantage of the XQuery for our study is allowing to write user defined functions to process our ontology instances.

In our CBIR system final similarity score between two masses is compose from three sub similarity scores. These are obtaining from high, mid and low-level features of the mass respectively. First, we show the similarity calculation for high-level features of two masses, `$massA` and `$massB`, in Figure 6.2. In the figure, the variables `$simShape`, `$simMargin` and `$simDensity` denote the

similarity score for each individual high-level image feature. Usage of these variables is given in Equation 8 of Chapter 4.

```
let $simShape :=
xs:double(data($similarityNodes[@from=xs:string(data($massA/ha
sShape/@isa)) and
@to=xs:string(data($massB/hasShape/@isa))]/@value))

let $simMargin :=
xs:double(data($similarityNodes[@from=xs:string(data($massA/ha
sMargin/@isa)) and
@to=xs:string(data($massB/hasMargin/@isa))]/@value))

let $simDensity :=
xs:double(data($similarityNodes[@from=xs:string(data($massA/ha
sDensity/@isa)) and
@to=xs:string(data($massB/hasDensity/@isa))]/@value))
```

Figure 6.2 High-level similarity calculation in XQuery.

Secondly, in Figure 6.3, we show the similarity calculation between `$massA` and `$massB` according to their mid-level image features, *MeanIntensity* and *Area*. As we mentioned in Equation 10 and 11 of Chapter 4, to normalize the middle level similarity scores we need to use their maximum value for the related dataset. For that purpose, we calculate the maximum values as `$maxMeanIntensity` and `$maxArea`. As a final point, to calculate the similarity scores for mid-level image features we perform Equation 10 and 11 of Chapter 4 and we set the variables `$simMeanIntensity` and `$simArea`.

```
let $maxMeanIntensity :=
max(data(doc($XMLFile)/onto/massCollection/mass/midLevelFeatur
eCollection/midLevelFeature[@name="MeanIntensity"]/@value))

let $simMeanIntensity := 1-
(abs(data($massB/midLevelFeatureCollection/midLevelFeature[@na
me="MeanIntensity"]/@value) -
data($massA/midLevelFeatureCollection/midLevelFeature[@name="M
eanIntensity"]/@value)) div $maxMeanIntensity)

let $maxArea :=
max(data(doc($XMLFile)/onto/massCollection/mass/midLevelFeatur
```

Figure 6.3 Mid-level similarity calculation in XQuery.

```
eCollection/midLevelFeature[@name="Area"]/@value))

let $simArea := 1-
(abs(data($massB/midLevelFeatureCollection/midLevelFeature[@na
me="Area"]/@value) -
data($massA/midLevelFeatureCollection/midLevelFeature[@name="A
rea"]/@value)) div $maxArea)
```

Figure 6.3 Mid-level similarity calculation in XQuery. (Cont.)

In our CBIR system, to calculate similarity score between low-level image features of the masses, we use Euclidean distance. Figure 6.4 shows the implementation of the function as `local:euclideanDistance` where the functions takes two double arrays and return a double value.

```
declare function local:euclideanDistance
($arg1 as xs:double*, $arg2 as xs:double*) as xs:anyAtomicType
{
 local:mySqrt(sum(local:euclideanPower($arg1, $arg2)))
};

declare function local:euclideanPower
($arg1 as xs:double*, $arg2 as xs:double*) as xs:double*
{
 for $i in 1 to count($arg1)
  return local:pow2($arg1[$i] - $arg2[$i])
};
```

Figure 6.4 Euclidean distance function in XQuery.

Thirdly, in Figure 6.5 we show the similarity calculation for low-level image features of the masses. Function `allSimScoresLowLevel` is used to get all similarity score between given massID and other masses for given low-level feature name. So, the function takes two parameters, mass identifier value (`$massID`) and low-level feature name (`$featureName`). Its result is required to determine the maximum similarity score for the related low-level feature. In the figure, we show the calculation of similarity score for Texture Browsing feature values. After obtaining maximum similarity score, we perform to calculate Equation 12 in Chapter 4 and set `$simTextureBrowsing` value.

```
declare function local:allSimScoresLowLevel($massID as
xs:string, $featureName as xs:string) as xs:double*
{
 for $mass in
doc($XMLFile)/onto/massCollection/mass[@name!=$massID]
 let $simLowLevel :=
local:euclideanDistance(data(doc($XMLFile)/onto/massCollection
/mass[@name=$massID]/lowLevelFeatureCollection/lowLevelFeature
[@isa=$featureName]/itemCollection/item/@value),data($mass/low
LevelFeatureCollection/lowLevelFeature[@isa=$featureName]/item
Collection/item/@value))
 return $simLowLevel
};

let $maxSimTB := max(local:allSimScoresLowLevel($massAMassID,
"TextureBrowsing"))

let $simTextureBrowsing := 1 -
(local:euclideanDistance(data($massA/lowLevelFeatureCollection
/lowLevelFeature[@isa="TextureBrowsing"]/itemCollection/item/@
value),
data($massB/lowLevelFeatureCollection/lowLevelFeature[@isa="Te
xtureBrowsing"]/itemCollection/item/@value)) div $maxSimTB)

let $ maxSimZernike :=
max(local:allSimScoresLowLevel($massAMassID, " Zernike"))

let $simZernike := 1 -
(local:euclideanDistance(data($query/lowLevelFeatureCollection
/lowLevelFeature[@isa="Zernike"]/itemCollection/item/@normaliz
edValue),data($current/lowLevelFeatureCollection/lowLevelFeatu
re[@isa="Zernike"]/itemCollection/item/@normalizedValue)))

let $ maxSimMarginMeanDifference:=
max(local:allSimScoresLowLevel($massAMassID, "
MarginMeanDifference"))

let $simMarginMeanDifference := 1 -
(local:euclideanDistance(data($query/lowLevelFeatureCollection
/lowLevelFeature[@isa="MarginMeanDifference"]/itemCollection/i
tem/@value),data($current/lowLevelFeatureCollection/lowLevelFe
ature[@isa="MarginMeanDifference"]/itemCollection/item/@value)
) div  $maxSimMarginMeanDifference)
```

Figure 6.5 Low-level similarity calculations in XQuery.

Finally, to test our XQuery methods we perform the our similarity calculation function between massA and massB whose details are given in Figure 4.2 in Chapter 4. As it shown in Figure 6.6 similarity scores for each level of image features and final similarity score are equal to sample similarity calculation of Figure 4.2.

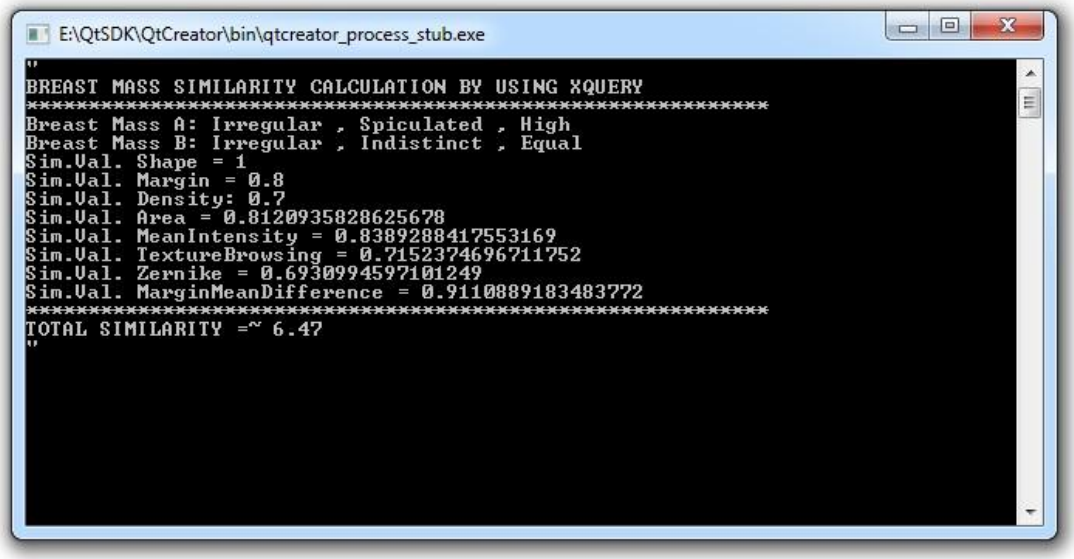



Figure 6.6 Example calculation result.

# CHAPTER SEVEN
# CONCLUSIONS

Aim of this thesis is to develop ontology-based content-based image retrieval system for breast masses. Hence, a successful implementation of radiological imaging system could be able to model and incorporate such knowledge into a more computable format. In this way, more complex tools such as case-based retrieval or evidence-based medicine can be possible in mammography. In order to achieve this goal, we have several improvements.

In this thesis, we describe a new ontology-based mammography annotation and retrieval tool (MART), and a mammography dataset (DEMS) which was created by using the MART. For that purpose, we have first developed mammography annotation ontology (MAO) which is a domain ontology and it provides shared vocabulary for mammography interpretation. During the development process of MAO we have used the 3th edition of the ACR BI-RADS Mammography Atlas. Main concepts and relations of the MAO are given in the thesis. Then we have developed MART, which is easy to use, and domain specific tool for mammography annotation. In other words, while developing the MART we have considered the requirements and habits of the mammography interpretation experts. All these three proposed components are is publicly available at "http://demir.cs.deu.edu.tr/index.php/downloads" and they are expected to be useful for researchers studying on mammography, for computer-aided diagnosis (CAD) studies as well as for medical student education.

Outputs of the MART are mammograms in lossless PNG and annotation in XML format. Hence, it is easily possible to share the outputs with other researches and to convert the annotations to any other format. Annotations of the MART are based on the MAO. In that respect, the tool save the annotations as OWL ontology instances and that make possible to publish them on Semantic Web and use Semantic Web technologies. Furthermore, MART is not only a mammography annotation tool, but it also supports CBR on mammography datasets. So, it may help experts in decision-making process and evidence-based medicine. MART also can be used to

evaluate performance of medical students if they annotate the mammography cases in the same dataset. Additionally, we provide DEMS with low-level feature set. DEMS contains 485 mammography cases where each of them has the mammograms in lossless PNG image format, and expert's annotations in XML format. The dataset is useful for benchmarking mammography related studies.

We develop a CBIR system where a breast mass is described with three sets of features: low, mid and high-level feature. High-level (HL) features are expert interpretation of a mass for shape, margin and density characteristics. Mid-level (ML) features are computer-calculated values for mass intensity and mass size. Both high- and mid-level features are human readable. For low-level (LL) features, we have first examined 25 different features and then choose the most three successful of them: Zernike Moments, Texture Browsing and Mean Margin Difference. We explain each feature level with their members in detail. Additionally, mathematical model of similarity calculation between two breast lesions and implementation of the model with SQWRL and XQuery explained in detail. Then, we compared the performance of individual feature set as well as different combination of them in terms of P@10 and precision-recall graph.

Additionally, in this thesis, we present an approach to model uncertainty in mammography, and perform SQWRL rules to infer BI-RADS scores for a given mass instance. Besides, mammography interpretation is a subjective and uncertain process itself and even a single expert may describe an abnormality with different annotations, in time (e.g., when describing the shape of a mass as oval or round). However, current state-of-the-art in ontology is based on crisp logic, which is incapable to handle uncertain condition. To overcome such uncertainty issues, we present an approach to model uncertainty in mammography.

We evaluated our proposal for uncertainty modeling with two mammography datasets, DEMS and DDSM, in terms of accuracy and sensitivity metrics. For BI-RADS 3, 4 and 5, we found that average sensitivity results are 75.33 and 67.13 for DEMS and DDSM, respectively. On the other hand, for DEMS, sensitivity value of the BI-RADS 6 is low, since masses with BI-RADS 5 and 6 generally have almost same mass descriptors. However, the only difference between BI-RADS 5 and 6 is

that BI-RADS 6 is to be pathologically proven malignancy while BI-RADS 5 is not. Furthermore, we have defined a new measure for level of uncertainty. We show that average level of uncertainty for crisp logic is clearly greater than our approach.

Major findings and contributions of this thesis can be summarized as follows;

1. We proposed new mammography annotation ontology based on the 3th edition of the ACR BI-RADS Mammography Atlas.
2. We proposed a new ontology based mammography annotation and retrieval tool which is a cross platform and anyone can easily create a new dataset based on our ontology by using the tool.
3. We proposed a new fully annotated mammography dataset called as DEMS which can be useful for benchmarking of mammography related works.
4. We proposed a new Content based Image Retrieval model for breast masses. The model uses high, mid and low-level image features of the breast.
5. We compared the performance of individual feature set as well as different combination of them. We show that using low-level features together with high and mid-level features improves the overall system performance and it is found statistically significant ($p<0.001$). This result can also be used to minimize the semantic gap between semantic descriptors and low-level features of breast masses. Also, the system helps to improve the performance of CADx system in mammography.
6. We gained that uncertainty exists in interpretation of BI-RADS scoring in mammography. However, current Semantic Web technologies are not sufficient to achieve modeling this uncertainty since they are based on crisp logic where the relations are binary. As a result, we show that use of Bayesian probability with description logic improves to handle uncertain relations between the concepts.

In future, to increase the performance of CADx system, more studies are needed on new low-level features for mammography, or different combination or fusion of different feature level. Furthermore, the findings of this study need to be experimented in other area of medical domain apart from mammography. We assume that Semantic Web technologies will be improved and they will be able to

handle uncertainty in their structure, and we hope that more built-in functions in SWRL and SQWRL will be developed and efficiency of existing functions will be improved.

## REFERENCES


Alto, H., & Rangayyan, R. M. (2005). Desautels JEL: Content-based retrieval and analysis of mammographic masses. *Journal of Electronic Imaging*, *14*(2).

Ashby, A. E., Hernandez, J. M., Logan, C. M., Mascio, L. N., Frankel, S., & Kegelmeyer, W. P (1995). UCSF/LLNL High resolution digital mammogram library. *Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society*, *1*, 539-540.

Berber, T. (2013). *Integration of content-based image retrieval and database management system: A case study with digital mammography*. Retrieved April 11, 2013 from http://www.fbe.deu.edu.tr/ArchieveDetail.aspx?Yayin=3166&Type=Thesis&Ogrenci=2006900796.

Bernardo, J. M., & Smith , A. F. M. (2001). Bayesian theory. *Measurement Science and Technology*, *12*, 221.

Breastcancer.org (2013). *Breast cancer statistics*. Retrieved October 25, 2013 from http://www.breastcancer.org/symptoms/understand_bc/statistics.jsp.

Bulu, H., Alpkocak, A., & Balci, P. (2012). Ontology-based mammography annotation and case based retrieval of breast masses. *Expert Systems with Applications*, *39*(12), 11194-11202.

Bulu, H., Alpkocak, A., & Balci, P. (2013). Uncertainty modeling for ontology-based mammography annotation with intelligent BI-RADS scoring. *Computers in Biology and Medicine*, *43*(4), 301-311.

Burnside, E., Rubin, D. L., & Shachter, R. (2000). A Bayesian network for mammography. *Proceedings of the AMIA Symposium, American Medical Informatics Association*, 106-10.

Castella, C., Kinkel, K., Eckstein, M.P., Sottas, P.E., Verdun, F.R., & Bochud, F.O. (2007). Semiautomatic mammographic parenchymal patterns classification using multiple statistical features. *Academic Radiology, 14*, 1486-1499.

Costa, P. C. G., & Laskey, K. B. (2006). PR-OWL: A framework for probabilistic ontologies. *Proceedings of the 4th International Conference (FOIS-2006)*, 237–249.

Davies, D. H. (1993). Digital mammography - the comparative evaluation of film digitizers. *British Journal of Radiology, 66(* 790), 930-933.

Delogu, P., Evelina Fantacci, M., Kasae, P., & Retico, A. (2007). Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier. *Computers in Biology and Medicine*, *37*(10), 1479-91.

Ding, Z., Peng, Y., & Pan, R. (2006). BayesOWL: Uncertainty modeling in semantic web ontologies. *Soft Computing in Ontologies and Semantic Web, 204*, 3–29.

Dokuz Eylul University (2012). *DEMS - Dokuz Eylul University mammography set*. Retrieved August 05, 2012 from http://demir.cs.deu.edu.tr/index.php/downloads.

Domeshek, E., & Kolodner, J. (1993). Using the point of large cases, artificial intelligence for engineering design. *Analysis and Manufacturing*, *7*(2), 87-96.

Dürig, M., & Studer, T. (2005). Probabilistic ABox reasoning: preliminary results. *Proceedings of the International Workshop on Description Logics (DL-2005)*, 104–111.

Dürig, M., & Studer, T. (2008). *PALC: Extending ALC ABoxes with probabilities*. Retrieved April 11, 2013 from http://www.iam.unibe.ch/ tstuder/papers/duerigstuder.pdf.

El Naqa, I., Yang ,Y., Galatsanos, N. P., & Wernick, M.N. (2002). Content-based image retrieval for digital mammography. *IEEE International Conference on Image Processing, 3*, 24–8.

Fernández-López, M. (1999). Overview of methodologies for building ontologies. *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem Solving Methods*, Stockholm, Sweden.

Fischer , E. A., Lo, J. Y., & Markey, M. K. (2004). Bayesian networks of BI-RADS descriptors for breast lesion classification. *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS-2004)*, San Francisco CA, USA, *4*, 3031–3034.

Friedman-Hill, E. (n.d.). *Jess rule engine*. Retrieved October 12, 2012 from http://www.jessrules.com/jess/index.shtml.

Fukushige, Y. (2004). Representing probabilistic knowledge in the Semantic Web. *Proceedings of the W3C Workshop on Semantic Web for Life Sciences*, Cambridge, MA, USA.

Hall, F. M., Storella, J. M., Silverstone, D. Z., & Wyshak, G. (1988). Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology*, *167*(2), 353–358.

Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2001). The digital database for screening mammography. *Proceedings of the Fifth International Workshop on Digital Mammography*, 212-218.

Heinsohn , J. (1994). Probabilistic description logics. *Proceedings of the 10th Conference on Uncertainty in Artificila Intelligence (UAI-1994)*, pp. 311-318.

Hu, B., Dasmahapatra, S., Lewis, P., & Shadbolt, N. (2003). Ontology-based medical image annotation with description logics. *Proceedings of The 15th IEEE International Conference on Tools with Artificial Intelligence*, New York, 77–82.

Hudelot, C., Atif, J., & Bloch, I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems Architecture*, *159*(15), 1929–1951.

Hung, S., & Chen, C. (2006). Mammographic case base applied for supporting image diagnosis of breast lesion. *Expert Systems with Applications*, *30*(1), 93-108.

Jaeger, M. (1994). Probabilistic reasoning in terminological logics. *Proceedings of the 5th International Conference on Principles of Knowledge Representation and Reasoning (KR-1994)*, 305–316.

Kahn, C. E., Roberts, L. M., Shaffer, K. A., & Haddawy, P. (1997). Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine*, *27*(1), 19-29.

Keles, A., & Yavuz, U. (2011). Expert system based on neuro-fuzzy rules for diagnosis breast cancer. *Expert Systems with Applications*, *38*(5), 5719-5726.

Khotanzad, A., & Hong, Y. (1990). Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 489–497.

Kinoshita, S. K., Azevedo-Marques, P. M., Pereira R. R. Jr., Rodrigues, J. A. H., & Rangayyan, R. M. (2007). Content-based retrieval of mammograms using visual features related to breast density patterns. *Journal of Digital Imaging*, *20*(2), 172-190.

Koller, D., Levy, A., & Pfeffer, A. (1997). P-CLASSIC: A tractable probabilistic description logic. *Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-1997)*, 390–397.

Lauria, A. (2006). GPCALMA: an Italian mammographic database of digitized images for research. *Digital Mammography, 4046*, 384-391.

Lauria, A. (2009). GPCALMA: implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination. *Physica*

*medica : PM : an International Journal Devoted to the Applications of Physics to Medicine and Biology : Official Journal of the Italian Association of Biomedical Physics (AIFB)*, *25*(2), 58-72.

Levy, M. A., O'Connor, M. J., & Rubin D. L. (2009). Semantic reasoning with image annotations for tumor assessment. *AMIA Annual Symposium Proceedings*, 359-363.

Lukasiewicz,T. (2008). Fuzzy description logic programs under the answer set semantics for the Semantic Web. *FundamentaInformaticae*, *82*(3), 289–310.

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast: Toward a full-field digital mammographic database. *Academic Radiology*, *19*(2), 236-48.

Mousa, R., Munib, Q., & Moussa, A. (2005). Breast cancer diagnosis system based on Wavelet analysis and fuzzy-neural. *Expert Systems with Applications*, *28*(4), 713-723.

Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004) A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. *International Journal of Medical Informatics, 73*, 1-23.

Müller, H., Rosset, A., Vallée, J.-P., Terrier, F., & Geissbuhler, A. (2004). A reference data set for the evaluation of medical image retrieval systems. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 28(6), 295-305.

Nair, P. C. (2011). Content Based Automatic Categorization and Annotation of Medical Images. *International Journal of Computing Technology and Information Security*, *1*(2), 35-39.

Nishikawa, R. M. (1997). Mammographic databases. *Breast Diseases*, *10*, (3-4).

Nottelmann, H., & Fuhr, N. (2006). Adding probabilities and rules to OWL lite subsets based on probabilistic datalog. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *14*(1), 17-42.

O'Connor, M., & Das, A. (2009). SQWRL: a query language for OWL. *Proceedings of OWL: Experiences and Directions(OWLED-2009).*

Oliveira, J. E. E., Machado, A. M. C., Chavez, G. C., Lopes, A. P. B., Deserno, T. M., & Araujo, A. (2010). MammoSys: a content-based image retrieval system using breast density patterns. *Computer Methods and Programs in Biomedicine*, *99*, 289–297.

Oliver, A., Freixenet, J., Marti, R., Pont, J., Pérez, E., Denton, E. R. E., & Zwiggelaar, R. (2008). A novel breast tissue density classification methodology. *Information Technology in Biomedicine*, *12*(1), 55–65.

Protege (2013). *ProtegeWiki: SWRL language FAQ*. Retrieved April 11, 2013 from http://protege.cim3.net/cgi-bin/wiki.pl?SWRLLanguageFAQ.

Qi ,D., Denton, E. R. E., & Zwiggelaar, R. (2006). Semantic analysis on medical images: a case study. *Proceedings of the 18th International Conference on Pattern Recognition*, *1*, 1260-1263.

QT Digia (n.d.). *QT: A cross-platform application and UI framework*. Retrieved January 10, 2013 from http://qt.digia.com/.

Rangayyan, R. M., Mudigonda, N. R., & Desautels, J. E. L. (2000). Boundary modelling and shape analysis methods for classification of mammographic masses. *Medical & Biological Engineering & Computing*, 38(5), 487-496.

Rapid-I (n.d.). *Rapid Miner*. Retrieved March 27, 2011 from http://rapid-i.com/content/view/181/196/.

Ren, L.J., & Barnaghi , P.M. (2007). Semantic media in medical information systems. *Proceedings of 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP-2007)*, 27-31.

Rosa, N., Felipe, J. C., Traina, A. J. M., Traina, C., Rangayyan, R. M., & Azevedo-Marques, P. M. (2008). Using relevance feedback to reduce the semantic gap in content-based image retrieval of mammographic masses. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 406-9.

Rubin , D. L., Mongkolwat, P., Kleper, V., Supekar, K., & Channin, D. (2008). Medical imaging on the Semantic Web: annotation and image markup. *2008 AAAI Spring Symposium Series, Semantic Scientific Knowledge Integration*, Stanford University.

Selvarani, A. G., & Annadurai, S. (2007). Medical image retrieval by combining low level features and dicom features, *IEEE International Conference on Computational Intelligent and Multimedia Applications*, *1*, 587 - 589.

Shadbolt, N., Lewis, P., Dasmahapatra, S., Dupplaw, D., Hu, B., & Lewis , H. (2007). MIAKT: Combining grid and web services for collaborative medical decision making. *Presented at the UK 2nd e-Science All Hands Meeting (AHM-2004)*, Nottingham, UK.

Stanford Center for Biomedical Informatics Research (n.d.). *Protégé ontology editor and knowledge-base framework* . Retrieved November 28, 2012 from http://protege.stanford.edu.

Stephan , P. (2013). *Mammograms - breast cancer screening and early detection*. Retrieved April 11, 2013 from http://breastcancer.about.com/od/mammograms/a/mammo_hub.htm.

The American College of Radiology (2012). *The American College of Radiology BI-RADS ATLAS and MQSA: Frequently asked questions*. Retrieved April 11, 2013 from

http://www.acr.org/~/media/ACR/Documents/PDF/QualitySafety/Resources/BIRADS/BIRADSFAQs.pdf

Udrea, O., Subrahmanian, V., & Majkic, Z. (2006). Probabilistic RDF. *Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration (IRI-2006)*, 172-177.

University of South Florida (n.d). *USF digital mammography: "other resources" page*. Retrieved January 10, 2013 from http://marathon.csee.usf.edu/Mammography/OtherResources.html.

Verma, B., McLeod, P., & Klevansky, A. (2010). Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert Systems with Applications*, *37(*4), 3344-3351.

W3C (2013). *OWL (web ontology language)*. Retrieved April 11, 2013 from http://www.w3.org/OWL.

W3C (2013). *RDF (resource description framework)*. Retrieved April 11, 2013 from http://www.w3.org/RDF.

Watson, I. (1999). *Survey of CBR application areas*. Retrieved July 07, 2012 from http://www.ai-cbr.org/iccbr99/presents/iwatson.pdf.

Wei, C. H., Li, C. T., & Wilson, R. (2005). A general framework for content-based medical image retrieval with its application to mammograms. *Medical Imaging 2005 Conference, FEB,* San Diego, CA.

Wei, C.H., Chen, S.Y., & Liu, X. (2012). Mammogram retrieval on similar mass lesions. *Computer Methods and Programs in Biomedicine*, *106*(3), 234–248.

Wu, P., Manjunath, B. S., Newsam, S., & Shin, H. D. (2000). A texture descriptor for browsing and similarity retrieval. *Signal Processing: Image Communication*, *16*(1-2), 33-43.

Yang, Y., & Calmet, J. (2005). Ontobayes: An ontology-driven uncertainty model. *Proceedings of the International Conference on Intelligent Agents Web Technologies and Internet Commerce (IAWTIC-2005)*, Vienna, Austria, 457–463.

Yu, S., & Huang, Y. (2010). Detection of microcalcifications in digital mammograms using combined model-based and statistical textural features. *Expert Systems with Applications*.

Zheng, B. (2009). Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future perspectives. *Algorithms*, *2*(2), 82