

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

**INTEGRATION OF CONTENT-BASED IMAGE
RETRIEVAL AND DATABASE MANAGEMENT
SYSTEM: A CASE STUDY WITH DIGITAL
MAMMOGRAPHY**

**by
Tolga BERBER**

**January, 2013
İZMİR**

**INTEGRATION OF CONTENT-BASED IMAGE
RETRIEVAL AND DATABASE MANAGEMENT
SYSTEM: A CASE STUDY WITH DIGITAL
MAMMOGRAPHY**

**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**

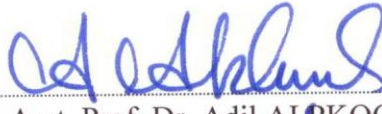
**by
Tolga BERBER**

January, 2013

İZMİR

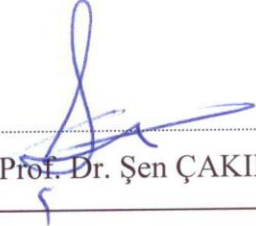
Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**INTEGRATION OF CONTENT-BASED IMAGE RETRIEVAL AND DATABASE MANAGEMENT SYSTEM: A CASE STUDY WITH DIGITAL MAMMOGRAPHY**” completed by **TOLGA BERBER** under supervision of **ASST. 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.



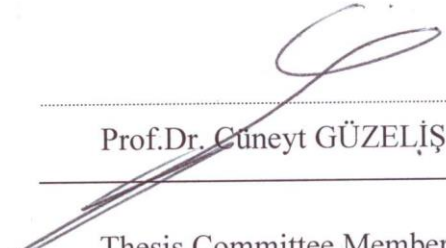
Asst. Prof. Dr. Adil ALPKOÇAK

Supervisor



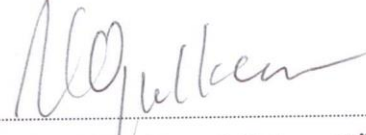
Asst. Prof. Dr. Şen ÇAKIR

Thesis Committee Member



Prof. Dr. Cüneyt GÜZELİŞ

Thesis Committee Member



Asst. Prof. Dr. Kemal Hakan GÜLKESEN

Examining Committee Member



Prof. Dr. Alp KUT

Examining Committee Member



Prof. Dr. Mustafa SABUNCU

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

At first, I would like to express my gratitude to my advisor Assistant Professor Dr. Adil ALPKOÇAK for his valuable suggestions, encouragement, patience and support during this study. Also, his academic perspective has very strong positive impact on 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 contribution to this study and sharing their ideas during development and writing of this thesis.

I would also thank to my friends and colleagues Assistant Professor Dr. Özlem AKTAŞ, Research Assistant Çağdaş Can BİRANT, Research Assistant Gökhan KARAKÜLAH for their support.

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

I have special thanks to my parents Ayşe and Yunus BERBER, my spouse Emine BERBER and my dear friend Zuhale MEHREKULA for their endless support, patience and encouraging me during preparation of this thesis.

Tolga BERBER

INTEGRATION OF CONTENT-BASED IMAGE RETRIEVAL AND DATABASE MANAGEMENT SYSTEM: A CASE STUDY WITH DIGITAL MAMMOGRAPHY

ABSTRACT

In this thesis, we proposed a new integration method for content-based image retrieval and database systems, and developed a case study on mammography retrieval to measure performance of our approach. Initially, we investigated 26 low-level features in total, 17 of them exist in the literature and rest of them is our proposal for mass contour description. Additionally, we proposed a new breast mass segmentation method called Breast Mass Contour Segmentation to determine accurate breast mass contours. Next, we reviewed available mammogram datasets to evaluate our proposal, and we also developed a new mammogram dataset due to insufficient annotation level of available datasets. During development of this dataset, we developed a new ontology-based annotation tool. Then, we performed series of experimentations on two different mammogram datasets to identify the best low-level features, machine learning and region selection methods for breast masses. Finally, we implemented our integration approach on PostgreSQL database management system using selected low-level features and evaluate the retrieval performance. Experimentation results showed that our integration approach of content-based image retrieval and Database Management Systems worked well and successfully applied to mammography mass retrieval system as case study.

Keywords: Content-based image retrieval, content-based mammography retrieval, multimedia database management systems, spatial access methods, metric access methods, image features, machine learning

İÇERİK TABANLI GÖRÜNTÜ GERİGETİRİM ve VERİ TABANI YÖNETİM SİSTEMİ BÜTÜNLEŞMESİ: SAYISAL MAMOGRAFYİYLE ÖRNEK BİR ÇALIŞMA

ÖZ

Bu tez kapsamında, içerik tabanlı görüntü gerigetirimi yaklaşımını veri tabanı sistemleri ile bütünleştirmek amacıyla yeni bir yöntem önerdik ve yöntemimizin başarımını ölçmek amacıyla bir mamografi gerigetirim sistemi geliştirdik. İlk olarak, 17'si literatürde mevcut, geri kalanı da meme kitlelerinin sınırını tanımlamak için önerdiğimiz, toplam 26 adet düşük seviyeli öznitelikleri inceledik. Ayrıca, meme kitlelerinin doğru sınırlarını bulmak için meme kitle sınır bölütlemesi adında yeni bir bölütleme algoritması önerdik. Mevcut mamografi veri kümelerini önerdiğimiz yaklaşımı sınamak için inceledik ve mevcut veri kümelerinin yetersiz betimleme seviyeleri sebebiyle yeni bir mamografi veri kümesi geliştirdik. Veri kümesinin geliştirilmesi sırasında, ontoloji tabanlı yeni bir betimleme aracı geliştirdik. Daha sonra, meme kitleleri için en iyi düşük seviyeli öznitelikleri, makine öğrenmesi yöntemlerini ve bölge seçim yöntemini belirlemek için iki farklı veri kümesi üzerinde bir dizi deneyler gerçekleştirdik. Son olarak, bütünleştirme yaklaşımımızı, seçilen düşük seviyeli öznitelikleri kullanarak PostgreSQL veri tabanı yönetim sistemi üzerinde gerçekleştirdik ve gerigetirim başarımını ölçtük. Deneylerden elde ettiğimiz sonuçlar, yaklaşımımız içerik tabanlı görüntü gerigetirimi ve veri tabanı yönetim sistemlerini birleştirilmesini başarı ile kullanıldığını gösterdi ve örnek bir çalışma olarak mamografi gerigetirim sistemi üzerine başarıyla uygulandı.

Anahtar sözcükler : İçerik tabanlı görüntü gerigetirimi, içerik tabanlı mamografi gerigetirimi, çokluortam veri tabanı yönetim sistemleri, uzaysal erişim yöntemleri, metrik erişim yöntemleri, görüntü öznitelikleri, makine öğrenmesi

CONTENTS

	Page
Ph.D. THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT.....	iv
ÖZ	v
 CHAPTER ONE – INTRODUCTION	 1
1.1 Overview.....	1
1.2 Aim of This Thesis	4
1.3 Thesis Organization	4
 CHAPTER TWO – CONTENT-BASED IMAGE RETRIEVAL	 5
2.1 Overview.....	5
2.2 Existing Content-Based Image Retrieval Systems	6
 CHAPTER THREE – LOW-LEVEL FEATURES FOR DIGITAL MAMMOGRAPHY	 11
3.1 Overview.....	11
3.2 Mammography.....	12
3.2.1 Analog Mammography.....	13
3.2.2 Digital Mammography	14
3.2.3 Mammography Views	15
3.2.4 Properties of Digital Mammography.....	16
3.2.5 Masses in Mammograms.....	16
3.3 Low-Level Features	17
3.3.1 Intensity Features	20
3.3.2 Shape Features.....	20
3.3.2.1 Statistical Shape Features.....	20

3.3.2.2 Moment Invariants	22
3.3.2.3 Fourier Features	23
3.3.2.4 Radial Distance Features.....	24
3.3.2.5 Region Based Shape Feature.....	25
3.3.2.6 Zernike Moments	26
3.3.3 Texture Features	27
3.3.3.1 Statistics of Gray Level Histogram (SGLH).....	27
3.3.3.2 Haralick-14.....	28
3.3.3.3 Gray Level Difference Matrix.....	31
3.3.3.4 Local Binary Patterns	32
3.3.3.5 Edge-Histogram	33
3.3.3.6 Gabor Filters	34
3.3.3.7 Homogenous Texture	35
3.3.3.8 Texture-Browsing	36
3.3.4 Margin Features.....	37
CHAPTER FOUR – BREAST MASS CONTOUR SEGMENTATION	40
4.1 Overview.....	40
4.2 Breast Mass Contour Segmentation (BMCS).....	42
4.2.1 Reference Segmentation Methods.....	42
4.2.1.1 Watershed Segmentation.....	43
4.2.1.2 Level-Set Segmentation	43
4.2.1.3 Seeded Region Growing Segmentation	43
4.2.2 Breast Mass Contour Segmentation Algorithm.....	44
4.3 Performance Evaluation.....	47
4.3.1 Segmentation Evaluation Metrics	48
4.3.1.1 Supervised Evaluation Metrics	48
4.3.1.1.1 Yasnoff Distance Metric.....	48
4.3.1.1.2 Distance Error Metric.	49
4.3.1.1.3 Classification-Based Metrics	49
4.3.1.1.4 Hausdorff Distance Metric.....	50
4.3.1.1.5 Scalable Discrepancy Measures.....	50

4.3.2 Experimentation	51
CHAPTER FIVE – DIGITAL MAMMOGRAPHY DATASETS	55
5.1 Overview	55
5.2 Available Mammography Datasets	55
5.2.1 Nijmegen Digital Mammogram Dataset	55
5.2.2 Washington University Digital Mammogram Dataset	55
5.2.3 OWH (Office of Women’s Health) Dataset	56
5.2.4 (Mini-)MIAS (Mammographic Image Analysis Society) Dataset	56
5.2.5 LLNL/UCSF Dataset	56
5.2.6 GPCALMA (Grid Platform for a Computer-Aided Library in Mammography) Dataset	57
5.2.7 INbreast Dataset	57
5.2.8 Digital Database for Screening Mammography (DDSM)	58
5.2.9 Dokuz Eylul University Mammogram Data Set (DEMS)	59
5.2.9.1 Mammography Annotation Tool (MAT)	61
5.2.9.2 Mammography Annotation Ontology (MAO)	61
5.2.9.3 DEMS Statistics	63
5.2.9.3.1 Abnormality Distribution	63
5.2.9.3.2 Breast Density Distribution.	63
5.2.9.3.3 Mass Distribution	64
5.2.9.3.4 Calcification Distribution.	65
5.2.9.3.5 Special Case Distribution	66
5.2.9.3.6 Associated Finding Distribution.	67
CHAPTER SIX – FEATURE SELECTION FOR CONTENT DESCRIPTION	68
6.1 Overview	68
6.2 Classifiers	68
6.2.1 k-Nearest Neighbor (k-NN)	69
6.2.2 Random Forests (RAF)	69

6.2.3 Naïve Bayes Classifier (BAY)	69
6.2.4 Artificial Neural Networks (ANN).....	70
6.2.5 Linear Discriminant Analysis (LDA).....	70
6.2.6 Support Vector Machines (SVM).....	71
6.3 Experimentation.....	71
6.3.1 Results of Shape Property Experiments	73
6.3.1.1 Low-Level Feature Performance Comparison	73
6.3.1.2 Dataset Performance Comparison.....	76
6.3.1.3 Mass Selection Method Performance Comparison	77
6.3.1.4 Classifier Performance Comparison	78
6.3.2 Results of Margin Property Experiments	81
6.3.2.1 Low-Level Feature Performance Comparison	81
6.3.2.2 Dataset Performance Comparison.....	85
6.3.2.3 Mass Selection Method Performance Comparison	85
6.3.2.4 Classifier Performance Comparison	86
6.3.3 Results of Density Property Experiments	89
6.3.3.1 Low-Level Feature Performance Comparison	90
6.3.3.2 Mass Selection Method Performance Comparison	92
6.3.3.3 Classifier Performance Comparison	93
6.3.4 Results of BI-RADS Property Experiments	95
6.3.4.1 Low-Level Feature Performance Comparison	95
6.3.4.2 Dataset Performance Comparison.....	101
6.3.4.3 Mass Selection Method Performance Comparison	101
6.3.4.4 Classifier Performance Comparison	102

CHAPTER SEVEN – INTEGRATING DATABASE AND CONTENT-BASED IMAGE RETRIEVAL: DIGITAL MAMMOGRAPHY MASS DATABASE. 109

7.1 Overview.....	109
7.2 Available Image Database Management Systems	110
7.2.1 SQL-MM:Still-image	110
7.2.2 Oracle InterMedia.....	111
7.2.2.1 CIRCE	112

7.2.3 IBM AIV Extender	113
7.2.4 IBM DB2 Still Image Extender	113
7.3 Multidimensional Indexing	113
7.3.1 k-d Tree	114
7.3.2 X-Tree	115
7.3.3 VP-Tree	115
7.3.3.1 M-Tree.....	116
7.4 PostgreSQL Extension	117
7.4.1 System Architecture	118
7.4.2 SQL Extension	119
7.4.2.1 Available Functions	120
7.4.2.1.1 create_mtree_field Function	120
7.4.2.1.2 mtree_handler Function	121
7.4.2.1.3 euclidean distance Function.....	121
7.4.2.1.4 containsoid Function.....	121
7.4.2.2 Range Queries	122
7.4.2.3 Nearest Neighbor Queries	123
7.4.2.4 Farthest Neighbor Queries	124
7.5 Performance Evaluation.....	125
7.5.1 Data Access Performance.....	126
7.5.2 Performance of Shape Property Queries	128
7.5.3 Performance of Margin Property Queries	129
7.5.4 Performance of Density Property Queries.....	130
7.5.5 Performance of BI-RADS Property Queries	130
CHAPTER EIGHT – CONCLUSIONS.....	132
REFERENCES.....	136
APPENDICES	155
A Low-Level Feature Performance Results.....	155
A.1 Low-Level Feature Performance Results of Shape Property	155

A.1.1 Low-Level Feature Performances of N/A Class	155
A.1.2 Low-Level Feature Performances of Round Class	156
A.1.3 Low-Level Feature Performances of Oval Class	157
A.1.4 Low-Level Feature Performances of Lobular Class	158
A.1.5 Low-Level Feature Performances of Irregular Class	159
A.2 Low-Level Feature Performance Results of Margin Property	160
A.2.1 Low-Level Feature Performances of N/A Class	160
A.2.2 Low-Level Feature Performances of Circumscribed Class.....	161
A.2.3 Low-Level Feature Performances of Microlobular Class	162
A.2.4 Low-Level Feature Performances of Obscured Class.....	163
A.2.5 Low-Level Feature Performances of Spiculated Class	164
A.2.6 Low-Level Feature Performances of Irregular Class	165
A.3 Low-Level Feature Performance Results of Density Property.....	166
A.3.1 Low-Level Feature Performances of Radiolucent Class.....	166
A.3.2 Low-Level Feature Performances of Low-Dense Class	167
A.3.3 Low-Level Feature Performances of Iso-Dense Class	168
A.3.4 Low-Level Feature Performances of High-Dense Class.....	169
A.4 Low-Level Feature Performance Results of BI-RADS Property	170
A.4.1 Low-Level Feature Performances of BI-RADS 0 Class	170
A.4.2 Low-Level Feature Performances of BI-RADS 1 Class	171
A.4.2 Low-Level Feature Performances of BI-RADS 2 Class	172
A.4.3 Low-Level Feature Performances of BI-RADS 3 Class	173
A.4.4 Low-Level Feature Performances of BI-RADS 4 Class	174
A.4.5 Low-Level Feature Performances of BI-RADS 4A Class	175
A.4.6 Low-Level Feature Performances of BI-RADS 4B Class	176
A.4.7 Low-Level Feature Performances of BI-RADS 4C Class	177
A.4.8 Low-Level Feature Performances of BI-RADS 5 Class	178
A.4.9 Low-Level Feature Performances of BI-RADS 6 Class	179
B Precision-Recall Graphs of Low-Level Features	180
B.1 Precision-Recall Graphs of Shape Property Queries	180
B.1.1 Precision-Recall Graph of Edge Histogram Feature	180
B.1.2 Precision-Recall Graph of Haralick-14 Feature	180

B.1.3 Precision-Recall Graph of Gray Level Difference Feature	181
B.1.4 Precision-Recall Graph of Gray Level Histogram Feature	181
B.1.5 Precision-Recall Graph of Homogeneous Texture Feature.....	182
B.1.6 Precision-Recall Graph of Invariant Moments Feature.....	182
B.1.7 Precision-Recall Graph of Local Binary Pattern Feature	183
B.1.8 Precision-Recall Graph of Global Margin Statistics Feature	183
B.1.9 Precision-Recall Graph of Radial Basis Signal Feature	184
B.1.10 Precision-Recall Graph of Texture Browsing Feature	184
B.1.11 Precision-Recall Graph of Zernike Moments Feature.....	185
B.2 Precision-Recall Graphs of Margin Property Queries	186
B.2.1 Precision-Recall Graph of Edge Histogram Feature	186
B.2.2 Precision-Recall Graph of Haralick-14 Feature	186
B.2.3 Precision-Recall Graph of Gray Level Difference Feature	187
B.2.4 Precision-Recall Graph of Gray Level Histogram Feature	187
B.2.5 Precision-Recall Graph of Homogeneous Texture Feature.....	188
B.2.6 Precision-Recall Graph of Invariant Moments Feature.....	188
B.2.7 Precision-Recall Graph of Local Binary Pattern Feature	189
B.2.8 Precision-Recall Graph of Global Margin Statistics Feature	189
B.2.9 Precision-Recall Graph of Radial Basis Signal Feature	190
B.2.10 Precision-Recall Graph of Texture Browsing Feature	190
B.2.11 Precision-Recall Graph of Zernike Moments Feature.....	191
B.3 Precision-Recall Graphs of Density Property Queries	192
B.3.1 Precision-Recall Graph of Edge Histogram Feature	192
B.3.2 Precision-Recall Graph of Haralick-14 Feature	192
B.3.3 Precision-Recall Graph of Gray Level Difference Feature	193
B.3.4 Precision-Recall Graph of Gray Level Histogram Feature	193
B.3.5 Precision-Recall Graph of Homogeneous Texture Feature.....	194
B.3.6 Precision-Recall Graph of Invariant Moments Feature.....	194
B.3.7 Precision-Recall Graph of Local Binary Pattern Feature	195
B.3.8 Precision-Recall Graph of Global Margin Statistics Feature	195
B.3.9 Precision-Recall Graph of Radial Basis Signal Feature	196
B.3.10 Precision-Recall Graph of Texture Browsing Feature	196

B.3.11 Precision-Recall Graph of Zernike Moments Feature.....	197
B.4 Information Retrieval Performance of BI-RADS Property Queries.....	198
B.4.1 Precision-Recall Graph of Edge Histogram Feature	198
B.4.2 Precision-Recall Graph of Haralick-14 Feature	198
B.4.3 Precision-Recall Graph of Gray Level Difference Feature	199
B.4.4 Precision-Recall Graph of Gray Level Histogram Feature	199
B.4.5 Precision-Recall Graph of Homogeneous Texture Feature.....	200
B.4.6 Precision-Recall Graph of Invariant Moments Feature.....	200
B.4.7 Precision-Recall Graph of Local Binary Pattern Feature	201
B.4.8 Precision-Recall Graph of Global Margin Statistics Feature	201
B.4.9 Precision-Recall Graph of Radial Basis Signal Feature	202
B.4.10 Precision-Recall Graph of Texture Browsing Feature	202
B.4.11 Precision-Recall Graph of Zernike Moments Feature.....	203

CHAPTER ONE

INTRODUCTION

1.1 Overview

Today, the amount of digital data is growing exponentially by various digital devices and on-line information systems. Need of software to manage such amount of data is increasing rapidly. Database management systems (DBMS) are the most preferred software solution for data management, since DBMS is software that handles storage and management of large collections of data that is called a database.

For modeling data, most of the DBMS use relational data model. Relational data model is based on grouping similar data into a logical data cluster, called *table*, and defining relations between those clusters. For example, a university database can contain a student table, a course table and a relation between students and courses. Today, DBMS provide almost same data manipulation and definition interfaces with the help of *Structured Query Language* (SQL) standard, which is proposed by International Organization for Standardization (ISO) (ISO, 2008). SQL has two sub parts, which are *Data Definition Language* (DDL) and *Data Manipulation Language* (DML). DDL is used to define tables and relations, while DML is used to add, remove, edit or retrieve data from tables. Thus, it is easy to say that any SQL compliant DBMS could easily define and manipulate data such that data properties could be defined declaratively.

DBMS has very high success rate for modeling data. On the contrary, they are not universal tools such that it is impossible for them to model all kind of data. Because, DBMS could only model and manipulate structures data. However, it is impossible to define a structure for any data. For instance, it is very hard to define a structure for a free-form text or an image data, since they have no predefined properties like name, surname. Hence, many information system designers use DBMS for storing unstructured data, and they prefer to use other data to manipulate them. From this point of view, DBMS are just storage engines for unstructured data.

Researches for modeling unstructured data become very important topic during last decades, because amount of available unstructured data is extremely high. Although unstructured data has no declarative properties, content of unstructured data could be used for modeling purposes. Therefore, unstructured data becomes an active object for any information system with the help of its content. Information Retrieval (IR) approach is a data retrieval method that is focusing on data content instead of its declarative properties. IR systems aim to rank documents in large collections according to user information need instead of finding a particular one. Though this seems improper from data management perspective, IR obtains very successful application areas dealing with unstructured data like web search engines.

IR systems usually represent the content of the data using multi-dimensional vectors. Hence, a document becomes a point in k -dimensional space. User information need is also represented in same space. As a result, IR systems rank documents according to their distances to user information need. Success of an IR system depends on multi-dimensional representation accuracy, since it is hard to convert a document to a multi-dimensional point. The most successful IR approach is textual IR that uses words in a collection as vector dimensions. Each textual document vector contains a weight factor for each word representing its importance for the document. Generally, importance factor is calculated by using the number of occurrences of the word in document. Similarly, user information need is generally expressed with keywords. Textual IR system converts user keywords to a query vector and ranks documents using distances of each document vector to query vector. But, contextual representation of other unstructured data is a non-trivial issue, since it is hard to split data content into small meaningful pieces like text documents.

After success of textual IR, researches on modeling other kinds of unstructured data accelerated. Especially, image IR, called content-based image retrieval (CBIR), is one of the most interesting and vivid topic. Like textual IR, CBIR converts images into multi-dimensional vectors, called low-level features, using mathematical operations. Each low-level feature aims to represent specific properties of image content like color, texture and shape. Moreover, representing user information need is another issue for CBIR. The most preferred and easiest way is query by example.

CBIR system calculates low-level features of sample image and ranks images according to low-level feature vector distances. Other kinds of unstructured data modeling methods use almost the same approach.

While there are methods for unstructured data access exists, information system designers are not taking them into account. The main reason of this situation is the lack of integration between DBMS and IR. However, two data modeling approaches have different data access perspectives. For a DBMS, finding exact data is a fundamental capability, while scoring all data according to user information need is what IR does. But, IR could help DBMS to model unstructured data. As a result, DBMS becomes a management tool for both structured and unstructured data.

Any retrieval system without a successful real life application will have no impact on literature. One of the application areas of CBIR is medical imaging archives (PACS). Although many hospitals have digital imaging devices and archiving systems, they cannot use image data for searching their image collection due to lack of integration between DBMS and CBIR. Besides, PACS are focused on storage and communication of medical images, and they support image retrieval by querying patient information. In this way, PACS are useless for researchers who want to search similar cases to a specific one. This kind of search need is crucial for uncertain cases. For instance, if a radiology expert can look into previous similar cases to examined one, he or she may find similar cases supporting his/her opinions and this makes decision making process much more easier for him/her. Since PACS are primarily used for storage of medical images, most of them store no medical annotations or do not use medical annotations for retrieval purpose. Hence, alternative access method for image retrieval approach is needed to find similar cases in a PACS archive.

Selecting a proper DBMS for CBIR integration is another decision for an integration approach designer such that DBMS should provide necessary functionality to define new data types and access methods. Although there are many commercial or open-source databases exist, only a small number of them supports necessary extensibility. One of them is PostgreSQL, which is an open source DBMS

that is originated from University of Berkeley. Today, it is being developed by community and supported by a commercial company, called Enterprise DB. It provides a very extensible programming interface that allows programmers to implement new data types and access methods. Thus, PostgreSQL is one of the most suitable solutions for CBIR integration.

1.2 Aim of This Thesis

Aim of this thesis is to integrate CBIR functionalities into a DBMS. Henceforth, DBMS provides necessary functionality to retrieve image data using common SQL interface. As a result, information system designers would be able to use images as active objects like other data types. Although CBIR approach has different perspective to retrieve images, it seems that CBIR is the only way to handle image data inside of a DBMS. Furthermore, usage of CBIR approach for image data access will increase with the help of DBMS integration. Aim of the thesis also involves to realize our integration approach on PostgreSQL, which is an open-source DBMS, and use it in mammography retrieval system as a case study.

1.3 Thesis Organization

This thesis is organized as follows. In chapter 2, a review of content based image retrieval systems are given. Chapter 3 introduces mammography and includes a review of low-level features used in or proposed to be used in mammogram images. Chapter 4 introduces a new mass contour segmentation approach capable of adjusting its parameters automatically. Detailed information about mammography datasets are given in chapter 5. Performance evaluation of mammography mass classification task is presented in chapter 6. Architecture and experimental results of our approach is given in chapter 7. Finally, chapter 8 concludes this thesis and provides future direction.

CHAPTER TWO

CONTENT-BASED IMAGE RETRIEVAL

2.1 Overview

Rapid advances in digital image production technology in recent years brought about issues related with effective way to find images, which have visual properties to answer user information need, in large image warehouses naturally. So, researchers have been trying to find a suitable image retrieval approach for large image warehouses for decades. At current stage, image retrieval approaches could be divided into three main groups. The first proposed approach in historical development of image retrieval is storing images in databases by adding a BLOB (binary large object) column and searching them by using other fields related with BLOB column. However, it is understood that this approach is insufficient since images do not directly integrated the search process.

Second approach proposed is to use image content to retrieve related ones in a warehouse and named Content-based image retrieval (CBIR). Content-based image retrieval (CBIR) is a technique, which uses visual contents to search images from large-scale image databases according to users' interests, and has been an active and fast advancing research area since the 1990s. During the past decade, remarkable progress has been made in both theoretical research and system development. However, there remain many challenging research problems that continue to attract researchers from multiple disciplines. CBIR uses the visual contents of an image such as color, shape and texture to represent and index images. In a typical CBIR system, the visual contents of images in the database are extracted priority and generally described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. Finally, feature databases are searched according to user information need or query. For instance, a typical CBIR query could be retrieving images that have visually similar shape or texture with a region or object like tumor in any image provided by user.

The third and final approach is to combine first and second approaches. Because, neither first nor second approach alone is capable of producing satisfying results. Using low-level features to represent images in CBIR arises several issues according to researches. First of all, users usually prefer to use high level concepts when describing their visual information need, i.e. query, while CBIR approach uses low-level features. The gap between low-level features like color, shape and texture and their semantic interpretations made by humans is named semantic gap. Researchers working on this field have tended to develop solutions to close this gap in CBIR systems. Many recent works have proposed integrated image retrieval approaches using both keywords assigned to images by humans and low-level features representing image content. Thus, some relationships between images that cannot be found by using keywords become possible to find with help of properties hidden in images.

This chapter is organized as follows. An in-depth review of current content-based image retrieval systems and their medical applications could be found in next section.

2.2 Existing Content-Based Image Retrieval Systems

Content-based image retrieval is one of the most vivid research fields. The first comprehensive pioneering work in this area is QBIC (Query By Image Content) project developed by IBM (Niblack et al., 1993). This system represents images by using low-level image features like color and texture descriptors. Moreover, QBIC is one of the first commercial image retrieval systems. After this stage, various content-based image retrieval systems developed either commercial or non-commercial, and they obtained widespread usage area (Niblack et al., 1993). Candid (Kelly, Cannon, & Hush, 1995), Photobook (Pentland, Picard, & Sclaroff, 1996), Netra (Ma & Manjunath, 1997) and BlobWorld (Carson, Thomas, Belongie, Hellerstein, & Malik, 1999) are examples of CBIR systems. Although it has wide usage area, CBIR technology has issues to be solved. These issues could be grouped into three major subjects; (1) what kind and how semantic layer is used, (2) which low-level features is used and which similarity metrics is used and (3) how data management and

organization becomes more effective (Smeulders, Worring, Santini, Gupta, & Jain, 2000).

Medical image archives (PACS, Picture Archiving and Communication System) is one of areas that CBIR is most needed. There are a lot of imaging sources exists in today's hospitals to be used for diagnosis and treatment purposes. Images could be originated from X-Ray, CT, MRI, Ultrasound, Nuclear Medicine, cardiology, pathology and gastroenterology. Fundamental goal of a PACS is running queries on images, retrieving them and presenting them in the way user wants. Yet, alphanumerical metadata is used to retrieve images in today's PACS. DICOM (Digital Image and Communications in Medicine) protocol has same issue. So, CBIR integration of the medical field will be an important contribution to enhancing quality of healthcare services, and may be extremely useful in evidence based medicine field. But, there are no PACS providing CBIR methods exist to be used in medical imaging archives.

Applications of CBIR on medical images are recently being developed, and still there is no fully available CBIR system exists (Müller, Michoux, Bandon, & Geissbuhler, 2004). However, a small number of studies has been performed to evaluate performance of CBIR technology on medical images and to define its medical use (Howarth, Yavlinsky, Heesch, & Rüger, 2004; Tsishkou, Kukharchik, Bovbel, Kheidorov, & Liventseva, 2003). IRMA (Content-based Image Retrieval in Medical Applications) is one of these works (Lehmann et al., 2004). In this work, images are categorized in four groups, which are acquisition direction, imaging technique, anatomy and biosystem of body region examined by radiology experts manually. Thus, each class could be represented by low-level features of its images. Moreover, performance of the system is improving with the help of users feedback continuously. Another work in this area is the ASSERT (Automated Search and Selection Engine with Retrieval Tools) project, which is aimed to work on high-resolution CT images and uses a semi-automatic way to extract low-level features of images (Aisen et al., 2003). Radiology experts select suspicious regions in high-resolution CT, and then low-level features of suspicious region are extracted. Thereby, a solution to the unresolved segmentation problem is being provided.

In addition to studies above, there are several works exist in the literature. For instance, some of the studies works on particular type of medical images (Kosch et al., 2001; Long, Antani, Lee, Krainak, & Thoma, 2003); some of them tries to represent images using objects (Chu, Cardenas, & Taira, 1998); some of them proposes performance of CBIR system with the help of metadata (Atnafu, Chbeir, & Brunie, 2002; Müller, Ruch, & Geissbuhler, 2004), and some of them aims to add CBIR capabilities to medical imaging archives (Bueno, Chino, Traina, Traina, & Azevedo-Marques, 2002; Güld, Thies, Fischer, & Lehmann, 2005). But none of the works defines a content-based medical image retrieval system; instead they measure performance of CBIR in medical field.

Another recent work on a content-based image retrieval system for digital mammography images uses positive and negative samples provided by user (El-Naqa, Yang, Galatsanos, Nishikawa, & Wernick, 2004). According to the results of this work sufficient results are obtained by using machine learning approaches like support vector machines and neural networks. In another recent work, relevance vector machine based on Bayes theory is proposed to retrieve calcification clusters in mammography images (Wei, Li, & Wilson, 2009). Additionally some recent works emphasize retrieval methods based on clustering theory (Greenspan & Pinhas, 2007; J. Z. Wang & Krovetz, 2005). Another important issue emphasized in recent works is the importance of relevance feedback integration to CBIR systems; and various suggestions made about this issue (Cho et al., 2012; Kherfi & Ziou, 2006; Rahman, Bhattacharya, & Desai, 2007; Yin, Pan, Chen, & Zhang, 2008).

Available PACS solutions provide retrieval of images in archive by using either demographic information of patient or metadata attached to images. Images and metadata of images are stores using DICOM (Digital Imaging and Communications in Medicine) standard in PACS. Although a lot of medical imaging device compatible with DICOM standard exist today, error rate of metadata could be very high. Whereat, undesirable results could be produced by queries using metadata. Hence, it is clear that using information acquired from image content alongside metadata in image retrieval in PACS produces more effective and more accurate results. In a work on this field (Müller, Ruch, et al., 2004) shows that using high-

level and low-level features collaboratively produces better results than using each one alone. Additionally, it is reported that CBIR enables retrieval of poorly annotated images with high visual similarity rate, and using MeSH terms in annotation decreases false positive rate.

On the other hand, another goal of the PACS is to find images in an acceptable period of time and in an effective way. For instance, finding visually similar images belonging the same body part in PACS to a patients current medical images helps to improve diagnosis of his/her. PACS supporting content-based retrieval capabilities is the most effective way to achieve this kind of diagnosis systems.

CBIR based systems use contents of images in archive during query and retrieval. In this way, images in the archive become active objects joining query and retrieval process, instead of being passive objects stored in databases. Moreover, CBIR approach enables usage of image content with external properties like case id, patient name, and surname in image retrieval. However, medical images could be produced from varying medical disciplines (chest, orthopedics, heart and vessel etc.) using different imaging methods (MRI, Magnetic Resonance Imaging, CT, Computerized Tomography, Ultrasound, X-Ray, etc.). For instance, a recent study presents a CBIR method using a combined feature vector of intensity and texture features for dental images (Ramamurthy, Chandran, Meenakshi, & Shilpa, 2012). Since each imaging method, medical discipline and disease may need different requirements, it is hard to implement a general CBIR method for all kind of medical images (Akgül et al., 2011).

Regardless of working fields, almost all experts dealing with images face with problems related with storage and retrieval of images. CBIR approach is the best solution developed to answer those problems. Although there are a lot of CBIR systems in development, there is no standard defined among them. MPEG-7 (Moving Picture Experts Group) is an ISO approved standard that is developed by MPEG group to answer this need. MPEG-7 is a standard aiming to define multimedia content using both low-level and semantic level. Hence, the success achieved by textual search engines could be moved to multimedia field; even all of the

multimedia data on the earth could be accessed by their content (Manjunath, Salembier, & Sikora, 2002). Additionally, low-level features included in MPEG-7 are designed for general-purpose images rather than medical images, but they are thought to be suitable for special purposes. Today, radiology experts are evaluating performance of low-level descriptors used in academic CBIR solutions and developers aim to find the most suitable low-level feature set for medical images. In a recent work, performance of shape features are tested on a liver lesions (Xu, Faruque, Beaulieu, Rubin, & Napel, 2012).

Regardless of low-level feature set, users of a CBIR system want to use high-level concepts to query image archive. So, it is suggested to use both low-level features and high-level concepts with external properties attached to the image. MeSH ontology, which is developed U.S. National Institute of Health (NIH) and commonly used, targets indexing high-level medical concepts and mapping relations of these concepts on a semantic map. This ontology and semantic network between its concepts is renewed every year. As a result, medical systems based on MeSH ontology easily integrate new medical concepts, so that systems stay up-to-date. Additionally, there are some approaches aiming to model high-level concepts by using automatic methods (Faruque et al., 2011; W. Yang, Feng, Lu, & Chen, 2011).

Today, DICOM protocol is used in medical imaging archives. This protocol puts standard on both storage of image including its metadata and communication of this data in a network environment. Although accuracy of metadata is open to discuss, they are important in terms of defining relationship between images and hospital information systems (HIS). Some of the metadata could also be used to classify images effectively. Thus, high-level medical concepts assigned according to examined body region, and images only related with examined body region could be filtered out. As a result, accuracy of both retrieval performance and clinical decisions given by using retrieval system is thought to increase. This approach is being used in another academic system, and it is reported that success rate is high (Lehmann et al., 2004).

CHAPTER THREE

LOW-LEVEL FEATURES FOR DIGITAL MAMMOGRAPHY

3.1 Overview

Up today, in the literature, many low-level image features have been suggested for CBIR systems for digital mammography, including CADx systems for mass or calcification classifications etc. The algorithms are mostly evaluated in publically available datasets, but the results are not comparable and, in some cases, irreproducible since almost every study uses a different or unknown subset of the datasets, rather than using the whole set. So, we need to evaluate low-level features on public datasets available to select adequate low-level features for mammography mass classification task.

Selecting qualified low-level features for mammography CAD system is the first step of developing a successful CAD system (Tang, Rangayyan, Xu, El Naqa, & Yang, 2009; Zheng, 2009). In a recent review, it is pointed out that several techniques are proposed to use in mammography CAD systems and stated that CAD systems help experts by detecting early stages of breast cancer (Tang et al., 2009). Although evolution of CAD systems in mammography takes about two decades, current performance of the CAD systems does not fully meet clinical needs. Performance improvement of such systems is expressed as a further study in this area. Each mass property must be assigned properly, since determining malignancy score of a mass depends on several mass properties. As a result, low level features which will be used in a mammography CAD system, must address not only one property of a mass but also they must be able to help measuring other aspects of a mammographic mass. Since selecting the most adequate feature set is the most important phase of CAD system development, an evaluation of low-level features is needed.

In this thesis, we provide an exhaustive literature search on low-level image features for mammographic mass classification. In total, we studied 26 low level features where 17 of them are already used in the CBIR literature; however, 5 of

them not used for mammographic mass classification before. Moreover, we proposed 9 new features describing margin of a mass. Total vector length is 578.

This chapter is organized as follows. Mammography is presented in section 3.2. We present an in-depth introduction of low-level features for mammography masses in terms of CBIR requirements in section 3.3.

3.2 Mammography

Mammography is a special type of x-ray based imaging that is used to obtain detailed internal structure of the breast. Mammography is a medical imaging system that is especially designed for breast imaging with capable of obtaining high-contrast and high-resolution breast images by using low-dose x-ray. Early diagnosis of a breast cancer is the key factor for successful treatment. Mammography plays a very important role in the early diagnosis of breast cancer. Total number of mammography screening performed in United States is 35.8 million in a year (Spelic, Kaczmarek, Hilohi, & Belella, 2007). According to reports of U.S. Food and Drug Administration (FDA), mammography helps early diagnosis of breast cancer of women aged 50 and above.

Mammography could determine changes in breast, before those changes are detected by women herself or her doctor. A mass could be identified by mammography about two years before it becomes palpable (Barclay, 2012). After discovering tuberosity in breast, mammography plays an important role for cancer diagnosis. If an abnormality is found during a mammography examination or a palpable mass is confirmed by mammography, additional imaging methods like ultrasound imaging or breast biopsy can be performed. Biopsy is a procedure that sample tissue taken from breast with a surgical procedure or a needle is evaluated under microscope whether it has cancer cells. Mammography and ultrasound imaging are used a guide during biopsy to ensure needle positioned correctly.

There are two kinds of mammography examination; *Screening mammography* and *Diagnostic Mammography*.

Screening Mammography is applied to women with no symptoms. Purpose of screening mammography is to detect very small masses, which could not be determined by women or medical expert, at early stages. Early diagnosis of breast cancer with mammography greatly improves the chances of successful treatment. So, every women older than 40 is suggested to perform a screening mammography examination in every year. In some cases, if woman has a family history of breast cancer, doctors recommend screening mammography before age 40. Today, screening mammography is widely used in many hospitals.

Diagnostic mammography is more detailed kind of mammography that is applied to patients, who discovers a lump in their breasts or has suspicious findings like nipple discharge or has a breast abnormality confirmed by screening mammography. Diagnostic mammography is more time consuming and is used to determine exact location of the mass, dimensions of the mass, relations with surrounding tissues and status of lymph nodes. After acquiring additional views of breast, mammography images are interpreted. Hence, diagnostic mammography is more time consuming and costly than screening mammography.

3.2.1 Analog Mammography

During a mammography examination, an x-ray source is fired and resulting x-rays are falls on a film cassette after passing compressed breast. X-Ray falling on phosphor layer on the film cassette creates brightness according to its amount. Those brightness levels forms the mammography image. Since x-rays pass through the tissues at different rates depending on structure and density of tissue, internal structure of the breast is imaged. This imaging technique provides very detailed image of breast with minimum radiation possible by using high sensitive photography films and special x-rays. After that, photofinishing of mammography image is same with photofinishing of ordinary image. Finally, radiologists evaluate mammography films.

Breast tissue includes fat, fiber and gland. Breast masses including benign and malign ones are seen as white regions (radiodense), while fat is seen as black

(radiolucent) on mammography film. Other tissues including glands, connective tissue, tumors and micro calcifications are seen as white regions at different levels on mammography film.

Breast should be flattened a little by compression like Figure 3.1 to view highest amount of breast tissue during mammography acquisition. Compression of the breast could cause discomfort in patients. But this discomfort ends in a short time period required to complete mammography acquisition. The main reason of compression of breast is to avoid overlapping breast tissues as much as possible so that anatomy of breast and possible abnormalities could be viewed better. For instance, insufficient amount of breast compression could cause low-detailed view of micro-calcifications, which are tiny calcium clusters and early sign of breast cancer. Moreover, lowering the x-ray dose and prevention of patient movement are another important reasons of breast compression.

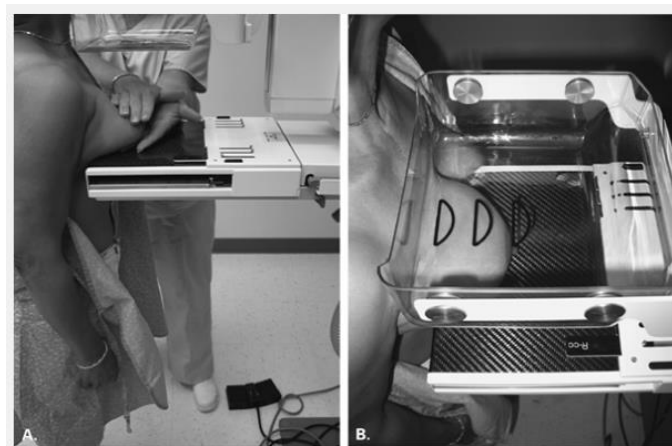


Figure 3.1 Mammography acquisition (DeParedes, 2007).

3.2.2 *Digital Mammography*

Digital mammography uses same structure of analog mammography. But a digital sensor matrix is used to acquire image instead of film cassette. So, breast image is digitally acquired and viewed on a workstation computer immediately. Several works show that digital mammography produces accurate results at least analog mammography. FDA approves full-field digital mammography to examine and

diagnose the breast cancer. Moreover, full-field digital mammography replaces traditional mammography devices rapidly.

3.2.3 Mammography Views

Left and right breast of the patient are imaged separately during screening mammography. In mammographic examinations, two views are commonly used, which are cranial-caudal (from top to bottom, CC) and mediolateral-oblique (from inner-top to outer using predefined angle, MLO). Hence, there are 4 images, which are two CC and two MLO of each breast in a typical screening mammography. Figure 3.2 includes sample MLO and CC view of a breast.

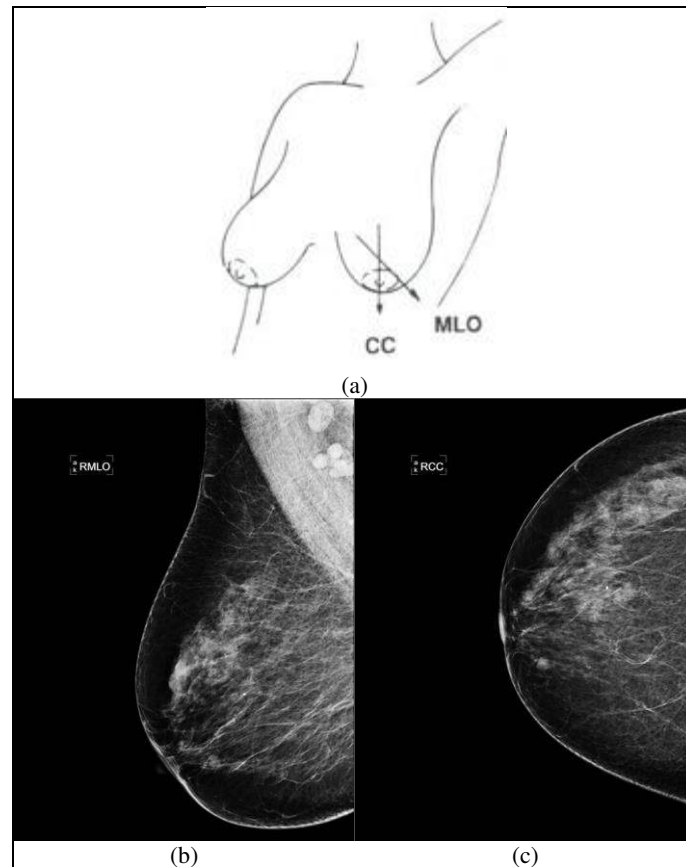


Figure 3.2 Mammography views. (a) MLO and CC direction of right breast (Imaginis, 2012). (b) MLO image of right breast. (c) CC image of right breast.

3.2.4 Properties of Digital Mammography

Since digital mammography images acquired directly digitally, they can be stored digitally in picture archiving and communication systems (PACS). Mammography images are the most detailed and has the largest size in a PACS compared to other imaging modalities. A screening mammography consists of four images generally. Resolution of each image is 50-100 μm and size of each image is 4096×4096 . Color depth of each pixel varies from 12 and 16 bits. So, a typical mammography image is represented in $4096 \times 4096 \times 2^{16}$ bits, and takes roughly in 30 MB file. As a result, a typical mammography screening case costs approximately 120 MB of disk space in PACS. So, mammography images consume a considerable amount of disk space in PACS of a hospital with those disk space requirements.

3.2.5 Masses in Mammograms

Both malign and benign masses show different textural and morphological characteristics from surrounding breast tissue. So, a mass can be distinguished from other breast tissues by using its texture and morphology. After a mass observed in a mammogram, the next important step is to determine malignancy of a mass.

Shape, contour and margin characteristics of a breast mass in digital mammograms have very important clues in discriminating malign and benign tumors. For instance, malign masses tend to spread other breast tissues, while benign masses remain stable. Moreover, malign masses forms very irregular shaped regions, as benign masses usually form very regular shapes. Figure 3.3 depicts relation of morphological and textural properties of breast masses with malignancy.

American College of Radiology (ACR) puts a standard on mammography reporting named Breast Imaging Reporting and Data System (BI-RADS, (D'Orsi, Bassett, & Berg, 2003)) which represents experts judgment about presence or absence of breast cancer. According to BI-RADS standard, each mammography mass has a BI-RADS score from 0 to 6 depending on its morphological and textural

properties depicted in Figure 3.3. BI-RADS scale values with diagnostic meanings are given in Table 3.1.

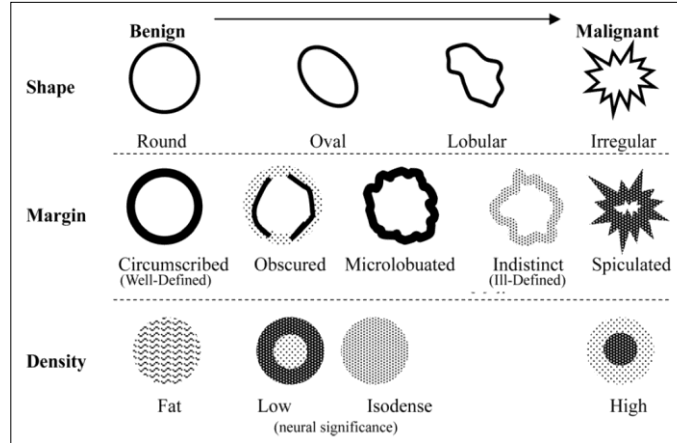


Figure 3.3 Samples of benign and malign masses according to their textural and morphological properties (Wei, Chen, & Liu, 2012).

Table 3.1 BI-RADS reporting scale for abnormalities (D’Orsi et al., 2003).

BI-RADS Score	Diagnosis	Criterion
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. Has three sub groups and biopsy should be considered.
		4A Finding needing intervention with a low suspicion for malignancy. Probability of being malignant (3 to 29%)
		4B Lesions with an intermediate suspicion of malignancy. Probability of being malignant (30 to 59%)
		4C Findings of moderate concern, but not classic for malignancy. Probability of being malignant (60 to 94%)
5	Highly Suspicious of Malignancy	Lesion that has a high probability of being malignant ($\geq 95\%$); 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.

3.3 Low-Level Features

In the literature, many low-level features and their combinations are used in Mammography CADx systems. In this work, we used 26 features belonging to four groups: intensity, shape, texture and margin features.

Table 3.2 shows a detailed list of low-level features, which are used in mammography CADx systems. First column of the table contains category of the feature, second column in Table 3.2 depicts feature name, an third column contains the legend of the feature, which is used throughout in this work. Fourth column contains number of elements in the feature vector, and the last two columns of the table contain original work in which the feature is proposed and works that uses this feature.

Table 3.2 List of all low level features included in this thesis.

Feature Group	Feature name	Legend	Length	Proposed by	Used by
Intensity	Mean Average Intensity	I-GEN	1	-	-
Shape	General Shape Properties	S-GEN	9	-	(Boninski & Przelaskowski, 2008; El-Naqa et al., 2004; Fan, Chang, Lin, & Hsieh, 2011; Golobardes, Llorca, Salamó, & others, 2002; Peng, Yao, & Jiang, 2006; Verma, McLeod, & Klevansky, 2010; X.-H. Wang, Park, & Zheng, 2009)
	Invariant Moments	S-INM	9	(Hu, 1962)	(El-Naqa et al., 2004; Kinoshita, Azevedo-Marques, Pereira, Rodrigues, & Rangayyan, 2007; Yin et al., 2008)
	Fourier Features of Complex Contour Representation	S-FDE	10	-	(Pourghassem & Ghassemian, 2008; Zheng, 2009)
	Fourier Features of Distance Contour Representation	S-DFD	10		
	Fourier Features of Curvature Contour Representation	S-CFD	10		
	Radial Distance Feature	S-RDD	7	(Georgiou, Mavroforakis, Dimitropoulos, Cavouras, & Theodoridis, 2007)	(Georgiou et al., 2007)
	Fourier Features of Radial Distance Signal	S-RDF	10	(Georgiou et al., 2007)	(Georgiou et al., 2007)
	MPEG-7 Region Based Shape Feature	S-RBS	36	(Ricard, Coeurjolly, & Baskurt, 2005)	-

Table 3.2 List of all low level features included in this thesis. (Cont.)

Feature Group	Feature name	Legend	Length	Proposed by	Used by
	Zernike Moments	S-ZER	15	(Khotanzad & Hong, 1990)	(N. A. Rosa et al., 2008)
Textures	Statistics of Gray Level Histogram	T-HIS	9	-	(Antonie, Zaïane, & Coman, 2003; Kinoshita et al., 2007; Müller, Rosset, Vallée, Terrier, & Geissbuhler, 2004; Subashini, Ramalingam, & Palanivel, 2010)
	Haralick-14	T-GLC	96	(Haralick, Shanmugam, & Dinstein, 1973)	(Kinoshita et al., 2007; Lauria, 2009; Pourghassem & Ghassemian, 2008; Rangaraj M Rangayyan, Nguyen, Ayres, & Nandi, 2010; Yin et al., 2008; Zheng, 2009)
	Gray-Level Difference	T-GLD	20	(Weszka, 1978)	(Kim & Park, 1999)
	Local Binary Patterns	T-LBP	18	(Timo Ojala, Pietikäinen, & Harwood, 1996)	(Dagan Feng, Fu, & Tian, 2008)
	Edge Histogram	T-EDH	80	(Park, Jeon, & Won, 2000)	(Dagan Feng et al., 2008; Timo Ojala, Mäenpää, Viertola, Kyllönen, & M, 2002)
	Homogeneous Texture	T-HOT	62	(Ro, Kim, Kang, Manjunath, & Kim, 2001)	-
	Texture Browsing	T-TEB	5	(Wu, Manjunath, Newsam, & Shin, 2000)	-
Margin	Column-wise Means	M-CWM	20	new	-
	Column-wise Standard Deviations	M-CWS	20	new	-
	Column-wise Skewness	M-CWW	20	new	-
	Column-wise Kurtosis	M-CWK	20	new	-
	Region Mean Differences	M-CMD	20	new	-
	Region Standard Deviation Differences	M-CSD	20	new	-
	Margin Mean Differences	M-RMD	20	new	-
	Margin Standard Deviation Differences	M-RSD	20	new	-
	Global Statistics of Inner and Outer Regions	M-GLS	8	new	-
Total			578		

3.3.1 Intensity Features

Color is the most extensively used visual content feature for CBIR. Color moments, which are basically the first order (mean), the second order (variance) and the third order (skewness), have been successfully used in many content-based retrieval systems particularly when the image contains just the object and have been proved to be efficient and effective in representing color distributions of the images (Stricker & Orengo, 1995). However, the mammography images are intensity based gray scale images, and color is not defined in mammography. Instead, color feature is represented with intensity feature. Therefore, intensity features group contains only one feature that is the mean of the gray level of a mass. Moreover, this feature is human readable and, hence, radiology experts can interpret the feature, and it makes the intensity feature to be considered as high-level feature, as well.

3.3.2 Shape Features

Shape features aims to identify object shape in an image, and rarely used in CBIR systems. In our case, shape of a mass is an important property which defines malignancy of a mass. Since shape features aims to describe shape of an object, we consider them as the most important low-level feature for mass shape classification task. There is several shape features proposed and used in mammography area in literature. Shape features evaluated in this work is suggested by literature, except Region Based Shape feature, which is a MPEG-7 shape feature. Shape features used in this work are described in following sections.

3.3.2.1 Statistical Shape Features

Statistical shape features represent contour information of an object in a segmented image. These features are all extracted from binary image. Table 3.3 shows the list of the statistical shape features.

Table 3.3 List of statistical shape features.

Feature	Formula	Explanation
Area	$F_{Area} = \sum_{x \in O} 1$	Number of pixels in the region, where O is the set of pixels in the segmented object
Perimeter	$F_{Per} = \sum_{x \in B} 1$	Total length of the boundary of object where B is the set of pixels on the boundary
Compactness (Perimeter Based)	$F_{CPer} = \frac{(F_{Per})^2}{4\pi F_{Area}}$	Determines compactness of a region
Compactness (Area Based)	$F_{Compact} = \frac{1}{n_b - 1} \sum_{p \in B} (dist(p, c) - F_{Rad})^2$	Variance of distances from the center of gravity and border pixels
Modified Compactness	$F_{ModCompact} = \frac{F_{per}}{F_{Area}}$	Simplified version of the compactness
Box min X,Y and max X,Y	-	Coordinates of extreme left, top, right and, respectively, bottom pixels of a region
Feret X,Y	-	Dimension of the minimum bounding box of the region in the horizontal and vertical directions
Roughness	$F_{Rough} = \frac{F_{Per}}{F_{CPer}}$	Roughness of a region
Length	$F_{Len} = F_{BoxMaxX} - F_{BoxMinX}$	Length of a region
Breadth	$F_{Bre} = F_{BoxMaxY} - F_{BoxMinY}$	Breadth of a region
Elongation	$F_{Elong} = \frac{F_{Len}}{F_{Bre}}$	
Centroid X,Y	$(x_c, y_c) = \sum_{(x,y) \in O} \left(\frac{1}{N}x, \frac{1}{N}y \right)$	Coordinates of the center of gravity of a region
Radius	$F_{Rad} = \frac{1}{n_B} \sum_{p \in B} \sqrt{(p_x - x_c)^2 + (p_y - y_c)^2}$	Mean of distances from the center of gravity and border points, where n_B is the number of boundary pixels

In the literature, Peng et al. (Peng et al., 2006) uses these features in a microcalcification classification (detection) system and obtained 96% true positive rate when FP/image is 20%. El-Naqa et al. (El-Naqa et al., 2004) used these features in a medical content-based retrieval system and their retrieval system obtained 100% of precision at 20% recall level. Verma et al. (Verma et al., 2010) used these shape features in a mammographic classification system and experimentations on DDSM dataset resulted with 97.5% of accuracy. Fan et al. (Fan et al., 2011) used these shape features in a medical classification system. They used a fuzzy decision tree and accuracy of the proposed technique is about 90%. Golobardes (Golobardes et al.,

2002) used these shape features in a microcalcification classification system and comments that performance of proposed system is equal to the other CAD systems. Wang et al. (X.-H. Wang et al., 2009) conducts experiments on medical content-based image retrieval systems with shape features. They state that multi feature systems outperform single featured ones. These features are also used in IShark (Boninski & Przelaskowski, 2008) medical CBIR system.

3.3.2.2 Moment Invariants

Moment Invariants, proposed by Hu (Hu, 1962), are the classical representation of shape information. If the image is represented as a binary image, then the central moments of order $p+q$ computed as follows;

$$\mu_{p,q} = \sum_{(x,y) \in R} (x - x_c)^p (y - y_c)^q$$

where (x_c, y_c) is the centroid of the object. This feature can be normalized to be scale invariant as follows.

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^{\gamma}}, \gamma = \frac{p+q+2}{2}$$

Based on these moments, scale, translation and rotation invariant properties of shapes can be extracted from binary images (L. Yang & Albregtsen, 1994). The features extracted by using central moments are shown in Table 3.4.

Yin et al. (Yin et al., 2008) used these moments in a medical image categorization system and accuracy of the system is measured as %97 on a small dataset. Kinoshita et al. (Kinoshita et al., 2007) uses these features in a CBIR system and proves that best performance of the system was obtained by using these features in conjunction with some other visual features. El-Naqa et al. (El-Naqa et al., 2004) used these features in a medical CBIR system and their retrieval system obtained 100% of precision at 20% recall level.

Table 3.4 List of moment invariant features.

Feature	Formula
7 rotational and scale invariant features	$M_1 = \mu_{20} + \mu_{02}$
	$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$
	$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2$
	$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2$
	$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2]$ $+ (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$
	$M_6 = M_1[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2]$ $+ 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})$
	$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2]$ $- (\mu_{30} - 3\mu_{12})(\mu_{03} + \mu_{12})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$
Principal axis	$F_\theta = \frac{1}{2} \tan^{-1} \left[\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right]$
Secondary axis	F_ϕ , where $F_\phi \perp F_\theta$
Eccentricity	$F_{Ecc} = \frac{M_1 + \sqrt{M_2}}{M_1 - \sqrt{M_2}}$
Axis ratio	$F_{AxisRatio} = \frac{F_\theta}{F_\phi}$
Majority	$F_{Majority} = \frac{F_\theta}{long}$, where F_{long} is the longest distance

3.3.2.3 Fourier Features

These features use boundary pixels of an object in an image, and transforms boundary pixels using Fourier transform. These features need a suitable contour representation to apply Fourier transform. There are three types of contour representations for these features; centroid distance, curvature and complex representations shown in Table 3.5.

Table 3.5 Boundary representation types.

Boundary Representation	Formal Definition
Centroid Distance Representation	$R_s = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2}$
Curvature Representation	$K(s) = \frac{d}{ds} \theta(s)$ where $\theta(s) = \tan^{-1} \left(\frac{y'_s}{x'_s} \right)$ and $y'_s = \frac{dy_s}{ds}$, $x'_s = \frac{dx_s}{ds}$
Complex Representation	$Z(s) = (x_s - x_c) + j(y_s - y_c)$

where, (x_c, y_c) is the centroid of the object, (x_s, y_s) are the successive pixel coordinates of object boundary.

To ensure feature size, contour representation is sampled to obtain M samples using a uniform sampling function before Fourier features are extracted. Each contour representation has its own Fourier feature extraction scheme given in Table 3.6.

Table 3.6 Fourier feature definitions of each contour representation.

Fourier Feature Name	Formal Definition
Centroid Distance Representation	$F_D = \left[\frac{ f_1 }{ f_0 }, \frac{ f_2 }{ f_0 }, \dots, \frac{ f_{M/2} }{ f_0 } \right]$
Curvature Representation	$F_C = [f_1 , f_2 , \dots, f_{M/2}]$
Complex Representation	$F_X = \left[\frac{ f_{-(M/2)-1} }{ f_1 }, \dots, \frac{ f_{-1} }{ f_1 }, \frac{ f_2 }{ f_1 }, \dots, \frac{ f_{M/2} }{ f_1 } \right]$

where, $|\cdot|$ operator denotes module of a complex number, f_0 and f_1 are the DC and first non-zero frequency components of the Fourier transform which are used for normalization, respectively (Persoon & Fu, 1977).

Classification performance of these features was measured in (Pourghassem & Ghassemian, 2008) by using the centroid distance and complex representation. According to this work, F_X feature has a classification accuracy of 42%. Zheng (Zheng, 2009) states that these visual features are one of the most preferred visual features in CAD systems.

3.3.2.4 Radial Distance Features

Radial distance signal represents distribution of contour pixels in means of distance to the centroid. Since margin is the one of the most important property of breast mass, features extracted from this signal carries very useful information to identify mass margin. Georgiou et al. (Georgiou et al., 2007) proposed 7 features based on radial distance signal whose formal definition is given in Table 3.7.

Table 3.7 List of radial distance features.

Feature Name	Formal Definition
Radial Distance Mean	$\mu = \frac{1}{N} \sum_{i=1}^N d(i)$
Radial Distance Standard Deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N \{d(i) - \mu\}^2}$
Mass Circularity	$C = \frac{P^2}{A}$
Entropy	$E = - \sum_{k=1}^N P_k \log(P_k)$
Area Ratio	$A_R = \frac{1}{m \cdot N} \sum_{i=1}^N \{d(i) - \mu\}$ where $A_R = 0, \forall d(i) \leq \mu$
Zero Crossing Count	$ZC = \sum_{i=1}^{N-1} \begin{cases} 1 & \text{sign}(d(i) - \mu) \neq \text{sign}(d(i+1) - \mu) \\ 0 & \text{otherwise} \end{cases}$
Mass Boundary Roughness	$R(j) = \sum_{i=j}^{L+j} d(i) - d(i+1) , \quad j = 1, \dots, \left\lfloor \frac{N}{L} \right\rfloor$

Performance of these 7 features is tested on a subset of DDSM dataset. Reported performance of these 7 features varies from 89.8% to 96.7% in AUC performance measure (Georgiou et al., 2007). Additionally in (Georgiou et al., 2007), Discrete Fourier and Wavelet Transforms of radial distance signal are also considered, and it is reported that Fourier representation of radial distance signal yield a classification performance better than the one by the original signal.

3.3.2.5 Region Based Shape Feature

Region-based shape feature (Ricard et al., 2005), which is an element of MPEG-7 standard, represents pixel distribution of a 2-D object. Feature uses a complex transform named Angular Radial Transform (ART). ART coefficients are defined by following formula.

$$F_{nm} = \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) I(\rho, \theta) \rho d\rho d\theta$$

where $V_{nm}(\rho, \theta) = A_m(\theta)R_n(\rho)$, $A_m(\theta) = \frac{1}{2\pi}e^{jm\theta}$ and

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2 \cos(\pi n \rho) & n \neq 0 \end{cases}. \text{ Here } F_{nm} \text{ denotes ART coefficient of order } n \text{ and } m.$$

Feature uses twelve angular and three radial functions. To the best of our knowledge, this shape feature has not been used in mammography imaging.

3.3.2.6 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, $|\cdot|$ 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,

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

Rosa et al. (N. A. Rosa et al., 2008) used Zernike Moments in a mammographic CBIR system. Experimentations on DDSM dataset show that the method achieves 90% of precision with respect to the recall.

3.3.3 Texture Features

Texture could be considered as repeating patterns of a local variation of pixel intensities (T. Ojala, Pietikainen, & Maenpaa, 2002) and widely used in CBIR systems. Similarly, these features are one of the most preferred low-level image features in mammography area. Most of the selected texture features in this work are already used in mammography. We also evaluate performance of four features in the literature, which is not used in mammography. Following sections give detailed information about texture features used in this work.

3.3.3.1 Statistics of Gray Level Histogram (SGLH)

Gray level histogram of an image includes useful information about it, especially about intensity distribution of the interested area. Statistical features of an area summarize pixel intensity of an image or object. Additionally, texture of an object could be defined as local intensity relationships among pixels, like Haralick-14 (Haralick et al., 1973), which will be described in next section. Let x_{ij} be the gray level pixel value at location (i, j) , N be the number of pixels in the image, L be the number of gray levels, $p(i)$ be the probability of intensity level i in the histogram and $P(x_{ij})$ be the probability of intensity level at x_{ij} . Features that are used to describe texture of a region or image are listed in Table 3.8.

Table 3.8 List of statistical histogram features.

Feature Name	Formal Definition	Measure
Mean	$\mu = \frac{\sum_{ij} x_{ij}}{N}$	Average Pixel Intensity
Standard Deviation	$\sigma = \sqrt{\frac{\sum_{ij} (x_{ij} - \mu)^2}{N}}$	Standard deviation of region of interest
Smoothness	$R = 1 - \frac{1}{(1 + \sigma^2)}$	Smoothness of a region
Skewness	$S = \frac{\sum_{ij} (x_{ij} - \mu)^3}{N\sigma^3}$	A measure of symmetry around average pixel intensity value

Table 3.8 List of statistical histogram features. (Cont.)

Feature Name	Formal Definition	Measure
Uniformity	$U = \sum_{i=0}^{L-1} p(i)^2$	Uniformity of histogram
Kurtosis	$K = \frac{\sum_{ij} (x_{ij} - \mu)^4}{(N - 1)\sigma^4}$	Measure to determine flatness of histogram distribution
Average Histogram	$AH_g = \frac{1}{L} \sum_{i=0}^{L-1} N(i)$	Intensity probabilities
Modified Standard Deviation	$\sigma_m = \sqrt{\sum_{i,j} (x_{ij} - \mu)^2 P(x_{ij})}$	Average contrast
Modified Skewness	$MS = \frac{1}{\sigma^3} \sum_{ij} (x_{ij} - \mu)^3 P(x_{ij})$	A measure of symmetry around average pixel intensity value

Subashini et al. (Subashini et al., 2010) used these statistical features in their automatic tissue density assessment approach. They report that with usage of pre-segmentation of breast tissue and using Radial Basis Function (RBF) kernel SVM classifier, average performance of these features is about 95%, evaluated by using 3-fold cross validation. Antonie et al. (Antonie et al., 2003) used these features in a rule-based classification system of mammography. In the proposed system, images are first segmented to remove unwanted regions. Then, a rule-learning algorithm is run and 80% of accuracy was reported from experimentation on MIAS dataset (Suckling et al., 1994). Kinoshita et al. (Kinoshita et al., 2007) measured performances of these features in conjunction with other features and the performance was obtained about 80%. MedGIFT system, which is used in the development of a reference medical dataset, also uses SGLH features (Müller, Rosset, et al., 2004).

3.3.3.2 Haralick-14

Haralick-14 features were proposed by Haralick et al. (Haralick et al., 1973) and intensively used in many texture related researches. These features aim to describe texture by using spatial intensity dependencies. In their work, texture was defined by calculating some statistical properties of Gray Level Co-occurrence Matrix (GLCM) representing intensity co-occurrence at d -distance and θ angle. GLCM is a square

matrix whose rows and columns represent a gray level intensity values. Hence, dimension of the GLCM matrix depends on the number of gray levels in the image. GLCM matrix is calculated by using the following formula.

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^m \sum_{q=1}^n \begin{cases} 1, & I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases}$$

Here, $I(x, y)$ is the image function, m and n are image width and height; Δx and Δy are distance parameters defined by d and θ , respectively. Figure 3.4 shows related pixels at $d=1$ and $\theta=(0^\circ, 45^\circ, 90^\circ, 135^\circ)$.

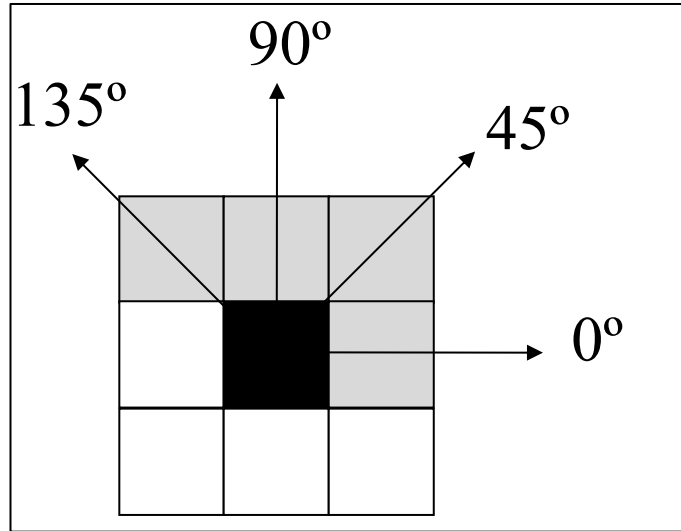


Figure 3.4 Pixels of 8-neighborhood at $d=1$ and $\theta=(0^\circ, 45^\circ, 90^\circ, 135^\circ)$.

After calculating co-occurrence matrix, each element of the matrix is normalized by using the following formula to create normalized co-occurrence matrix ($p(i, j)$).

$$p(i, j) = \frac{C(i, j)}{\sum_{k=1}^L \sum_{l=1}^L C(k, l)}$$

where, L denotes the number of gray levels in the image, $p(i, j)$, which is also denoted as p_{ij} , is the (i, j) -th element of the normalized co-occurrence matrix. Table 3.9 shows the list of features extracted from GLCM.

Table 3.9 Haralick-14 features extracted from GLCM.

Feature	Formal Definition
Angular second moment (f_1)	$\sum_i \sum_j \{p_{ij}\}^2$
Contrast (f_2)	$\sum_{n=0}^{N_g} n^2 \left\{ \sum_{\substack{i=1 \\ i-j =n}}^{N_g} \sum_{j=1}^{N_g} p_{ij} \right\}$
Correlation (f_3)	$\frac{\sum_i \sum_j (ij) p_{ij} + \mu_x \mu_y}{\sigma_x \sigma_y}$ where μ_x and σ_x are row-wise average and standard deviation of co-occurrence matrix; μ_y and σ_y are column-wise average and standard deviation of co-occurrence matrix
Sum of Squares (Variance) (f_4)	$\sum_i \sum_j (i - \mu)^2 p_{ij}$
Inverse Difference Moment (f_5)	$\sum_i \sum_j \frac{1}{1 + (i - j)^2} p_{ij}$
Sum Average (f_6)	$\sum_{i=2}^{2N_g} i \left\{ \sum_{\substack{j=1 \\ j+k=i}}^{N_g} \sum_{k=1}^{N_g} p_{jk} \right\}$
Sum Variance (f_7)	$\sum_{i=2}^{N_g} (i - f_8) \left\{ \sum_{\substack{j=1 \\ j+k=i}}^{N_g} \sum_{k=1}^{N_g} p_{jk} \right\}$
Sum Entropy (f_8)	$- \sum_{i=2}^{N_g} \left\{ \sum_{\substack{j=1 \\ j+k=i}}^{N_g} \sum_{k=1}^{N_g} p_{jk} \right\} \log \left\{ \sum_{\substack{j=1 \\ j+k=i}}^{N_g} \sum_{k=1}^{N_g} p_{jk} \right\}$
Entropy (f_9)	$- \sum_i \sum_j p_{ij} \log(p_{ij})$
Difference Variance (f_{10})	$\text{variance of } \left\{ \sum_{\substack{i=1 \\ i-j =n}}^{N_g} \sum_{j=1}^{N_g} p_{ij} \right\}$
Difference Entropy (f_{11})	$- \sum_{i=0}^{N_g-1} \left\{ \sum_{\substack{j=1 \\ j-k =i}}^{N_g} \sum_{k=1}^{N_g} p_{jk} \right\} \log \left\{ \sum_{\substack{j=1 \\ j-k =i}}^{N_g} \sum_{k=1}^{N_g} p_{jk} \right\}$
Information Measures of Correlation (f_{12}, f_{13})	$f_{12} = \frac{f_9 - HXY1}{\max\{HX, HY\}}$ $f_{13} = 1 - e^{\sqrt{-2(HXY2 - f_9)}}$ where HX and HY are entropies of p_x and p_y , and $HXY1 = - \sum_i \sum_j p_{ij} \log(p_{x,i} p_{y,j})$ $HXY2 = - \sum_i \sum_j p_{x,i} p_{y,j} \log(p_{x,i} p_{y,j})$

Table 3.9 Haralick-14 features extracted from GLCM. (Cont.)

Feature	Formal Definition
Maximal Correlation Coefficient (f_{14})	$f_{14} = \sqrt{\text{Second largest eigenvalue of } Q}$ where; $Q(i, j) = \sum_k \frac{p_{ik}p_{jk}}{p_{x,i} \cdot p_{y,j}}$

Pourghassem and Ghassemian (Pourghassem & Ghassemian, 2008) use these features to classify medical image type and obtain 22% of accuracy. It is also reported that accuracy of these feature-set increases to 54% when used in conjunction with some other visual features. Lauria (Lauria, 2009) uses these features in a distributed mass and microcalcification detection system, and obtains and reports sensitivity rates are 80% and 96%, respectively. Rangayyan et al. (Rangaraj M Rangayyan et al., 2010) investigates the effects of mammographic pixel resolution on performances of these features and finds that these features perform the best at images, which have 100 μ m and 200 μ m pixel resolution. Yin et al. (Fan et al., 2011) uses these features in a medical image categorization system and obtains 97% of accuracy in a small dataset. As mentioned before, Kinoshita (Kinoshita et al., 2007) uses these features in conjunction with other visual features in a mammographic classification system and obtained 80% of accuracy. Zheng (Zheng, 2009) states that Haralick's texture features are the one of the most commonly used visual feature in CAD system.

3.3.3.3 Gray Level Difference Matrix

Gray Level Difference Matrix (GLDM), which is a texture representation of images, is proposed by Weszka et al. (Kim & Park, 1999). These features consist of statistical properties of absolute intensity differences of two pixels at a given displacement. In other words, unlike Haralick's texture features, this feature aims to measure intensity deviation between two pixels whose distance is defined by vector length and angle between them. A difference image is calculated using the following formula.

$$f_{\Delta x, \Delta y}(x, y) = |f(x, y) - f(x + \Delta x, y + \Delta y)|$$

where, $(\Delta x, \Delta y)$ denotes displacement between two pixels, $f_{\Delta x, \Delta y}$ and f represents difference image and original image, respectively.

Kim and Park (Kim & Park, 1999) used this texture representation in their microcalcification detection system and report that their system has 74% of AUC value in a dataset, which has 120 X-Ray Mammograms.

3.3.3.4 Local Binary Patterns

Local Binary Patterns (LBP) was proposed by Ojala et al. (Timo Ojala et al., 1996) to represent texture of a region or the whole image. Extraction of this feature uses 8-neighbour of each pixel. Neighbors of a pixel are threshold by using pixel value. Each neighbor above the threshold is multiplied by weight factor as shown in Figure 3.5.

1	6	3
9	5	7
8	4	2

0	1	0
1		1
1	0	0

1	2	4
128		8
64	32	16

$$\mathbf{LBP} = 2 + 8 + 64 + 128 = 202$$

$$\mathbf{C} = (6+7+8+9)/4 - (1+2+3+4)/4 = 5$$

Figure 3.5 Calculation of LBP and C value.

Ojala et al. (T. Ojala et al., 2002) extends LBP to include rotation invariance and uniformity by adding right shift operation to the LBP calculation. In medical field, performance of this feature is measured by Feng et al. (Dagan Feng et al., 2008) in conjunction with MPEG-7 Edge Histogram Feature. Accuracy of method is calculated as 80%. Although there are medical applications of this feature, there is no work in mammography area that uses this feature to the best of our knowledge.

3.3.3.5 Edge-Histogram

Edges are important textural information for image retrieval and similarity calculation. Edge histogram feature aims to represent spatial edge information of an image (Park et al., 2000) and it is an element of MPEG-7 standard. Extraction procedure of this texture feature is implemented in spatial domain unlike other two MPEG-7 texture features. First, the image is divided into 4×4 sub images. A histogram of 5 standardized edge directions, shown in Figure 3, is computed for each sub image.

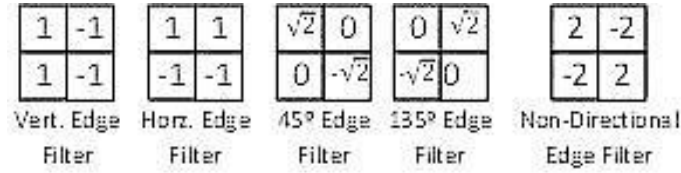


Figure 3.6 Filters used to determine edges.

To calculate edge histogram, edge filter responses at each 2×2 non-overlapping blocks of each sub-image is calculated and a block is assigned to the edge whenever filter response is maximum. Hence, the number of elements in edge histogram feature is 80 (union of 5 bin histograms of each 16 sub-image).

In medical image field, usage of edge histogram feature is not very often. Feng et al. (Dagan Feng et al., 2008) uses this feature on imageCLEFmed dataset. Accuracy of classification is measured 80.5%. A small performance improvement is also reported when edge histogram is used in conjunction with local binary patterns feature (Dagan Feng et al., 2008). Ojala et al. (Timo Ojala et al., 2002) proposes an extension on edge histogram feature by adding edge histograms for row-wise, column-wise and 2×2 groups of sub images. Hence, total number of elements becomes 150. No research on mammography area using this feature exists in literature as we know.

3.3.3.6 Gabor Filters

Gabor functions are Gaussian functions that are modulated by complex sinusoids. In image processing, these functions are used for edge and bar detections. Gabor filters are defined by the following functions;

$$g_{mn}(x', y') = a^{-m} \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) e^{-\frac{1}{2} \left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2} \right) + 2\pi j W x}$$

where $a > 1$, m, n are integers, $x' = a^{-m}(x \cos \theta + y \sin \theta)$, $y' = a^{-m}(-x \sin \theta + y \cos \theta)$, $\theta = \frac{n\pi}{K}$, K is the number of orientations (Ma & Manjunath, 1996). As a texture feature, the following representation is used. For an image $I(x, y)$, its Gabor transform is computed as:

$$W_{mn}(x, y) = \int I(x_1, y_1) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1$$

where, $*$ denotes the complex conjugate. Then, a texture of a region is now characterized by means and standard deviation of W_{mn} . So, a Gabor feature vector is denoted as follows:

$$F = [\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, \dots, \mu_{SK}, \sigma_{SK}]^T$$

where, $\mu_{mn} = \iint |W_{mn}(x, y)| dx dy$ and $\sigma_{mn} = \sqrt{\iint (|W_{mn}(x, y)| - \mu_{mn})^2 dx dy}$, S is the number of scales in the multiresolution decomposition.

In medical image field, this feature set is used in some works to represent texture information of the image. Müller et al. (Müller, Rosset, 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 features” 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.

3.3.3.7 Homogenous Texture

Homogeneous Texture feature is given in MPEG-7 standard as low-level feature aiming to characterize texture of a region. However, to the best of our knowledge, it is not used in medial image retrieval. The feature is based on energy and energy deviation from set of frequency channels of image. Frequency channels are determined by Gabor functions in frequency domain and are indexed by (s, r) where s is the radial index and r is the angular index. Hence, the following formula is used to model the (s, r) -th frequency channel.

$$G_{s,r}(\omega, \theta) = e^{\frac{-(\omega-\omega_s)^2}{2\sigma_s^2}} \cdot e^{\frac{-(\theta-\theta_r)^2}{2\tau_r^2}}$$

where σ_s and τ_r are the standard deviation of the Gaussian in the radial direction and angular direction, respectively. Angular length of each frequency channel is 30° . So, $\theta_r = 30^\circ \cdot r$ with $r \in \{0,1,2,3,4,5\}$. In radial direction, centers of each frequency channels are spaced by an octave scale which is $\omega_s = \omega_0 2^{-s}$ with $s \in \{0,1,2,3,4\}$ and $\omega_0 = 3/4$. Additionally, bandwidth of a frequency channel is denoted as $B_s = B_0 2^{-s}$ with $s \in \{0,1,2,3,4\}$ and $B_0 = 1/2$. Under this setting, σ_s and τ_r parameters of Gabor function above could be estimated as follows.

$$\tau_r = \frac{15^\circ}{\sqrt{2 \ln 2}}$$

$$\sigma_s = \frac{B_s}{2\sqrt{2 \ln 2}}$$

Energy of a frequency channel is the log-scaled sum of the square of the Gabor-filtered Fourier transform coefficients of an image.

$$e_i = \log_{10}[1 + p_i]$$

where, p_i is sum of the squares of the Gabor-filtered Fourier transform coefficients of frequency channel i and defined as follows.

$$p_i = \sum_{\omega=0}^1 \sum_{\theta=0^\circ}^{360^\circ} \left[G_{s,r}(\omega, \theta) \left| \sqrt{\omega_x^2 + \omega_y^2} \right| F(\omega \cos \theta, \omega \sin \theta) \right]^2$$

where, ω_x and ω_y are elements of Jacobian between Cartesian and Polar coordinates, $i = 6 \times s + r + 1$, and $F(u, v)$ is the Fourier transform of image $f(x, y)$. Similarly, energy distribution of a frequency channel is log-scaled standard deviation of the squares of the Gabor filtered Fourier transform coefficients of an image.

$$d_i = \log_{10}[1 + q_i]$$

where, q_i is the standard deviation of the squares of the Gabor filtered Fourier transform coefficients of frequency channels i and defined as follows;

$$q_i = \sqrt{\sum_{\omega=0}^1 \sum_{\theta=0^\circ}^{360^\circ} \left\{ \left[G_{s,r}(\omega, \theta) \left| \sqrt{\omega_x^2 + \omega_y^2} \right| F(\omega \cos \theta, \omega \sin \theta) \right] - p_i \right\}^2}$$

Finally, mean intensity (f_{DC}) and standard deviation of intensity (f_{SD}) is calculated. As a result, homogeneous texture feature is shown in the following formula.

$$HTD = [f_{DC}, f_{SD}, e_1, e_2, \dots, e_{30}, d_1, d_2, \dots, d_{30}]$$

To the best of our knowledge, this MPEG-7 feature is not used in mammography area.

3.3.3.8 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 feature, like Homogeneous texture feature, is not used in medical field. Representation of this feature is defined as the following feature vector.

$$TBD = [v_1, v_2, v_3, v_4, v_5]$$

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 feature uses Gabor filter functions with 6 orientations and 4 scales as in the Homogeneous Texture feature. Like Homogeneous Texture feature, there is no CADx system using this feature in mammography area to the best of our knowledge.

3.3.4 Margin Features

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 (Delogu, Evelina Fantacci, Kasae, & Retico, 2007; Mudigonda, Rangayyan, & Desautels, 2000; R. M. Rangayyan, Mudigonda, & Desautels, 2000; Varela, Timp, & Karssemeijer, 2006). Although shape features are useful for margin characterization, intensity difference between inner and outer object areas is another important feature.

In this work, we propose a new set of low-level features aiming 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 formulas.

$$r = \sqrt{x^2 + y^2}$$

$$\theta = \tan^{-1} \left(\frac{y}{x} \right)$$

where (x, y) is the coordinates of original image, (r, θ) are length and angle axis of the polar coordinate system. Figure 3.7 contains both original and segmented regions and their polar representations.

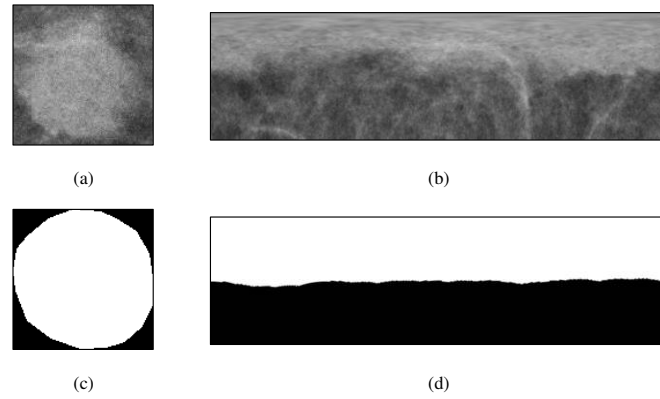


Figure 3.7 (a) Original ROI (b) Polar representation of original ROI (c) Binary segmentation of the mass (d) Polar representation of the segmented ROI

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 the statistical features in Table 3.10.

Table 3.10 List of the margin features.

Feature Name	Formal Definition
Angular Means	$\mu(i) = \frac{1}{N} \sum_{j=1}^M P(i, j)$
Angular Standard Deviation	$\sigma(i) = \sqrt{\frac{1}{N} \sum_{j=1}^M (P(i, j) - \mu(i))^2}$

Table 3.10 List of the margin features. (Cont.)

Feature Name	Formal Definition
Angular Skewness	$\gamma_1(i) = \frac{\frac{1}{N} \sum_{j=1}^M (P(i,j) - \mu(i))^3}{\sigma(i)^3}$
Angular Kurtosis	$\gamma_2(i) = \frac{\frac{1}{N} \sum_{j=1}^M (P(i,j) - \mu(i))^4}{\sigma(i)^4}$
Angular Mean Differences	$\mu_D(i) = \mu_{IR}(i) - \mu_{OR}(i)$
Angular Standard Deviation Differences	$\sigma_D(i) = \sigma_{IR}(i) - \sigma_{OR}(i)$
Angular mean differences of inner and outer margin area	$\mu_{MD}(i) = \mu_{IMA}(i) - \mu_{OMA}(i)$
Angular standard deviation differences of inner and outer margin area	$\sigma_{MD}(i) = \sigma_{IMA}(i) - \sigma_{OMA}(i)$
Global region statistics	$GRS = \{\mu_{IR}, \sigma_{IR}, \gamma_{1,IR}, \gamma_{2,IR}, \mu_{OR}, \sigma_{OR}, \gamma_{1,OR}, \gamma_{2,OR}\}$

where i parameter denotes i -th column of specified region, $P(i,j)$ denotes (i,j) element of polar representation of the region. We calculate column-wise statistics in each case, since each column carries angular data of original region. For instance, statistical properties of inner and outer region areas will have very high differences if mass has a sharp margin. To support scale independency, first 20 Fourier features of each feature, shown in Table 3.10, except global statistics feature are used as features.

CHAPTER FOUR

BREAST MASS CONTOUR SEGMENTATION

4.1 Overview

Precision of breast mass segmentation methods play an important role in CAD systems for mass classification point of view, since most valuable properties of breast masses that define malignancy is related to mass morphology. For instance, malign masses tend to spread 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. Rangayyan et al. (Rangaraj M. Rangayyan & Nguyen, 2005) states importance of shape characteristics while defining mass malignancy and proposed several low-level features to determine mass characteristics by using region shape. Thus, diagnosing performance of a CAD system depends on accuracy of segmentation method. But segmentation algorithms that have high accuracy could be useless in some situations. At this point, a useful segmentation means that segmentation algorithm is successful if segmentation results not only have high accuracy, but also have high consistency with real mass area. Finding real mass area is a crucial task even for a radiology expert. Therefore success rate of a segmentation algorithm can be measured by using accuracy and conformance with selection of expert. In other words, a segmentation algorithm can produce useful diagnosis suggestions if consistency between estimated mass area and mass selection of radiology expert increases.

Recent studies show that machine-learning methods are frequently used for this segmentation issue. For instance, Tao et al. (Tao, Lo, Freedman, Makariou, & Xuan, 2010) propose a classification system to identify spiculation of a mass with help of automatic segmentation approach. In their study, the ROI of a candidate mass is splitted into sub regions and each sub region is labeled using machine learning techniques. Then, the final segmentation result is obtained by using graph-cut algorithm and optimization methods. Song et al. (Song et al., 2009) propose a segmentation method, which uses plane-fitting method based on dynamic

programming optimization approach. Proposed method includes a preprocessing step in which edge information is enhanced by using gradients. Experimentation results of the proposed method on a subset of publicly available dataset give an acceptable level of performance. Meanwhile, in another recent study by Song et al. (Song et al., 2010), performance of dynamic programming with combination of template matching is tested and found that the template matching approach performs better than plane fitting approach. A very recent an in-depth review on this topic with respect to mass detection can be found in (Oliver et al., 2010).

Some of the studies focus on contour tracing approaches since margin characteristics of masses includes important clues about for malignancy of masses. Elter et al. (Elter, Held, & Wittenberg, 2010) proposed a contour tracing approach to extract shape of the region. To extract contour information, the ROI of the candidate mass is transformed into polar coordinate system, and then contour of estimated mass is calculated by using a shortest path algorithm. Proposed segmentation method tested on 60 ROI images of DDSM dataset and proposed method obtains the highest performance scores. Hong et al. (Hong & Sohn, 2010), propose different representation using the iso-contour maps for topographic, in which a salient region forms a dense quasi-concentric pattern of contours. The topological and geometrical structure of the image is analyzed using an inclusion tree that is a hierarchical representation of the enclosure relationships between contours. They have evaluated their approaches only for salient masses and show that a topographic representation is largely invariant to brightness and contrast, and it provides a robust and efficient representation for the characterization of mammographic features. Some other researchers use well-known segmentation methods to identify mass region. Domínguez and Nandi (Rojas-Domínguez & Nandi, 2008) propose a new thresholding approach, which uses combination of several threshold levels. Wei et al. (Wei et al., 2012) uses a seeded region growing approach to extract mass region in their retrieval system. Dubey et al. (Dubey, Hanmandlu, & Gupta, 2010) measure performance of Level-set and watershed segmentation methods on mass segmentation.

In this thesis, a new approach to the problem of segmentation of breast masses from the surrounding background tissue in a given ROI is presented. This approach is based on the concept of region growing with a capability of selecting optimum threshold value adaptively. Hence, we called this new approach Breast Mass Contour Segmentation (BMCS) method. The proposed approach is evaluated on DEMS dataset. The performance of the proposed approach is directly compared to the performance of the implementations of three state-of-the-art region growing approaches solutions, in terms of a set of different measurement metrics, such as Yassnoff Distance, Balanced Accuracy. Experimental results show that BMCS outperformed other three segmentation methods. Additionally, we use BMCS to measure performance of region selection method over mass classification task.

4.2 Breast Mass Contour Segmentation (BMCS)

Breast mass detection using a single mammography image relies on the fact that pixels inside a mass have different characteristics from the other pixels within the breast area. These characteristics can be simply related to grey-level intensity, texture or morphological characteristics.

Breast mass segmentation methods can be roughly classified into three groups: Region based, contour-based and clustering methods. Region-based methods divide the image into spatially connected homogeneous regions while contour-based methods rely on the boundary of regions. On the other hand, clustering methods group together the pixels having the same properties and might result in non-connected regions. Since goal of this study is to enhance boundary of a mass from a given region, we proposed an improvement on region growing approach, which is one of the well-known region based method.

4.2.1 Reference Segmentation Methods

We choose three segmentation methods to measure performance of our segmentation approach. All of the segmentation approaches are region based, since we focused on mass boundary enhancement instead of mass detection. And there are

some of the works (Dubey et al., 2010; Wei et al., 2012) available, which uses region-based methods to enhance initial detection results. Selected segmentation approaches are discussed below, shortly.

4.2.1.1 Watershed Segmentation

Watershed Segmentation method is based on Watershed transform (Beucher & Lantuejoul, 1979; Vincent & Soille, 1991). This method aims to find catchment basins, which define border between two objects. If water falls into these basins, level of the water rises until neighbor basins share the same level. So output of the algorithm is a hierarchy of catchment basins. The key point is to find most discriminative basins, since most discriminative basins are the basins that separate two different objects.

4.2.1.2 Level-Set Segmentation

Level-set segmentation approach is region enhancement approach, which evolves an initial region according to an energy field (Osher & Sethian, 1988). In other words, level-set methods start with an initial region and evolve this region while minimizing region energy. Energy of a region is calculated using the level-set equation, which is a partial differential equation, in each step. Finally algorithm stops when difference between energy of region in consecutive steps falls under a threshold value.

4.2.1.3 Seeded Region Growing Segmentation

Our approach, BMCS is based on Seeded Region Growing (SRG) method (Adams & Bischof, 1994). SRG segmentation approach expands an initial region or point by similar neighbor pixels. Similar term means a pixel whose intensity is in predefined range with seed region/point in this scope. So, we include seeded region growing method in performance comparison as basis.

4.2.2 Breast Mass Contour Segmentation Algorithm

In this thesis, we propose a Breast Mass Contour Segmentation (BMCS) approach for a given ROI in an image, which is based on seeded region growing (SRG) algorithm and it is a variant of region based segmentation methods. SRG algorithm starts with a seed point and extends it by adding neighbor pixels that intensity value exceeds a predefined threshold value. But this approach is not appropriate for mass contour segmentation, because it is hard to choose a general threshold value for all kind of masses. Thus, SRG segmentation method needs to be tuned for each kind of masses. Otherwise, SRG segmentation method result too small or too large mass segments than reference mass segment, and these results are commonly referenced as *under-* and *over-segmentation*, respectively. We proposed a new thresholding approach for SRG, which adaptively adjusts the threshold value based on mass size estimation to prevent *under-* and *over-segmentation*.

BMCS contains four steps; (1) ROI detection, (2) preprocessing, (3) mass size estimation and (4) segmentation. At first step, BMCS needs a rough bounding rectangle of a mass. Since BMCS focus on mass boundary enhancement, bounding rectangle selection method is discarded. In the experiments, we prefer to use bounding rectangles of a masses region drawn by user and given to system as input. We avoid using result of a machine segmentation method, since false positives may occur in machine segmentation techniques (Rojas-Domínguez & Nandi, 2009).

In preprocessing step, the goal is to increase contrast level between mass and background tissue; as well as emphasizing contour information. Initially, we trimmed out the pixels with extreme intensity values, which are highest and lowest 5% of intensity histogram. Although we eliminated extreme pixel intensities from ROI, there are still exceptional circumstances must be solved, like calcifications appearing in front of the masses. So, we applied a median filter with window size of 5% of the bigger ROI dimension (width or height) to ROI for eliminating undesirable situations. However, applying a median filter to a ROI may produce artificial contours. Hence, we applied an averaging filter with the same size of median filter to overcome false contour problem. Finally, a Laplacian edge enhancement filter with

same size of previously applied filters is used to enhance edges of ROI to emphasize edge information.

In mass size estimation step, we estimate mass size empirically by using an intensity histogram based segmentation approach named OTSU. OTSU segmentation algorithm aims to find one or more split points on intensity histogram which separates whole intensity histogram into two or more groups whose intra-class variances are minimum (Smith et al., 1979). In this step, we initially applied OTSU histogram segmentation approach to divide ROI into two groups namely large and small segment. Therefore, we could say that size of a segment in ROI should be larger than small segment and smaller than large segment. As a result, we have obtained an estimation of size interval for real mass segment. Moreover, we used division point of intensity histogram found in this step as initial threshold value in forth step.

BMCS is based on Seeded Region Growing (SRG) segmentation algorithm (Adams & Bischof, 1994), which requires a predefined seed points or seed area in an image and expands initial seed by adding neighbor pixels while they satisfy following condition.

$$|\mu - I_{x',y'}| < \theta$$

where μ is the average intensity level of segmented region, $I_{x',y'}$ is the intensity level at point (x', y') which is a neighbor of current segmented region and θ is the intensity threshold. μ value is recalculated when a new pixel is added to evolving region. We proposed some changes in SRG to apply breast mass segmentation problem. First, we select the brighter pixels in center of ROI as initial seed points, since we know that breast masses contain the brighter pixels than their neighbor pixels and there is a real mass in center of ROI. We also expand initial seed with neighbor pixels satisfying following condition.

$$I_{x',y'} > \theta$$

where $I_{x',y'}$ is the intensity of neighbor pixel to be added to current evolving region and θ is the intensity threshold. Another crucial problem of SRG is finding a proper threshold value, because it is hard to find a general threshold value for each mass. Therefore, we proposed an adaptive threshold modification approach to solve this problem. Initially, we select split point calculated in mass size estimation step as preliminary threshold value. When no neighbor pixels left to satisfy expansion condition above, we check the area of the current evolving mass region whether its area is greater than area of small segment. If the area of current region is smaller than *small segment*, we decrease threshold value and continue to expand region. Similarly, we stop segmentation process if area of the evolving region exceeds the area of the *large segment*. Consequently, result of our segmentation algorithm produces segments whose area stays in predefined interval. Furthermore, our approach proposes a solution for *over-* and *under-segmentation* problems. Finally, we dilate final segment using a sphere-shaped morphological to filter out remaining edge artifacts. Figure 4.1 and Figure 4.2 contains details of BMCS.

```

// Input: ROI
// Output: Set Segmented Points
function BMCS(ROI)
Width, Height  $\leftarrow$  Size of ROI
H  $\leftarrow$  Intensity Histogram of ROI
Eliminate Pixels of ROI with highest and lowest 5% of H
Apply Median Filter To ROI with 5% of MAX(Width, Height)
Apply Smoothing Filter To ROI with 5% of MAX(Width, Height)
Apply Laplacian Filter to ROI with 5% of MAX(Width, Height)
T  $\leftarrow$  OTSU division point of ROI
MinSize  $\leftarrow$  Count of Pixels whose intensity value above T
MaxSize  $\leftarrow$  Count of Pixels whose intensity value below T
if MaxSize < MinSize then
    Exchange Values of MaxSize and MinSize
end
MassRegion  $\leftarrow$  {}
NeighbourPixels  $\leftarrow$  {}
CurrentPixel  $\leftarrow$  Pixel whose value is Maximum
while Intensity(CurrentPixel) > T && SizeOf(MassRegion) <= MaxSize
    NeighbourPixels  $\leftarrow$  NeighbourPixels  $\cup$  Neighbours of CurrentPixel
    MassRegion  $\leftarrow$  MassRegion  $\cup$  CurrentPixel
    CurrentPixel  $\leftarrow$  Best of NeighbourPixels
    if Intensity(CurrentPixel) <= T
        if SizeOf(MassRegion) < MinSize
            T  $\leftarrow$  T - 1
        end
    end
end
Dilate ROI using a sphere-filter with
    size of 5% of MAX(Width, Height)
return MassRegion

```

Figure 4.1 Algorithm of BMCS.

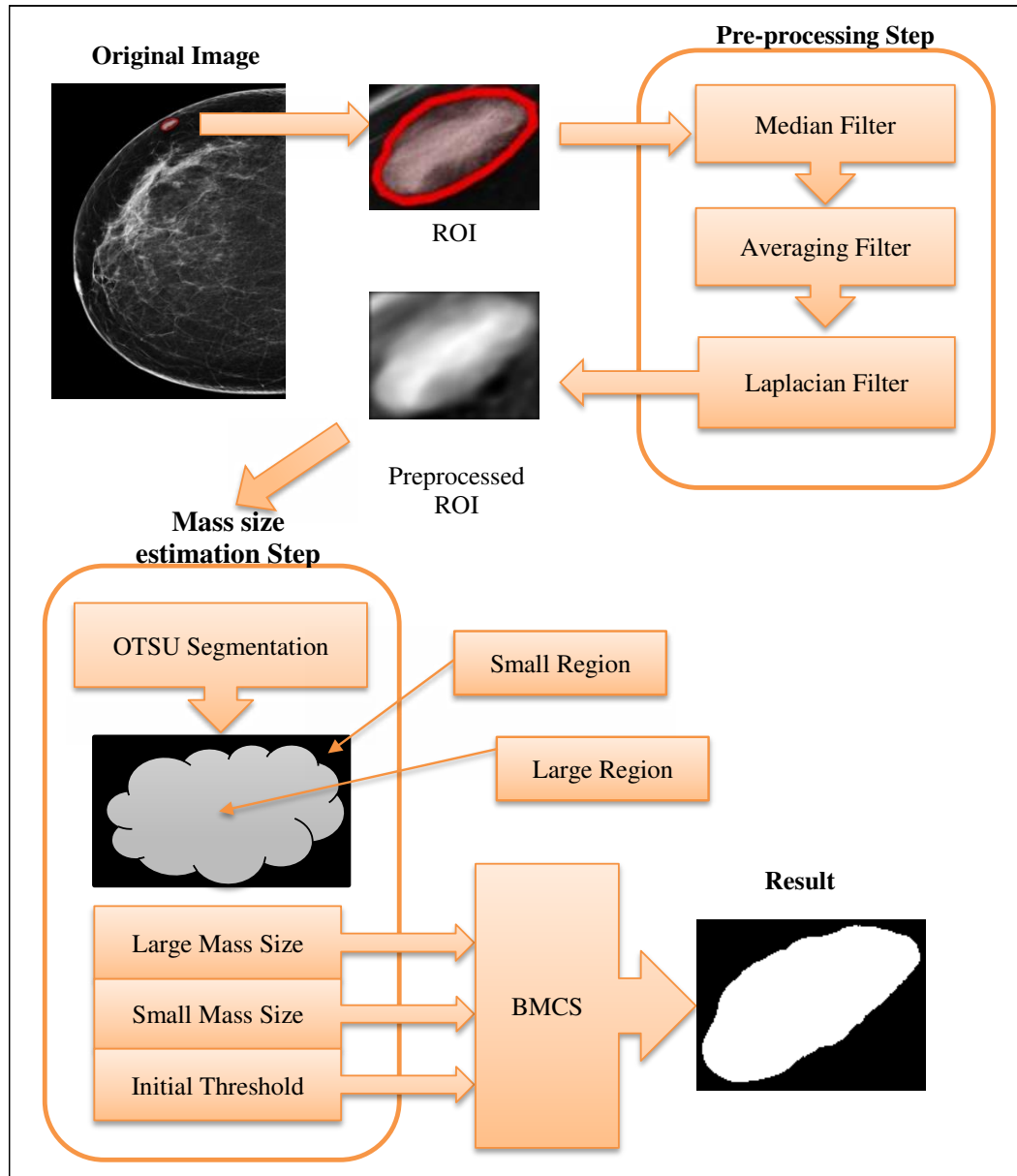


Figure 4.2 Block diagram of BMCS segmentation method.

4.3 Performance Evaluation

We run a set of experimentations on DEMS dataset and compared performance of BMCS with different region based segmentation algorithms. Each algorithm is implemented in C++ by using QT UI Framework and OpenCV image processing library. However, we need to select suitable evaluation metrics at first for a successful quantitative comparison. Thus, we first provide a survey on segmentation evaluation metrics.

4.3.1 Segmentation Evaluation Metrics

For a successful comparison, evaluation metrics are crucial for researchers to choose the right segmentation algorithm for their application needs or to compare their segmentation methods with other ones in the literature (Unnikrishnan, Pantofaru, & Hebert, 2007). Segmentation evaluation methods for this purpose can be classified into two major groups similarly to the machine-learning algorithms. The first group is supervised methods, which need previously defined region boundaries, and the other one is unsupervised methods that do not need previous region boundaries. Since we already have region boundary information, supervised evaluation metrics are more useful ones for our case.

4.3.1.1 Supervised Evaluation Metrics

Supervised evaluation metrics measures accuracy of segmentation method by using conformance of machine segmented region (MSR) and reference region (RR) (Hui Zhang, Fritts, & Goldman, 2008). Conformance between MSR and RR is determined in two ways; area conformance and edge conformance. Similarly, we can classify supervised segmentation evaluation metrics into two groups; area-based metrics (ABM), which measures conformance between MSR and RR, and edge-based metrics (EBM), which measures, edge conformance between MSR and RR.

4.3.1.1.1 Yasnoff Distance Metric. The one of the oldest and most commonly used ABM has been proposed by Yasnoff et al. (Yasnoff, Mui, & Bacus, 1977). Yasnoff proposes segmentation error metric, E_Y , which represents the ratio of misclassified pixels to the whole ROI. E_Y is defined by following formula.

$$E_Y = 1 - \frac{|B_S \cap B_R| + |M_S \cap M_R|}{|B_R + M_R|}$$

where B and M denotes background and mass pixel sets, subscript S and R denotes segmentation and reference, respectively. By the way, the symbol $|\cdot|$ represents the element count operator over a set. Second term in definition of E_Y is known as

Accuracy metric, which is commonly used in classification evaluation. Hence formula could also be rewritten as follows.

$$E_Y = 1 - \frac{TP + TN}{Total} = 1 - Accuracy$$

where TP and TN denotes successfully segmented mass and background pixels respectively. However, quality of accuracy measure is open to discuss, because of its weak predictive power. For instance, consider a ROI containing a very small mass only, a segmentation algorithm can achieve high accuracy rate if it assigns all pixels as the background. As a result, E_Y of the segmentation algorithm indicates very small error rate, even though algorithm could find no region in the image.

4.3.1.1.2 Distance Error Metric. Yasnoff proposed another error metric called *Distance Error* metric to overcome weak predictive power problem (Yasnoff et al., 1977). Distance error metric considers the spatial distribution of incorrectly segmented pixels to regions. Formal definition of distance error metric is given in following formula.

$$D_Y = \frac{100}{A} \sqrt{\sum_{i=1}^k d_i^2}$$

where A is the number of pixels in reference region and d_i is the distance between i -th pixel of $F_S \cap B_R$ and its nearest neighbor in F_R .

4.3.1.1.3 Classification-Based Metrics These metrics are well-known metrics from classification evaluation task, and Rosa et al. (B. Rosa, Mozer, & Szewczyk, 2011) proposed usage of these metrics for this kind of evaluation. In original work, a confusion matrix of segmented pixels are created; and *precision*, *recall* (a.k.a. *sensitivity*) and *F-Measure* metrics are calculated to measure performance of calculi segmentation approach. Formal definitions of metrics are given in following formulas.

$$Precision = \frac{|F_S \cap F_G|}{|F_S \cap F_G| + |F_S \cap B_G|}$$

$$Recall = \frac{|F_S \cap F_G|}{|F_S \cap F_G| + |B_S \cap F_G|}$$

$$F - Measure = 2 \cdot \frac{P \cdot R}{P + R}$$

Since these classification evaluation metrics is used for segmentation evaluation, following metrics could also be used in segmentation evaluation.

$$Specificity = \frac{|F_S \cap F_G|}{(|B_S \cap F_G| + |B_S \cap B_G|)}$$

$$Balanced Accuracy = \frac{Sensitivity + Specificity}{2}$$

4.3.1.1.4 Hausdorff Distance Metric. Hausdorff distance (Chalana & Kim, 1997) is an edge based segmentation evaluation metric. Let the edge of two objects represented by two sets; $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$ and $\mathcal{B} = \{b_1, b_2, \dots, b_m\}$, where a_i and b_i are the ordered pair of x and y points. Error metric between two edge point sets is called distance to closest point (DCP) and could be defined as following formula.

$$d(a_i, \mathcal{B}) = \min_j \|b_j - a_i\|$$

Hausdorff distance metric measures the maximum DCP between two edge point sets. Formal definition of Hausdorff distance is given below.

$$h(\mathcal{A}, \mathcal{B}) = \max \left(\max_i \{d(a_i, \mathcal{B})\}, \max_j \{d(b_j, \mathcal{A})\} \right)$$

4.3.1.1.5 Scalable Discrepancy Measures. Scalable discrepancy measures are

proposed to measure error of *over-* and *under-segmented* edge pixels (Odet, Belaroussi, & Benoit-Cattin, 2002). Formal definitions of these metrics are given below.

$$ODI = \frac{1}{N_O} \sum_{i=0}^{N_O} \left(\frac{d_O(i)}{d_{TH}} \right)^n$$

$$UDI = \frac{1}{N_U} \sum_{i=0}^{N_U} \left(\frac{d_U(i)}{d_{TH}} \right)^n$$

where N_O and N_U are the number of *over-* and *under-segmented* edge pixels, d_O and d_U are the distances of *over-* and *under-segmented* pixels, d_{TH} is the maximum allowed distance threshold, and n is the scale parameter for distances below d_{TH} .

4.3.2 Experimentation

We evaluated the performance of our segmentation approach, BMCS, against to the three other region based segmentation methods. The first method is proposed by Dominguez and Nandi (Rojas-Domínguez & Nandi, 2008), which uses level-set segmentation algorithm with Chan-Vese energy fields. Method uses an image contrast enhancement method and applies a median filter, since median filtering is used in several mammographic mass detection algorithms for reducing noise while preserving edge information (Dubey et al., 2010; Rojas-Domínguez & Nandi, 2008; Subashini et al., 2010). In this algorithm, we used manually selected region instead of using a predefined zero level. In other words, level-set segmentation algorithm is used to enhance or fine-tune a roughly selected region. Finally, region that has the largest area is labeled as mass.

The second method is proposed by Wei et al. (Wei et al., 2012), and uses seeded region-growing approach to enhance mass boundary. Proposed method includes a preprocessing step including a linear contrast enhancement method and median filtering to enhance distinction between mass region and background. Finally, multi-

seeded region growing method with empirically defined constant area threshold is used to find mass boundary.

The third method uses linear mapping and histogram equalization method for image enhancement (Dubey et al., 2010). Then, watershed segmentation algorithm is applied on enhanced image using edge pixels for background region marker and centroid of highest valued pixels for foreground region marker. Figure 4.3 depicts sample segmentation results of four methods and Table 4.1 has corresponding error metrics of each sample.

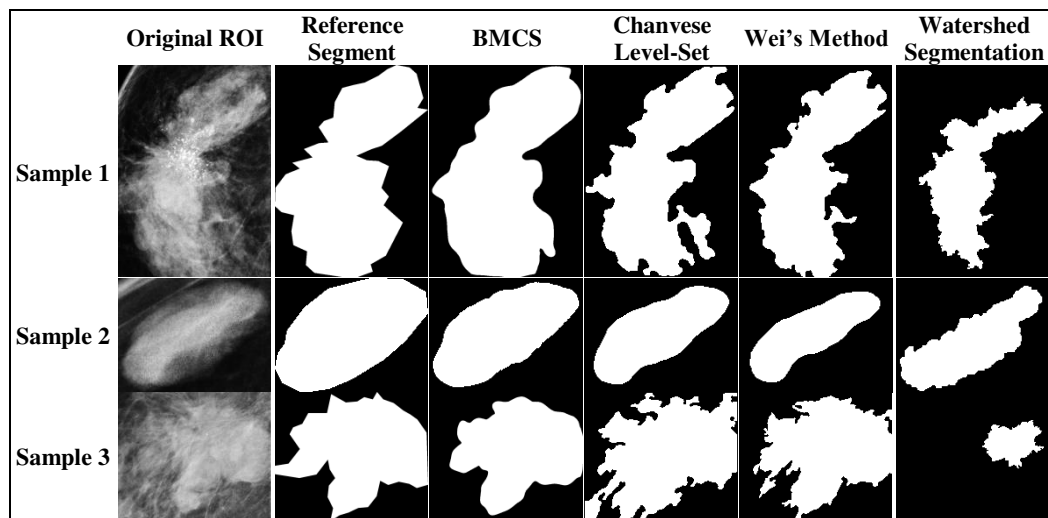


Figure 4.3 Original images and segmentation results with compared algorithms. Images represents from left to right as follows; original region image, expert selection (reference), BMCS results, level set segmentation results, Wei's segmentation method results, and watershed segmentation results.

BMCS algorithm has the highest True Positive Rate, which means that segmentation result cover maximum area of manually selected region among other methods for the first sample. On the other hand, FPR rate of proposed segmentation method is the highest score for the same sample. But produced false pixels distance to the manual selections has no significant difference according to the Yasnoff Error (E_y) metric comparing to the other methods. Similarly, BMCS produces the closest segment to the reference segment. BMCS outperforms all the other methods in all metrics for second sample segmentation result. For last sample, Level-Set Segmentation method and Wei's segmentation method produces higher true positive

values than our segmentation algorithm but they produce false positives at least two times higher than BMCS. Additionally, false positives produced by BMCS method have no statistical significance with LevelSet and Wei's segmentation methods. Moreover, Balanced accuracy of our method outperforms other methods. This means that BMCS is good at finding true pixels adequately, while eliminating false pixels.

Table 4.1 Performance metrics of samples given in Figure 4.3.

		Accuracy	Balanced Accuracy	E_y	D_y	FPR	TPR	H	(UDI,ODI)
BMCS	Sample 1	95,09%	94,98%	4,91%	0,58	6,34%	96,30%	55	(0.0, 0.0)
	Sample 2	85,93%	89,23%	14,07%	0,00	0,00%	78,46%	11.18	(0.0, 0.0)
	Sample 3	91,11%	91,80%	8,89%	0,19	3,07%	86,66%	154.65	(0.0, 0.0)
Levelset Segmentation	Sample 1	89,96%	90,49%	10,04%	0,70	2,91%	83,90%	124.17	(0.0, 0.21)
	Sample 2	74,16%	80,22%	25,84%	0,00	0,00%	60,44%	20.13	(0.0, 0.0)
	Sample 3	82,39%	81,48%	17,61%	6,63	25,38%	88,34%	317.29	(0.0, 0.10)
Wei's Method	Sample 1	86,95%	87,91%	13,05%	0,02	0,18%	76,00%	114.24	(0.0, 0.0)
	Sample 2	65,43%	73,54%	34,57%	0,00	0,00%	47,07%	27.30	(0.0, 0.0)
	Sample 3	90,56%	90,96%	9,44%	1,00	5,99%	87,92%	142.39	(0.0, 0.0)
Watershed Segmentation	Sample 1	70,11%	72,35%	29,89%	0,03	0,17%	44,86%	154.56	(0.18, 0.0)
	Sample 2	72,26%	78,76%	27,74%	0,00	0,00%	57,53%	36.80	(0.22, 0.0)
	Sample 3	52,00%	57,65%	48,00%	0,00	0,00%	15,30%	633.54	(0.70, 0.0)

Overall performance of all methods is given in Table 4.2. Watershed segmentation method has the lowest false positive rate and D_y values. So we can infer that watershed method produces very accurate and consistent with manual selection. But watershed segmentation method has the lowest true positive rate and accuracy values, so watershed segmentation could not cover mass area as accurate as other methods. Additionally difference between BMCS and Watershed in false positive rate is statistically insignificant ($p=0.83$). Watershed segmentation approach obtains the highest Hausdorff distance value. Moreover, difference between BMCS and watershed segmentation in accuracy and D_y is statistically significant ($p=0$ for both cases). This means that BMCS and Watershed segmentation algorithms produces same amount of False Positive pixels but results obtained from Watershed

segmentation method fits better to reference data. On the contrary, BMCS produces more accurate results than Watershed segmentation algorithm. These results indicate that BMCS segmentation approach reduces false positive rate while producing high accuracy values comparing to Watershed segmentation approach.

Table 4.2 Performance results of different segmentation methods over DEMS dataset.

Method	Accuracy	Balanced Accuracy	Ey	Dy	FPR	TPR	H	(UDI,ODI)
BMCS	80,04%	83,15%	19,96%	1,73	10.86%	77,16%	95.99	(0.0051, 0.0019)
Chanvese Levelset	81,15%	82,47%	18,85%	2,27	14.45%	79,38%	100.07	(0.0076, 0.0045)
Wei's Method	78,22%	78,51%	21,78%	2,60	20.97%	78,00%	118.76	(0.0363, 0.0023)
WaterShed Segmentation	59,15%	67,30%	40,85%	1,01	10.51%	45,10%	188.48	(0.1418, 0.0020)

Level-set segmentation approach has the highest accuracy and true positive rates. But its false positive rate, D_y , H and (UDI, ODI) metrics are higher than BMCS. Statistically; difference between BMCS and Level-set segmentation is significant in all metrics for 10% confidence interval. Hence, we can say that BMCS produces less false positive pixels with more suitable regions to reference data than Level-set method, and produces more optimum segmentation result comparing with Level-set segmentation method.

Finally, BMCS outperforms other methods in terms of balanced accuracy, Hausdorff distance and Scalable discrepancy metrics. So, we can infer that BMCS segmentation approach produces acceptable mass segments with less possibility of producing *over-* and *under-segmented* masses.

CHAPTER FIVE

DIGITAL MAMMOGRAPHY DATASETS

5.1 Overview

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. In this thesis, DDSM and DEMS datasets are used in evaluation. Before introducing DEMS dataset, let us summarize other available mammogram datasets, shortly.

5.2 Available Mammography Datasets

5.2.1 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 (“USF Digital Mammography: ‘Other Resources’ Page,” n.d.).

5.2.2 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, $100 \mu\text{m}$ 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, 1998).

5.2.3 OWH (Office of Women's Health) Dataset

According to the Nishikawa's article (Nishikawa, 1998), this dataset developed by Office of Women Health under U.S. Ministry of Health and it is not freely available to everyone. 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 includes CC and MLO view of both right and left breast acquired using Lumiscan 85 film scanner at $50 \mu\text{m}$ 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.

5.2.4 (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), and 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 $50 \mu\text{m}$ pixel resolution with 8-bit color depth, but this set of data is not available now (Nishikawa, 1998). Moreover, a new dataset named mini-MIAS containing cropped versions of original images at 1024×1024 image size and $200 \mu\text{m}$ pixel resolutions were created according to intensive demand.

5.2.5 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, 1998).

5.2.6 GPCALMA (Grid Platform for a Computer-Aided Library in Mammography) Dataset

This dataset 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 2067×2657 size with 85 μm effective resolution and 12 bit intensity depth and stored using CALMA format (Lauria et al., 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, and lesion location and type. Moreover, dataset includes some demographical information and follow-up studies.

5.2.7 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).

5.2.8 Digital Database for Screening Mammography (DDSM)

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, Bowyer, Kopans, Moore, & Kegelmeyer, 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, breast density assessment made by a radiology expert, with the degree of difficulty. This dataset is the de-facto standard for mammographic researches. In this thesis we used the whole DDSM dataset available in our experimental setup.

Statistics of masses in DDSM dataset is shown in Table 5.1. Since ACR put BI-RADS standard after this dataset being prepared, annotations of cases in this dataset misses some properties defined in BI-RADS standard. For instance, masses in DDSM have no density property, which is defined in BI-RADS standard. Moreover, some properties have no values assigned. Such values are assigned to a special value names Miscellaneous.

Table 5.1 Statistics of DDSM dataset.

Mass Property	Value	Lesion Count
Shape	Misc	5
	Round	228
	Lobular	587
	Irregular	1308
	Oval	613
BI-RADS Score	0	185
	1	3
	2	131
	3	537
	4	1179
	5	706

Table 5.1 Statistics of DDSM dataset. (Cont'd.).

Mass Property	Value	Lesion Count
Margin	Misc	105
	Circumscribed	618
	Microlobulated	170
	Obscured	408
	Indistinct /ill-defined	753
	Spiculated	687
Total		2741

5.2.9 Dokuz Eylul University Mammogram DataSet (DEMS)

This thesis study also includes development of a mammography dataset named DEMS, which is also for low-level feature evaluation. Firstly, DEMS mammogram dataset is a fully annotated digital mammogram dataset for computer-aided diagnosis (CAD) studies, and it is compliant with the state of the art semantic web knowledge representation technologies. Case selection performed in two stages. The first, candidate cases are 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. All of the patients and physicians data is removed and the whole dataset were anonymized. More than 50K radiology reports of the cases are obtained from Hospital's reporting server in text format. To select initial candidate cases, a textual Boolean information retrieval system is developed to speed up selection process. Additionally, developed system helps us to select candidate cases for each concept in the ontology by assigning initial annotations. Finally, DEMS contains 485 mammographic cases where 255 of them contain one or more lesion, where expert radiologists annotated each case in three phases using the MAT.

Screening mammography generally involves two views of each breast: one from above (Cranial-Caudal view, CC) and the other from oblique or angled views (Mediolateral-Oblique, MLO). Therefore, a typical screening mammography case contains four mammograms, MLO and CC views for two breasts. In this respect,

each mammography case in DEMS contains four images in lossless PNG format and one XML file called as “DEMS Annotation XML”. Each image converted from DICOM file by using a third party tool. The name of the 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, 2560x3328 or 3328x4096 size. Figure 5.1 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 (Green contour) and skin thickening (Blue contour) 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.

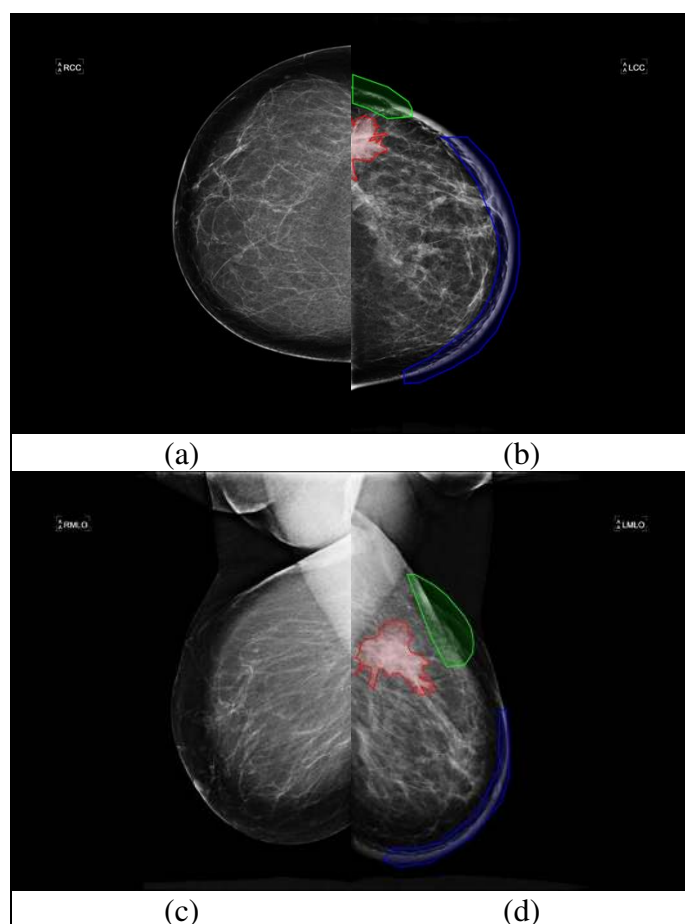


Figure 5.1 Sample mammography case with its ROI's in DEMS, (a) RCC view, (b) LCC view, (c) RMLO view (d) LMLO view

5.2.9.1 Mammography Annotation Tool (MAT)

Mammography annotation tool is developed to annotate mammograms in DEMS. MAT allows radiologists to examine four images in total, CC and MLO projection of the right and left breasts, for a typical mammography case. During this examination, the radiologists mark and annotate the abnormalities on images by using a variety of tools, which are easily done by clicking on it, and can also add the breast type. MAT stores all annotations in XML format, which is then easily converted into OWL format. The annotation file can also produce classical radiology reports in natural language, or to any other desired format. MAT is developed MAT using C++ programming language with QT framework with a cross-platform support. Figure 5.2 depicts a sample screen shot of the MAT.

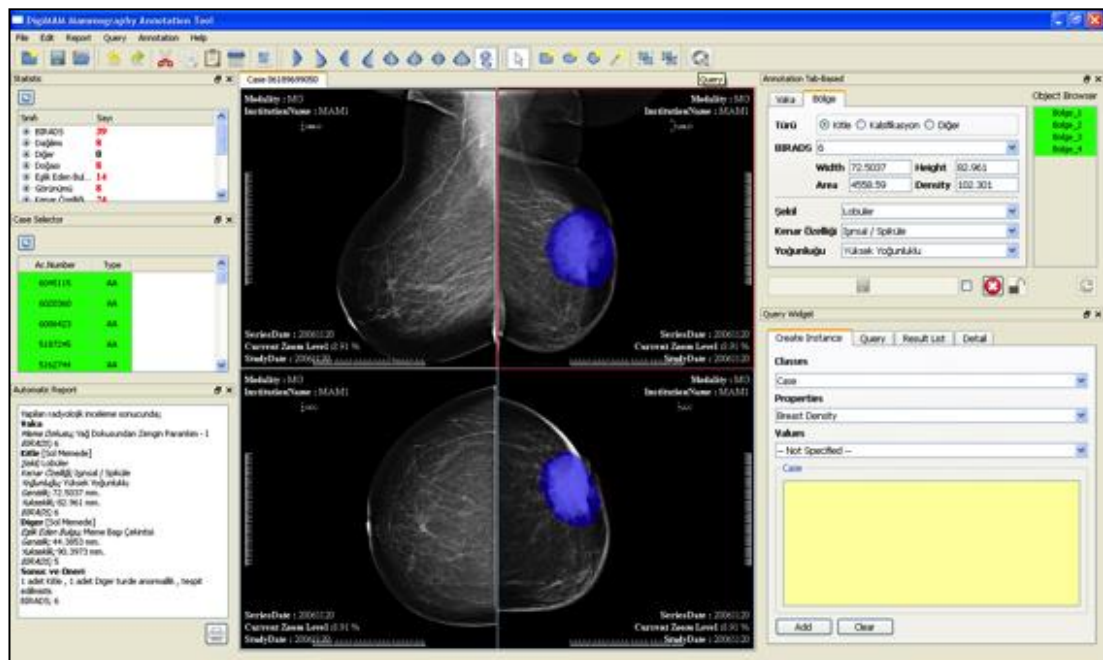


Figure 5.2 Screen capture of mammography annotation tool.

5.2.9.2 Mammography Annotation Ontology (MAO)

DEMS has ontology-based annotations. Therefore, an ontology called Mammography Annotation Ontology is developed (Bulu, Alpkocak, & Balci, 2012). MAO is built on the 3rd edition of BI-RADS Mammography Atlas (Liberian & Menell, 2002), and is used to annotate any abnormality observed in mammograms. Principally, the MAO provides a shared vocabulary and knowledge that makes

annotations understandable and computable by computers. Prominently it makes reasoning of any other information possible and forms a knowledge base. MAO is developed considering the domain covered with intended use of the ontology and using middle-out strategy as ontology development methodology (Fernández López, 1999). Figure 5.3 shows major concepts of the MAO with their relationships.

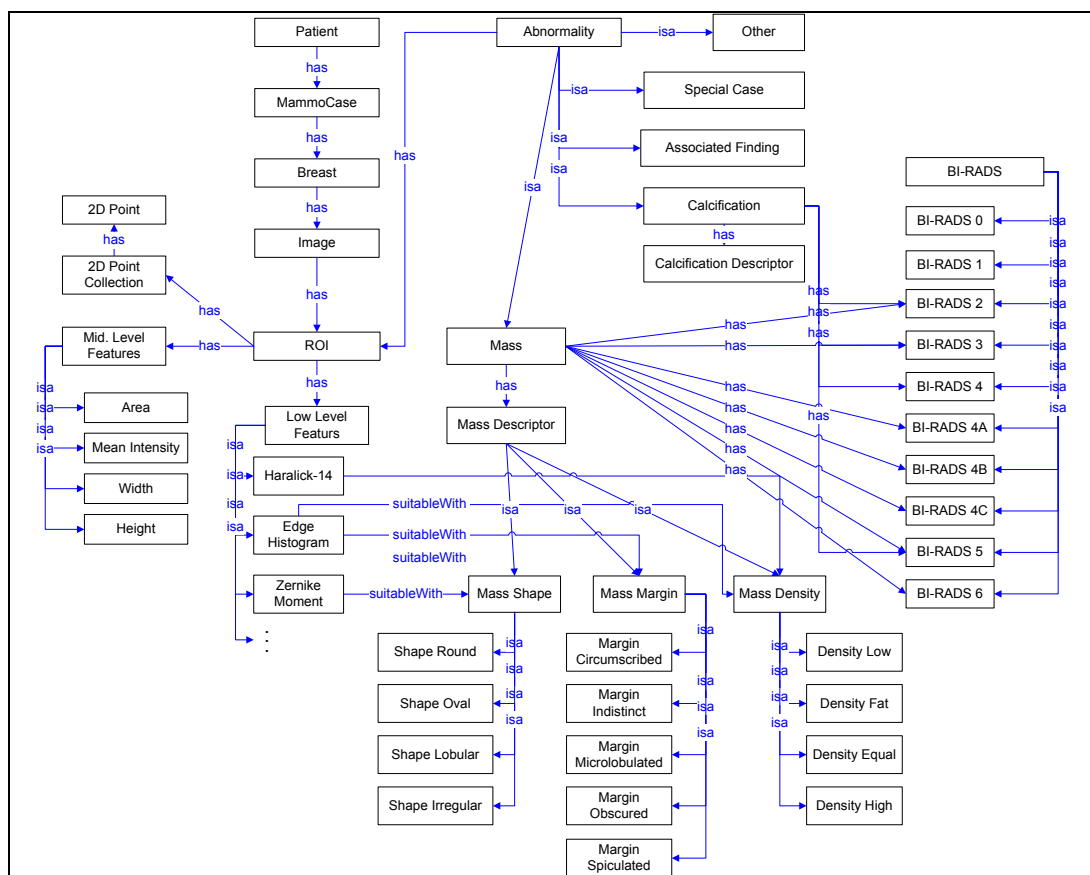


Figure 5.3 Major classes and their relationships in Mammography Annotation Ontology (MAO).

DEMS Annotation XML file contains Patient and Case tags. For privacy reasons only birth date of patient is kept. On the other hand Case tag includes all image and annotation data with date of study, which is important to calculate age of the patient at the examination date. Images are described by Image tag, which contains important DICOM headers and lesion annotation denoted by GraphicItem tag. A sample GraphicItem tag is shown in Figure 5.4.

```

<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 5.4 Sample GraphicItem tag in DEMS Annotation XML

5.2.9.3 DEMS Statistics

5.2.9.3.1 Abnormality Distribution. There are four types of lesions in DEMS, which are mass, calcification, special case and associated finding. Additionally, some breast may contain metallic clip(s) after surgical operations and these are seen in the mammograms clearly. To be able to distinguish them from any other lesions, another lesion type named other is created and they are annotated in these group. Table 5.2 contains abnormality distribution of DEMS.

Table 5.2 Abnormality distribution of DEMS.

Abnormality Type	Case Count	Lesion Count	Total Number of Annotations	
			Count	(%)
Mass	116	136	260	33
Calcification	119	144	274	35
Special Cases	19	20	32	5
Associated Findings	65	97	146	24
Other	10	13	23	3

where, Case Count column shows the number of unique mammographic case contains related lesion type/property/value, Lesion Count column shows the number of unique lesion marked with related lesion/property/value and Total Annotation column shows the number of annotation instances totally created for related lesion/property/value with their percentages.

5.2.9.3.2 Breast Density Distribution. The BI-RADS mammography atlas

identifies four types of breast densities, namely, Almost Entirely Fat, Scattered Fibroglandular Tissue, Heterogeneously Dense and Extremely Dense. During the mammographic examinations expert should add the breast density of the patient to the annotation. In DEMS all cases has their own breast density annotation. Table 5.3 shows count of the breast densities in DEMS, i.e. Extremely Dense breast type has the lowest percentage.

Table 5.3 Distribution of breast densities in DEMS.

Breast Density	Case	
	Count	(%)
Almost Entirely Fat (BI-RADS Density I)	182	38
Scattered Fibroglandular Tissue (BI-RADS Density II)	160	33
Heterogeneously Dense (BI-RADS Density III)	104	21
Extremely Dense (BI-RADS Density IV)	39	8

5.2.9.3.3 Mass Distribution. The one of the major lesion type of the DEMS is mass. According to BI-RADS mammography atlas each mass has three attributes; shape, margin and density. Furthermore, each attribute has its set of allowed values. For example, mass shape can be round, lobular, oval or irregular. Figure 5.5 shows one of the masses in DEMS with its boundary, where the mass has irregular shape, spiculated margin and equal density. Table 5.4 shows the distribution of masses in DEMS according to their features in detail.

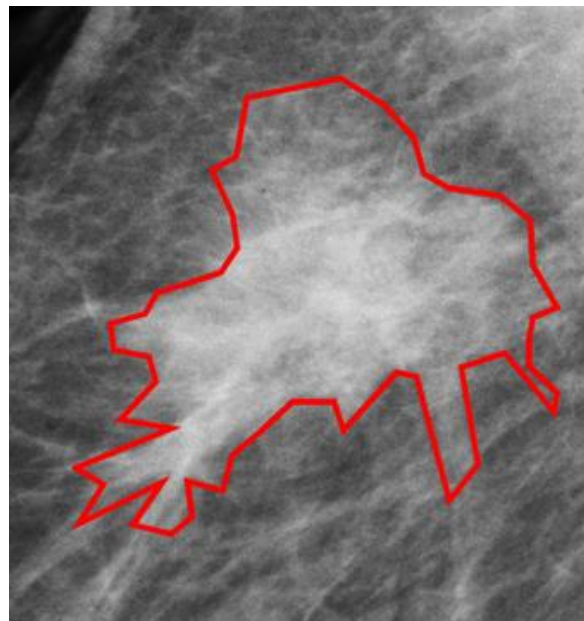


Figure 5.5 Sample mass in DEMS

In DEMS, all masses annotated as BI-RADS category 6 are pathologically proven malignant lesion. On the other hand, BI-RADS category 2 indicates benign lesions. Total percentage of BI-RADS 2 and 6 masses is 32. For all other masses pathologic examination is required to determine if they are benign or malign.

Table 5.4 Mass distributions with respect to properties in DEMS.

Property	Value	Case Count	Lesion Count	Annotation	
				Count	(%)
BI-RADS					
	2	23	27	50	20
	3	26	29	55	21
	4A	9	9	18	7
	4B	6	6	12	4
	4C	10	10	19	7
	5	37	39	75	29
	6	14	16	31	12
Shape					
	Round	21	27	50	19
	Lobular	28	29	57	22
	Irregular	56	59	113	44
	Oval	21	21	40	15
Margin					
	Circumscribed	44	52	98	38
	Microlobular	5	5	9	3
	Obscured	16	16	31	12
	Illdefined / Illdefined	22	26	49	19
	Spiculated	37	37	73	28
Density					
	High	51	62	119	46
	Equal / Isodence	55	59	111	43
	Low / Not Fat Containing	3	3	6	2
	Fat Containing Radiolucent	11	12	24	9

5.2.9.3.4 Calcification Distribution. The second major abnormality type in DEMS is calcification. Like masses, according to BI-RADS mammography atlas annotation of calcifications is determined. So, each calcification has its category, type and distribution attribute with their allowed values. Table 5.5 shows the distribution of calcifications in DEMS with respect to their features in detail.

Table 5.5 Calcification distributions with respect to calcification properties in DEMS.

Property	Value	Case Count	Lesion Count	Annotation	
				Count	(%)
BI-RADS					
2		77	95	177	66
3		9	9	18	6
4		24	24	47	17
5		16	16	32	11

Table 5.5 Calcification distributions with respect to calcification properties in DEMS (Cont'd.).

Property	Value	Case Count	Lesion Count	Annotation	
				Count	(%)
Category					
	Typically Benign	79	97	181	67
	Intermediate Concern	28	28	55	19
	Higher Probability of Malignancy	21	21	42	14
Type					
	Skin	5	6	10	4
	Vascular	12	12	24	8
	Coarse or Popcorn Like	16	17	32	12
	Large Rod Like	6	6	11	4
	Round	17	18	34	12
	Lucent Centered	10	10	20	7
	Eggshell or Rim	7	8	14	6
	Milk of Calcium	3	3	6	2
	Suture	1	4	4	3
	Dystrophic	9	9	18	6
	Punctuate	12	12	24	8
	Amorphous or Indistinct	9	9	18	6
	Coarse Heterogeneous	7	7	13	5
	Fine Pleomorphic	22	22	44	15
	Fine Linear of Fine Linear Branching	3	3	6	2
Distribution					
	Grouped or Clustered	35	38	70	26
	Segmental	7	7	14	5
	Regional	12	12	24	8
	Diffuse Scattered	48	52	97	36
	Single None	33	33	65	22
	Linear	4	4	8	3

5.2.9.3.5 Special Case Distribution. The third abnormality type in DEMS is special cases. There are six types of special cases in mammography. Types and BI-RADS scores of the special cases are given in Table 5.6 with their counts. DEMS has no special case whose BI RADS score is higher than 3.

Table 5.6 Special Case distribution with respect to properties in DEMS.

Property		Value	Case Count	Lesion Count	Annotation	
					Count	(%)
BIRADS						
	1		4	4	6	20
	2		7	8	12	40
	3		8	8	14	40
Type						
	Asymmetric Tubular Structure / Solitary Dilated Duct		1	1	2	5
	Intramammary Lymph Node		7	7	13	35
	Global Asymmetry		5	5	9	25
	Focal Asymmetry		3	3	4	15
	Bilateral Accessories Breast Tissue		1	2	2	10
	Unilateral Accessories Breast Tissue		2	2	2	10

5.2.9.3.6 Associated Finding Distribution. The last abnormality type in DEMS is associated findings. In mammography seven types of associated findings are exist. Different from the other abnormalities in DEMS, some associated findings may not have BI-RADS scores. For these types of lesions, an N/A value added BI-RADS score. Types and BI-RADS scores of all the special cases are given in Table 5.7 with their counts.

Table 5.7 Associated Finding distribution with respect to properties in DEMS.

Property	Value	Case Count	Lesion Count	Annotation	
				Count	(%)
BIRADS					
	N/A	15	21	29	22
	1	5	5	10	5
	2	20	30	41	31
	3	1	1	1	1
	4	5	6	12	6
	5	23	32	49	33
	6	1	2	4	2
Type					
	Skin Retraction	12	13	16	14
	Nipple Retraction	7	7	14	7
	Skin Thickening	18	18	35	19
	Trebecular Thickening	11	11	22	11
	Skin Lesion	5	5	10	5
	Axillary Adenopathy	25	37	37	38
	Architectural Distortion	6	6	12	6

CHAPTER SIX

FEATURE SELECTION FOR CONTENT DESCRIPTION

6.1 Overview

Selecting proper low-level features for breast mass property classification task needs a set of experimentation to be conducted. Moreover, it is also important to measure effects of other parameters; e.g. classifier performances, dataset differences etc. Therefore, we conducted a series of experimentations on two datasets which are DDSM (Heath et al., 1998, 2001) and DEMS (Akçay, Alpkoçak, Balci, & Dicle, 2009). Details of both datasets are given in chapter 5. Unlike similar works in this topic, we include all cases of DDSM dataset in our experiments setup instead of selecting a subset of it. Finally, we made publicly available of classification results and low-level features on both datasets, so that any researcher in this area can repeat our findings or compare with any other method or features easily.

In this thesis, we measure performance of four parameters, which are low-level feature, dataset, mass selection type and classifier on mammography mass property classification task. Total number of experiments we conducted is 11544; and only one parameter changed in each experiment.

This chapter organized as follows. Section 6.2 introduces classifiers used in experiments. Finally, Results of all experiments are presented in section 6.3.

6.2 Classifiers

In this thesis, we use several classifiers to measure performance of low-level features. Each classifier is tested using 10-fold cross validation to ensure consistency and estimation quality using data mining software called RapidMiner (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006). We used k -Nearest Neighbor (Bremner et al., 2005), Random Forests (Breiman, 2001), (Prinzie & Van den Poel, 2007), Naïve Bayes (Harry Zhang, 2004), Artificial Neural Networks (Bhadeshia, 1999),

Linear Discriminant Analysis (Mika, Ratsch, Weston, Scholkopf, & Mullers, 1999) and Support Vector Machine (Cortes & Vapnik, 1995) classifiers in this work.

6.2.1 *k*-Nearest Neighbor (*k*-NN)

k-Nearest Neighbor (*k*-NN) classifier is a simple voting classifier which assigns object to a class if most of the *k*-neighbors of the object belongs to that class. Here, neighbors of an object are determined by selecting objects in dataset having minimum distance to the object. In this thesis, we used 5-NN classifier only.

6.2.2 *Random Forests (RAF)*

Decision Trees are usually highly efficient classifiers. Unfortunately, it is reported that decision trees has over-fitting problem (Breiman, 2001). To overcome this problem, a group of decision trees are grouped and used as a classifier instead of one decision tree. This approach is called Random Forests, and it has been reported by Ho (Ho, 2000) and Breiman (Breiman, 2001) that Random forest with adequate number of Decision Trees deals with over-fitting problem. In our experimentation setup, we prefer to use Random Forests with 10 trees where each tree is based on information gain criteria.

6.2.3 *Naïve Bayes Classifier (BAY)*

This classifier is a probabilistic classifier based on Bayes theorem. This classifier assumes that features are independent from each other; an each feature vector is associated to a class using conditional probability between feature vector and each class, as shown in the following formula.

$$P(y|\vec{x}) = \frac{P(y)P(\vec{x}|y)}{\sum_{y_i \in Y} P(y_i)P(\vec{x}|y)}$$

where, $P(y)$ is the prior probability of class y and $P(\vec{x}|y)$ is the probability of generating feature vector $\vec{x}$ from class y . In our case, we assume that features are independent from each other. As a result, formula could be rewritten as follows;

$$P(\vec{x}|y_i) = \prod_{j=1}^n P(\vec{x}[j]|y_i)$$

where, $\vec{x}[j]$ is the value of j th element of $\vec{x}$. Each value for this formula can be calculated using Maximum Likelihood Estimation technique.

6.2.4 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are mathematical models simulating biological neural networks. Algebraic ANNs, also called as feed-forward ANNs, consist of neurons, which are usually grouped into layers and usually perform a weighted sum followed by a nonlinear mapping. Such algebraic ANNs used as classifiers contain at least two layers named the input layer having the same number of neurons with input data vector size and the output layer, which determines result of the classification. If an ANN has more than two layers, then the rest of the layers are called hidden layers and there is no restriction of number of neurons in these layers. In this thesis, we construct the most widely used feed-forward ANN, i.e. Multi-Layer Perceptron (MLP) such that it has 3 layers and the number of the neurons in the hidden layer is chosen dynamically to adopt different input sizes.

6.2.5 Linear Discriminant Analysis (LDA)

This classifier aims to find a hyper-plane that splits feature space in a way that classes are separated in an optimal way. Formal definition of such a hyper-plane is given in following formula.

$$\forall_{y \in Y} \forall_{\langle \vec{x}_i, y_i \rangle \in O, y=y_i} H_{y_i} : \vec{w} \cdot \vec{x}_i + b = y_i$$

where O is the set of observation pairs, $y \in Y$ is the numeric label associated with the observation. This formula defines a set of hyper-planes for each class. The parameters for each hyper-plane are calculated by using least-squares minimization.

6.2.6 Support Vector Machines (SVM)

Support Vector Machines (SVMs) is a kind of Linear Discriminant Analyzers, but instead of finding hyper-planes for each class, SVM searches for linear boundaries that separate given classes from each other. Hence, SVMs are also referred as max-margin learning algorithms (Guo, Sun, Deng, & Li, 1994). In SVM, input vectors are mapped into a high (may be infinite) dimensional feature space by a nonlinear mapping implicitly defined by a kernel in order to transform an originally nonlinear separation problem into a linearly separable one. Besides, SVMs could operate also on linearly non-separable feature spaces as a consequence of the associated optimization formulations allowing for a (desirably) small number of non-separable vectors. SVMs usually possess high generalization abilities as provided by their high separation margins and by their sparse representations defined with a small set of so-called support vectors. In our experiments, SVM is used with radial basis kernel.

6.3 Experimentation

We conducted total number of 11544 experiments to measure performance of individual low-level feature performance on mass property classification task. In each experiment we change only one variable, which could be low-level feature, classifier and property class. Results of each experiment are measured using *accuracy*, *specificity*, *sensitivity* and *balanced accuracy* metrics. All of these metrics are based on confusion matrix including number of correctly and incorrectly classified samples. Table 6.1 shows a sample confusion matrix.

Table 6.1 Sample confusion matrix.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positives (TP) True samples that are correctly classified	False Negatives (FN) True samples that are incorrectly classified
	Negative	False Positives (FP) False samples that are incorrectly classified	True Negatives (TN) False samples that are correctly classified

Sensitivity metric aims to measure classification ability to identify positive results and defined as following formula.

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity metric is used to measure system performance of negative examples. In following formula calculation of *Specificity* metric is given.

$$Specificity = \frac{TN}{TN + FP}$$

Accuracy metric aims to measure system performance of correctly classified examples either they are positive or negative. *Accuracy* can be calculated by using following formula.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Since we used more than one metric, it is hard to define which experiment produces better results than other one. Although *accuracy* metric measures overall classification performance, it has very weak predictive power in some cases, e.g. classes with very little number of true samples. So, we use *balanced accuracy* metric, which is average of *sensitivity* and *specificity* metrics, to rank experiment results. Additionally, we present other three metrics to present performance of experiment on true and false samples. Only results of top 10 successful experiments

are presented to summarize. More detailed performance values could be found in appendices.

6.3.1 Results of Shape Property Experiments

Shape property has four classes defined in ACR BI-RADS atlas namely *round*, *oval*, *lobular* and *irregular*. Here, round masses has low malignancy rate while irregular masses have the highest malignancy rate. Unfortunately, some masses have no shape property, so those masses assigned to *N/A* class. Detailed performance results of shape property experimentations are discussed below.

6.3.1.1 Low-Level Feature Performance Comparison

N/A class contains masses with no shape information given. Moreover, masses of this class are originated from the DDSM dataset. Figure 6.1 shows performance graph of this class. Shape features obtain top three performance results. But, sensitivity values are too low for all low-level features. We can say that results of this class are successful for negative examples.

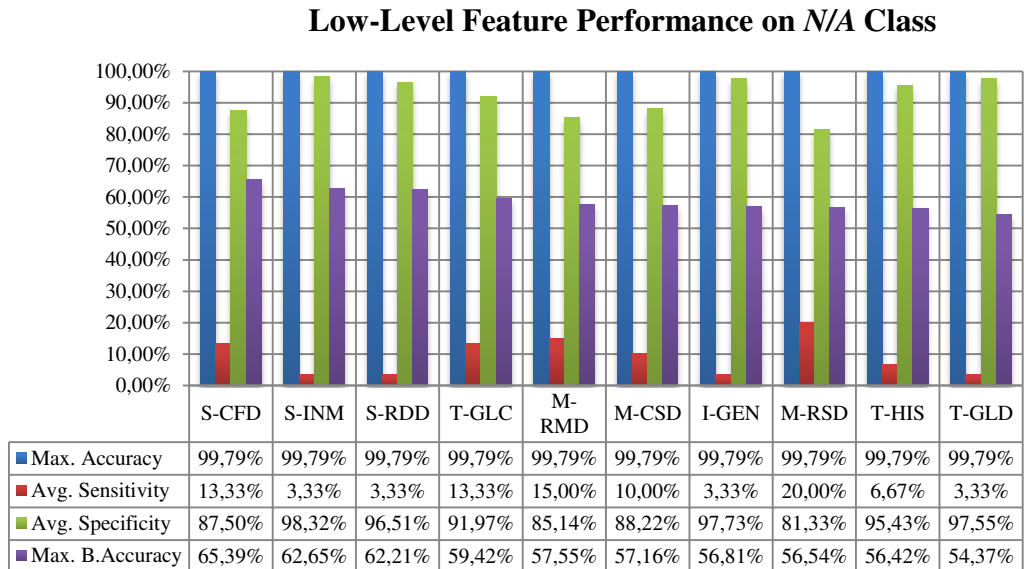


Figure 6.1 Low-Level feature performances of *N/A* class.

Low-level feature performances for round class are given in Figure 6.2. Round class is dominated by shape features, where 9 out of 10 top results belong to shape

feature type. Although, radial basis signal feature obtains the highest balanced accuracy score, Zernike moment feature is the most successful for identifying true samples. Additionally, only texture feature enters top 10 list is gray level difference feature.

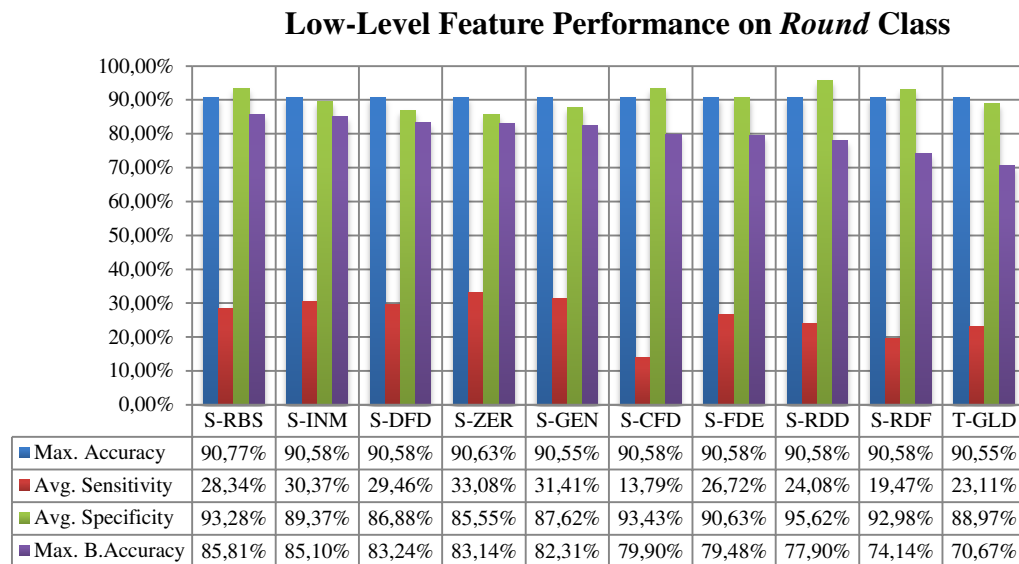


Figure 6.2 Low-Level feature performances of Round class.

Low-level feature performance of oval class is given in Figure 6.3. Similar to results of round class, shape features dominate classification results of oval class also. The most successful low-level feature for this task is Fourier Distance Descriptor of radial distance signal. According to the results, low-level features could only retrieve about 30% of oval classes at maximum.

Figure 6.4 shows performance results of low-level features on Lobular class. Shape features are the most successful ones for this class similar to performance of other classes. Moreover, texture and contour features produces good results. But, sensitivity of all results is too low for this class.

Low-Level Feature Performance on *Oval* Class

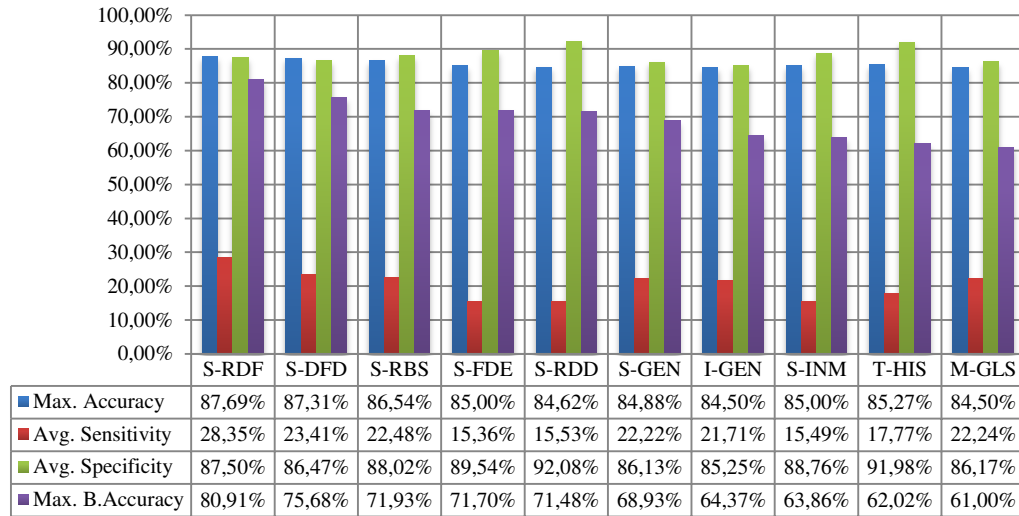


Figure 6.3 Low-Level feature performances of Oval class.

Low-Level Feature Performance on *Lobular* Class

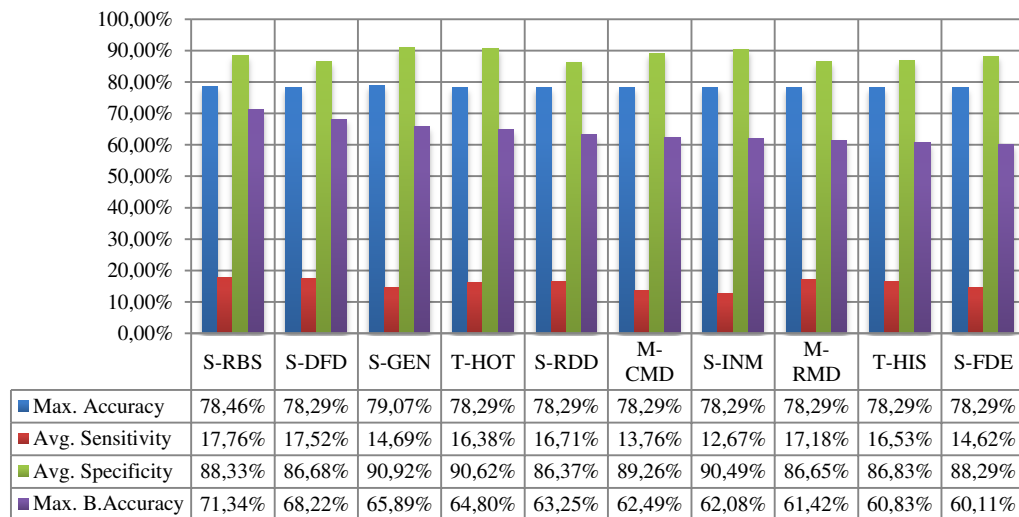


Figure 6.4 Low-Level feature performances of Lobular class.

Finally, Figure 6.5 depicts low-level feature performance results on Irregular class. According to the results, experiments on Irregular class show better sensitivity results than other classes and shape features dominate this class similar to previous results. Additionally, there are two texture features in top 10 which are edge histogram feature and homogeneous texture feature. Average sensitivity of almost all low-level features is above 70%. Although it is considered as texture feature, average sensitivity of edge histogram feature is about 75%. But, average specificity of only two low-level features is above 50%, which are general shape features and Fourier

coefficients of distance contour representation. Hence, we can say that Irregular class is the most successful one in shape classification task in terms of sensitivity.

Low-Level Feature Performance on *Irregular* Class

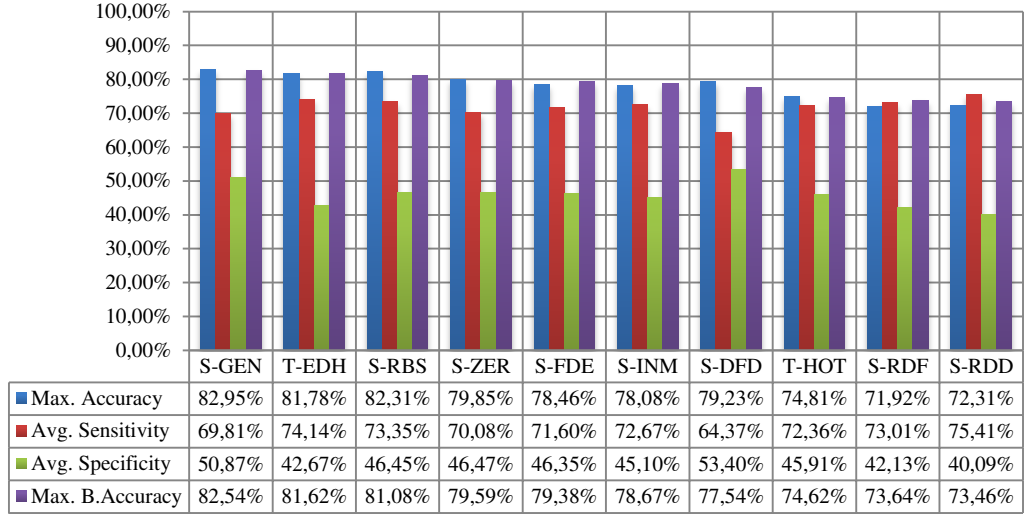


Figure 6.5 Low-Level feature performances of Irregular class.

In sum, Irregular class is the best among other class considering all metrics. Especially, 75% of all true samples are correctly classified using shape and texture features. Besides, accuracy of experiments achieved maximum accuracy rate between 71% and 82%.

6.3.1.2 Dataset Performance Comparison

Figure 6.6 shows the results of experiments on two dataset for mass shape classification task. According to the results, experiment on DEMS show better results than DDSM almost in all metrics. Sensitivity scores of some experiments on DDSM only shows great scores than ones conducted on DEMS. Accuracy and sensitivity values of DEMS outperform DDSM on the other hand.

Similar to low-level feature experimentation results, experiments on irregular class shows better performance results than other ones in terms of sensitivity scores. Moreover, Irregular class outperforms other classes in terms of accuracy and balanced accuracy metrics, if N/A class is discarded.

We can conclude that DEMS dataset has more accurate dataset to perform experimentations on mass shape classification task and Irregular class is the best class for mass classification task independent from dataset difference.

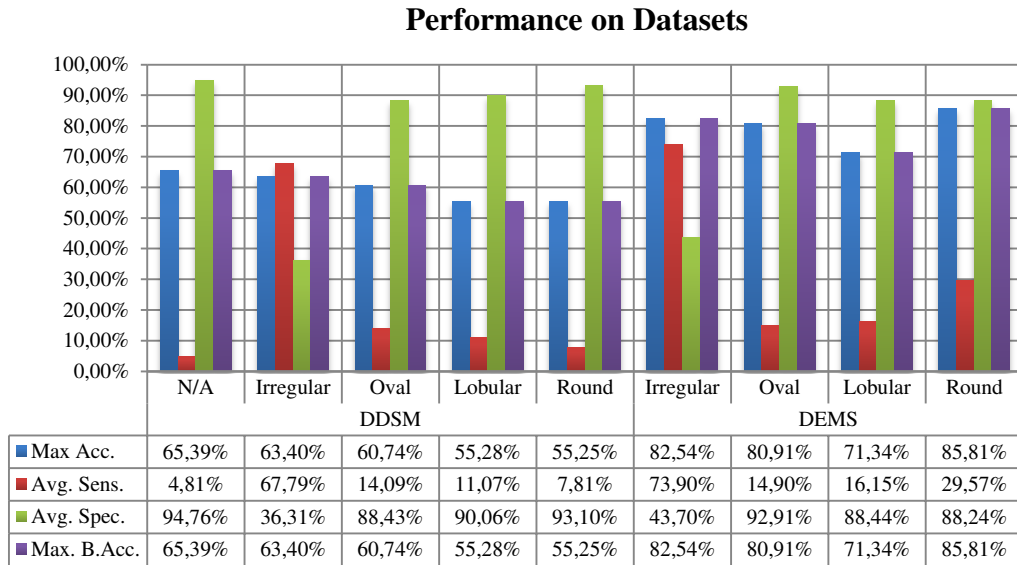


Figure 6.6 Dataset performance comparison.

6.3.1.3 Mass Selection Method Performance Comparison

Figure 6.7 shows effect of region selection method on classification performance for shape classification task. Here, ASR indicates automatically segmented regions using BMCS algorithm and MSR indicates reference regions. According to the results, automatically segmented regions shows almost same performances with manually segmented regions. Sensitivity and balanced accuracy scores of oval and lobular classes on MSR is greater than ASR. Additionally, experiments on irregular class obtain the highest performance scores on both ASR and MSR like previous experiments. Hence, it is clear that ASR and MSR shows almost same performance results on mass shape classification task.

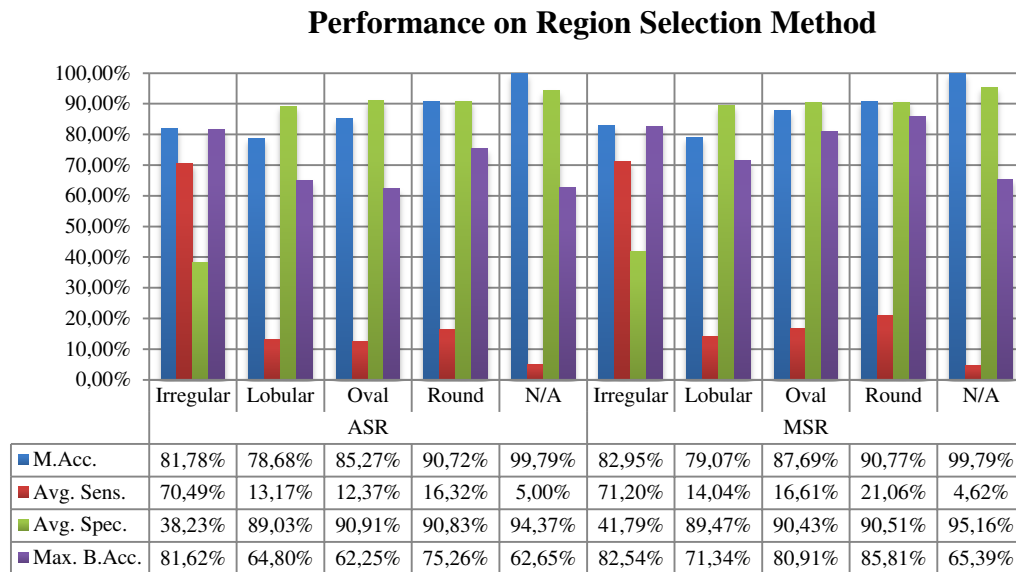


Figure 6.7 Region selection method performance comparison.

6.3.1.4 Classifier Performance Comparison

Figure 6.8 includes classifier performances of *N/A* shape class. According to the results, only Bayes and SVM classifiers could identify positive examples, since *N/A* class includes masses with no shape information.

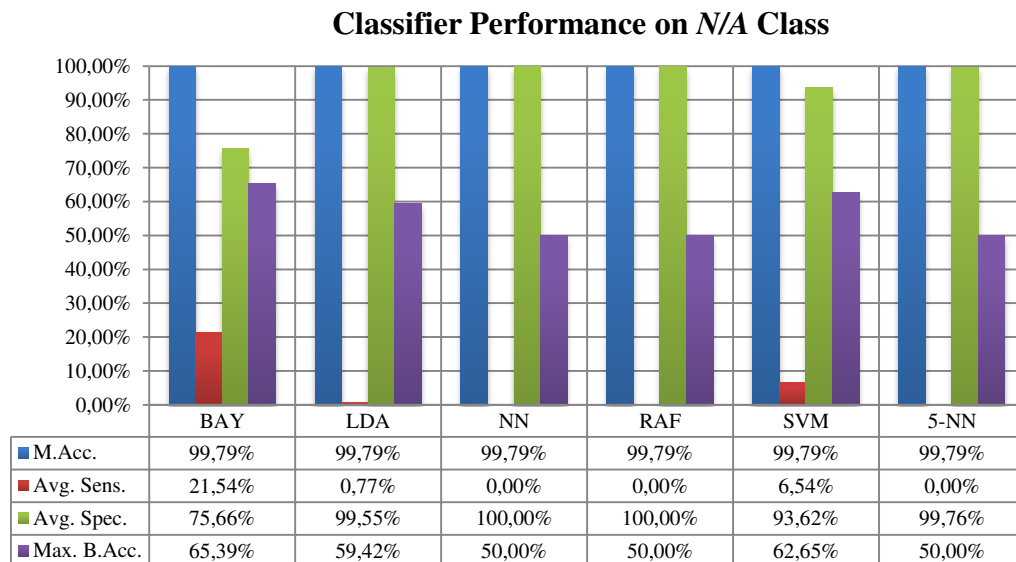


Figure 6.8 Classifier performance on *N/A* class.

Figure 6.9 shows classifier performances of experiments conducted on masses belonging to *Round* class. Although *Random Forest* classifier obtained the highest maximum balanced accuracy score, *Naïve Bayes* classifier is the best classifier

considering sensitivity value. Moreover balanced accuracy difference between *Random Forest* and *Naïve Bayes* classifiers is so small. It is also interesting that *k*-NN classifier obtains the second highest sensitivity score.

Classifier Performance on *Round* Class

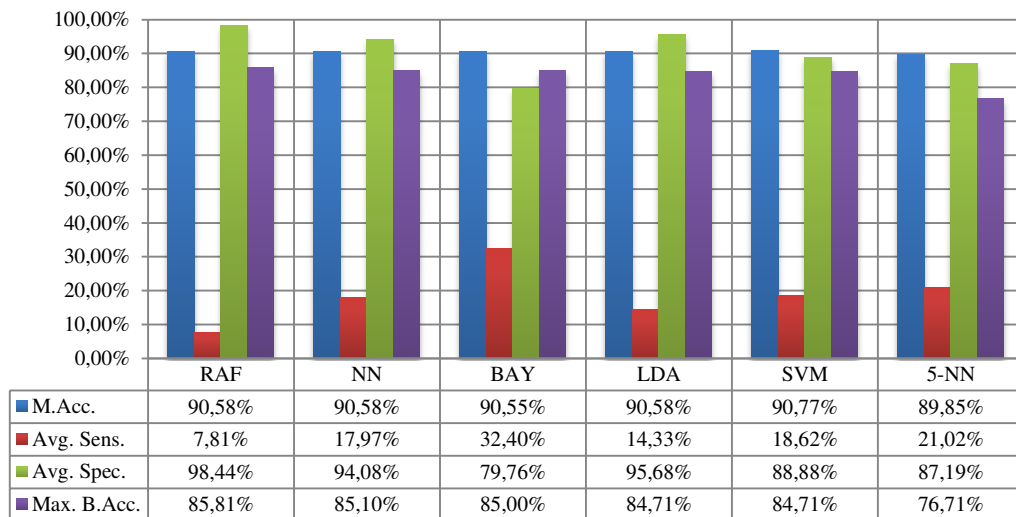


Figure 6.9 Classifier performance on *Round* class.

Classifier Performance on *Oval* Class

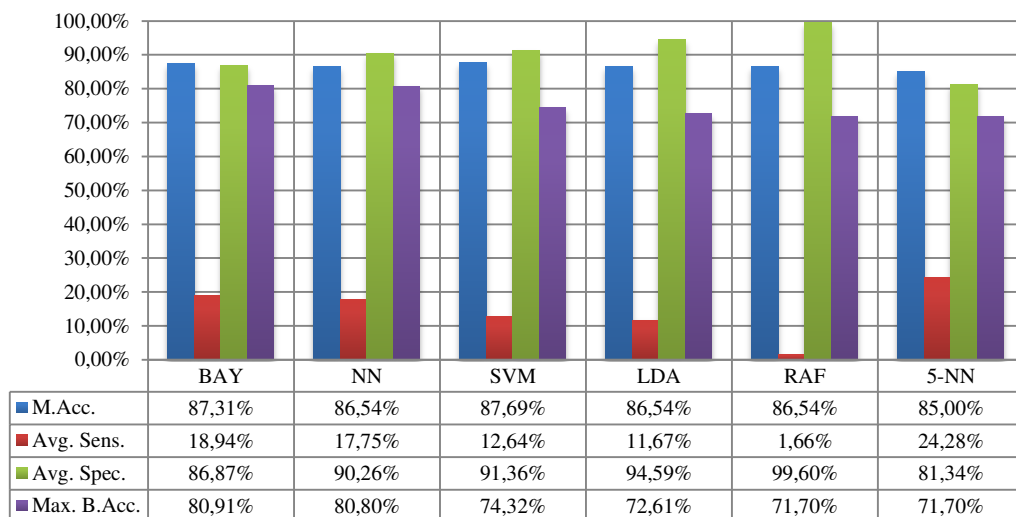


Figure 6.10 Classifier performance on *Oval* class.

Performance evaluation results of experiments conducted on *Oval* shaped masses are given in Figure 6.10. According to the results *Naïve Bayes* classifier obtains the highest balanced accuracy score. But *k*-NN classifier obtains the highest sensitivity score value, although it has the least specificity score.

Performance of classifiers on Lobular class is depicted on Figure 6.11. Similar to *Oval* class experiments, *Naïve Bayes* classifier obtains the highest *Balanced Accuracy* score although *k-NN* classifier has the highest *Sensitivity* scores.

Classifier Performance on Lobular Class

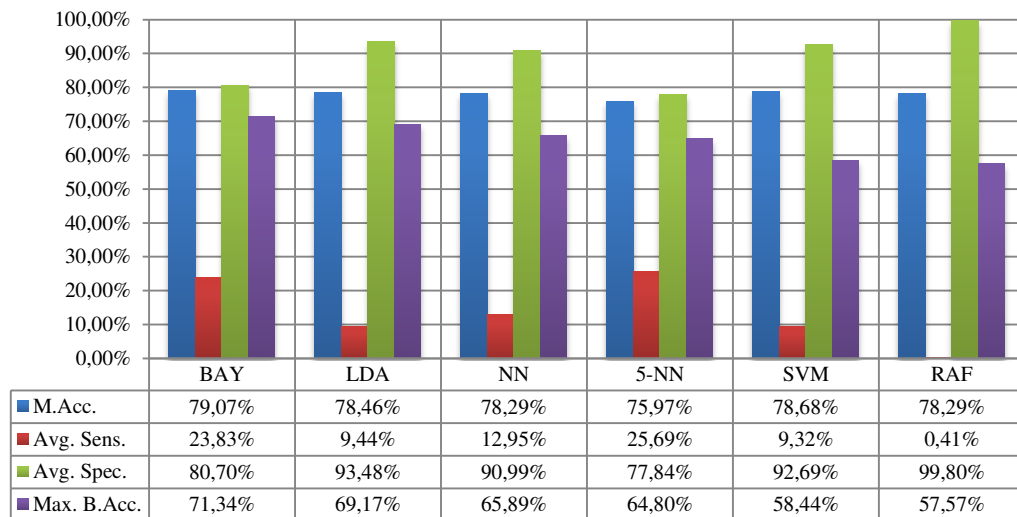


Figure 6.11 Classifier performance on *Lobular* class.

Classifier Performance on Irregular Class

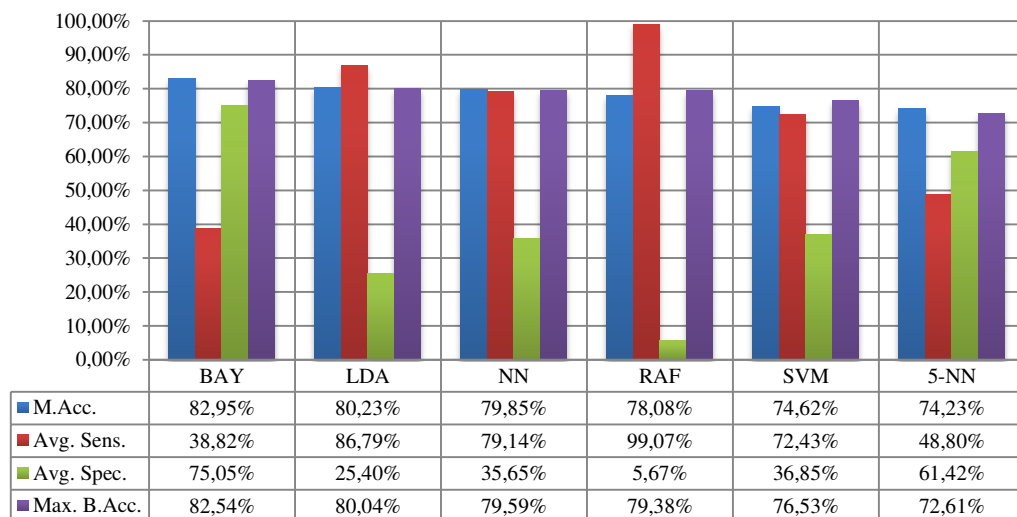


Figure 6.12 Classifier performance on *Irregular* class.

Finally, Figure 6.12 includes results of experiments conducted on *Irregular* class. Similar to previous experiments, *Naïve Bayes* classifier obtains the highest *Balanced Accuracy* result. Although *Random Forest* classifier obtains the highest *Sensitivity* score, it has the lowest *Specificity* score.

As a result, We can say that Naïve Bayes classifier performs well on almost all classes. Similar to previous experiment results, best performance scores obtained from experiments conducted on *Irregular* class.

6.3.2 Results of Margin Property Experiments

Margin property has five classes defined in ACR BI-RADS atlas namely *circumscribed*, *microlobular*, *obscured*, *spiculated* and *irregular*. Here, circumscribed masses has low malignancy rate while irregular masses have the highest malignancy rate. *N/A* class represents masses with no margin property. Detailed performance results of margin property experimentations are discussed below.

6.3.2.1 Low-Level Feature Performance Comparison

Figure 6.13 includes low-level feature performances of experiments conducted on *N/A* class. *Local Binary Pattern* feature obtains the highest balanced accuracy score and second sensitivity score. Although *gray level co-occurrence matrix* obtains very low *balanced accuracy* score, it has the highest *sensitivity* score.

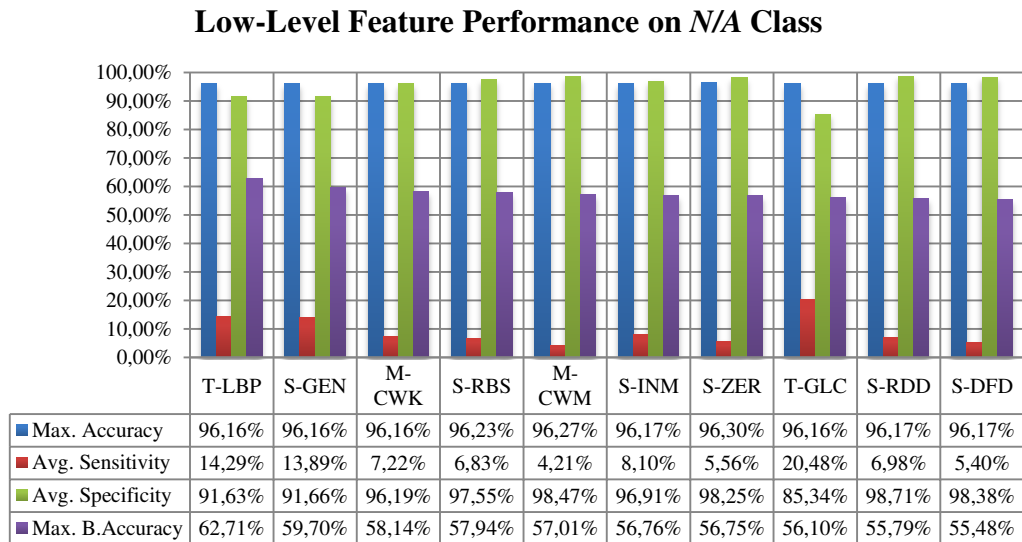


Figure 6.13 Low-Level feature performances of *N/A* class.

Figure 6.14 shows low-level feature performances of experiments conducted on *Circumscribed* class. Most of the successful features are shape features. Especially distance based shape representations (DFD and RBS) obtain the highest scores. Only edge histogram and histogram properties features enter top ten list.

Low-Level Feature Performance on *Circumscribed* Class

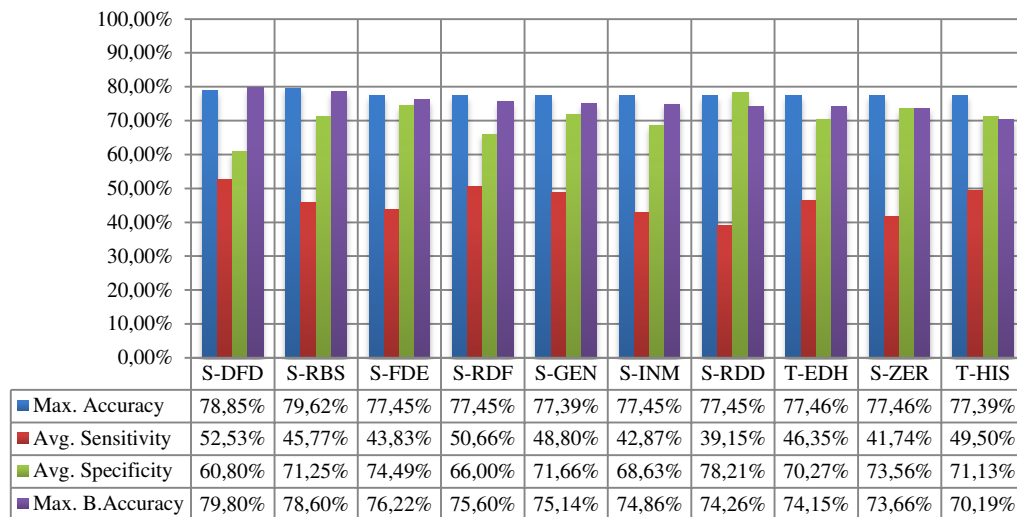


Figure 6.14 Low-Level feature performances of *Circumscribed* class.

Low-Level Feature Performance on *Microlobular* Class

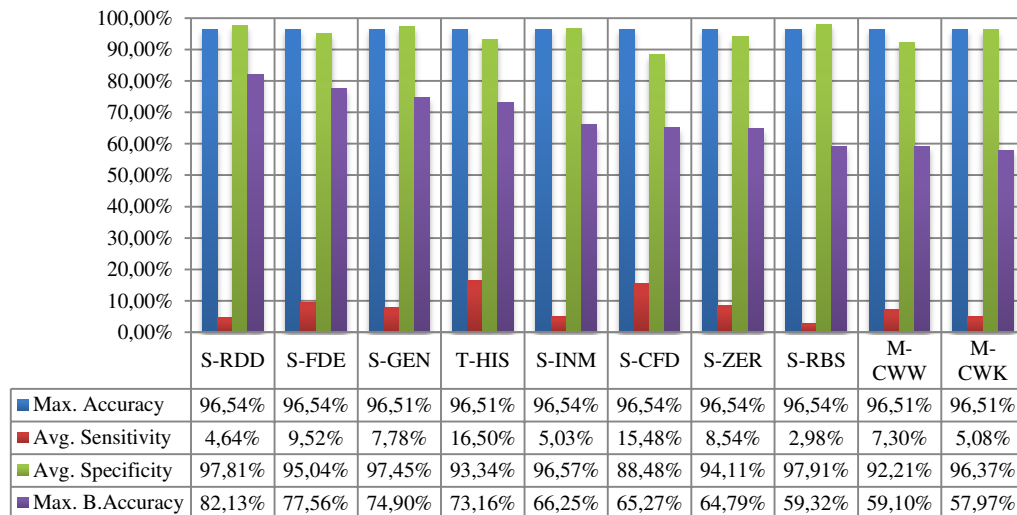


Figure 6.15 Low-Level feature performances of *Microlobular* class.

Figure 6.15 includes performance results of experiments conducted on *Microlobular* class. Although most of the successful features belong to shape group,

histogram properties texture feature and *fourier representation of curvature contour representation* shape feature obtain the highest two *sensitivity* score.

Figure 6.16 includes performance results of experiments conducted on *Obscured* class. According to the results, shape features obtain 5 out of 10 most successful balanced accuracy score, while texture features obtains 4 of them. Hence, we can infer that *Obscure* masses could be better identified by using both shape and texture features.

Low-Level Feature Performance on *Obscured* Class

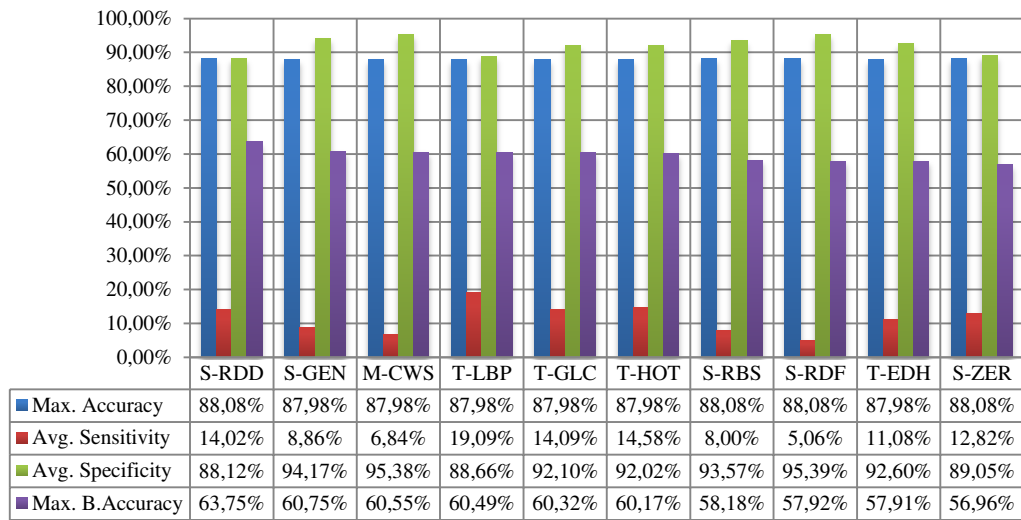


Figure 6.16 Low-Level feature performances of *Obscured* class.

Low-Level Feature Performance on *Spiculated* Class

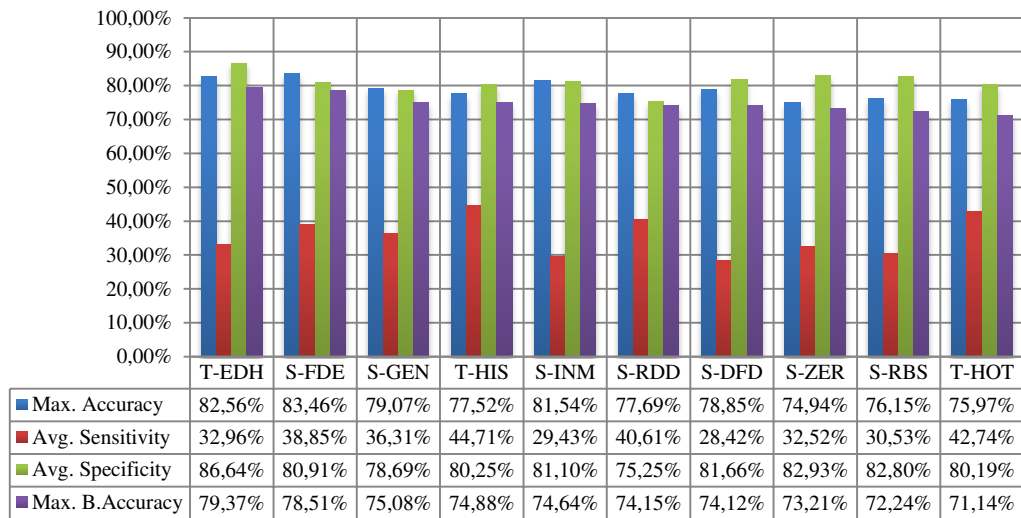


Figure 6.17 Low-Level feature performances of *Spiculated* class.

Figure 6.17 includes results of experiments conducted on *Spiculated* class. Although highest balanced accuracy score belongs to a texture feature, shape features obtain 7 of 10 highest score. Two of the successful features are MPEG-7 texture descriptors, which are edge histogram and homogeneous texture features.

Figure 6.18 includes performance results of experiments on *Irregular* class. Similar to previous experiments shape features shows better performance than other feature groups.

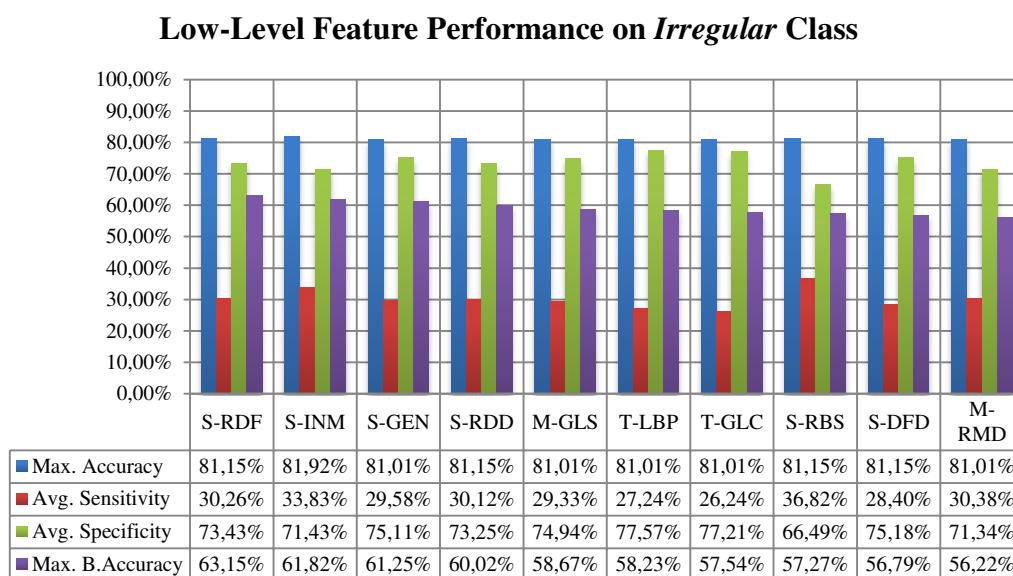


Figure 6.18 Low-Level feature performances of *Irregular* class.

In sum, we can say that shape features are more suitable for margin classification task than other feature groups. Although *Obscured* class is an exception for that situation, combination of shape features and texture features could produce better results for that class. It is also interesting that most successful classification performances are obtained from experiments that are conducted on *Circumscribed*, *Spiculated* and *Irregular* classes, since malignancy level of a mass increases from *Circumscribed* masses to *Irregular* masses. In other words, experiments show that benign and malign classes could be identified correctly than intermediate classes.

6.3.2.2 Dataset Performance Comparison

Figure 6.19 includes results of experiments aiming to test dataset performance on *Margin* property. Similar to shape experiments, balanced accuracy results of experiments conducted on DEMS dataset are higher than ones conducted on DDSM dataset. Results indicate that successful classes in each dataset are different. Performance results of *Irregular* class are the highest ones on DDSM dataset while Performance results of *Circumscribed* and *Spiculated* classes are highest one on DEMS dataset.

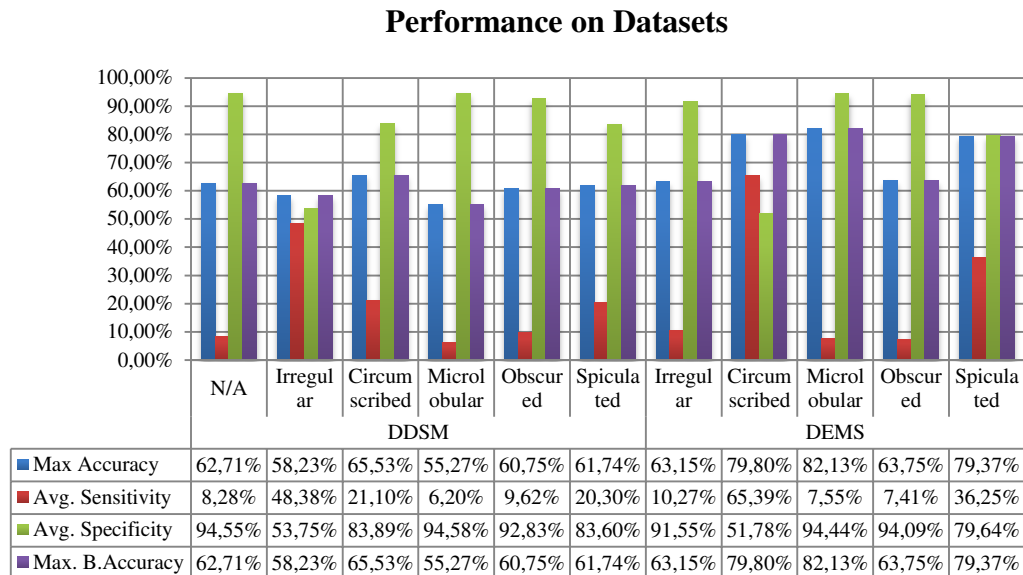


Figure 6.19 Dataset performance comparison.

6.3.2.3 Mass Selection Method Performance Comparison

Figure 6.20 includes region selection performance comparison of margin classification task. Unlike shape experimentation results, experiments show different performance characteristics. Only experiments on *Spiculated* class show no major differences on each region selection method. On the other hand, experiments on *Microlobular* class shows that performance of ASR method is better than MSR method. Performance of MSR method is better than ASR method in other classes. Hence, we could say that margin of a mass is more accurately identified by machine learning methods, when manually selected region boundaries is used.

Performance on Region Selection Method

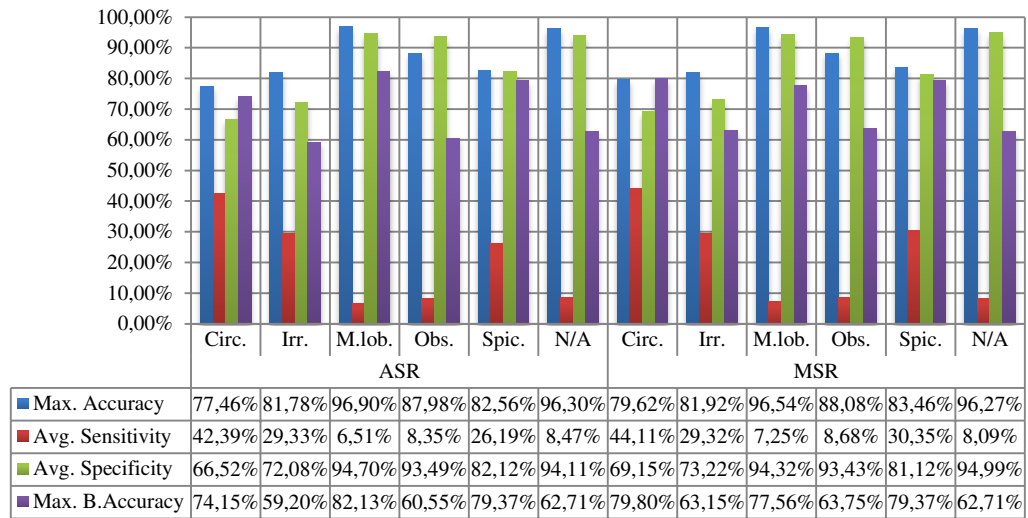


Figure 6.20 Region selection method performance comparison.

6.3.2.4 Classifier Performance Comparison

Figure 6.21 includes classifier performances of N/A class. According to the results, Naïve Bayes classifier outperforms other classifiers in terms of sensitivity and balanced accuracy metrics.

Classifier Performance on N/A Class

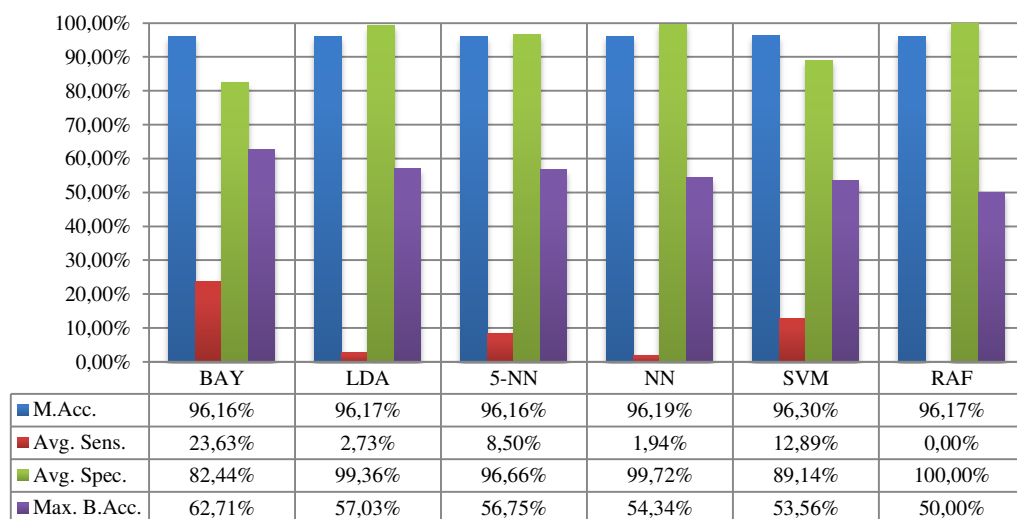


Figure 6.21 Classifier performance on N/A class.

Figure 6.22 includes classifier performances of experiments on *Circumscribed* class. Although *Naïve Bayes* classifier obtains the best balanced accuracy score,

Neural Network and *Support Vector Machine* classifiers obtain the highest sensitivity scores.

Classifier Performance on *Circumscribed* Class

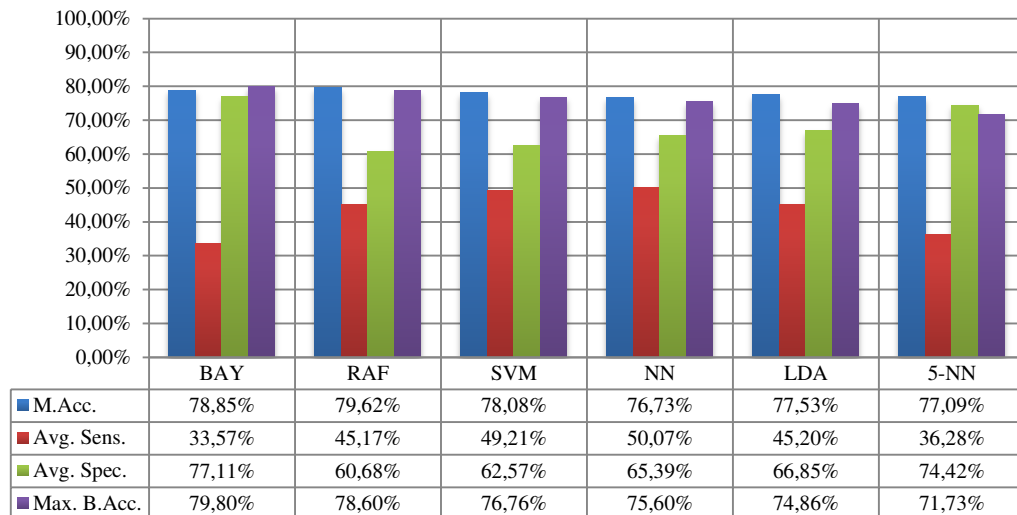


Figure 6.22 Classifier performance on *Circumscribed* class.

Figure 6.23 includes performance results of experiments on *Microlobular* class. It is interesting that *k-NN* classifier obtains the highest balanced accuracy score. Additionally *Naïve Bayes* classifier obtains the highest sensitivity score, which is nearly three times better than second one.

Classifier Performance on *Microlobular* Class

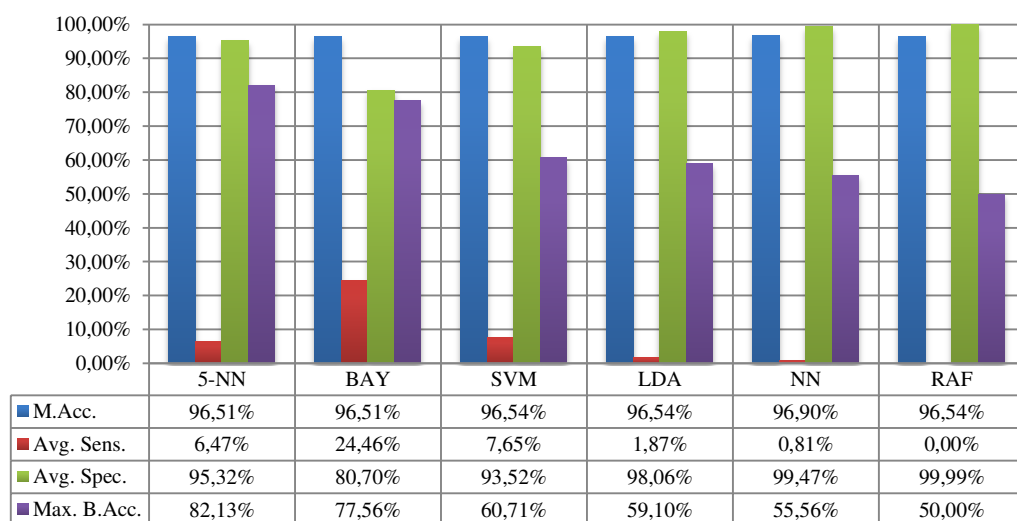


Figure 6.23 Classifier performance on *Microlobular* class.

Figure 6.24 includes performance results of experiments on *Obscured* class. According to the results, *Naïve Bayes* classifier obtains the highest balanced accuracy score. Besides, *k-NN* classifier obtains the second sensitivity score.

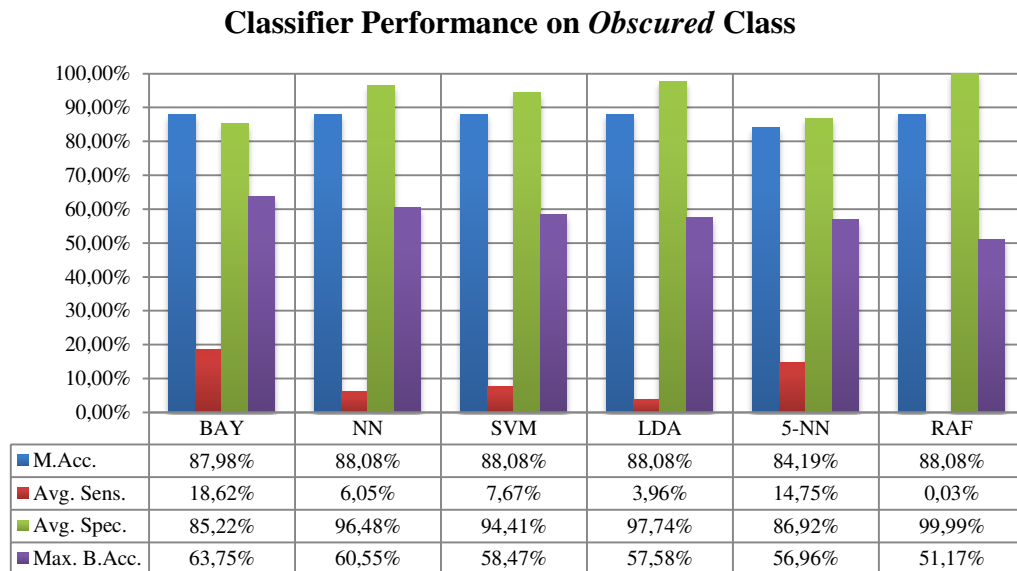


Figure 6.24 Classifier performance on *Obscured* class.

Figure 6.25 includes classifier performance of experiments on *Spiculated* class. *Naïve Bayes* classifier obtains the highest balanced accuracy score, while *neural network* classifier obtains the highest sensitivity score.

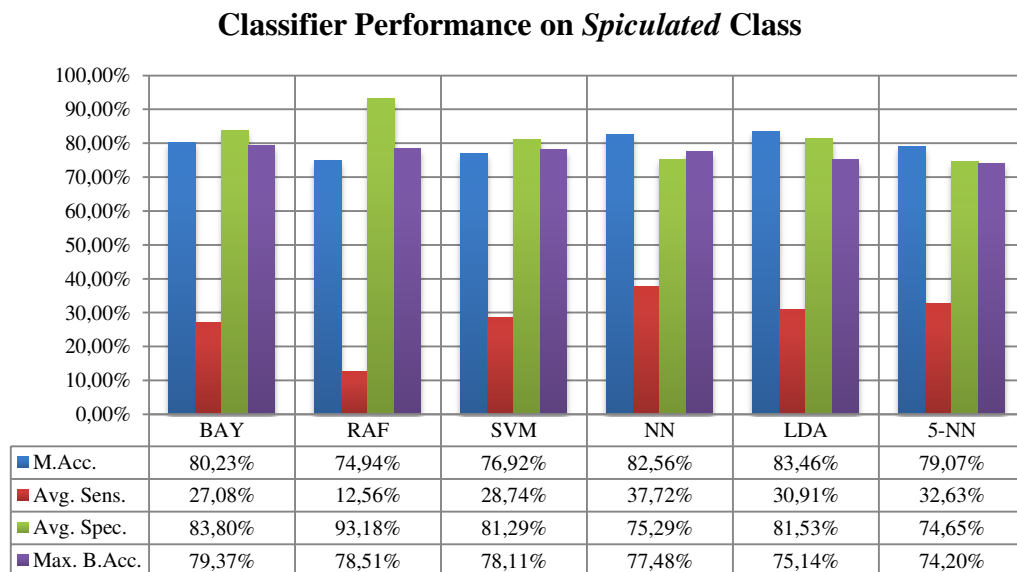


Figure 6.25 Classifier performance on *Spiculated* class.

Figure 6.26 includes classifier performance results of experiments on *Irregular* class. Like previous experimental results, Naïve Bayes classifier obtains the highest balanced accuracy score. On the other hand, *Random Forest* classifier obtains the highest sensitivity score.

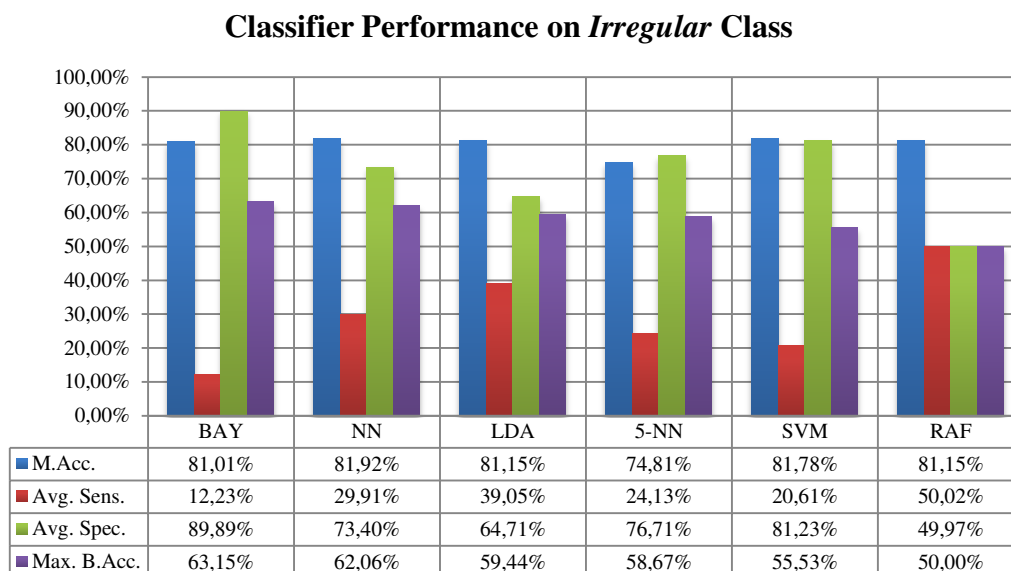


Figure 6.26 Classifier performance on *Irregular* class.

Finally, we could say that Naïve Bayes classifier is the most suitable one for margin classification task. Like previous experimental results, best performance scores obtained from *Circumscribed*, *Spiculated* and *Irregular* classes.

6.3.3 Results of Density Property Experiments

Density property has four classes defined in ACR BI-RADS atlas namely *high-dense*, *iso-dense*, *low-dense* and *radiolucent*. Here, *radiolucent* masses have low malignancy rate while *high-dense* masses have the highest malignancy rate. Detailed performance results of margin property experimentations are discussed below. Since only DEMS dataset contains density annotation information, no dataset performance comparison is given in this section.

6.3.3.1 Low-Level Feature Performance Comparison

Figure 6.27 includes performance evaluation results of experiments on *High-Dense* class. According to the results, Global Margin Features obtain the highest balanced accuracy score. Additionally, other two margin features obtains second highest sensitivity scores with 73,59%, while Color feature obtains the highest sensitivity score. Besides, texture features obtain remarkably better evaluation scores.

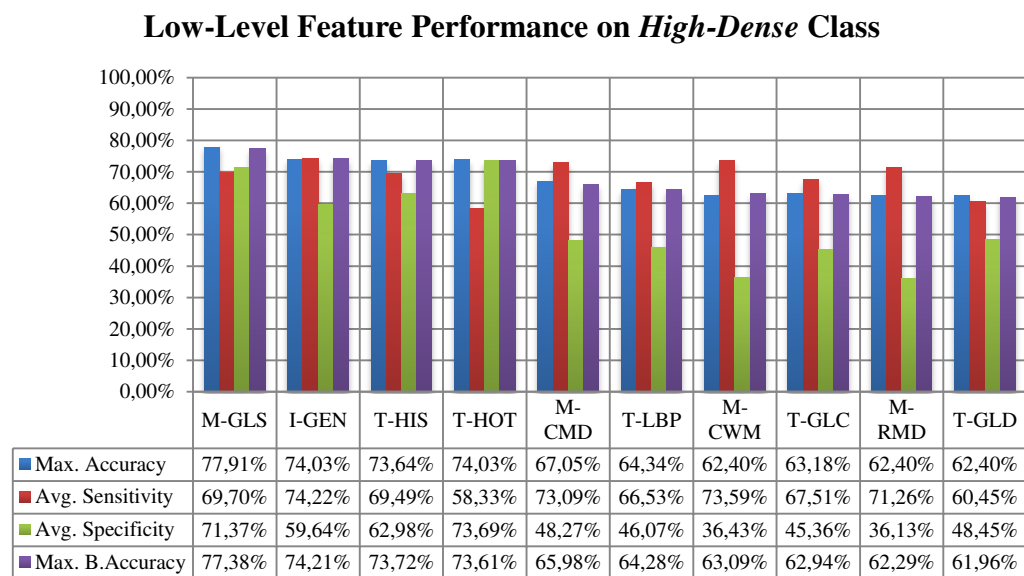


Figure 6.27 Low-Level feature performances of *High-Dense* class.

Figure 6.28 includes performance evaluation results of experiments on *Iso-Dense* class. Similar to previous experiment, global marginal properties feature obtains the highest balanced accuracy score. Moreover, same feature obtains the highest sensitivity score. Color and texture features show considerable balanced accuracy values. But, sensitivity scores of them are low. It is interesting that radial basis signal feature enter the top ten lists.

Figure 6.29 includes performance evaluation results of experiments on Low-Dense class. According to the results, radial distance representation of shape features obtains the highest balanced accuracy score. But sensitivity scores of all experiments are insufficient. Only homogeneous texture feature could retrieve 22.22% of positive examples.

Low-Level Feature Performance on *Iso-Dense* Class

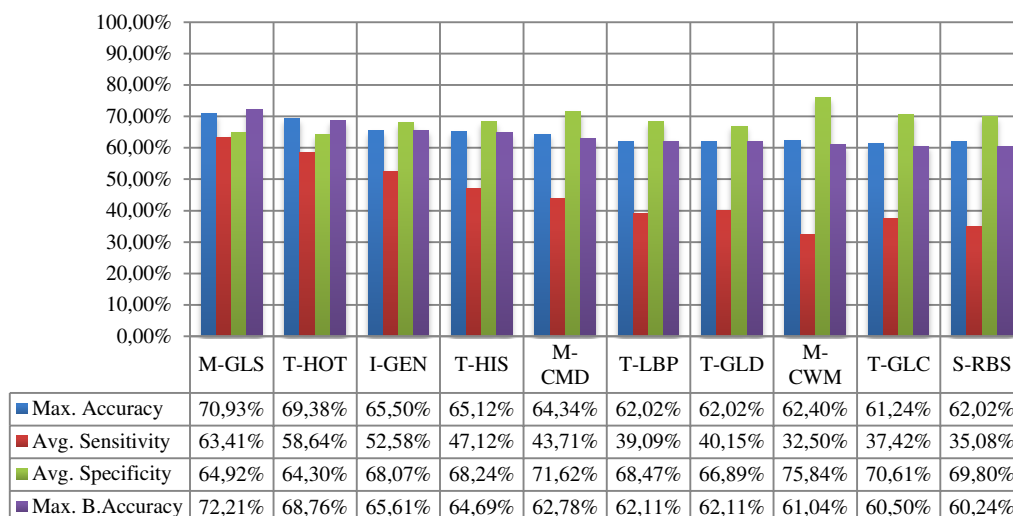


Figure 6.28 Low-Level feature performances on *Iso-Dense* class.

Low-Level Feature Performance on *Low-Dense* Class

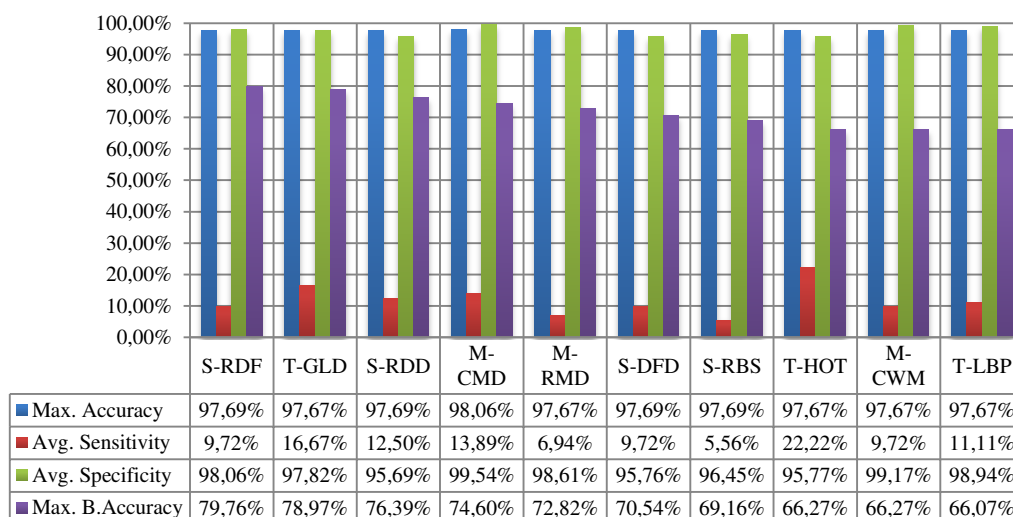


Figure 6.29 Low-Level feature performances *Low-Dense* class.

Figure 6.30 includes performance evaluation results of experiments on *Radiolucent* class. Results indicate that texture features perform better than other feature groups. It is also interesting that performance of shape features enter top 10 results for density classification task.

Low-Level Feature Performance on *Radiolucent* Class

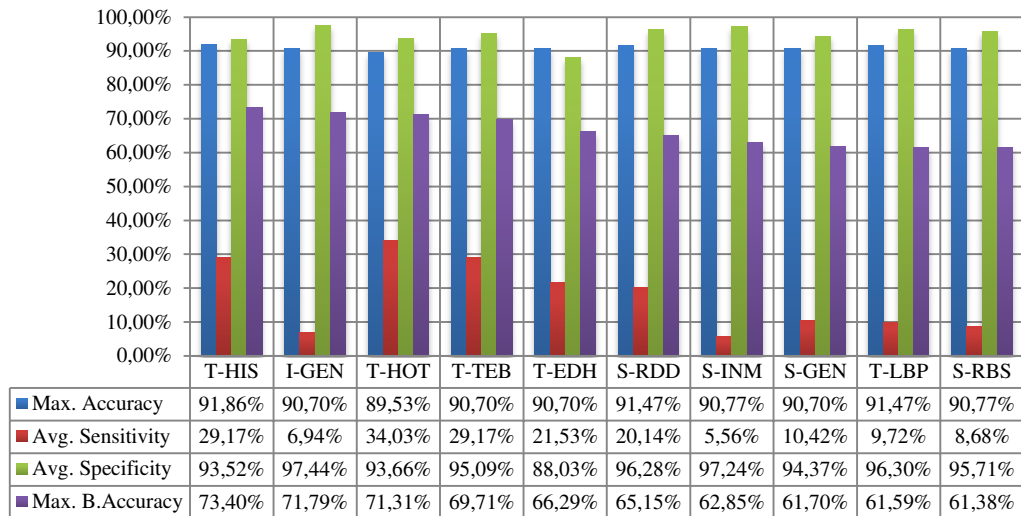


Figure 6.30 Low-Level feature performances *Radiolucent* class.

Finally, we could say that margin and texture features could be used for density classification task. *High-Dense* and *Iso-Dense* class obtains the highest performance evaluation scores.

6.3.3.2 Mass Selection Method Performance Comparison

Performance on Region Selection Method

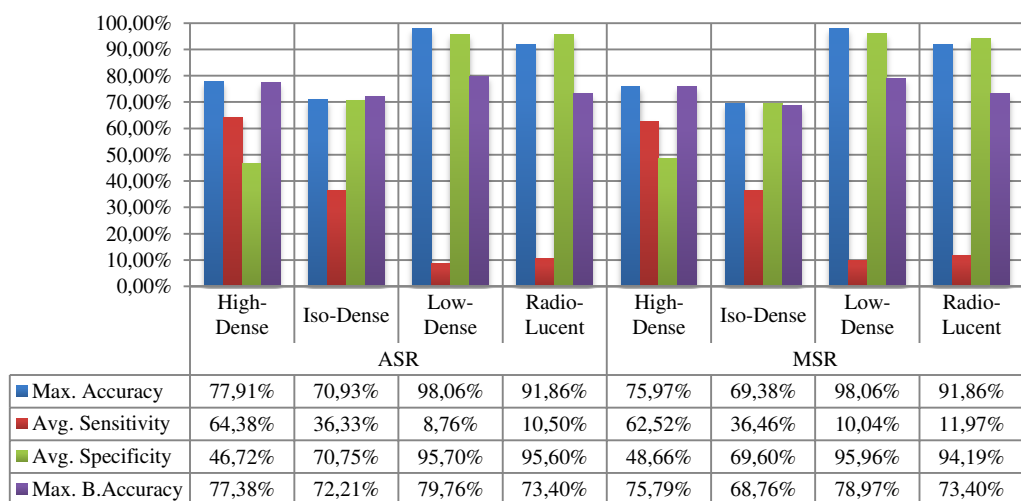


Figure 6.31 Region selection method performance comparison.

Figure 6.31 includes performance evaluation results of region selection method on margin classification task. According to the results, ASR method produces slightly better balanced accuracy results. Additionally, sensitivity scores of both method

produces are almost same. Hence, we can say that performance of two region selection methods show no difference in terms of evaluation metrics used.

6.3.3.3 Classifier Performance Comparison

Figure 6.32 includes classifier performance evaluation results of experiments on High-Dense class. *Linear Discriminant Analysis* classifier obtains the highest balanced accuracy score, while *Random Forest* classifier obtains the highest sensitivity score.

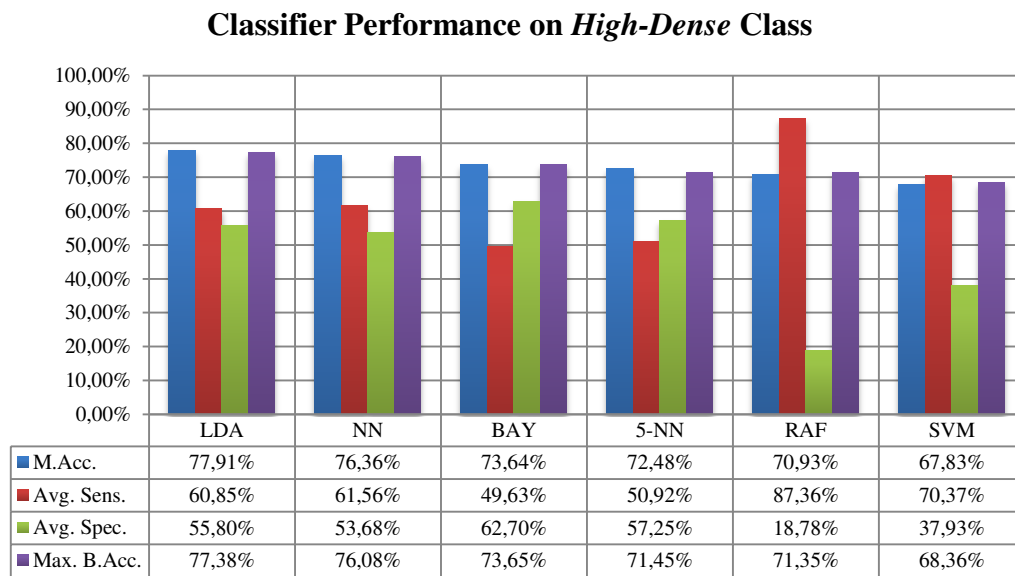


Figure 6.32 Classifier performance on *High-Dense* class.

Figure 6.33 includes classifier performance evaluation results of experiments on *Iso-Dense* class. Similar to High-Dense class experiments, linear discriminant analysis classifier obtains the highest balanced accuracy score. Moreover, it obtains second highest sensitivity score in this case. Neural network classifier obtains the highest sensitivity score, on the other hand. It is also noticeable that k-NN classifier obtains third highest sensitivity score in this experimentation setup.

Figure 6.34 includes classifier performance results of experiments on Low-Dense class. According to the results, Naïve Bayes classifier obtains the highest balanced accuracy and sensitivity score. Results of other classifier performances are insufficient.

Classifier Performance on *Iso-Dense* Class

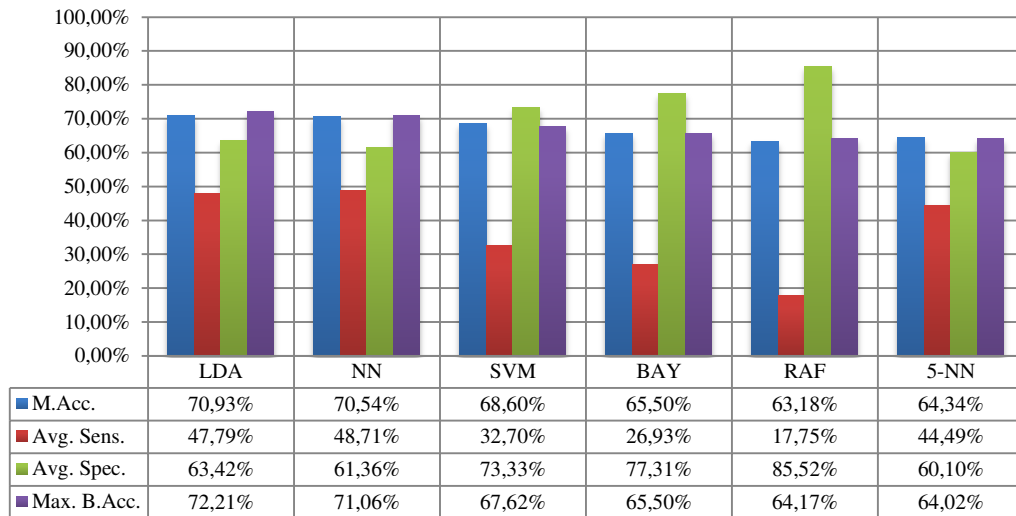


Figure 6.33 Classifier Performance on *Iso-Dense* class.

Classifier Performance on *Low-Dense* Class

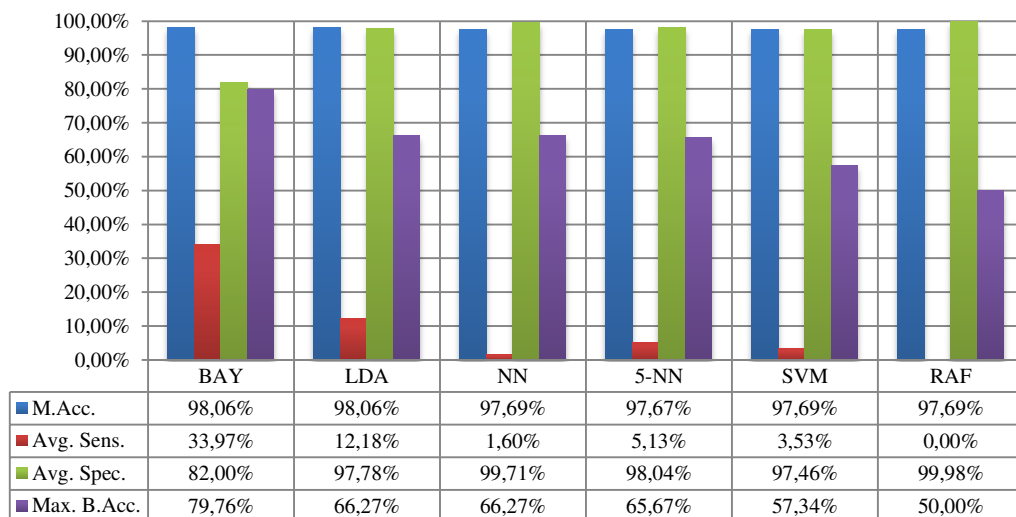


Figure 6.34 Classifier performance on *Low-Dense* class.

Figure 6.35 includes classifier performance results of experiments on *Radiolucent* class. According to the results, naïve Bayes classifier obtains the highest sensitivity and balanced accuracy scores.

In sum, we could say that high-dense and iso-dense classes could be identified by LDA classifier and the other classes could be identified by naïve Bayes classifier. Additionally, high-dense and iso-dense classes obtain better performance scores than other two classes.

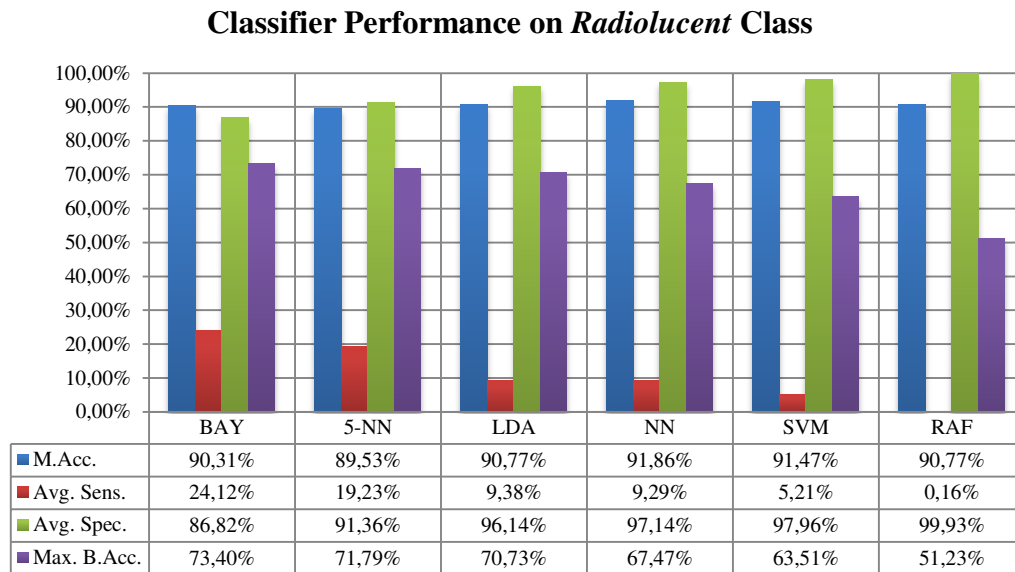


Figure 6.35 Classifier performance on *Radiolucent* class.

6.3.4 Results of BI-RADS Property Experiments

BI-RADS property determines malignancy level of a mass and varies from 0 to 6. Similar to other properties, BI-RADS 0 means low malignancy while 6 means high malignancy rate. Here, only BI-RADS 2, BI-RADS 3 and BI-RADS 5 classes are common between two datasets. BI-RADS 0, BI-RADS 1 and BI-RADS 4 classes only exists in DDSM dataset, while BI-RADS 4A, BI-RADS 4B, BI-RADS 4C and BI-RADS 6 classes only exists in DEMS dataset. Detailed performance results of BI-RADS property experimentations are discussed below.

6.3.4.1 Low-Level Feature Performance Comparison

Figure 6.36 includes low-level feature performances of experiments on BI-RADS 0 class. According to the results, shape features obtain the highest balanced accuracy results. But, texture features obtain highest sensitivity scores.

Figure 6.37 includes low-level feature performance results of experiments on BI-RADS 1 class. Although performance scores especially sensitivity scores are not good enough, texture browsing feature obtains both highest balanced accuracy and sensitivity scores.

Low-Level Feature Performance on *BI-RADS 0* Class

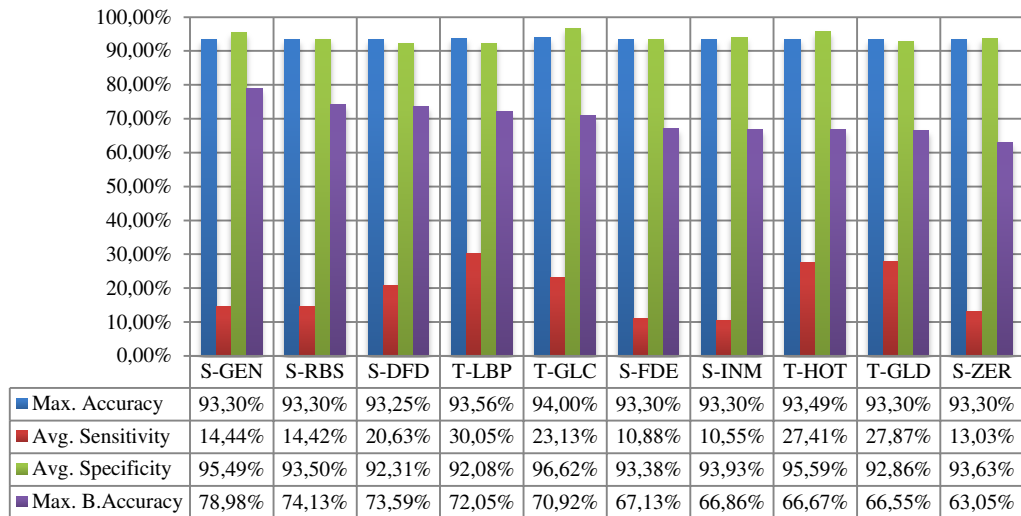


Figure 6.36 Low-Level feature performance on *BI-RADS 0* class.

Low-Level Feature Performance on *BI-RADS 1* Class

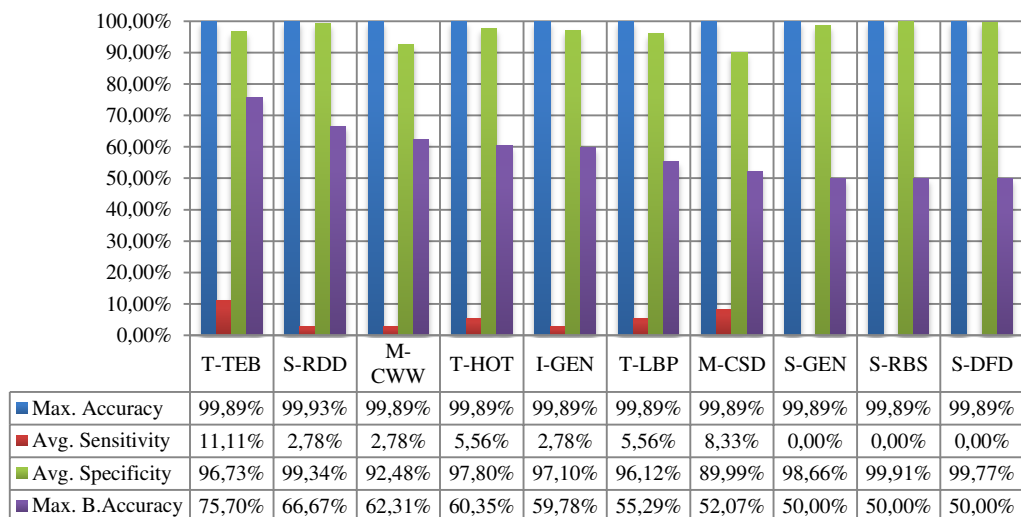


Figure 6.37 Low-Level feature performance on *BI-RADS 1* class.

Figure 6.38 includes low-level feature performances of experiments on BI-RADS 2 class. Both shape and texture features get 5 of 10 results. Hence we could say that shape and texture features could be used to identify BI-RADS 2 class. Similar to previous experimental results, sensitivity scores of all experiments are too low.

Figure 6.39 includes low-level feature performances of experiments on BI-RADS 3 class. Shape features outperform other feature groups of experiments on this class, although histogram properties texture feature obtains highest sensitivity score.

Low-Level Feature Performance on *BI-RADS* 2 Class

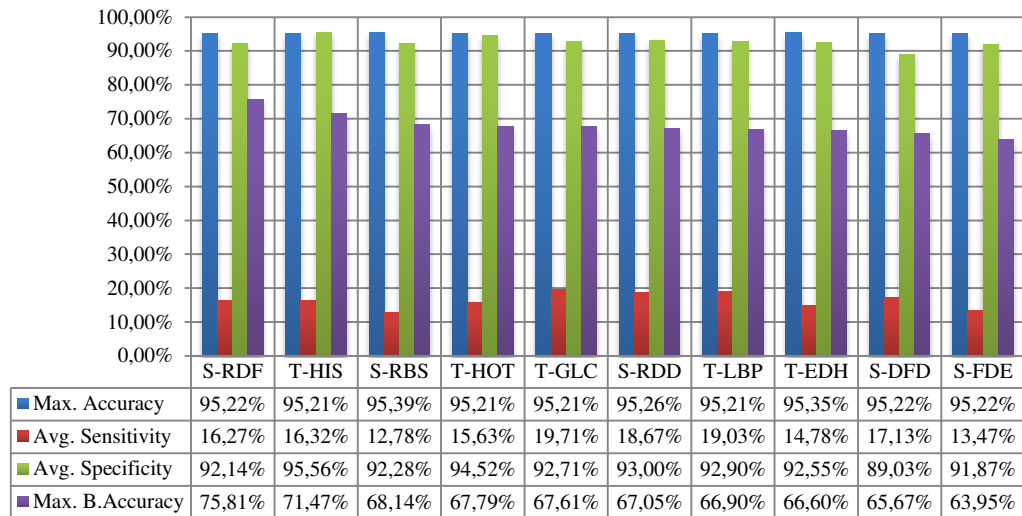


Figure 6.38 Low-Level feature performance on *BI-RADS* 2 class.

Low-Level Feature Performance on *BI-RADS* 3 Class

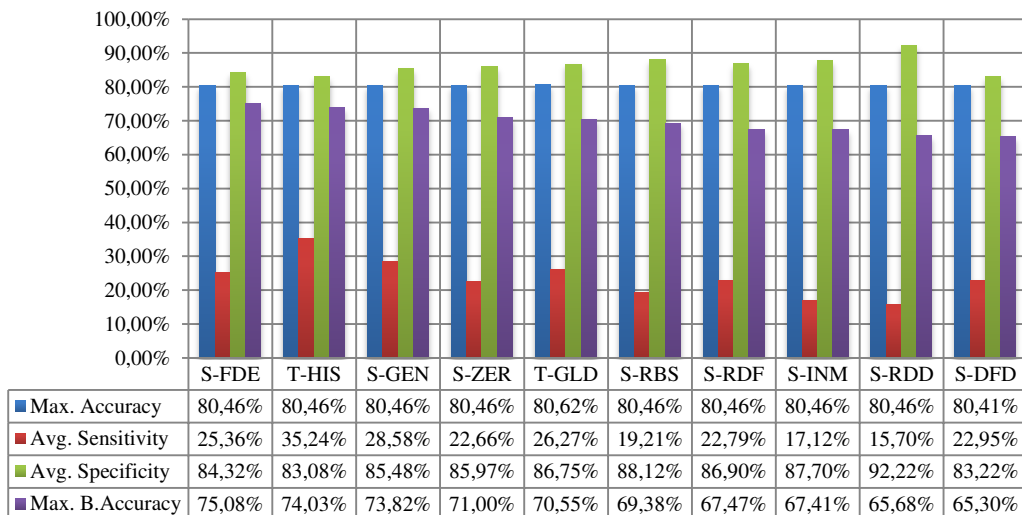


Figure 6.39 Low-Level feature performance on *BI-RADS* 3 class.

Figure 6.40 includes low-level feature performances of experiments on *BI-RADS* 4 class. Texture features obtain 5 of the 10 top balanced accuracy results, while shape features obtain 4 of them. But highest scores belong to texture features. Sensitivity scores of all experiments are above 60%.

Figure 6.41 includes low-level feature performances of experiments on *BI-RADS* 4A class. Margin features obtain 7 of the top 10 balanced accuracy results. But sensitivity scores of all experiments are below 10%. Since samples of this class

belongs to DEMS dataset and number of the samples are too low, degree of sensitivity scores is acceptable.

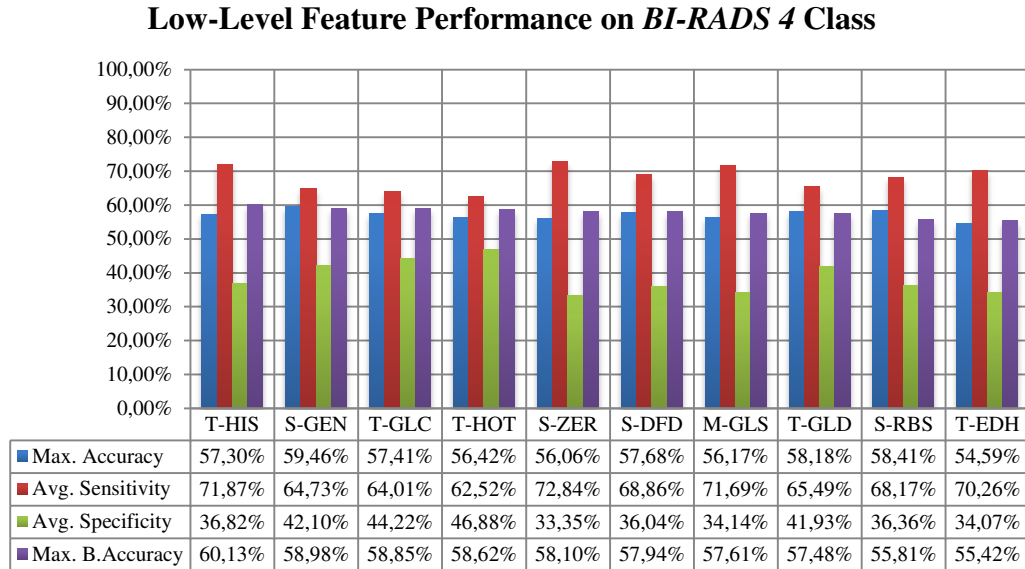


Figure 6.40 Low-Level feature performance on *BI-RADS* 4 class.

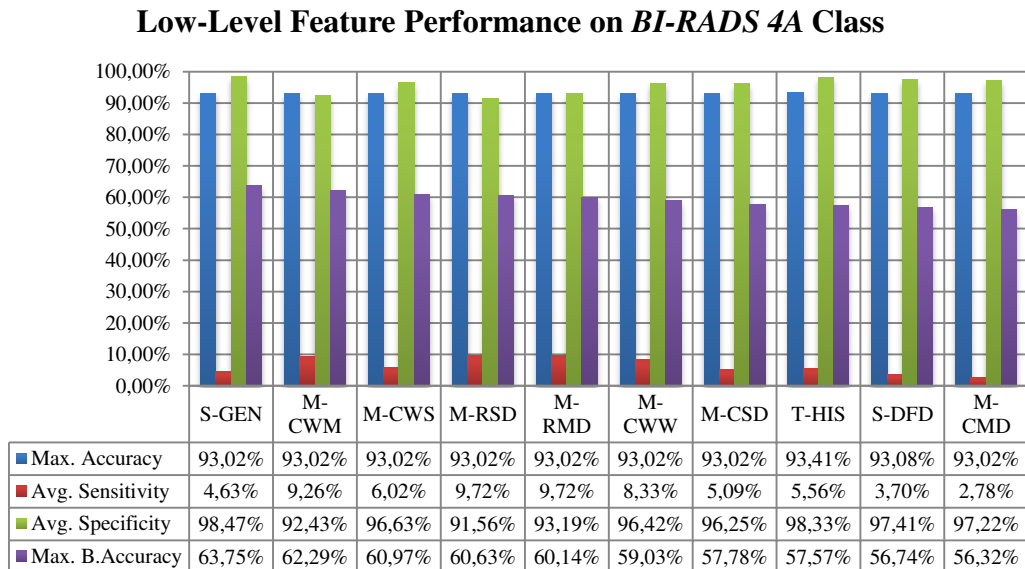


Figure 6.41 Low-Level feature performance on *BI-RADS* 4A class.

Figure 6.42 includes low-level feature performances of experiments on BI-RADS 4B class. Texture features obtain 4 of the 10 results. However, sensitivity scores of all experiments are below from 10% like BI-RADS 4A class.

Figure 6.43 includes low-level feature performances of experiments on BI-RADS 4C class. Shape features obtain 5 of the 10 results. Although sensitivity scores are

better than other two BI-RADS 4 subsets, sensitivity scores are not good enough as evaluation results of other properties.

Low-Level Feature Performance on *BI-RADS 4B* Class

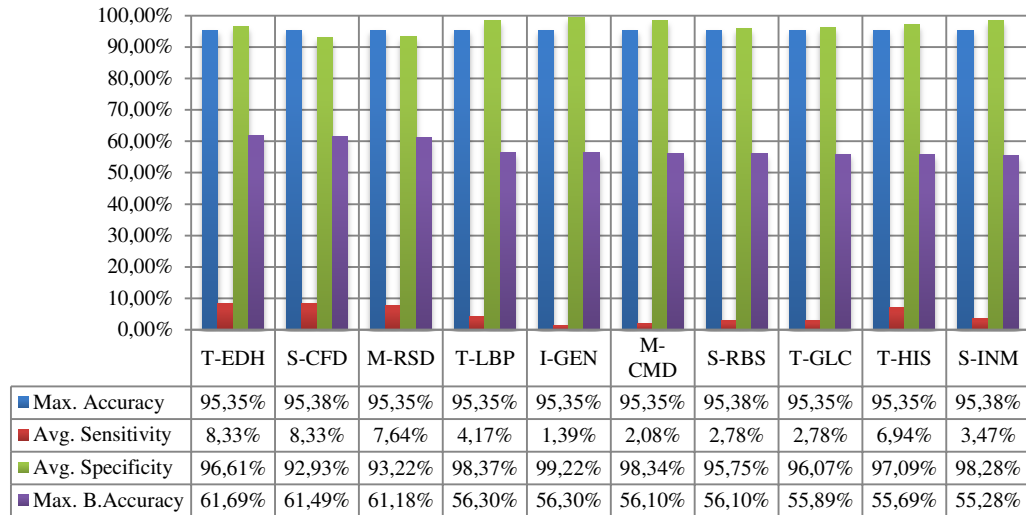


Figure 6.42 Low-Level feature performance on *BI-RADS 4B* class.

Low-Level Feature Performance on *BI-RADS 4C* Class

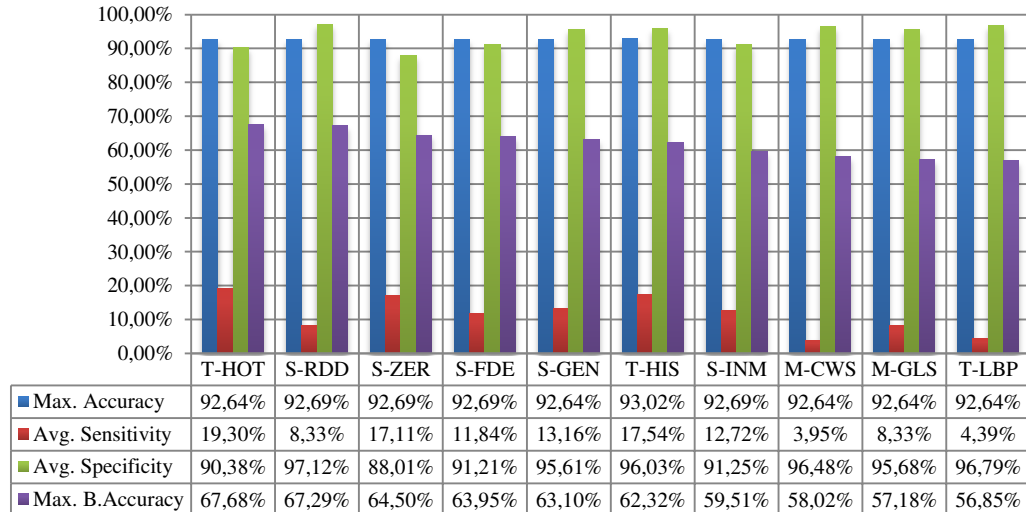


Figure 6.43 Low-Level feature performance on *BI-RADS 4C* class.

Low-level feature performance evaluation results of BI-RADS 5 class are given in Figure 6.44. Performance results of shape features are better than other features, although edge histogram texture feature obtains the highest balanced accuracy score.

Figure 6.45 includes low-level feature performance evaluation results of experiments on BI-RADS 6 class. According to the results, shape, texture and margin features show almost equal balanced accuracy scores.

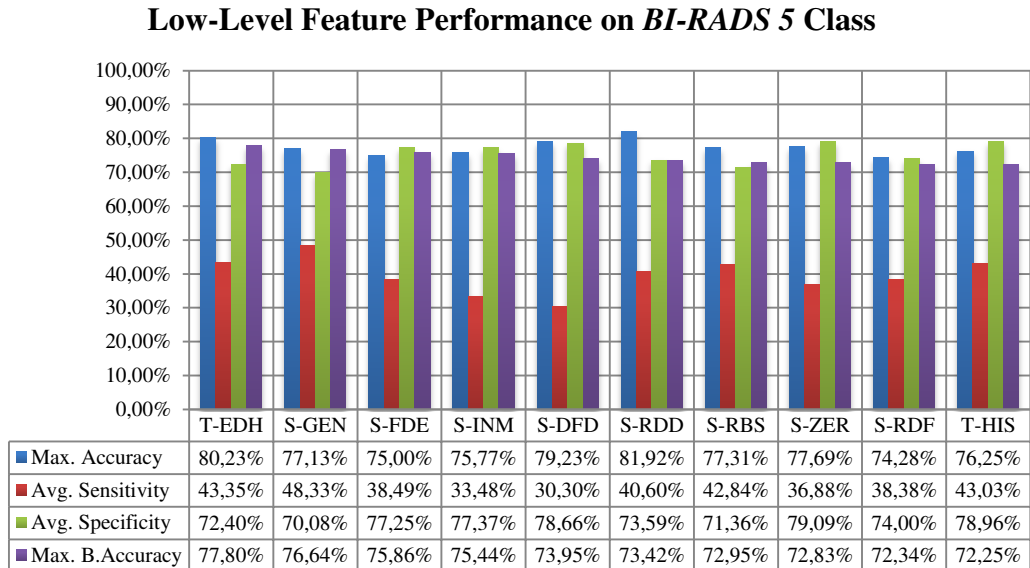


Figure 6.44 Low-Level feature performance on *BI-RADS* 5 class.

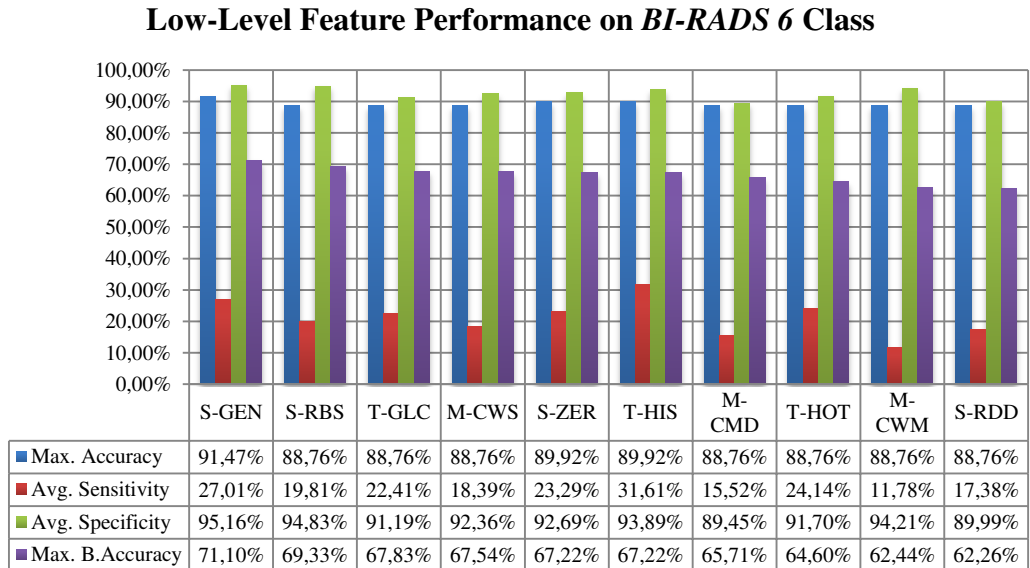


Figure 6.45 Low-Level feature performance on *BI-RADS* 6 class.

In sum, shape and texture features shows better performance than other feature groups. But experiments on BI-RADS 3, BI-RADS 4, BI-RADS 5 and BI-RADS 6 classes obtain acceptable sensitivity scores in terms of sensitivity metric. Especially performance results of BI-RADS 4 and BI-RADS 5 class are noticeable.

Nevertheless, scores of all experiments will be increased by using high level annotations, since BI-RADS score is based on other properties like shape, margin and density.

6.3.4.2 Dataset Performance Comparison

Figure 6.46 includes dataset performance comparison of experiments on BI-RADS property classification task. In common classes scores of DEMS class is higher than DDSM scores. Besides BI-RADS 4 is the most successful class in DDSM experiments while BI-RADS 5 is the most successful class in DEMS experiments.

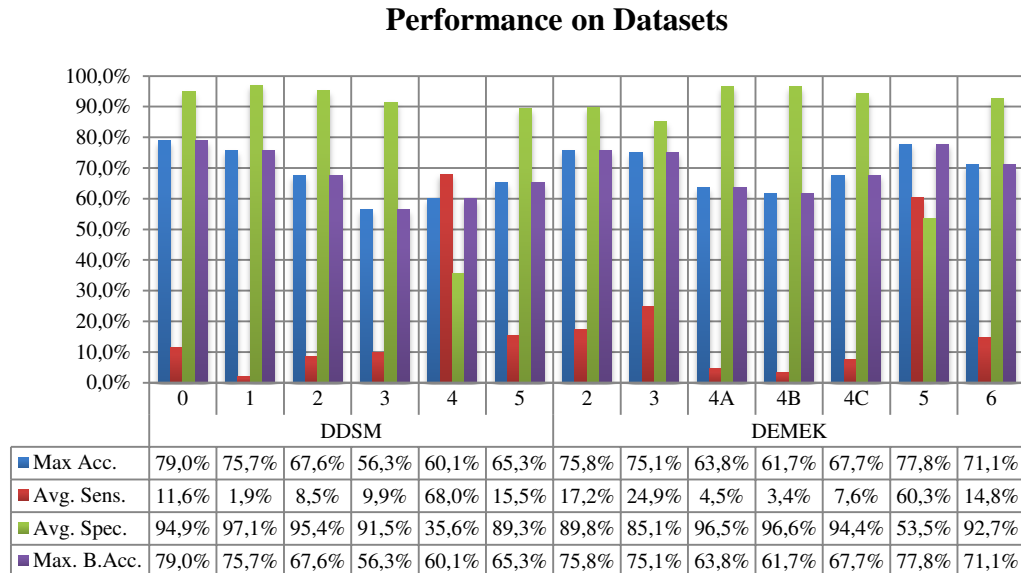


Figure 6.46 Dataset performance comparison.

6.3.4.3 Mass Selection Method Performance Comparison

Figure 6.47 and Figure 6.48 include region selection method performance evaluation results of experiments on BI-RADS classification task. Most of the results are almost same except BI-RADS 0, BI-RADS 2 and BI-RADS 4A. ASR performs better than MSR experiments on BI-RADS 4A class while MSR performs better than

ASR on other two classes. BI-RADS 4 and BI-RADS 5 is the most successful classes of experiments on both ASR and MSR method.

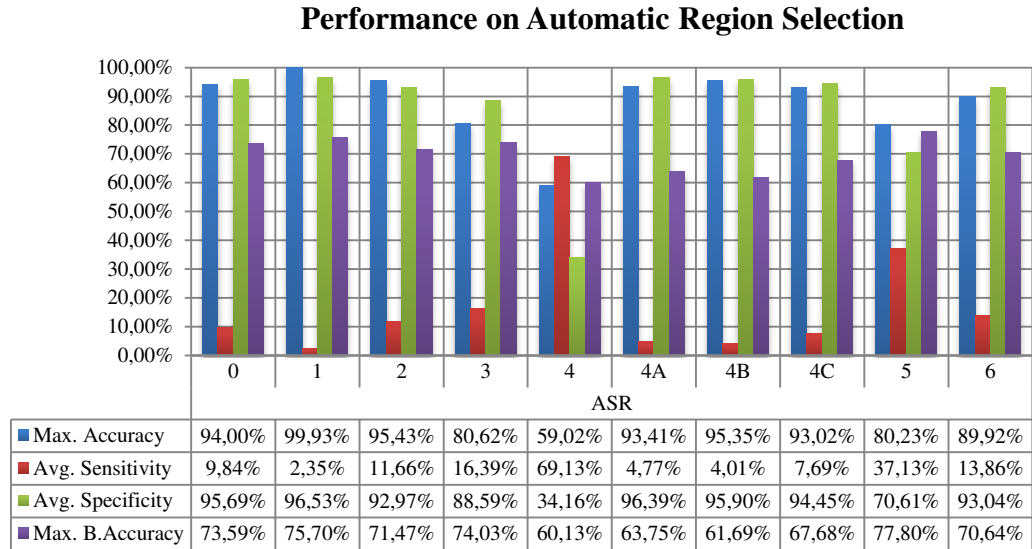


Figure 6.47 ASR performance results.

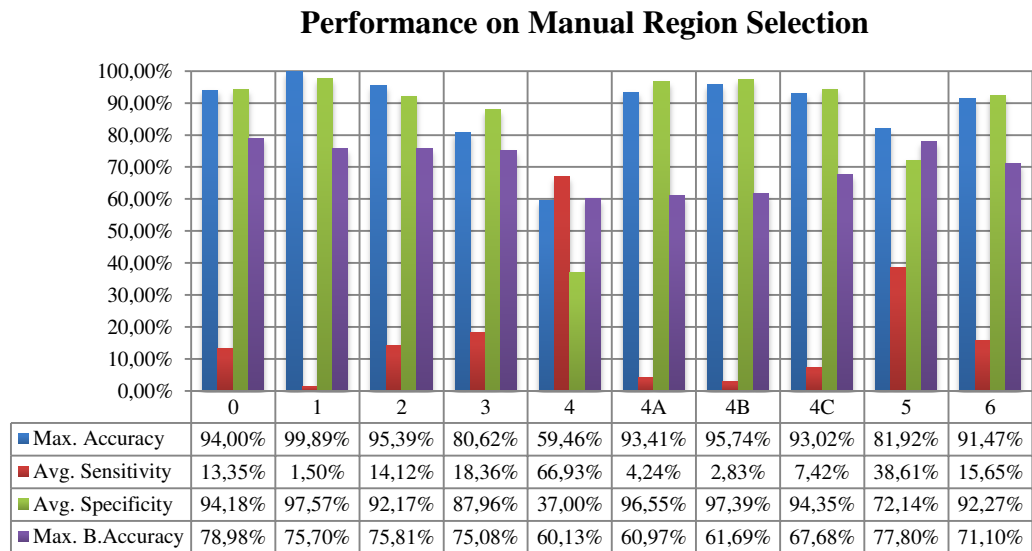


Figure 6.48 MSR performance results.

6.3.4.4 Classifier Performance Comparison

Figure 6.49 includes classifier performances of experiments on BI-RADS 0 class. Naïve Bayes classifier obtains the highest balanced accuracy and sensitivity scores. Results of the other classifiers are twice times worse than naïve bayes classifier in terms of sensitivity metric.

Classifier Performance on *BI-RADS 0* Class

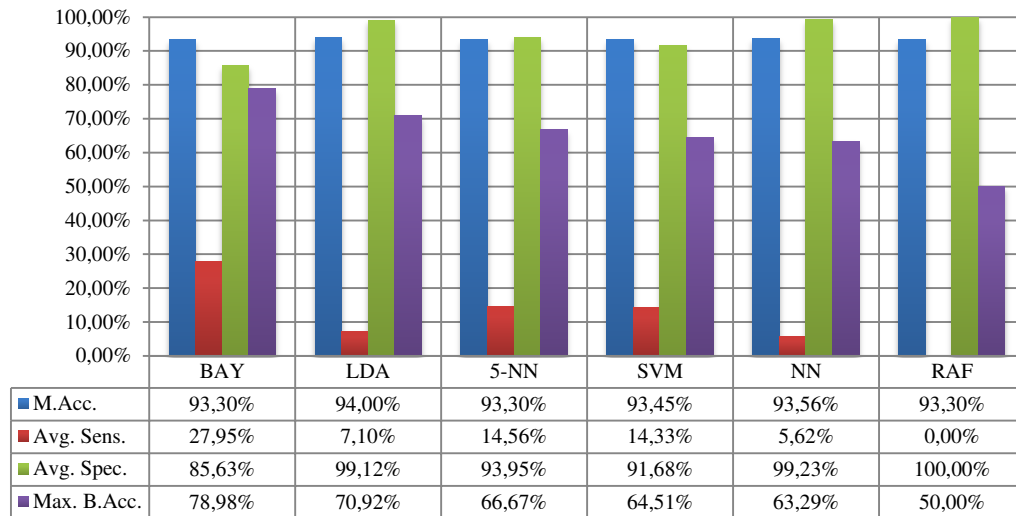


Figure 6.49 Classifier performance on *BI-RADS 0* class.

Figure 6.50 includes classifier performance results of experiments on BI-RADS 1 class. Although naïve Bayes classifier outperforms other classifiers, sensitivity scores of all experiments are below 10%. However, sensitivity score of naïve Bayes classifier is 5 times better than SVM classifier, which is second successful classifier for these experiments.

Classifier Performance on *BI-RADS 1* Class

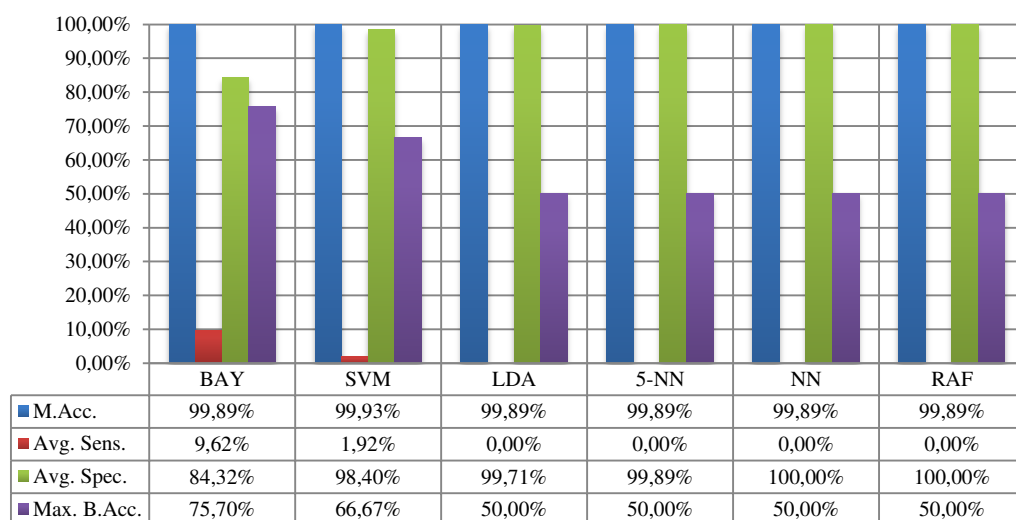


Figure 6.50 Classifier performance on *BI-RADS 1* class.

Figure 6.51 includes classifier performances of experiments on BI-RADS 2 class. Naïve Bayes classifier obtains the highest balanced accuracy score like previous

classes. Additionally, k-NN classifier obtains the highest sensitivity score, which is slightly better than naïve Bayes score.

Classifier Performance on *BI-RADS* 2 Class

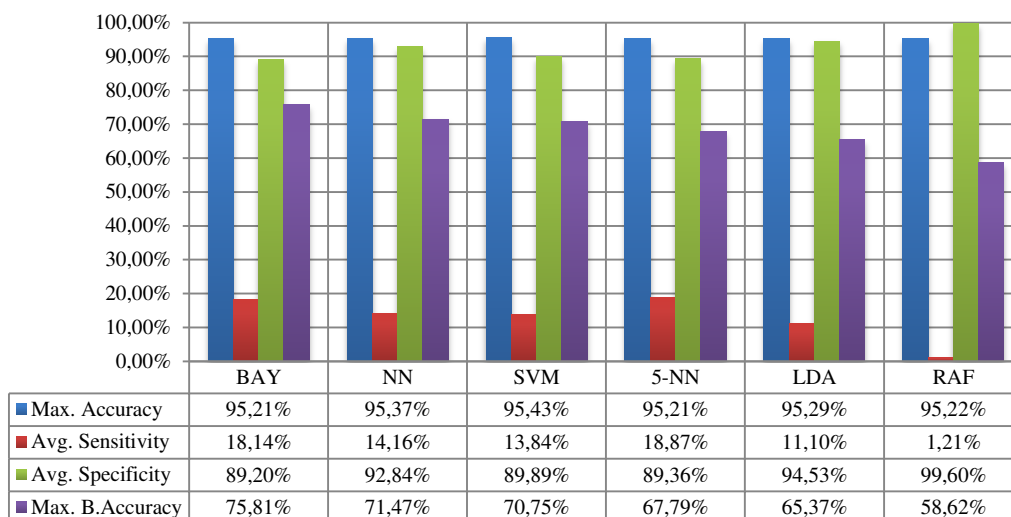


Figure 6.51 Classifier performance on *BI-RADS* 2 class.

Figure 6.52 includes classifier performance of experiments on BI-RADS 3 class. Although random forest classifier obtains the highest balanced accuracy score, naïve Bayes classifier obtains the highest sensitivity score, which is followed by k-NN classifier.

Classifier Performance on *BI-RADS* 3 Class

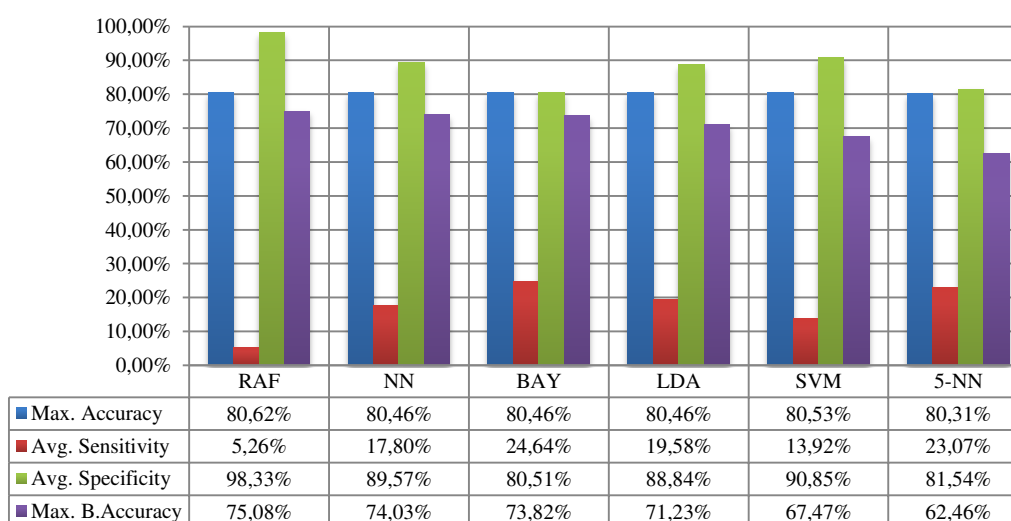


Figure 6.52 Classifier performance on *BI-RADS* 3 class.

Figure 6.53 includes classifier performance evaluation results of experiments on BI-RADS 4 class. According to the results, SVM classifier obtains the highest balanced accuracy score. On the other hand, random forest classifier obtains the highest sensitivity score. However, its specificity score is 0%. So we could say that random forest classifier assigns all samples to BI-RADS 4 class.

Classifier Performance on *BI-RADS 4* Class

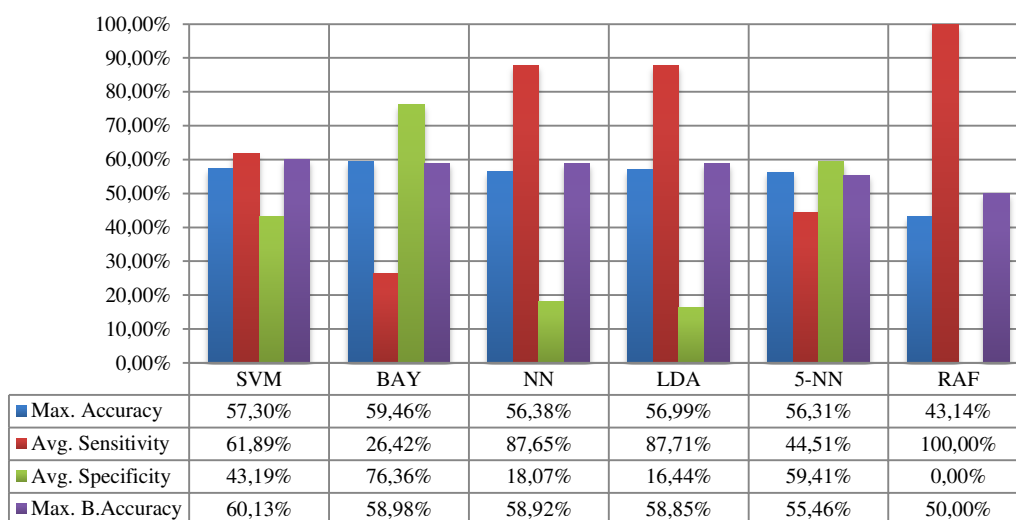


Figure 6.53 Classifier performance on *BI-RADS 4* class.

Classifier Performance on *BI-RADS 4A* Class

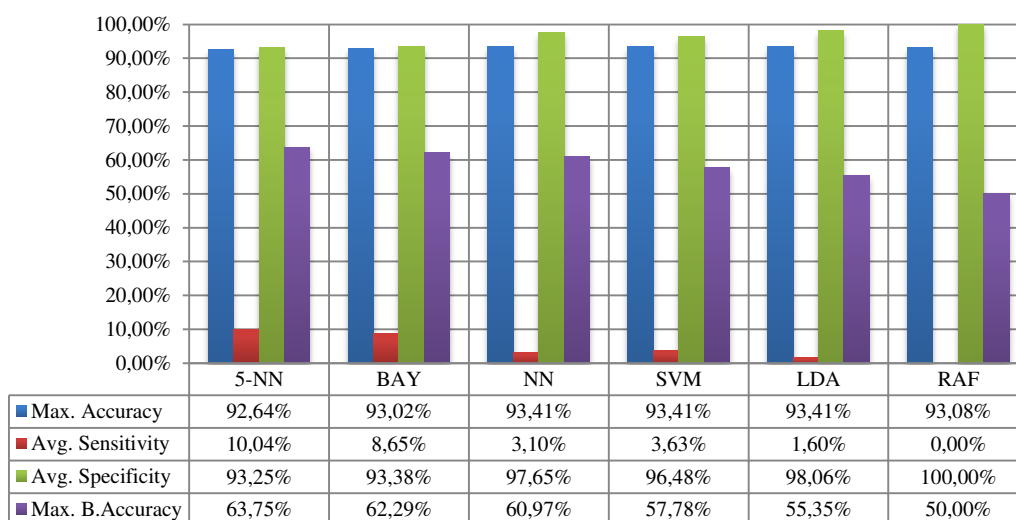


Figure 6.54 Classifier performance on *BI-RADS 4A* class.

Figure 6.54 includes classifier performance results of experiments on BI-RADS 4A class. According to the results, k-NN classifier obtains the highest balanced

accuracy score. Naïve Bayes classifier obtains the second rank in terms of both sensitivity and balanced accuracy metrics.

Classifier Performance on *BI-RADS 4B* Class

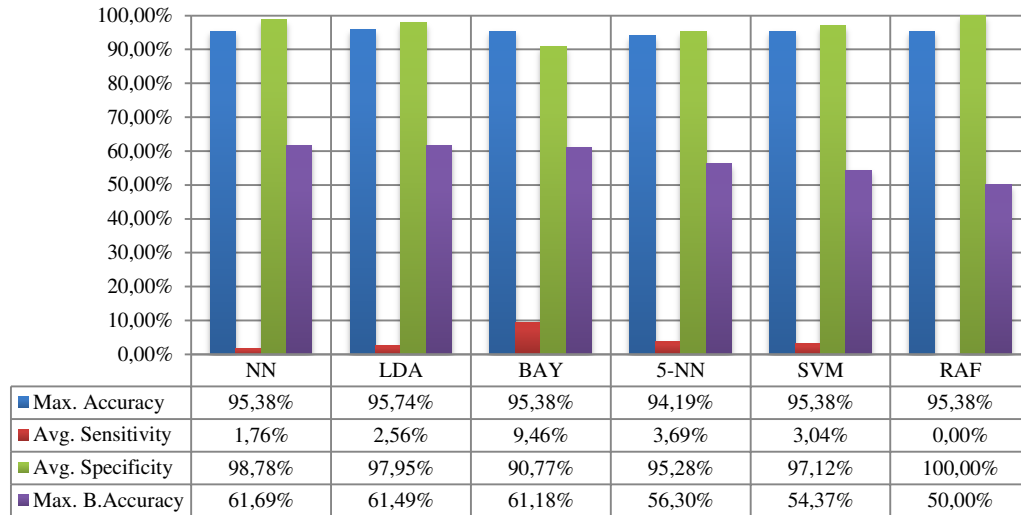


Figure 6.55 Classifier performance on *BI-RADS 4B* class.

Figure 6.55 includes classifier performance evaluation results of experiments on BI-RADS 4B class. Here, neural network classifier obtains the highest balanced accuracy scores. Though, naïve Bayes classifier obtains the highest sensitivity score, which is below 10%.

Classifier Performance on *BI-RADS 4C* Class

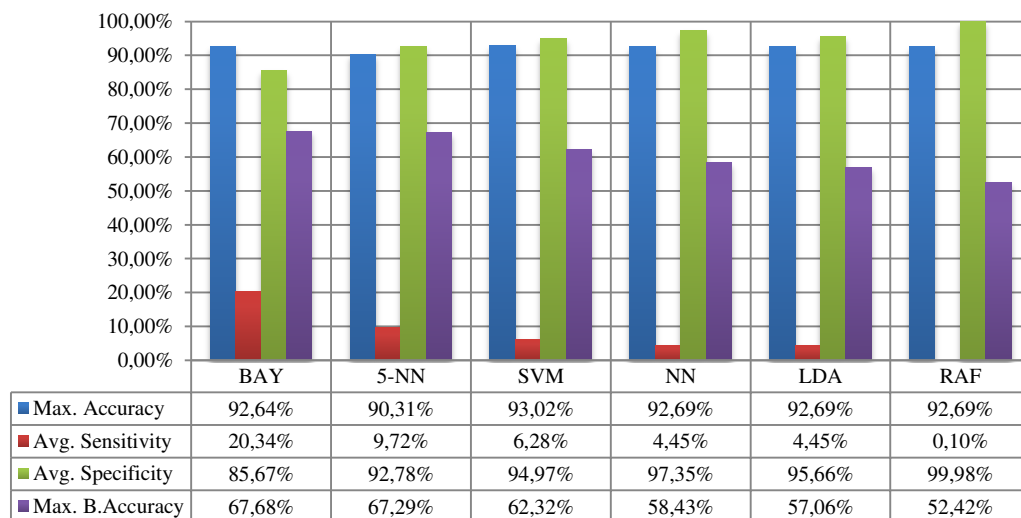


Figure 6.56 Classifier performance on *BI-RADS 4C* class.

Figure 6.56 includes classifier performance evaluation results of experiments on BI-RADS 4C class. According to the results, naïve Bayes classifier obtains the highest evaluation scores in terms of both balanced accuracy and sensitivity scores.

Classifier Performance on *BI-RADS* 5 Class

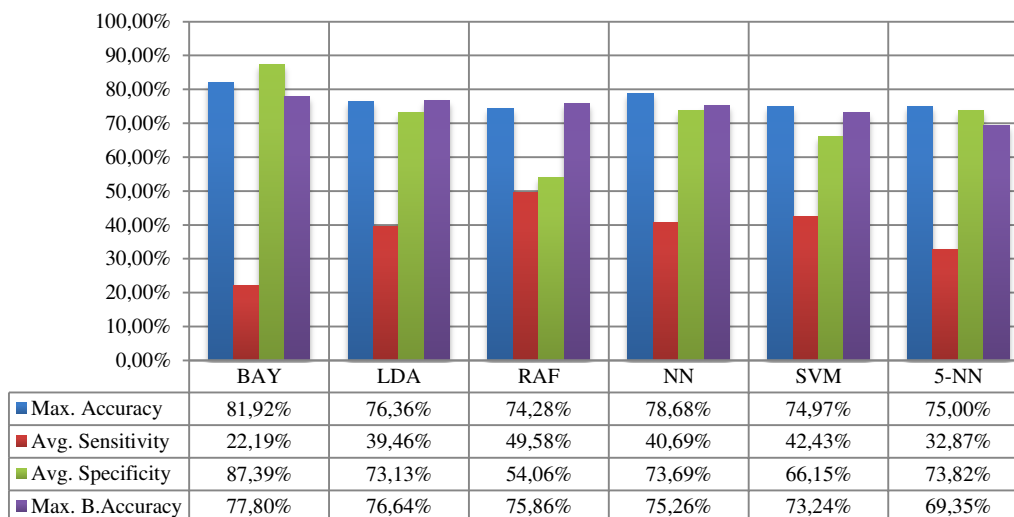


Figure 6.57 Classifier performance on *BI-RADS* 5 class.

Figure 6.57 includes classifier performance evaluation results of experiments on BI-RADS 5 class. Even though naïve Bayes classifier obtains the highest balanced accuracy score, random forest classifier obtains the highest sensitivity score.

Classifier Performance on *BI-RADS* 6 Class

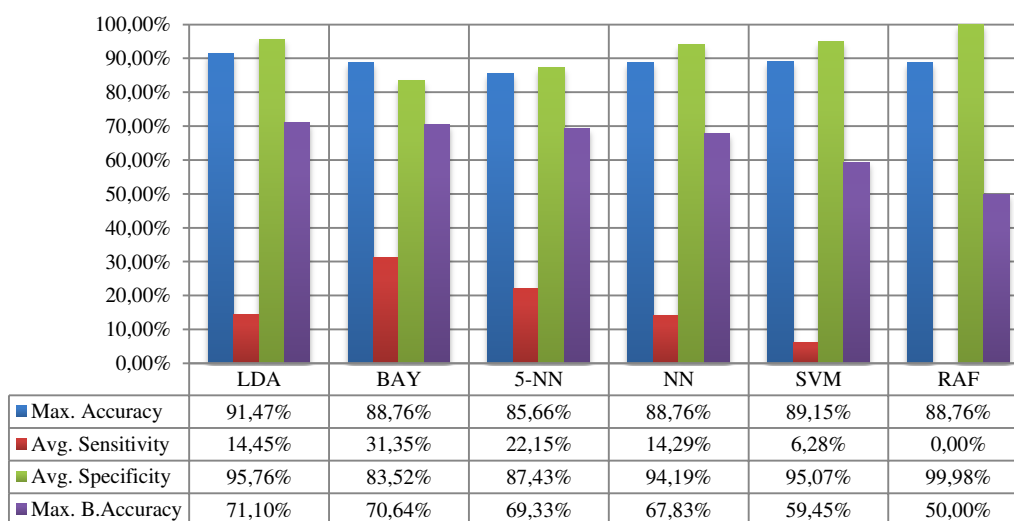


Figure 6.58 Classifier performance on *BI-RADS* 6 class.

Figure 6.58 includes classifier performance evaluation results of experiments on BI-RADS 6 class. LDA classifier obtains the highest balanced accuracy score. However, naïve Bayes classifier obtains the highest sensitivity score. Additionally k-NN classifier obtains second best sensitivity score.

Overall, we could say that Naïve Bayes classifier could be used to identify BI-RADS property of masses. Moreover, BI-RADS 5 and BI-RADS 4 are the most successful classes for BI-RADS classification task.

CHAPTER SEVEN
INTEGRATING DATABASE AND CONTENT-BASED IMAGE
RETRIEVAL:
DIGITAL MAMMOGRAPHY MASS DATABASE

7.1 Overview

Database management systems are used to store and retrieve structured data in a uniform way. However, data storage and retrieval need changes during last decades, because of increasing rate of multimedia production. Currently, images are static and non-queryable objects from database management systems point of view. This makes hard to search large image collections for a particular image. Although some approaches propose using manually attached textual tags to search images, image search using a sample produces are expected be produce better results. Furthermore, manually attached tags are too subjective.

CBIR approach aims to find a particular image in large image collections. But, all of the experimental CBIR systems use their own data representation format. Moreover, each system uses different subset of low-level image features. So, it is hard to integrate CBIR approach into any system related with image data. For instance, FIRE (Deselaers, Keysers, & Ney, 2005), which is a flexible image retrieval engine, uses its own server and data format, and uses HTTP protocol to communicate with its clients. Hence, if someone needs to use FIRE, he/she needs to learn HTTP protocol and usage of FIRE. Moreover, FIRE uses its own subset of low-level features. Although FIRE is built on extensible infrastructure, C/C++ knowledge is needed to implement additional features. Finally, we could conclude that current CBIR systems need to be integrated to a data system that could be used by other systems easily.

In this thesis, we propose a database management system extension aiming to integrate CBIR capabilities to a DBMS, which is well-known open source database management system named PostgreSQL. Although there are similar systems exist in

the literature, some of them are discontinued or inaccessible for public use and the others are too restricted to be used in any system.

This chapter is organized as follows. Section 7.2 provides a literature overview for available database management systems that integrates CBIR approach. Multidimensional indexing and indexing methods is presented in section 7.3. Architecture of our extension is presented in section 7.4. Finally, section 7.5 includes performance evaluation of our extension on DEMS dataset.

7.2 Available Image Database Management Systems

Researchers have been trying to integrate CBIR methods into DBMS for last decade, this is ultimate goal of multimedia retrieval. Hence, some products or extension to DBMS have been developed and released on well-known relational database management systems. Furthermore, an ISO standard called SQL/MM:StillImage is being developed to define image related capabilities for relational databases and described in following sections. Although such efforts exist, there is no commonly accepted standard or product exists to be used for image access or index.

This section includes commercial or non-commercial database management systems that we found in the literature that support image access methods.

7.2.1 SQL-MM:Still-image

SQL-MM is a standard for database management system aiming to define multimedia capabilities and access methods (ISO, 2003). Part 5 of the standard focuses on image object, its properties, related functions and access methods. According to the standard, an SQL/MM compliant database management system should store binary representation on the image with its height, width and image format metadata.

The standard includes some image manipulation functions like zooming, rotating, format changing and thumbnail creation. These methods are all related to StillImage object and should exist in a DBMS supporting SQL/MM standard.

SQL/MM provides four types of low-level features for image access, which are average color, color histogram, positional color, and texture feature. There is no implementation restriction or details given in the standard. Consequently, DBMS developers are free to choose any color space for color features or any texture feature exists in the literature. Standard provides no support for new feature implementations.

7.2.2 Oracle InterMedia

Oracle interMedia (Pelski, 2005) is one of the well-known image extensions exists. It supports SQL/MM StillImage standard and provides three kinds of features, which are shape, texture and color. System uses java advanced imaging (JAI) technology to process images and oracle database management system as image storage. Overview of the system is given in Figure 7.1.

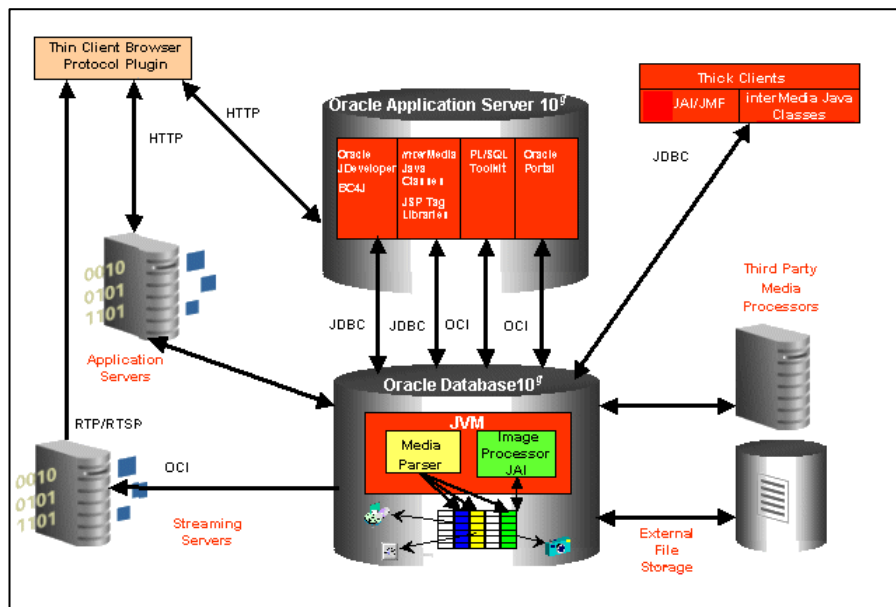


Figure 7.1 Architecture of oracle interMedia (Pelski, 2005).

Although oracle interMedia provides an extensible infrastructure, it seems main goal of the extension is to provide metadata access for images, especially for DICOM images. Moreover, oracle interMedia is tightly integrated with other oracle products like Oracle Application Server. Additionally, interMedia only supports range queries, which aims to find images whose distance is below a predefined distance score.

7.2.2.1 CIRCE

This is an experimental CBIR database management system that is built on Oracle Database 9i. To the best of our knowledge, system was being used in a PACS system that is called cbPACS (Traina et al., 2005) and no other image related system is using CIRCE. System architecture is given in Figure 7.2.

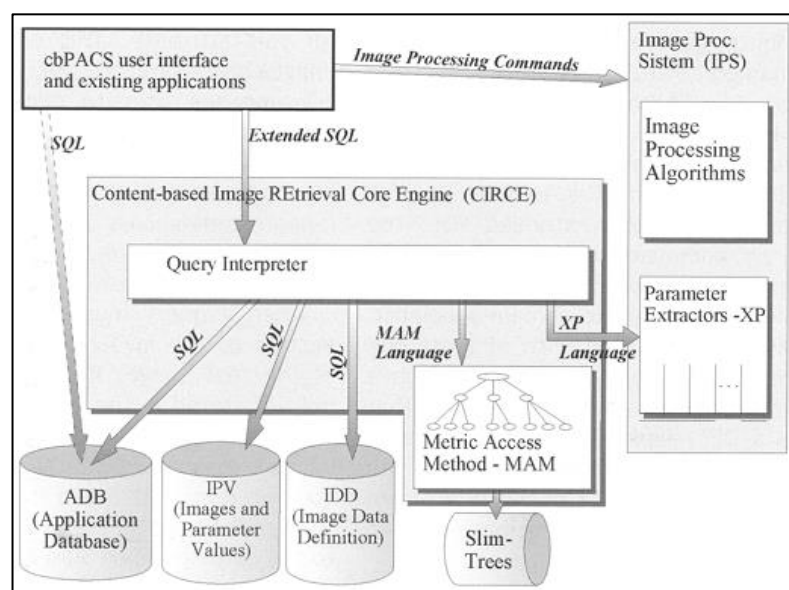


Figure 7.2 Architecture of CIRCE (Traina et al., 2005).

CIRCE uses a non-standard SQL syntax. So, it has its own SQL processor, which converts CIRCE SQL into traditional SQL commands. CIRCE uses slim-tree (Traina Jr., Traina, Seeger, & Faloutsos, 2000) which is an enhanced version of M-Tree (Ciaccia & Zezula, 1997). According to the experimental results, CIRCE provides a noticeable time reduce for querying a medical image archive. Moreover, SQL language syntax provided is so simple to understand and easy to use. However,

system is not publicly available and only used in a PACS system of Brazilian hospital.

7.2.3 IBM AIV Extender

This extension is based on QBIC project (Niblack et al., 1993) and implemented on IBM DB2 universal database. Similar to oracle interMedia, system supports CBIR mechanism. However, IBM discontinues the product, and does not provide any support for AIV extender.

7.2.4 IBM DB2 Still Image Extender

This extension (Stolze, 2005) aims to manipulate images stored in IBM DB2 universal database. Extension supports SQL/MM functionality and manipulation functions only. No content-based access method is provided by this extension. So, this third party extension for IBM DB2 universal database aims only to provide metadata extraction and manipulation functions for images.

7.3 Multidimensional Indexing

Efficient processing and accessing to multimedia data is an important requirement for CBIR systems. Because all of the CBIR tools represent images in high-dimensional feature vectors, hence, they need to efficient access methods for high-dimensional data. For example, FIRE uses a memory based k-d tree implementation to improve search performance. In total, CBIR systems need a data storage mechanism that supports multi-dimensional data processing capabilities.

Multi-dimensional indexing structures are a very hot topic during last decades. Most of the methods aim to work on geographical data like R-Tree (Guttman, 1984) to index two dimensional data. But query performance often worsens while dimensionality increases. Many specialized indexing structures proposed to overcome dimensionality problems.

Query types of multi-dimensional structures are different from traditional query types. There are two types of multi-dimensional query types exists which are nearest neighbor query and range query. Range queries aim to find multi-dimensional data whose distance to particular data is below a predefined threshold value. Nearest neighbor queries deals with number of retrieved documents instead of distances. In other words, nearest neighbor queries aim to find k multidimensional data points nearest to particular multi-dimensional point.

In this section, we present an overview of multi-dimensional indexing structures in literature.

7.3.1 k -d Tree

This multidimensional access structure is an extended version of binary search tree (Bentley, 1975). The k - term denotes the dimension of data is being indexed. So, k -d tree is capable to index any multi-dimensional data.

k -d tree uses as split algorithm which uses different dimensions at each level. A sample 2-d tree is presented in Figure 7.3. As you can see from the figure, each level is splitted by using different dimension.

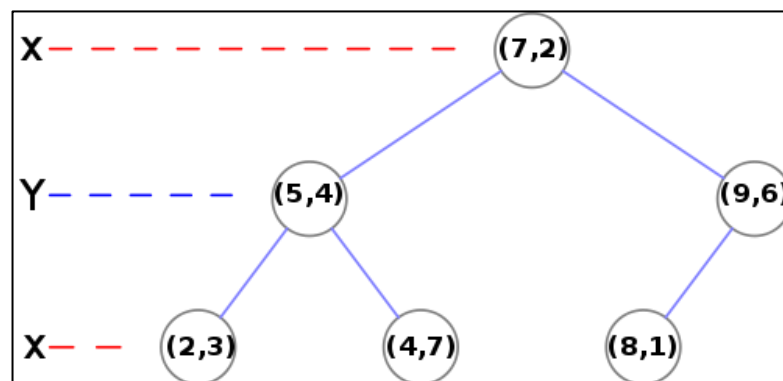


Figure 7.3 Sample k -d tree (Wikipedia, 2012).

k -d tree supports both range and nearest neighbor queries. But performance of k -d tree decreases while dimension of data increased is increased. According to the experiments, it is shown that number of the elements must be greater than 2^k to ensure optimal query times.

7.3.2 X-Tree

X-Tree (Berchtold, Keim, & Kriegel, 1996) is a hybrid approach that uses a hierarchical R-Tree like (Guttman, 1984) and linear array like dictionary. Aim of X-Tree is minimizing overlapping regions by using super nodes, which store history of page splits occurred before. General overview of an X-Tree is given in Figure 7.4.

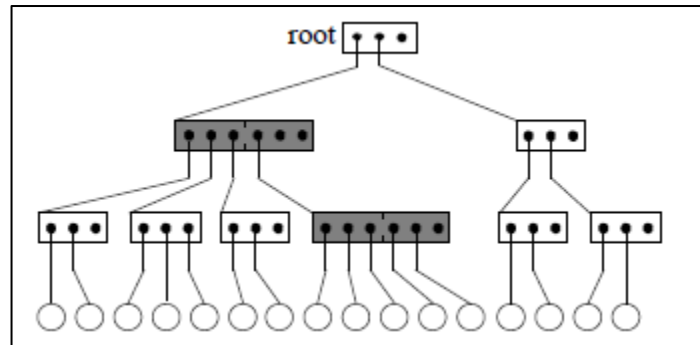


Figure 7.4 General overview of X-Tree (Berchtold et al., 1996).

Darker nodes of the Figure 7.4 represent super nodes. A typical X-Tree could be in a heterogeneous structure, so that non-leaf nodes could be either normal nodes or super nodes. Method uses most of algorithms from other multidimensional approaches like R*-Tree. Only insert method differs from other indexing approaches. During data insertion, method chooses a minimum-bounding rectangle (MBR), and tries to add data to that MBR region. If adding causes a split, method checks an overlapping region exists. If so, current node is enlarged and become a super node. So, X-Tree guarantees minimum number of overlapping regions exists in dataset.

7.3.3 VP-Tree

A vantage point tree (Yianilos, 1993) (vp-tree) is metric space partitioning tree based on distance metrics. Unlike k-d tree, VP-tree partitions the space into circular structures instead of rectilinear structures. Figure 7.5 shows examples both k-d tree space partitioning and VP tree space partitioning.

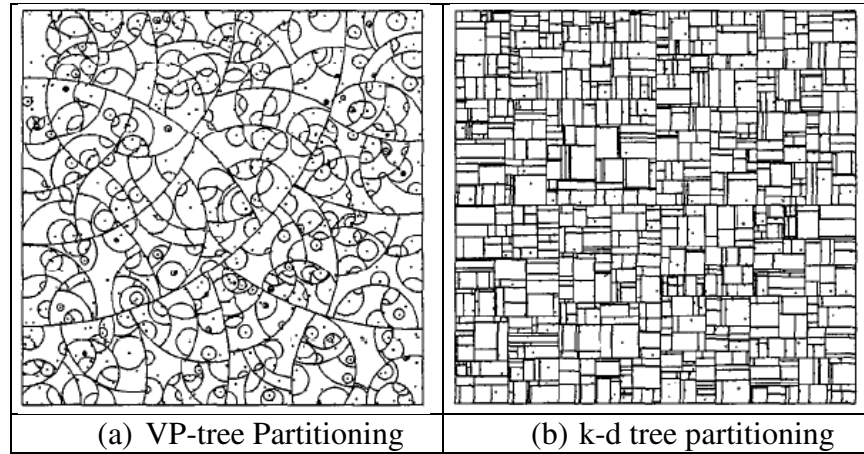


Figure 7.5 Space partitioning examples of (a) VP-tree and (b) k-d tree (Yianilos, 1993)

VP tree aims to split metric space into two groups using a distance threshold to a particular point (vantage point). In other words, points are categorized into two groups, which are close points whose distance is below distance threshold and distance points whose distance is above distance threshold. If the procedure is applied recursively, metric space is divided into small regions of neighbor data.

7.3.3.1 M-Tree

M-tree (Ciaccia & Zezula, 1997) is a specialized metric access method aiming to reduce access time for multi-dimensional data. Similar to vantage point tree, M-tree partitions data space by using center points determined at insertion time. A Sample M-tree node includes entries containing a routing object (O_r) which is the centroid of its neighbor data points, pointer to tree covered by O_r ($T(O_r)$), covering radius of all child data ($r(O_r)$), and distance to the parent routing object ($d(O_r, P(O_r))$). Graphical and hierarchical representation of M-tree is given in Figure 7.6.

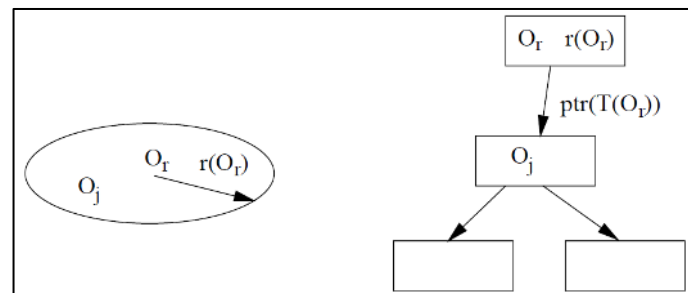


Figure 7.6 Routing object and child objects of M-tree (Patella, 1999)

M-tree uses triangular inequality property of metric spaces to both truncate search tree and eliminate unnecessary distance calculation steps, which is depicted in Figure 7.7. During query execution process, entries whose parent distance is higher than distance between query object and parent object are discarded, because of triangular inequality rule of metric spaces.

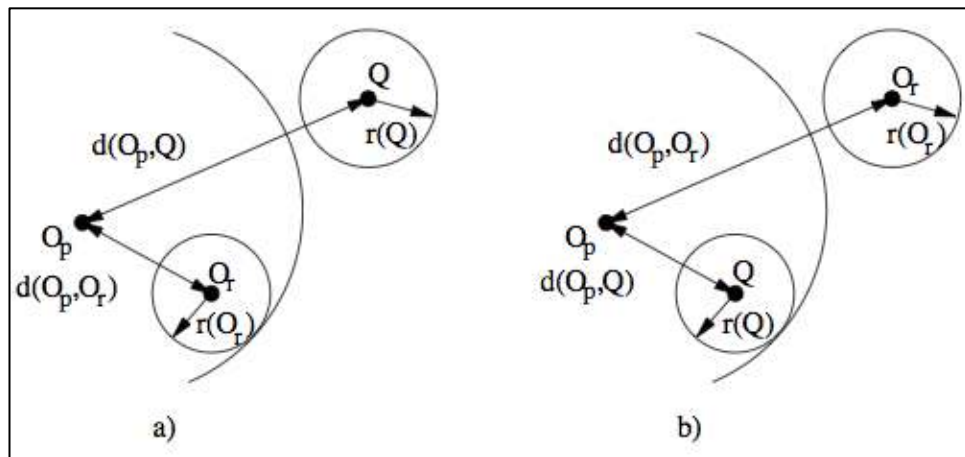


Figure 7.7 Triangular inequality used in M-tree (Patella, 1999).

In literature there is a number of works aiming to extend M-tree exists. Some of them aims to balance tree by reorganizing existing one (Traina Jr. et al., 2000), while an some of the proposes an efficient split algorithm with a second phase of tree reconstruction (Vieira, Jr, Chino, & Traina, 2004). Although such extensions exist, we prefer to use M-tree implementation for this thesis due to its single phase of construction.

7.4 PostgreSQL Extension

PostgreSQL is one of the open source database management systems exist in the literature. It was developed in University of California at Berkeley. Michael Stonebraker and his graduate students developed initial Postgres95 (Momjian, 2000). After Postgres95 is released, it become a community project and named PostgreSQL.

PostgreSQL is written in C and includes many features that do not even exist in commercial database management systems. For instance, PostgreSQL supports object relational relations between tables, array storage, partial indexing etc.

Furthermore, PostgreSQL provides a very extensible programming API to implement new data types, indexing mechanisms, and user defined functions. It is crucial to implement new indexing structures to achieve goal of this thesis. However, current indexing API of PostgreSQL does not support space-partitioning trees. Hence, we design our own indexing structure on PostgreSQL database server. Another important property of PostgreSQL for this thesis is support of array types. Low-level descriptors of a CBIR system could be represented by numeric vector, and they can be stored and retrieved by using numerical array type of PostgreSQL.

7.4.1 System Architecture

Our extension uses two tables to maintain meta-data and indexing structure. All of the fields to be indexed using our extension should be registered for creation of meta-data and indexing table structure. Then all of the operations on user data are manipulated by a trigger function to maintain indexing structure. Overview of our extension is given in Figure 7.8.

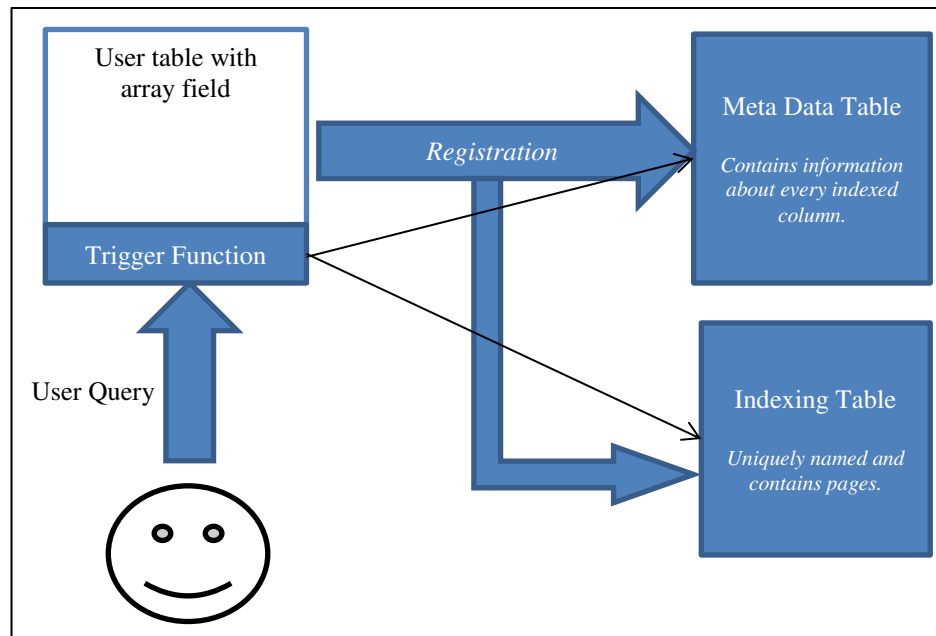


Figure 7.8 Overview of proposed system architecture.

Currently our approach requires a field with *id* in user table to link index tuple to user tuple and uses *Euclidean* distance to index data. Metadata table contains crucial

information like index of root tuple, sequence names for both user id field and index table etc. Structure of metadata table is given in Table 7.1.

Table 7.1 Structure of metadata table.

Column Name	Description
schemaname	Schema of user table
tblname	Name of the user table
fieldname	Field name to be indexed
idxtablename	Name of the index table
rootnode	Id of the root node
data_seq_name	Name of the id sequence of user table
idx_seq_name	Name of the id sequence of index table
dim	Dimension of the data

Indexing table contains actual M-tree structure. A new data type named mnode is used for all tuples in index table. Currently, a node can contain data up to 8K due to size restrictions of PostgreSQL. Structure of index table is given in Table 7.2.

Table 7.2 Structure of index table.

Column Name	Description
id	Unique identifier of index node.
node	Actual node data.

As a result, any application that can produce low-level feature descriptors or user defined functions that return numerical array could use our extension to model their data.

7.4.2 SQL Extension

We implemented several user-defined functions to achieve M-tree functionality, since PostgreSQL does not provide any extensibility feature for its SQL engine. On the other hand, users of our extension need only to learn M-tree related functions instead of a new SQL-like language in this way. Currently, system does not involve with feature extraction from images. Meanwhile system supports any numerical arrays; any C or SQL function could be used for this purpose.

Extension is built on PostgreSQL v9.1, which provides two kind of C API. We prefer to use newest calling convention for our extension, which is also suggested by PostgreSQL developers.

Running multi-dimensional queries on PostgreSQL is another issue, since we cannot modify the query results being produced. In ideal case, typical m-tree query could be able to return ids of user tuples and calculated distance values. Since PostgreSQL uses indexes to eliminate unnecessary tree lookup, modifying the query results is impossible. Hence, we have used set returning function method to create our own query engine, which simply returns ids and distance scores of user tuples.

7.4.2.1 Available Functions

We developed four utility functions to ensure data and index integrity, as well as debugging purposes. Functions are compiled and tested on a Core i5 machine with 4 GB of ram. Besides, code could be compiled on other operating systems like windows or *nix. Detailed information about utility functions is given in following sections.

7.4.2.1.1 create_mtree_field Function This function registers a user field to be indexed using M-tree. Function takes four parameters. First three parameters are strings and determine schema, table and field name to be indexed respectively. Forth parameter is integer and denotes dimension of the numeric array to be indexed. Figure 7.9 includes prototype and sample call of the function.

```
create_mtree_field(cstring, cstring, cstring, integer)

select
create_mtree_field('public','edgehistogram','featureVector', 80);
```

Figure 7.9 Prototype and sample function call of *create_mtree_field* function.

Example creates an m-tree on field *featureVector*, which contains 80 elements. Function creates a metadata tuple in metadata table. After successful creation of meta-data table-indexing table is created and an empty root node is added to index

data. Existing data is processed and initial creation of m-tree is completed. Finally, a trigger function is attached to user table to provide synchronization between user table and index table.

7.4.2.1.2 mtree_handler Function This function is used as trigger function that provides synchronization between user table and index table. Functions take no arguments and shouldn't be called by user. `create_mtree_field` function automatically registers this function to user table on insert, update and delete operations. Hence, when user inserts updates or deletes a tuple from table, this function applies necessary modifications to the indexing table.

7.4.2.1.3 euclidean distance Function This function calculates Euclidean distance between two numeric arrays. This function takes two arguments, which are source and destination numeric array. Return value of this function is the Euclidean distance between two numeric arrays. Figure 7.10 shows function prototype and sample usage.

```
euclidean distance(float8[], float8[])

select euclidean distance('{1,1,1,1}', '{2,2,2,2}');
```

Figure 7.10 Prototype and sample function call of *euclidean distance* function.

Here, example calculates distance between two four dimensional vectors which are $\langle 1,1,1,1 \rangle$ and $\langle 2,2,2,2 \rangle$ and returns distance value.

7.4.2.1.4 containsoid Function This function checks whether an m-tree node contains a particular object id. Function takes two arguments, which are m-tree node to check and an integer value that is id of the object. Figure 7.11 shows function prototype and sample usage.

```
containsoid(mnode, integer)

select containsoid(node1, 1);
```

Figure 7.11 Prototype and sample function call of *containsoid* function.

Here, example checks whether *node1* contains object id 1. We have implemented this function for debugging purposes.

7.4.2.2 Range Queries

This query type aims to find tuples whose distance to particular query point is below a predefined distance threshold. We have implemented a special function named `rangequery` to run range queries on our system. Since this function is a part of our query execution system, this function returns a set of tuples containing ids and distance scores. Prototype of this function is given in Figure 7.12.

```
rangequery(cstring, cstring, cstring, float8[], float8)
```

Figure 7.12 Prototype of *rangequery* function.

Similar to `create_mree_field`, first three parameters denote schema, table and field names of user data, forth parameter denotes query object, and finally fifth parameter denotes threshold value.

We have used range search algorithm that is provided with original M-Tree implementation. Algorithm uses triangular inequality narrow down search space. Range query algorithm of M-tree is given in Figure 7.13.

```
RangeSearch(N: MTreeNode, range(Q, rQ): query)
  Op ← Parent of N
  if N is not Leaf
    foreach Or in N
      if abs(dist(Op, Q) - dist(Or, Op)) < rQ + r(Or)
        if d(Or, Q) < rQ + r(Or)
          RangeSearch(T(Or), range(Q, rQ))
        end
      end
    end
  else
    foreach Oi in N
      if abs(dist(Op, Q) - dist(Oi, Op)) ≤ rQ
        if dist(Oi, Op) ≤ rQ
          add Oi to resultset
        end
      end
    end
  end
end
```

Figure 7.13 Range search algorithm of M-Tree (Patella, 1999).

Here $dist(O_p, Q)$ is calculated in previous nodes, so there is no need to calculate parent distance at each step. Sample execution of range query is depicted in Figure 7.14.

```
select * from rangequery('public', 'edgehistogram', 'feature',
'{1,2,3,4}', 0.7) as rng(id integer, distance float8)
```

Figure 7.14 Sample usage of range query.

Here, query finds tuples whose distance to vector $\langle 1, 2, 3, 4 \rangle$ is below 0.7. Result set of this function contains ids and distances to query vector.

7.4.2.3 Nearest Neighbor Queries

This queries aims to find k nearest tuples to a specific point. Similar to range query type a set returning function named *knnquery* executes nearest neighbor queries on our indexing mechanism. Prototype of this function is given in Figure 7.15.

```
knnquery(cstring, cstring, cstring, float8[],
int)
```

Figure 7.15 Prototype of *knnquery* function.

All of the parameters except the last one is same as range query function. Last parameter denotes k which is the number of nearest tuple to return.

Algorithm of *knnquery* is similar to one used in R-Tree. Algorithm uses a priority queue define current maximal range value. Algorithm of *knnsearch* is presented in Figure 7.16.

NN_Update function updates current result array and return maximum distance score, d_k donates current level of maximum distance function which is defined as maximum distance score of current result array; d_{min} and d_{max} donates minimum and maximum distance limits of tree node $T(O_r)$ to query respectively. Formal definition of d_{min} and d_{max} is given following formula.

$$d_{min}(T(O_r)) = \max\{d(O_r, Q) - r(O_r), 0\}$$

$$d_{max}(T(O_r)) = d(O_r, Q) + r(O_r)$$

```

KnnNodeSearch(N: MTreeNode, NN(Q, k): query)
  Op ← Parent of N
  if N is not leaf node
    foreach Or in N
      if abs(dist(Op, Q) - dist(Or, Op)) ≤ dk + r(Or)
        if dmin(T(Or)) ≤ dk
          PR ← PR ∪ [T(Or), dmin(T(Or))]
          if dmax(T(Or)) < dk
            dk ← NN_Update(_, dmax(T(Or)))
            Remove all entries in PR which dmin(T(Or)) > dk
          end
        end
      end
    end
  else
    foreach Oj in N
      if dist(Oj, Q) ≤ dk
        dk ← NN_Update(oid(Oj), dist(Oj, Q))
        Remove all entries in PR which dmin(T(Or)) > dk
      end
    end
  end
end

```

Figure 7.16 Algorithm of knn search (Patella, 1999).

Sample execution of nearest neighbor query is given in Figure 7.17.

```

select * from knnquery('public', 'edgehistogram', 'feature',
'{1,2,3,4}', 100) as knn(id integer, distance float8)

```

Figure 7.17 Sample execution of nearest neighbor queries.

Sample query finds first 100 nearest tuples to vector <1, 2, 3, 4>. Result set of this function contains ids and distances to query vector.

7.4.2.4 Farthest Neighbor Queries

Although this type of query is rarely used in CBIR systems, we have implemented this query type to find k farthest tuples to particular point. This kind of query will be helpful, when user needs to normalize distance scores of a result set. Similar to previous query types, a set returning function named *kfnquery* executes this query

and returns ids and distance scores. Figure 7.19 depicts prototype of *kfnquery* function.

```
kfnquery(cstring, cstring, cstring, float8[],
        int)
```

Figure 7.18 Prototype of *kfnquery* function.

All parameters are same to *knnquery* function. Function uses same algorithm with *knnquery* function. But reverse of the distance scores is used instead of distance scores. We used logarithmic inversion of distance values which is given in following formula.

$$\text{Reverse Distance} = -\ln(1 - e^{-\text{distance}})$$

Sample execution of nearest neighbor query is given in Figure 7.19.

```
select * from kfnquery('public', 'edgehistogram', 'feature',
'{1,2,3,4}', 100) as kfn(id integer, distance float8)
```

Figure 7.19 Sample execution of farthest neighbor queries.

Sample query finds first 100 farthest tuples to vector <1, 2, 3, 4>. Result set of this function contains ids and distances to query vector.

7.5 Performance Evaluation

We conducted a set of experiments on two datasets to measure performance of our system in both access time and information retrieval perspectives. We have executed all types of queries on a very huge dataset and calculated total execution time, number of page access and number of distance calculations to measure data access performance of our system. We generated *Precision-Recall* graphs to measure information retrieval performance of our system. However, we could only present the most successful Precision-Recall graphs. The rest of the precision recall graphs are presented in appendices.

7.5.1 Data Access Performance

We used the ImageCLEFMed dataset to measure data access performance of M-Tree. We run two kinds of queries to measure speed-up. First query type is sequential scan (SEQ) which is a typical database query executing multidimensional query without using M-Tree indexing method. SEQ queries are using distance function on all tuples in user data and query object. Then, database system eliminates the tuples that does not satisfy query condition. After running SEQ query, we run the same query with our extension and observed query parameters, which are number of page access, number of distance calculations and query execution time. We run each query type with two distinct points one of which belong to dataset and one of which does not. Hence we could compare data access parameters objectively.

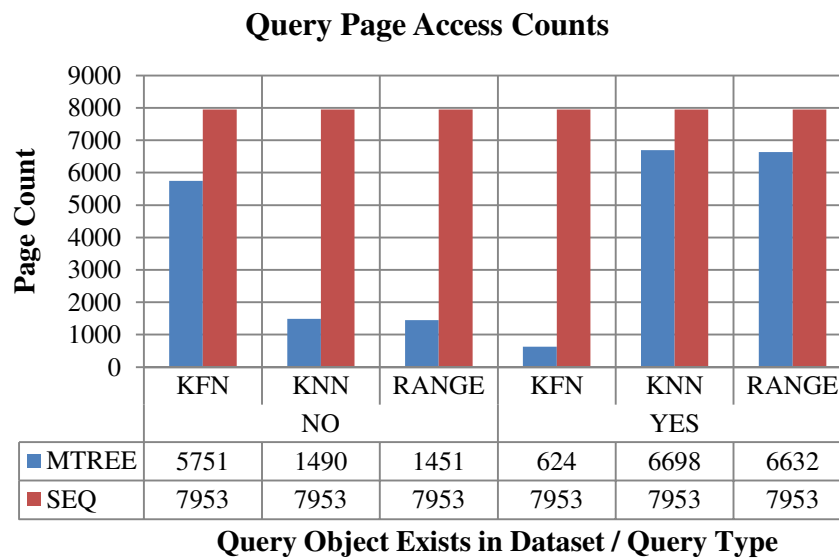


Figure 7.20 Number of page accesses of M-Tree and SEQ.

Figure 7.20 describes total number of page accesses for each multidimensional query type. According to the results, M-tree reduces number of page accesses in all query types. But, it is interesting that M-tree performs better for KNN and range query types using a foreign data. On the contrary, performance of KFN increases using data point belonging to dataset.

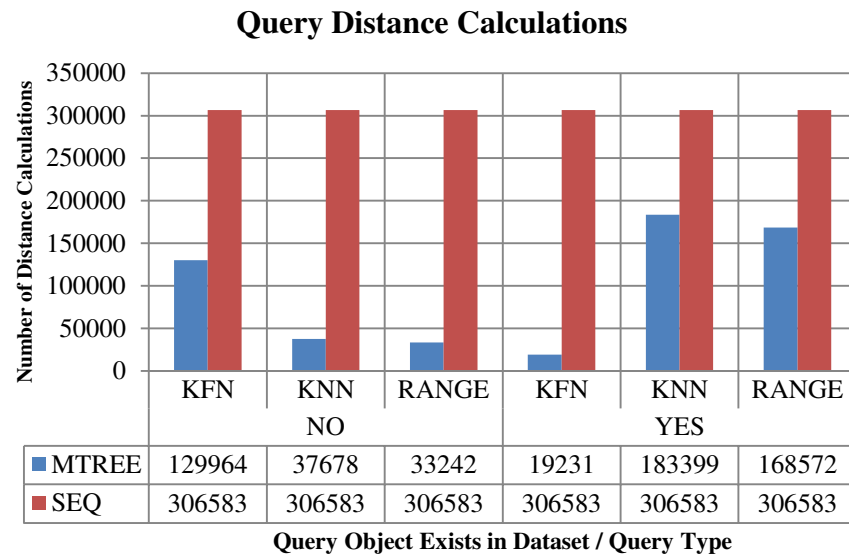


Figure 7.21 Number of distance calculations of M-Tree and SEQ.

Figure 7.21 includes total number of distance calculations performed of both SEQ and M-tree. According to the results, M-tree decreases the total number of results about 40% in the worst case. Similar to page access time results, query types show different performance characteristics in terms of query type and query object origin.

Figure 7.22 shows query execution times of M-tree and SEQ. Although M-tree reduces number of page access and distance calculation, some SEQ queries run faster than M-tree. Various factors could cause this difference. First, PostgreSQL uses a very efficient cache mechanism. Hence, SEQ queries that fit in memory executed faster. Furthermore, we implemented our system on top of PostgreSQL's query execution engine. So, SEQ could access tuple data faster than our system does. Additionally, our system is also using PostgreSQL's query execution engine. But our system has very good access times comparing to other systems (Bueno et al., 2002).

In sum, M-tree improves multidimensional access methods comparing to typical query execution. Although query execution times are worse than typical queries, we think that M-tree will outperform typical query in all metrics.

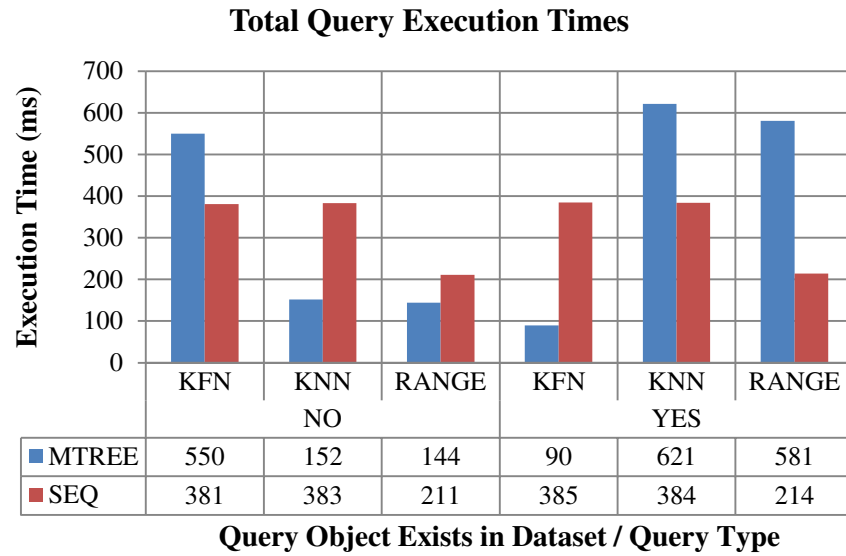


Figure 7.22 Total query execution times of MTree and SEQ.

7.5.2 Performance of Shape Property Queries

We used masses of DEMS dataset to measure image retrieval performance of our integration approach. First, we extract features of breast masses and stored them in a PostgreSQL DBMS using our extension. Then, we executed CBIR queries using each of the masses. As a result, we executed total number of 260 queries. Finally, we presented average of all performance queries.

Figure 7.23 includes PR graph of Zernike moments feature on shape property queries. According to the results, *Irregular* class has the best precision results followed by *Round* class. Precision levels of *Irregular* class vary between 56% and 37%.

Irregular class also obtains the best PR scores using other features. This result supports previous findings. Unlike previous experimental results, *Round* class obtains the second score. So we could say that our system performs better on the most benign and malign shape classes.

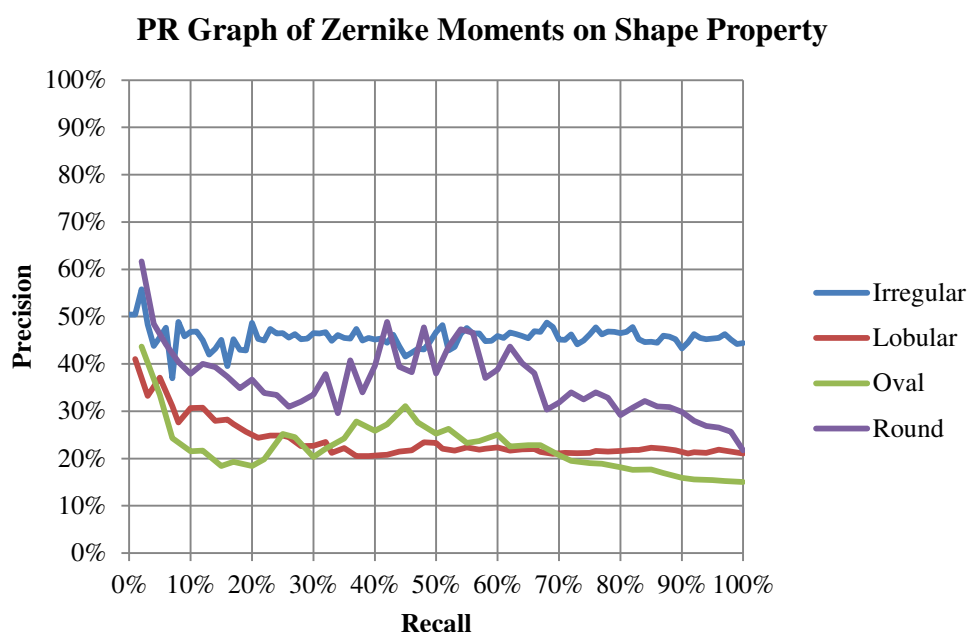


Figure 7.23 PR graph of zernike moments on shape property.

7.5.3 Performance of Margin Property Queries

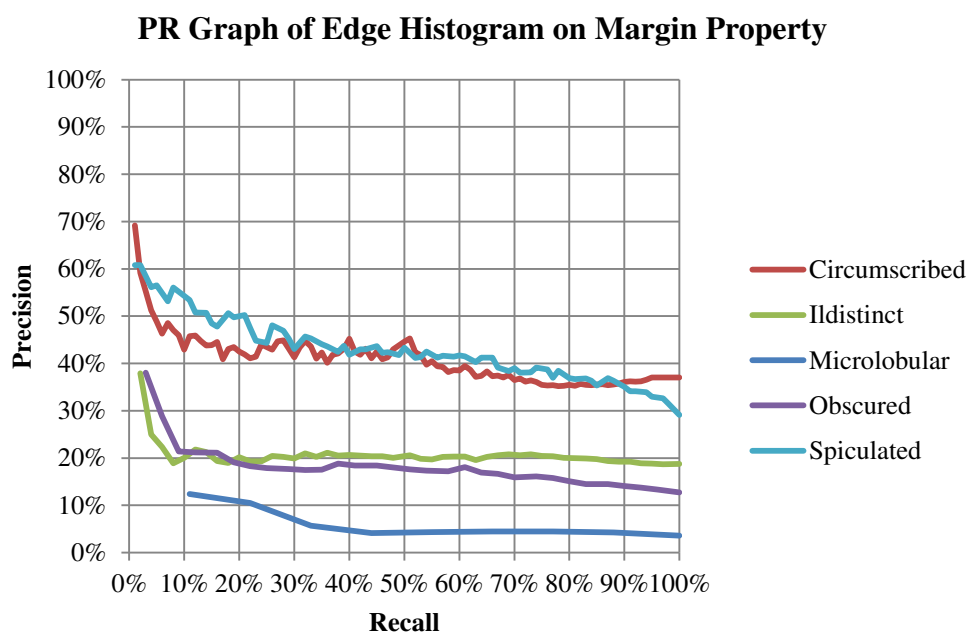


Figure 7.24 PR graph of edge histogram on margin property

Edge Histogram feature achieves the best precision recall scores of experiments on margin property among other features. Figure 7.24 includes PR graph of margin property experiments. According to the results, performance of *circumscribed* and

spiculated class is higher than other classes, like experimental results of low-level features. In sum, we could say that our system is also capable to identify most benign and malign margin classes.

7.5.4 Performance of Density Property Queries

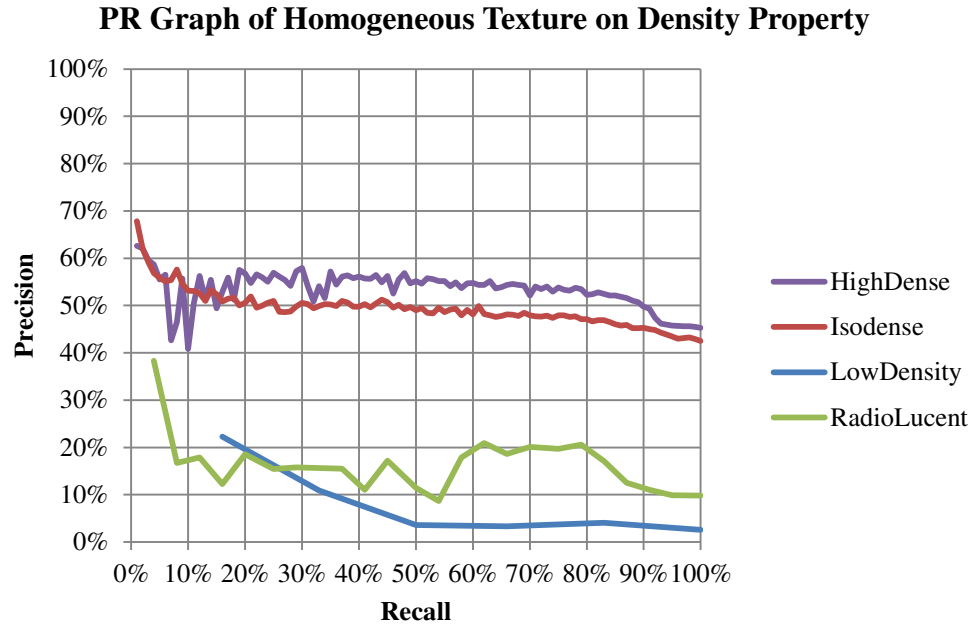


Figure 7.25 PR graph of homogeneous texture on density property.

Homogeneous texture is the best feature according to the experimental results. Figure 7.25 includes *precision recall* graph of homogeneous texture feature of queries on margin property. According to the results, *high-* and *iso-dense* classes obtain the best precision results. Moreover, these results are consistent with low-level feature comparison results.

7.5.5 Performance of BI-RADS Property Queries

Homogeneous texture property also obtains the best performance results for BI-RADS property queries. According to the low-level feature experiments, BI-RADS 5 is the best identified class. Figure 7.26 includes PR graph of homogeneous texture property. Performance of queries on BI-RADS 5 class obtains

the highest precision scores among other classes. This result is also supports previous findings.

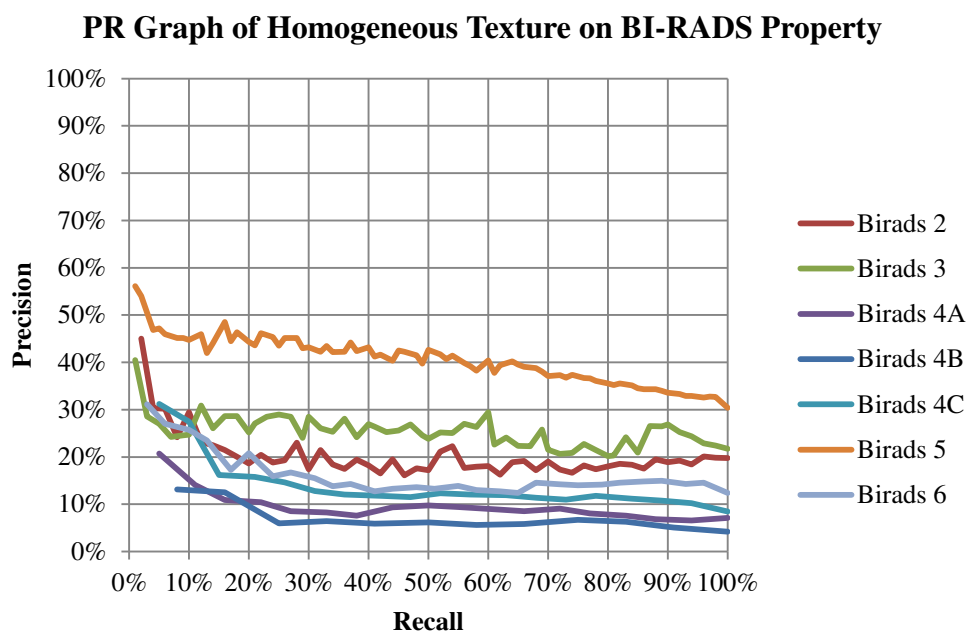


Figure 7.26 PR graph of homogeneous texture on BI-RADS property.

CHAPTER EIGHT

CONCLUSIONS

Aim of this thesis, in terms of broader perspective, is to integrate CBIR methods into traditional database management system. Therefore, large-scale image collections could be searched using sample image. However, in order to achieve this goal, we have several improvements.

This thesis presents a new mammogram dataset named DEMS to be used in computer aided diagnosis and interpretation of mammography. DEMS includes only FFDM images since they are gaining importance with technological advances on image acquisition devices and do not include artificial artifacts like digitized mammograms. Additionally, FFDM images provide both sufficient contrast rates and intensity depth for mammogram interpretation. Lesions in DEMS have well defined contour information. In other words, instead of marking the lesions by using minimum bounding rectangle, expert determined the boundary of the lesions, thoroughly. DEMS uses Mammography Annotation Ontology, and all annotations are in XML format compliant with the state-of-the-art semantic web technologies. Thus, semantic relationships in mammography could be easily investigated and discovered by using any inference engines and query tools like SPARQL and SQWRL. Hence, DEMS is a crucial dataset for intelligent CADx with semantic interpretation of mammograms. Moreover, it is easy to parse and convert DEMS Annotation XML files to different annotation formats and any further work on this area could use our ontology for annotation or semantic research.

DEMS dataset contains 485 mammography cases in total. In case selection stage, we considered all possible lesion types seen in mammographic examinations. As a result, DEMS contains 136 masses in 116 cases, 144 calcification in 119 cases, 20 special cases in 19 cases and 97 associated findings in 65 cases. We believe that many researchers will use DEMS as a reference set in their researches on computer-based detection and diagnosis of mammary lesions and hope to see it will help to close the semantic gap in interpretations of mammograms.

During development of DEMS, a new mammography annotation tool (MAT) based on our ontology is developed. MAT is developed using Qt, which is a cross platform C++ toolkit. Anyone using MAT could easily create a new dataset based on our ontology. Furthermore, MAT supports querying previously annotated cases using a sample one. We believe that MAT would be a fundamental tool for mammogram dataset developers and it would be a powerful information utility for mammography researchers.

We proposed a new breast mass contour segmentation algorithm, called breast mass contour segmentation (BMCS). The approach is an extended version of classical seeded region growing algorithm with additional capability of dynamic threshold adjustment and proper stopping conditions for the size of segments to compensate for under and over segmentation.

In order to evaluate BMCS, we first provide an extensive summary of segmentation evaluation methods available in literature, discussing descriptive powers of each evaluation metric individually, since segmentation accuracy or simple overlapping rate is not enough to express segmentation performance. The performance of BMCS is tested against the leading three region-based segmentation approaches, namely, level-set segmentation, seeded region growing segmentation and watershed segmentation. In order to perform a comparison, we have used a data set containing 260 masses. We have showed that BMCS mostly outperforms all the other methods tested and BMCS minimizes false positive pixels while obtaining high accuracy rates and least distance values to reference region. In other words, BMCS that is using adaptive threshold value and two stopping condition to prevent under and over segmentation, works well and improves the segmentation results in terms of several metrics.

Additionally, we evaluated individual performances of low-level image features on breast mass classification task on four different properties; shape, margin, density and BIRADS score. Each property, except mass density, is tested on two datasets DEMS and DDSM. The DDSM dataset is very common and publically available dataset including scanned images of analog film mammography. Unlike the most of

the studies based on DDSM dataset, we try to use the whole dataset instead of selecting a subset. However, we had to discard 42 cases from DDSM dataset due to corrupted images file and/or incorrect annotation data. A complete set of all level image features used in this study, both from DDSM and DEMS, are available to researchers for further improvement and comparison at <http://digimam.cs.deu.edu.tr/>. We believe that this data set of low level image features can be a reference for future studies related to breast mass classifications.

According to low-level feature performance evaluation results, we could say that shape and texture features are the best feature groups to be used to identify mass properties. Shape features identify shape and margin property of a mass better than other feature groups, while texture features identify density of a mass better than other feature groups. On the other hand BI-RADS property of a mass could be identified by using both texture and shape features. However, we can suggest no common features to be used in this task. Additionally, combination of the low-level features is still needed to be tested, since we only consider individual feature performances.

This thesis is also contains performance evaluation of experiments on mass property classification task on two different dataset: DDSM and DEMS. Images in two dataset differ from each other, where DDSM dataset contains digitized mammography films while DEMS dataset contains only digital mammography images. Furthermore, some of the mass annotations of DDSM dataset are missing while DEMS contains fully annotated mass regions according to latest BI-RADS standard. Hence, we had some troubles on comparing performances of both datasets. Particularly, DDSM mass annotations have no density property and have different BI-RADS score properties from DEMS annotations. Although there are such differences, experimentations showed that DEMS dataset is more suitable than DDSM for performance evaluation purposes.

We also measured the performance of manual region selections (MSR) and automatic segmented regions (ASR) on mass property classification task. We

observe that there is no significant difference between ASR and MSR region selection method in terms of all metrics.

During development of this thesis, we also aim to choose best machine learning algorithm for mass classification task. We found that Naïve Bayes classifier is the most successful classifier for mass property classification task, according to the experimental results.

Ultimate goal of this thesis is integration of content-based image retrieval methods and a relational database management system. We have implemented an extension for PostgreSQL database management system to accomplish this goal. Any multi-dimensional data could be efficiently searched and retrieved with the help of our extension. Furthermore, experimental results encourage us to improve our system by including new distance functions and low-level features etc.

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

- 1) We proposed nine new low-level image features for mammogram masses as well as extensive investigation of the existing features for CBIR performance.
- 2) We proposed a new segmentation algorithm for breast masses, called breast mass contour segmentation (BMCS), and showed that it increases segmentation accuracy of breast masses.
- 3) We developed a new fully annotated mammography dataset called DEMS and a tool for mammogram annotation and retrieval.
- 4) We integrated CBIR and DBMS together in PostgreSQL.

All in all, this thesis shows that a database management system could become an essential part of a content-based image retrieval system. But, we think that a database could become a complete content-based image retrieval system by integrating other CBIR tools like relevance feedback, low-level feature fusion, weighting etc.

REFERENCES

- Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641–647. doi:10.1109/34.295913
- Aisen, A. M., Broderick, L. S., Winer-Muram, H., Brodley, C. E., Kak, A. C., Pavlopoulou, C., Dy, J., et al. (2003). Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. *Radiology*, 228(1), 265–70. doi:10.1148/radiol.2281020126
- Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. *Journal of digital imaging: the official journal of the Society for Computer Applications in Radiology*, 24(2), 208–22. doi:10.1007/s10278-010-9290-9
- Akçay, O., Alpkoçak, A., Balcı, P., & Dicle, O. (2009). MuDMDs: A Multipurpose Digital Mammography Reference Dataset. *Proceedings of 6th National Medical Informatics Conference*. Antalya, Turkey.
- Antonie, M. L., Zaiane, O. R., & Coman, A. (2003). Associative classifiers for medical images. *Mining Multimedia and Complex Data*, 2797(2003), 68–83. doi:10.1007/978-3-540-39666-6_5
- 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* (Vol. 1, pp. 539–540). IEEE. doi:10.1109/IEMBS.1995.575239
- Atnafu, S., Chbeir, R., & Brunie, L. (2002). Content-based and metadata retrieval in medical image database. *Proceedings of 15th IEEE Symposium on Computer-*

- Based Medical Systems (CBMS 2002)* (pp. 327–332). IEEE Computer. Soc.
doi:10.1109/CBMS.2002.1011398
- Barclay, L. (2012). ACOG Statement: Begin Annual Mammograms at Age 40 Years.
- Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9), 509–517.
doi:10.1145/361002.361007
- Berchtold, S., Keim, D., & Kriegel, H. (1996). The X-tree: An index structure for high-dimensional data. *Proceedings of the 22nd VLDB Conference* (pp. 28–39). Mumbai (Bombay), India.
- Beucher, S., & Lantuejoul, C. (1979). Use of Watersheds in Contour Detection. *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*. Rennes, France.
- Bhadeshia, H. K. D. H. (1999). Neural networks in materials science. *ISIJ International*, 39(10), 966–979.
- Boninski, P., & Przelaskowski, A. (2008). Content-Based Indexing of Medical Images for Digital Radiology Applications. *Information Technologies in Biomedicine*, 47, 113–120. doi:10.1007/978-3-540-68168-7_11
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
doi:10.1023/A:1010933404324
- Bremner, D., Demaine, E., Erickson, J., Iacono, J., Langerman, S., Morin, P., & Toussaint, G. (2005). Output-Sensitive Algorithms for Computing Nearest-Neighbour Decision Boundaries. *Discrete & Computational Geometry*, 33(4), 593–604. doi:10.1007/s00454-004-1152-0

- Bueno, J. M., Chino, F., Traina, A. J. M., Traina, C., & Azevedo-Marques, P. M. (2002). How to add content-based image retrieval capability in a PACS. *Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002)* (pp. 321–326). IEEE Computer. Soc. doi:10.1109/CBMS.2002.1011397
- 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. doi:10.1016/j.eswa.2012.03.058
- Carson, C., Thomas, M., Belongie, S., Hellerstein, J., & Malik, J. (1999). Blobworld: A system for region-based image indexing and retrieval. *Visual Information and Information Systems* (pp. 509–516). Springer Berlin Heidelberg. doi:10.1007/3-540-48762-X_63
- Chalana, V., & Kim, Y. (1997). A methodology for evaluation of boundary detection algorithms on medical images. *IEEE transactions on medical imaging*, 16(5), 642–52. doi:10.1109/42.640755
- Cho, H., Hadjiiski, L., Sahiner, B., Chan, H.-P., Paramagul, C., Helvie, M., & Nees, A. V. (2012). Interactive content-based image retrieval (CBIR) computer-aided diagnosis (CADx) system for ultrasound breast masses using relevance feedback. In C. K. Abbey & C. R. Mello-Thoms (Eds.), *Proceedings of SPIE*. doi:10.1117/12.912164
- Chu, W. W., Cardenas, A. F., & Taira, R. K. (1998). Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge and Data Engineering*, 10(6), 872–888. doi:10.1109/69.738355
- Ciaccia, P., & Zezula, P. (1997). M-tree : An Efficient Access Method for Similarity Search in Metric Spaces. *Proceedings of the 23rd International Conference on Very Large* (pp. 426 – 435). San Francisco, CA: Morgan Kaufmann Publishers Inc.

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. doi:10.1007/BF00994018
- Dagan Feng, D., Fu, H., & Tian, G. (2008). Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms. *2008 International Conference on Technology and Applications in Biomedicine* (pp. 51–53). Ieee. doi:10.1109/ITAB.2008.4570523
- Davies, D. H. (1993). Digital mammography--the comparative evaluation of film digitizers. *British Journal of Radiology*, 66(790), 930–933. doi:10.1259/0007-1285-66-790-930
- 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), 1479–1491. doi:10.1016/j.compbiomed.2007.01.009
- DeParedes, E. S. (2007). *Atlas of Mammography* (3rd Editio.). Lippincott Williams & Wilkins.
- Deselaers, T., Keysers, D., & Ney, H. (2005). FIRE – Flexible Image Retrieval Engine: ImageCLEF 2004 Evaluation. In C. Peters, P. Clough, J. Gonzalo, G. F. Jones, M. Kluck, & B. Magnini (Eds.), *Multilingual Information Access for Text, Speech and Images* (Vol. 3491, pp. 688–698). Springer Berlin Heidelberg. doi:10.1007/11519645_67
- Dubey, R. B., Hanmandlu, M., & Gupta, S. K. (2010). A comparison of two methods for the segmentation of masses in the digital mammograms. *Computerized medical imaging and graphics*, 34(3), 185–91. doi:10.1016/j.compmedimag.2009.09.002
- D’Orsi, C. J., Bassett, L. W., & Berg, W. A. (2003). BI-RADS: Mammography. In C. J. D’Orsi, E. B. Mendelson, & D. M. Ikeda (Eds.), *Breast Imaging Reporting*

and Data System: ACR BI-RADS – Breast Imaging Atlas (4th ed.). Reston, VA: American College of Radiology.

- El-Naqa, I., Yang, Y., Galatsanos, N., Nishikawa, R., & Wernick, M. (2004). A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Transactions on Medical Imaging*, 23(10), 1233–1244.
- Elter, M., Held, C., & Wittenberg, T. (2010). Contour tracing for segmentation of mammographic masses. *Physics in medicine and biology*, 55(18), 5299–315. doi:10.1088/0031-9155/55/18/004
- Fan, C.-Y., Chang, P.-C., Lin, J.-J., & Hsieh, J. C. (2011). A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Applied Soft Computing*, 11(1), 632–644. doi:10.1016/j.asoc.2009.12.023
- Faruque, J. S., Rubin, D. L., Beaulieu, C. F., Rosenberg, J., Kamaya, A., Tye, G., Napel, S., et al. (2011). A Scalable Reference Standard of Visual Similarity for a Content-Based Image Retrieval System. *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, 158–165. doi:10.1109/HISB.2011.9
- Fernández López, M. (1999). Overview of methodologies for building ontologies. *Proceedings of the IJCAI99 Workshop on Ontologies and ProblemSolving Methods Lessons Learned and Future Trends* (pp. 1–13). Stockholm, Sweden.
- Georgiou, H., Mavroforakis, M., Dimitropoulos, N., Cavouras, D., & Theodoridis, S. (2007). Multi-scaled morphological features for the characterization of mammographic masses using statistical classification schemes. *Artificial Intelligence in Medicine*, 41(1), 39–55. doi:10.1016/j.artmed.2007.06.004

- Golobardes, E., Llorca, X., Salamó, M., & others. (2002). Computer aided diagnosis with case-based reasoning and genetic algorithms. *Knowledge-Based Systems*, 15(1-2), 45–52. doi:10.1016/S0950-7051(01)00120-4
- Greenspan, H., & Pinhas, A. T. (2007). Medical Image Categorization and Retrieval for PACS Using the GMM-KL Framework. *IEEE Transactions on Information Technology in Biomedicine*, 11(2), 190–202. doi:10.1109/TITB.2006.874191
- Guo, S., Sun, W., Deng, Y., & Li, W. (1994). Panther: an inexpensive and integrated multimedia environment. *Proceedings of IEEE International Conference on Multimedia Computing and Systems MMCS-94* (pp. 382–391). IEEE Comput. Soc. Press. doi:10.1109/MMCS.1994.292483
- Guttman, A. (1984). R-trees. *ACM SIGMOD Record*, 14(2), 47. doi:10.1145/971697.602266
- Göld, M. O., Thies, C., Fischer, B., & Lehmann, T. M. (2005). Combining global features for content-based retrieval of medical images. *Cross Language Evaluation Forum 2005 Workshop*. Citeseer.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), 610–621.
- Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., & Munishkumaran, S. (1998). Current status of the digital database for screening mammography. *Digital Mammography* (Vol. 13, pp. 457–460). Kluwer Academic Publishers.
- Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2001). The Digital Database for Screening Mammography. In M. J. Yaffe (Ed.), *Proceedings of the Fifth International Workshop on Digital Mammography* (pp. 212–218). Medical Physics Publishing.

- Ho, T. (2000). Complexity of Classification Problems and Comparative Advantages of Combined Classifiers. *Multiple Classifier Systems* (pp. 97–106). London, UK: Springer Berlin / Heidelberg. doi:10.1007/3-540-45014-9_9
- Hong, B.-W., & Sohn, B.-S. (2010). Segmentation of regions of interest in mammograms in a topographic approach. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 14(1), 129–39. doi:10.1109/TITB.2009.2033269
- Howarth, P., Yavlinsky, A., Heesch, D., & Rüger, S. (2004). Visual features for content-based medical image retrieval. *Proceedings of Cross Language Evaluation Forum (CLEF) Workshop*.
- Hu, M. K. (1962). Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, 8(2), 179–187.
- Imaginis. (2012). General Information on Mammography.
- ISO. (2003). Information Technology - Database Languages - SQL Multimedia and Application Packages - Part 5: Still Image. International Organization For Standardization.
- ISO. (2008). Information Technology - Database Languages - SQL - Part 1: Framework (SQL/Framework). International Organization For Standardization.
- Kelly, P. M., Cannon, T. M., & Hush, D. R. (1995). Query by image example: the comparison algorithm for navigating digital image databases (CANDID) approach. In S. A. Benton, M. H. Wu, R. Shashidhar, B. E. Rogowitz, S. S. Fisher, G. G. Grinstein, T. Wilson, et al. (Eds.), (pp. 238–248). doi:10.1117/12.205289
- Kherfi, M. L., & Ziou, D. (2006). Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples.

IEEE Transactions on Image Processing, 15(4), 1017–1030.
doi:10.1109/TIP.2005.863969

Khotanzad, A., & Hong, Y. (1990). Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(4), 489–497.

Kim, J. K., & Park, H. W. (1999). Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Transactions on Medical Imaging*, 18(3), 231–238. doi:10.1109/42.764896

Kinoshita, S. K., Azevedo-Marques, P. M. de, Pereira, R. R., 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), 172–90. doi:10.1007/s10278-007-9004-0

Kosch, H., Boszormenyi, L., Bachlechner, A., Hanin, C., Hofbauer, C., Lang, M., Riedler, C., et al. (2001). SMOOTH-A distributed multimedia database system. *Proceedings of the 27th International Conference on Very Large Data Bases* (pp. 713–714). Citeseer.

Lauria, A. (2009). GPCALMA: implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination. *Physica medica*, 25(2), 58–72. doi:10.1016/j.ejmp.2008.05.002

Lauria, A., Massafra, R., Tangaro, S., Bellotti, R., Fantacci, M. E., Delogu, P., Torres, E., et al. (2006). GPCALMA: An Italian Mammographic Database of Digitized Images for Research. (S. M. Astley, M. Brady, C. Rose, & R. Zwiggelaar, Eds.) *Digital Mammography*, 4046, 384–391. doi:10.1007/11783237_52

- Lehmann, T. M., Güld, M. O., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., et al. (2004). Content-based image retrieval in medical applications. *Methods of information in medicine*, 43(4), 354–61. doi:10.1267/METH04040354
- Liberman, L., & Menell, J. H. (2002). Breast imaging reporting and data system (BI-RADS). *Radiologic Clinics of North America*, 40(3), 409–430. doi:10.1016/S0033-8389(01)00017-3
- Long, L. R., Antani, S., Lee, D.-J., Krainak, D. M., & Thoma, G. R. (2003). Biomedical Information from a National Collection of Spine X-rays: Film to Content-based Retrieval. In R. L. Galloway, Jr., M. J. Yaffe, A. V. Clough, M. Sonka, D. P. Chakraborty, H. K. Huang, W. F. Walker, et al. (Eds.), (pp. 70–84). doi:10.1117/12.487798
- Ma, W. Y., & Manjunath, B. S. (1996). Texture features and learning similarity. *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 425–430). IEEE Comput. Soc. Press. doi:10.1109/CVPR.1996.517107
- Ma, W. Y., & Manjunath, B. S. (1997). NeTra: a toolbox for navigating large image databases. *Proceedings of International Conference on Image Processing* (Vol. 1, pp. 568–571). IEEE Comput. Soc. doi:10.1109/ICIP.1997.647976
- Manjunath, B. S., Salembier, P., & Sikora, T. (2002). *Introduction to MPEG 7: Multimedia Content Description Language*. (B. S. Manjunath, P. Salembier, & T. Sikora, Eds.). Wiley Interscience.
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: Rapid prototyping for complex data mining tasks. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 935–940). ACM.

- Mika, S., Ratsch, G., Weston, J., Scholkopf, B., & Mullers, K. (1999). Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX, Proceedings of the 1999 IEEE Signal Processing Society Workshop* (pp. 41–48). IEEE. doi:10.1109/NNSP.1999.788121
- Momjian, B. (2000). How PostgreSQL Rose to Fame. Retrieved November 23, 2012, from www.oreillynet.com/pub/a/network/2000/06/16/magazine/postgresql_history.html
- 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. doi:10.1016/j.acra.2011.09.014
- Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. E. (2000). Gradient and texture analysis for the classification of mammographic masses. *IEEE transactions on medical imaging*, 19(10), 1032–1043. doi:10.1109/42.887618
- 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), 1–23. doi:10.1016/j.ijmedinf.2003.11.024
- 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*, 28(6), 295–305. doi:10.1016/j.compmedimag.2004.04.005
- Müller, H., Ruch, P., & Geissbuhler, A. (2004). Enriching content-based medical image retrieval with automatically extracted MeSH-terms. doi:10.1.1.66.5309
- Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D., Yanker, P., et al. (1993). QBIC project: querying images by content, using color, texture, and shape. In M. M. Blouke, H. Marz, E. R. Dougherty, M.

- Rabbani, L. A. Ray, R. S. Acharya, F. Y. Wu, et al. (Eds.), *Proceedings of SPIE* (Vol. 1908, pp. 173–187). doi:10.1117/12.143648
- Nishikawa, R. M. (1998). Mammographic databases. *Breast disease*, 10(3-4), 137–50.
- Odet, C., Belaroussi, B., & Benoit-Cattin, H. (2002). Scalable discrepancy measures for segmentation evaluation. *Proceedings. International Conference on Image Processing* (Vol. 1, pp. I–785–I–788). IEEE. doi:10.1109/ICIP.2002.1038142
- Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987. doi:10.1109/TPAMI.2002.1017623
- Ojala, Timo, Mäenpää, T., Viertola, J., Kyllönen, J., & M. (2002). Empirical evaluation of MPEG-7 texture descriptors with a large-scale experiment. *Proceedings of 2nd International Workshop on Texture Analysis and Synthesis* (Vol. 2, pp. 99–102).
- Ojala, Timo, Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1), 51–59. doi:10.1016/0031-3203(95)00067-4
- Oliver, A., Freixenet, J., Martí, J., Pérez, E., Pont, J., Denton, E. R. E., & Zwiggelaar, R. (2010). A review of automatic mass detection and segmentation in mammographic images. *Medical image analysis*, 14(2), 87–110. doi:10.1016/j.media.2009.12.005
- Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1), 12–49. doi:10.1016/0021-9991(88)90002-2

- Park, D. K., Jeon, Y. S., & Won, C. S. (2000). Efficient use of local edge histogram descriptor. *Proceedings of the 2000 ACM workshops on Multimedia* (p. 54). New York, New York, USA: ACM. doi:10.1145/357744.357758
- Patella, M. (1999). *Similarity search in multimedia databases*. Dipartimento di Elettronica Informatica e Sistemistica, Bologna. Università degli Studi di Bologna.
- Pelski, S. (2005). *Oracle interMedia User's Guide* (Vol. 2). Oracle.
- Peng, Y., Yao, B., & Jiang, J. (2006). Knowledge-discovery incorporated evolutionary search for microcalcification detection in breast cancer diagnosis. *Artificial Intelligence in Medicine*, 37(1), 43–53. doi:10.1016/j.artmed.2005.09.001
- Pentland, A., Picard, R. W., & Sclaroff, S. (1996). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3), 233–254. doi:10.1007/BF00123143
- Persoon, E., & Fu, K.-S. (1977). Shape Discrimination Using Fourier Descriptors. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(3), 170–179. doi:10.1109/TSMC.1977.4309681
- Pourghassem, H., & Ghassemian, H. (2008). Content-based medical image classification using a new hierarchical merging scheme. *Computerized Medical Imaging and Graphics*, 32(8), 651–661.
- Prinzie, A., & Van den Poel, D. (2007). Random Multiclass Classification: Generalizing Random Forests to Random MNL and Random NB. *Database and Expert Systems Applications*, 4653, 349–358. doi:10.1007/978-3-540-74469-6_35
- Rahman, M. M., Bhattacharya, P., & Desai, B. C. (2007). A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching

- Techniques With Relevance Feedback. *IEEE Transactions on Information Technology in Biomedicine*, 11(1), 58–69. doi:10.1109/TITB.2006.884364
- Ramamurthy, B., Chandran, K. R., Meenakshi, V. R., & Shilpa, V. (2012). CBMIR: Content Based Medical Image Retrieval System Using Texture and Intensity for Dental Images. In J. Mathew, P. Patra, D. K. Pradhan, & A. J. Kuttyamma (Eds.), *Proceedings of ICECCS 2012* (pp. 125–134). Springer Berlin Heidelberg. doi:10.1007/978-3-642-32112-2_16
- 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. doi:10.1007/BF02345742
- Rangayyan, Rangaraj M, Nguyen, T. M., Ayres, F. J., & Nandi, A. K. (2010). Effect of Pixel Resolution on Texture Features of Breast Masses in Mammograms. *Journal of Digital Imaging*, 23(5), 547–553. doi:10.1007/s10278-009-9238-0
- Rangayyan, Rangaraj M., & Nguyen, T. M. (2005). Pattern classification of breast masses via fractal analysis of their contours. *International Congress Series*, 1281, 1041–1046. doi:10.1016/j.ics.2005.03.329
- Ricard, J., Coeurjolly, D., & Baskurt, A. (2005). Generalizations of angular radial transform for 2D and 3D shape retrieval. *Pattern Recognition Letters*, 26(14), 2174–2186. doi:10.1016/j.patrec.2005.03.030
- Ro, Y. M. R., Kim, M. K., Kang, H. K. K., Manjunath, B. S. M., & Kim, J. K. (2001). MPEG-7 Homogeneous Texture Descriptor. *ETRI Journal*, 23(2), 41–51. doi:10.4218/etrij.01.0101.0201
- Rojas-Domínguez, A., & Nandi, A. K. (2008). Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection. *Computerized medical imaging and graphics: the official*

- journal of the Computerized Medical Imaging Society*, 32(4), 304–15.
doi:10.1016/j.compmedimag.2008.01.006
- Rojas-Domínguez, A., & Nandi, A. K. (2009). Development of tolerant features for characterization of masses in mammograms. *Computers in biology and medicine*, 39(8), 678–88. doi:10.1016/j.compbimed.2009.05.002
- Rosa, B., Mozer, P., & Szewczyk, J. (2011). An algorithm for calculi segmentation on ureteroscopic images. *International journal of computer assisted radiology and surgery*, 6(2), 237–46. doi:10.1007/s11548-010-0504-x
- Rosa, N. A., 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. *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (Vol. 2008, pp. 406–409). Vancouver, BC. doi:10.1109/IEMBS.2008.4649176
- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380. doi:10.1109/34.895972
- Smith, P., Reid, D. B., Environment, C., Palo, L., Alto, P., & Smith, P. L. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. doi:10.1109/TSMC.1979.4310076
- Song, E., Jiang, L., Jin, R., Zhang, L., Yuan, Y., & Li, Q. (2009). Breast mass segmentation in mammography using plane fitting and dynamic programming. *Academic radiology*, 16(7), 826–35. doi:10.1016/j.acra.2008.11.014
- Song, E., Xu, S., Xu, X., Zeng, J., Lan, Y., Zhang, S., & Hung, C.-C. (2010). Hybrid segmentation of mass in mammograms using template matching and dynamic

programming. *Academic radiology*, 17(11), 1414–24.
doi:10.1016/j.acra.2010.07.008

Spelic, D., Kaczmarek, R., Hilohi, M., & Belella, S. (2007). United States radiological health activities: inspection results of mammography facilities. *Biomedical imaging and intervention journal*, 3(2), e35. doi:10.2349/bij.3.2.e35

Stolze, K. (2005). A DB2 UDB still image extender. Retrieved November 22, 2012, from <http://www.ibm.com/developerworks/data/library/techarticle/dm-0504stolze/>

Stricker, M., & Orengo, M. (1995). Similarity of color images. *Proc. SPIE Storage and Retrieval for Image and Video Databases III* (Vol. 2420, pp. 381–392). Citeseer.

Subashini, T. S., Ramalingam, V., & Palanivel, S. (2010). Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, 114(1), 33–43. doi:10.1016/j.cviu.2009.09.009

Suckling, J., Parker, J., Dance, D. R., Astley, S., Hutt, I., Boggis, C., Ricketts, I., et al. (1994). The Mammographic Image Analysis Society Digital Mammogram Database. In A. G. Gale, S. M. Astley, D. R. Dance, & A. Y. Cairns (Eds.), *Proceedings of 2nd International Workshop on Digital Mammography* (pp. 375–378). Amsterdam.

Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I., & Yang, Y. (2009). Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *IEEE Transactions on Information Technology in Biomedicine*, 13(2), 236–251. doi:10.1109/TITB.2008.2009441

Tao, Y., Lo, S.-C. B., Freedman, M. T., Makariou, E., & Xuan, J. (2010). Multilevel learning-based segmentation of ill-defined and spiculated masses in mammograms. *Medical Physics*, 37(11), 5993. doi:10.1118/1.3490477

- Traina, C., Traina, A. J. M., Araújo, M. R. B., Bueno, J. M., Chino, F. J. T., Razente, H., & Azevedo-Marques, P. M. (2005). Using an image-extended relational database to support content-based image retrieval in a PACS. *Computer methods and programs in biomedicine*, 80 Suppl 1, S71–83.
- Traina Jr., C., Traina, A., Seeger, B., & Faloutsos, C. (2000). Slim-Trees: High Performance Metric Trees Minimizing Overlap between Nodes. In C. Zaniolo, P. Lockemann, M. Scholl, & T. Grust (Eds.), *Advances in Database Technology* (Vol. 1777, pp. 51–65). Springer Berlin Heidelberg. doi:10.1007/3-540-46439-5_4
- Tsishkou, D. V., Kukharchik, P. D., Bovbel, E. J., Kheidorov, I. E., & Liventseva, M. M. (2003). Boosting biomedical images indexing. *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering*, 2003. (pp. 74–75). IEEE. doi:10.1109/APBME.2003.1302590
- Unnikrishnan, R., Pantofaru, C., & Hebert, M. (2007). Toward objective evaluation of image segmentation algorithms. *IEEE transactions on pattern analysis and machine intelligence*, 29(6), 929–44. doi:10.1109/TPAMI.2007.1046
- USF Digital Mammography: “Other Resources” Page. (n.d.). Retrieved from <http://marathon.csee.usf.edu/Mammography/OtherResources.html>
- Varela, C., Timp, S., & Karssemeijer, N. (2006). Use of border information in the classification of mammographic masses. *Physics in medicine and biology*, 51(2), 425–441. doi:10.1088/0031-9155/51/2/016
- 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. doi:10.1016/j.eswa.2009.10.016

- Vieira, M. R., Jr, C. T., Chino, F. J. T., & Traina, A. J. M. (2004). DBM-Tree : A Dynamic Metric Access Method Sensitive to Local Density Data. *SBB D* (pp. 163–177). doi:10.1.1.94.552
- Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), 583–598. doi:10.1109/34.87344
- Wang, J. Z., & Krovetz, R. (2005). CLUE: cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing*, 14(8), 1187–1201. doi:10.1109/TIP.2005.849770
- Wang, X.-H., Park, S. C., & Zheng, B. (2009). Improving performance of content-based image retrieval schemes in searching for similar breast mass regions: an assessment. *Physics in Medicine and Biology*, 54(4), 949–961. doi:10.1088/0031-9155/54/4/009
- Wei, C.-H., Chen, S. Y., & Liu, X. (2012). Mammogram retrieval on similar mass lesions. *Computer methods and programs in biomedicine*, 106(3), 234–48. doi:10.1016/j.cmpb.2010.09.002
- Wei, C.-H., Li, C.-T., & Wilson, R. (2009). A Content-Based Approach to Medical Image Database Retrieval. In J. Erickson (Ed.), *Database Technologies* (pp. 1062–1083). IGI Global. doi:10.4018/978-1-60566-058-5.ch062
- Weszka, J. S. (1978). Threshold Evaluation Techniques. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(8), 622–629. doi:10.1109/TSMC.1978.4310038
- Wikipedia. (2012). k-d Tree. Retrieved from http://en.wikipedia.org/wiki/K-d_tree
- 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. doi:10.1016/S0923-5965(00)00016-3

- Xu, J., Faruque, J., Beaulieu, C. F., Rubin, D., & Napel, S. (2012). A comprehensive descriptor of shape: method and application to content-based retrieval of similar appearing lesions in medical images. *Journal of digital imaging: the official journal of the Society for Computer Applications in Radiology*, 25(1), 121–8. doi:10.1007/s10278-011-9388-8
- Yang, L., & Albrechtsen, F. (1994). Fast computation of invariant geometric moments: a new method giving correct results. *Proceedings of 12th International Conference on Pattern Recognition* (pp. 201–204). IEEE Comput. Soc. Press. doi:10.1109/ICPR.1994.576257
- Yang, W., Feng, Q., Lu, Z., & Chen, W. (2011). Metric learning for maximizing MAP and its application to content-based medical image retrieval. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 1901–1904). IEEE. doi:10.1109/ISBI.2011.5872780
- Yasnoff, W. A., Mui, J. K., & Bacus, J. W. (1977). Error measures for scene segmentation. *Pattern Recognition*, 9(4), 217–231. doi:10.1016/0031-3203(77)90006-1
- Yianilos, P. N. (1993). Data Structures and Algorithms for Nearest Neighbor Search in General Metric Spaces. *Proceeding SODA '93 Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms* (pp. 311–321).
- Yin, D., Pan, J., Chen, P., & Zhang, R. (2008). Medical Image Categorization Based on Gaussian Mixture Model. *2008 International Conference on BioMedical Engineering and Informatics*, 128–131. doi:10.1109/BMEI.2008.210
- Yu, S.-N., & Huang, Y.-K. (2010). Detection of microcalcifications in digital mammograms using combined model-based and statistical textural features. *Expert Systems with Applications*, 37(7), 5461–5469. doi:10.1016/j.eswa.2010.02.066

- Zhang, Harry. (2004). The Optimality of Naive Bayes. *Proceedings of the 17th International FLAIRS conference* (Vol. 1). Menlo Park, CA: AAAI Press.
- Zhang, Hui, Fritts, J. E., & Goldman, S. A. (2008). Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2), 260–280. doi:10.1016/j.cviu.2007.08.003
- Zheng, B. (2009). Computer-Aided Diagnosis in Mammography Using Content-Based Image Retrieval Approaches: Current Status and Future Perspectives. *Algorithms*, 2(2), 828–849. doi:10.3390/a2020828.Computer-Aided

APPENDICES

A Low-Level Feature Performance Results.

A.1 Low-Level Feature Performance Results of Shape Property

A.1.1 Low-Level Feature Performances of N/A Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-CFD</i>	99,79%	13,33%	87,50%	65,39%
<i>S-INM</i>	99,79%	3,33%	98,32%	62,65%
<i>S-RDD</i>	99,79%	3,33%	96,51%	62,21%
<i>T-GLC</i>	99,79%	13,33%	91,97%	59,42%
<i>M-RMD</i>	99,79%	15,00%	85,14%	57,55%
<i>M-CSD</i>	99,79%	10,00%	88,22%	57,16%
<i>I-GEN</i>	99,79%	3,33%	97,73%	56,81%
<i>M-RSD</i>	99,79%	20,00%	81,33%	56,54%
<i>T-HIS</i>	99,79%	6,67%	95,43%	56,42%
<i>T-GLD</i>	99,79%	3,33%	97,55%	54,37%
<i>M-CMD</i>	99,79%	13,33%	86,77%	52,50%
<i>S-FDE</i>	99,79%	5,00%	95,39%	52,49%
<i>M-CWM</i>	99,79%	0,00%	99,47%	50,00%
<i>T-EDH</i>	99,79%	0,00%	99,75%	50,00%
<i>S-ZER</i>	99,79%	1,67%	95,19%	50,00%
<i>S-RDF</i>	99,79%	0,00%	99,88%	50,00%
<i>S-GEN</i>	99,79%	0,00%	99,49%	50,00%
<i>M-CWK</i>	99,79%	0,00%	99,06%	50,00%
<i>T-HOT</i>	99,79%	0,00%	96,50%	50,00%
<i>M-CWS</i>	99,79%	0,00%	99,54%	50,00%
<i>M-CWW</i>	99,79%	13,33%	81,98%	50,00%
<i>M-GLS</i>	99,79%	0,00%	98,47%	50,00%
<i>T-LBP</i>	99,79%	0,00%	97,02%	50,00%
<i>T-TEB</i>	99,79%	0,00%	96,14%	50,00%
<i>S-RBS</i>	99,79%	0,00%	99,58%	50,00%
<i>S-DFD</i>	99,79%	0,00%	99,96%	50,00%

A.1.2 Low-Level Feature Performances of Round Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RBS</i>	90,77%	28,34%	93,28%	85,81%
<i>S-INM</i>	90,58%	30,37%	89,37%	85,10%
<i>S-DFD</i>	90,58%	29,46%	86,88%	83,24%
<i>S-ZER</i>	90,63%	33,08%	85,55%	83,14%
<i>S-GEN</i>	90,55%	31,41%	87,62%	82,31%
<i>S-CFD</i>	90,58%	13,79%	93,43%	79,90%
<i>S-FDE</i>	90,58%	26,72%	90,63%	79,48%
<i>S-RDD</i>	90,58%	24,08%	95,62%	77,90%
<i>S-RDF</i>	90,58%	19,47%	92,98%	74,14%
<i>T-GLD</i>	90,55%	23,11%	88,97%	70,67%
<i>T-HIS</i>	90,55%	28,40%	88,30%	69,38%
<i>T-EDH</i>	90,63%	17,33%	92,04%	68,55%
<i>T-LBP</i>	90,55%	13,05%	94,88%	68,27%
<i>T-HOT</i>	90,55%	25,63%	87,59%	67,75%
<i>T-GLC</i>	90,55%	16,32%	93,23%	66,15%
<i>T-TEB</i>	90,55%	15,43%	93,34%	64,84%
<i>M-GLS</i>	90,55%	11,71%	93,11%	63,59%
<i>M-RSD</i>	90,55%	16,91%	86,78%	62,87%
<i>M-RMD</i>	90,55%	13,65%	87,57%	55,66%
<i>M-CSD</i>	90,55%	16,46%	84,40%	55,42%
<i>M-CMD</i>	90,55%	10,17%	90,10%	55,38%
<i>M-CWM</i>	90,55%	8,67%	93,05%	55,38%
<i>M-CWS</i>	90,63%	7,91%	93,76%	54,55%
<i>M-CWK</i>	90,55%	9,18%	90,19%	53,48%
<i>I-GEN</i>	90,55%	3,40%	97,19%	53,07%
<i>M-CWW</i>	90,55%	11,92%	87,57%	51,14%

A.1.3 Low-Level Feature Performances of Oval Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RDF</i>	87,69%	28,35%	87,50%	80,91%
<i>S-DFD</i>	87,31%	23,41%	86,47%	75,68%
<i>S-RBS</i>	86,54%	22,48%	88,02%	71,93%
<i>S-FDE</i>	85,00%	15,36%	89,54%	71,70%
<i>S-RDD</i>	84,62%	15,53%	92,08%	71,48%
<i>S-GEN</i>	84,88%	22,22%	86,13%	68,93%
<i>I-GEN</i>	84,50%	21,71%	85,25%	64,37%
<i>S-INM</i>	85,00%	15,49%	88,76%	63,86%
<i>T-HIS</i>	85,27%	17,77%	91,98%	62,02%
<i>M-GLS</i>	84,50%	22,24%	86,17%	61,00%
<i>S-ZER</i>	84,88%	11,20%	91,24%	60,80%
<i>T-HOT</i>	84,50%	16,61%	90,92%	60,31%
<i>T-GLD</i>	84,50%	16,49%	90,81%	59,60%
<i>M-CWS</i>	84,50%	10,71%	92,40%	59,50%
<i>M-RMD</i>	84,50%	8,30%	94,82%	58,96%
<i>M-CSD</i>	84,50%	5,63%	94,18%	58,91%
<i>S-CFD</i>	85,00%	11,45%	90,44%	58,41%
<i>M-CMD</i>	84,50%	8,64%	94,51%	58,27%
<i>T-EDH</i>	84,50%	15,74%	88,39%	57,81%
<i>T-GLC</i>	84,50%	15,61%	91,19%	57,58%
<i>M-CWK</i>	84,50%	9,10%	92,45%	57,20%
<i>M-RSD</i>	84,50%	6,19%	94,38%	54,14%
<i>T-LBP</i>	84,50%	13,35%	90,87%	54,08%
<i>M-CWM</i>	84,50%	9,58%	91,64%	53,50%
<i>M-CWW</i>	84,50%	8,93%	90,79%	53,25%
<i>T-TEB</i>	84,50%	4,68%	96,49%	52,79%

A.1.4 Low-Level Feature Performances of Lobular Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RBS</i>	78,46%	17,76%	88,33%	71,34%
<i>S-DFD</i>	78,29%	17,52%	86,68%	68,22%
<i>S-GEN</i>	79,07%	14,69%	90,92%	65,89%
<i>T-HOT</i>	78,29%	16,38%	90,62%	64,80%
<i>S-RDD</i>	78,29%	16,71%	86,37%	63,25%
<i>M-CMD</i>	78,29%	13,76%	89,26%	62,49%
<i>S-INM</i>	78,29%	12,67%	90,49%	62,08%
<i>M-RMD</i>	78,29%	17,18%	86,65%	61,42%
<i>T-HIS</i>	78,29%	16,53%	86,83%	60,83%
<i>S-FDE</i>	78,29%	14,62%	88,29%	60,11%
<i>T-LBP</i>	78,29%	15,00%	90,19%	59,14%
<i>T-GLC</i>	78,29%	14,98%	88,53%	58,65%
<i>M-GLS</i>	78,29%	13,11%	91,11%	58,55%
<i>M-CWM</i>	78,29%	18,54%	84,42%	58,55%
<i>M-CSD</i>	78,29%	10,03%	91,29%	58,16%
<i>S-ZER</i>	78,29%	11,83%	90,83%	57,20%
<i>S-RDF</i>	78,29%	11,41%	89,20%	56,78%
<i>T-TEB</i>	78,29%	9,48%	92,66%	56,26%
<i>I-GEN</i>	78,29%	6,80%	93,91%	55,52%
<i>M-CWS</i>	78,29%	14,70%	85,98%	55,28%
<i>T-EDH</i>	78,29%	16,07%	87,57%	55,18%
<i>T-GLD</i>	78,29%	12,50%	89,77%	55,17%
<i>M-RSD</i>	78,29%	8,39%	91,65%	54,78%
<i>M-CWK</i>	78,68%	16,63%	85,21%	54,74%
<i>M-CWW</i>	78,29%	5,70%	93,95%	54,52%
<i>S-CFD</i>	78,29%	10,79%	89,80%	53,88%

A.1.5 Low-Level Feature Performances of Irregular Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-GEN</i>	82,95%	69,81%	50,87%	82,54%
<i>T-EDH</i>	81,78%	74,14%	42,67%	81,62%
<i>S-RBS</i>	82,31%	73,35%	46,45%	81,08%
<i>S-ZER</i>	79,85%	70,08%	46,47%	79,59%
<i>S-FDE</i>	78,46%	71,60%	46,35%	79,38%
<i>S-INM</i>	78,08%	72,67%	45,10%	78,67%
<i>S-DFD</i>	79,23%	64,37%	53,40%	77,54%
<i>T-HOT</i>	74,81%	72,36%	45,91%	74,62%
<i>S-RDF</i>	71,92%	73,01%	42,13%	73,64%
<i>S-RDD</i>	72,31%	75,41%	40,09%	73,46%
<i>T-GLD</i>	69,77%	71,77%	41,43%	71,42%
<i>S-CFD</i>	68,46%	69,06%	37,03%	70,07%
<i>T-HIS</i>	68,99%	69,40%	48,43%	69,69%
<i>M-GLS</i>	67,83%	73,45%	39,03%	69,39%
<i>T-GLC</i>	67,44%	73,82%	41,38%	68,22%
<i>T-LBP</i>	60,85%	78,77%	34,83%	64,27%
<i>M-RSD</i>	61,24%	64,26%	38,62%	62,76%
<i>M-CWM</i>	62,79%	70,27%	34,31%	60,89%
<i>M-CWS</i>	58,14%	73,19%	30,83%	60,83%
<i>M-RMD</i>	58,81%	62,07%	41,46%	60,13%
<i>I-GEN</i>	60,55%	76,41%	28,12%	58,70%
<i>T-TEB</i>	58,14%	81,52%	24,35%	58,23%
<i>M-CSD</i>	59,43%	64,08%	36,77%	58,06%
<i>M-CWK</i>	58,23%	66,97%	33,23%	55,57%
<i>M-CWW</i>	58,77%	63,74%	36,65%	55,44%
<i>M-CMD</i>	58,64%	66,34%	34,25%	53,95%

A.2 Low-Level Feature Performance Results of Margin Property

A.2.1 Low-Level Feature Performances of N/A Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-LBP</i>	96,16%	14,29%	91,63%	62,71%
<i>S-GEN</i>	96,16%	13,89%	91,66%	59,70%
<i>M-CWK</i>	96,16%	7,22%	96,19%	58,14%
<i>S-RBS</i>	96,23%	6,83%	97,55%	57,94%
<i>M-CWM</i>	96,27%	4,21%	98,47%	57,01%
<i>S-INM</i>	96,17%	8,10%	96,91%	56,76%
<i>S-ZER</i>	96,30%	5,56%	98,25%	56,75%
<i>T-GLC</i>	96,16%	20,48%	85,34%	56,10%
<i>S-RDD</i>	96,17%	6,98%	98,71%	55,79%
<i>S-DFD</i>	96,17%	5,40%	98,38%	55,48%
<i>S-RDF</i>	96,17%	4,84%	98,60%	55,04%
<i>M-GLS</i>	96,16%	8,33%	93,90%	54,86%
<i>T-HOT</i>	96,16%	4,76%	98,86%	54,59%
<i>T-HIS</i>	96,16%	3,97%	98,95%	54,48%
<i>T-EDH</i>	96,23%	6,19%	98,21%	54,34%
<i>S-FDE</i>	96,17%	3,89%	98,18%	53,80%
<i>T-GLD</i>	96,16%	4,60%	98,49%	53,51%
<i>T-TEB</i>	96,16%	5,08%	95,66%	53,20%
<i>M-CWS</i>	96,27%	3,49%	98,01%	52,73%
<i>S-CFD</i>	96,19%	4,60%	96,07%	52,69%
<i>M-RSD</i>	96,16%	13,73%	86,59%	52,46%
<i>M-RMD</i>	96,16%	14,92%	85,65%	52,12%
<i>M-CSD</i>	96,16%	12,70%	87,99%	51,98%
<i>M-CMD</i>	96,16%	11,67%	89,29%	51,80%
<i>M-CWW</i>	96,16%	15,63%	84,54%	51,11%
<i>I-GEN</i>	96,16%	3,97%	96,31%	51,01%

A.2.2 Low-Level Feature Performances of Circumscribed Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-DFD</i>	78,85%	52,53%	60,80%	79,80%
<i>S-RBS</i>	79,62%	45,77%	71,25%	78,60%
<i>S-FDE</i>	77,45%	43,83%	74,49%	76,22%
<i>S-RDF</i>	77,45%	50,66%	66,00%	75,60%
<i>S-GEN</i>	77,39%	48,80%	71,66%	75,14%
<i>S-INM</i>	77,45%	42,87%	68,63%	74,86%
<i>S-RDD</i>	77,45%	39,15%	78,21%	74,26%
<i>T-EDH</i>	77,46%	46,35%	70,27%	74,15%
<i>S-ZER</i>	77,46%	41,74%	73,56%	73,66%
<i>T-HIS</i>	77,39%	49,50%	71,13%	70,19%
<i>T-HOT</i>	77,39%	46,82%	75,85%	69,90%
<i>S-CFD</i>	77,53%	42,19%	63,58%	68,83%
<i>T-GLD</i>	77,39%	42,55%	76,94%	68,67%
<i>T-GLC</i>	77,39%	42,98%	73,18%	68,31%
<i>M-RSD</i>	77,39%	38,81%	64,77%	64,42%
<i>T-LBP</i>	77,39%	44,63%	70,48%	64,08%
<i>I-GEN</i>	77,39%	49,62%	57,80%	63,39%
<i>T-TEB</i>	77,39%	32,99%	73,63%	61,96%
<i>M-GLS</i>	77,39%	46,55%	62,90%	60,81%
<i>M-CWS</i>	77,39%	44,25%	58,80%	60,32%
<i>M-CWM</i>	77,39%	43,77%	59,96%	59,69%
<i>M-CSD</i>	77,39%	41,89%	60,54%	59,30%
<i>M-RMD</i>	77,39%	34,24%	68,26%	58,46%
<i>M-CWK</i>	77,39%	42,13%	58,98%	56,71%
<i>M-CWW</i>	77,39%	36,42%	65,16%	54,71%
<i>M-CMD</i>	77,39%	33,43%	66,91%	52,77%

A.2.3 Low-Level Feature Performances of Microlobular Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RDD</i>	96,54%	4,64%	97,81%	82,13%
<i>S-FDE</i>	96,54%	9,52%	95,04%	77,56%
<i>S-GEN</i>	96,51%	7,78%	97,45%	74,90%
<i>T-HIS</i>	96,51%	16,50%	93,34%	73,16%
<i>S-INM</i>	96,54%	5,03%	96,57%	66,25%
<i>S-CFD</i>	96,54%	15,48%	88,48%	65,27%
<i>S-ZER</i>	96,54%	8,54%	94,11%	64,79%
<i>S-RBS</i>	96,54%	2,98%	97,91%	59,32%
<i>M-CWW</i>	96,51%	7,30%	92,21%	59,10%
<i>M-CWK</i>	96,51%	5,08%	96,37%	57,97%
<i>M-GLS</i>	96,51%	2,83%	98,29%	57,50%
<i>T-GLD</i>	96,12%	9,79%	91,40%	56,63%
<i>T-TEB</i>	96,51%	10,44%	90,62%	55,89%
<i>S-DFD</i>	96,90%	3,75%	96,74%	55,56%
<i>T-HOT</i>	96,51%	13,02%	90,60%	55,27%
<i>M-RSD</i>	96,51%	9,72%	88,99%	55,15%
<i>T-GLC</i>	96,51%	5,37%	96,05%	54,82%
<i>M-RMD</i>	96,51%	10,16%	89,94%	54,48%
<i>M-CWM</i>	96,51%	3,52%	96,63%	54,35%
<i>T-LBP</i>	96,51%	3,81%	96,83%	53,68%
<i>S-RDF</i>	96,54%	1,25%	98,47%	53,38%
<i>T-EDH</i>	96,51%	5,76%	92,99%	52,49%
<i>M-CMD</i>	96,51%	8,35%	91,51%	51,87%
<i>M-CSD</i>	96,51%	7,15%	91,47%	51,43%
<i>M-CWS</i>	96,51%	0,71%	98,35%	50,67%
<i>I-GEN</i>	96,51%	0,34%	99,06%	50,00%

A.2.4 Low-Level Feature Performances of Obscured Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RDD</i>	88,08%	14,02%	88,12%	63,75%
<i>S-GEN</i>	87,98%	8,86%	94,17%	60,75%
<i>M-CWS</i>	87,98%	6,84%	95,38%	60,55%
<i>T-LBP</i>	87,98%	19,09%	88,66%	60,49%
<i>T-GLC</i>	87,98%	14,09%	92,10%	60,32%
<i>T-HOT</i>	87,98%	14,58%	92,02%	60,17%
<i>S-RBS</i>	88,08%	8,00%	93,57%	58,18%
<i>S-RDF</i>	88,08%	5,06%	95,39%	57,92%
<i>T-EDH</i>	87,98%	11,08%	92,60%	57,91%
<i>S-ZER</i>	88,08%	12,82%	89,05%	56,96%
<i>M-CSD</i>	87,98%	7,56%	92,93%	56,89%
<i>M-GLS</i>	87,98%	8,09%	95,16%	56,51%
<i>T-HIS</i>	87,98%	8,50%	94,41%	56,30%
<i>S-FDE</i>	88,08%	9,36%	91,68%	55,96%
<i>M-CMD</i>	87,98%	6,39%	95,10%	55,42%
<i>T-GLD</i>	87,98%	9,72%	93,94%	55,40%
<i>M-RMD</i>	87,98%	9,26%	91,62%	55,34%
<i>S-INM</i>	88,08%	8,92%	92,48%	55,13%
<i>M-CWK</i>	87,98%	7,00%	93,57%	54,90%
<i>S-CFD</i>	88,08%	4,90%	96,55%	53,96%
<i>S-DFD</i>	88,08%	4,95%	95,15%	53,70%
<i>M-CWM</i>	87,98%	5,17%	94,29%	53,09%
<i>M-CWW</i>	87,98%	6,86%	93,73%	52,86%
<i>T-TEB</i>	87,98%	3,46%	96,72%	52,42%
<i>I-GEN</i>	87,98%	1,75%	97,47%	51,30%
<i>M-RSD</i>	87,98%	5,07%	94,14%	50,29%

A.2.5 Low-Level Feature Performances of Spiculated Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-EDH</i>	82,56%	32,96%	86,64%	79,37%
<i>S-FDE</i>	83,46%	38,85%	80,91%	78,51%
<i>S-GEN</i>	79,07%	36,31%	78,69%	75,08%
<i>T-HIS</i>	77,52%	44,71%	80,25%	74,88%
<i>S-INM</i>	81,54%	29,43%	81,10%	74,64%
<i>S-RDD</i>	77,69%	40,61%	75,25%	74,15%
<i>S-DFD</i>	78,85%	28,42%	81,66%	74,12%
<i>S-ZER</i>	74,94%	32,52%	82,93%	73,21%
<i>S-RBS</i>	76,15%	30,53%	82,80%	72,24%
<i>T-HOT</i>	75,97%	42,74%	80,19%	71,14%
<i>T-GLC</i>	75,45%	35,98%	80,13%	70,94%
<i>S-RDF</i>	76,54%	35,27%	78,17%	70,09%
<i>T-GLD</i>	74,94%	41,65%	74,70%	67,96%
<i>S-CFD</i>	74,94%	15,74%	87,50%	65,41%
<i>M-GLS</i>	74,94%	32,32%	79,54%	65,14%
<i>M-CWS</i>	74,94%	23,69%	82,58%	63,47%
<i>I-GEN</i>	75,30%	26,93%	78,18%	62,48%
<i>T-LBP</i>	74,94%	29,45%	80,34%	60,93%
<i>M-CSD</i>	74,94%	9,75%	90,74%	56,29%
<i>M-CWM</i>	74,94%	21,00%	80,85%	55,80%
<i>M-CMD</i>	74,94%	18,09%	82,99%	55,04%
<i>M-RSD</i>	74,94%	14,18%	87,13%	54,88%
<i>M-RMD</i>	74,94%	12,63%	87,39%	54,26%
<i>M-CWK</i>	74,94%	19,50%	80,89%	54,00%
<i>T-TEB</i>	74,94%	27,10%	75,13%	53,85%
<i>M-CWW</i>	75,01%	14,74%	85,48%	53,37%

A.2.6 Low-Level Feature Performances of Irregular Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RDF</i>	81,15%	30,26%	73,43%	63,15%
<i>S-INM</i>	81,92%	33,83%	71,43%	61,82%
<i>S-GEN</i>	81,01%	29,58%	75,11%	61,25%
<i>S-RDD</i>	81,15%	30,12%	73,25%	60,02%
<i>M-GLS</i>	81,01%	29,33%	74,94%	58,67%
<i>T-LBP</i>	81,01%	27,24%	77,57%	58,23%
<i>T-GLC</i>	81,01%	26,24%	77,21%	57,54%
<i>S-RBS</i>	81,15%	36,82%	66,49%	57,27%
<i>S-DFD</i>	81,15%	28,40%	75,18%	56,79%
<i>M-RMD</i>	81,01%	30,38%	71,34%	56,22%
<i>T-EDH</i>	81,01%	32,11%	70,31%	55,37%
<i>M-CWK</i>	81,01%	27,83%	72,79%	54,96%
<i>T-HOT</i>	81,01%	25,34%	77,36%	54,89%
<i>I-GEN</i>	81,78%	28,00%	73,47%	54,65%
<i>T-GLD</i>	80,62%	27,19%	76,14%	54,64%
<i>T-HIS</i>	81,01%	26,20%	76,45%	54,57%
<i>S-FDE</i>	81,01%	29,65%	71,38%	54,54%
<i>M-CWS</i>	81,01%	31,62%	68,93%	54,53%
<i>M-CWM</i>	81,01%	29,99%	71,00%	54,29%
<i>S-CFD</i>	81,15%	30,99%	69,28%	54,23%
<i>T-TEB</i>	81,01%	30,57%	69,03%	54,11%
<i>S-ZER</i>	81,15%	29,68%	71,74%	53,08%
<i>M-CWW</i>	81,78%	27,37%	71,64%	52,04%
<i>M-CSD</i>	81,01%	28,31%	71,29%	51,93%
<i>M-CMD</i>	81,01%	30,01%	69,29%	51,33%
<i>M-RSD</i>	81,01%	25,38%	72,90%	50,30%

A.3 Low-Level Feature Performance Results of Density Property

A.3.1 Low-Level Feature Performances of Radiolucent Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-HIS</i>	91,86%	29,17%	93,52%	73,40%
<i>I-GEN</i>	90,70%	6,94%	97,44%	71,79%
<i>T-HOT</i>	89,53%	34,03%	93,66%	71,31%
<i>T-TEB</i>	90,70%	29,17%	95,09%	69,71%
<i>T-EDH</i>	90,70%	21,53%	88,03%	66,29%
<i>S-RDD</i>	91,47%	20,14%	96,28%	65,15%
<i>S-INM</i>	90,77%	5,56%	97,24%	62,85%
<i>S-GEN</i>	90,70%	10,42%	94,37%	61,70%
<i>T-LBP</i>	91,47%	9,72%	96,30%	61,59%
<i>S-RBS</i>	90,77%	8,68%	95,71%	61,38%
<i>M-RSD</i>	90,70%	11,81%	91,45%	61,38%
<i>T-GLC</i>	90,70%	11,11%	95,44%	61,32%
<i>M-GLS</i>	91,47%	12,85%	96,44%	61,16%
<i>S-RDF</i>	90,77%	7,99%	95,68%	61,16%
<i>T-GLD</i>	91,09%	12,50%	94,94%	60,26%
<i>M-RMD</i>	90,70%	4,51%	97,65%	59,46%
<i>S-DFD</i>	90,77%	6,25%	94,90%	59,25%
<i>S-ZER</i>	90,77%	15,97%	88,47%	58,90%
<i>S-FDE</i>	90,77%	5,90%	95,18%	58,47%
<i>S-CFD</i>	90,77%	6,25%	93,10%	55,29%
<i>M-CWM</i>	90,70%	3,47%	96,94%	54,91%
<i>M-CWS</i>	90,70%	3,47%	96,33%	54,33%
<i>M-CWK</i>	90,70%	2,43%	96,15%	54,06%
<i>M-CMD</i>	90,70%	2,78%	97,19%	53,21%
<i>M-CSD</i>	90,70%	5,21%	94,62%	52,03%
<i>M-CWW</i>	90,70%	4,17%	95,05%	51,98%

A.3.2 Low-Level Feature Performances of Low-Dense Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RDF</i>	97,69%	9,72%	98,06%	79,76%
<i>T-GLD</i>	97,67%	16,67%	97,82%	78,97%
<i>S-RDD</i>	97,69%	12,50%	95,69%	76,39%
<i>M-CMD</i>	98,06%	13,89%	99,54%	74,60%
<i>M-RMD</i>	97,67%	6,94%	98,61%	72,82%
<i>S-DFD</i>	97,69%	9,72%	95,76%	70,54%
<i>S-RBS</i>	97,69%	5,56%	96,45%	69,16%
<i>T-HOT</i>	97,67%	22,22%	95,77%	66,27%
<i>M-CWM</i>	97,67%	9,72%	99,17%	66,27%
<i>M-CWS</i>	97,67%	4,17%	98,88%	66,07%
<i>T-LBP</i>	97,67%	11,11%	98,94%	66,07%
<i>S-ZER</i>	97,69%	9,72%	96,57%	65,88%
<i>M-GLS</i>	98,06%	8,33%	99,17%	65,87%
<i>T-HIS</i>	97,67%	11,11%	99,01%	65,87%
<i>T-GLC</i>	97,67%	8,33%	98,68%	65,67%
<i>T-EDH</i>	97,67%	11,11%	95,97%	64,29%
<i>M-CSD</i>	97,67%	15,28%	87,33%	59,92%
<i>S-CFD</i>	97,69%	13,89%	86,81%	59,91%
<i>M-CWK</i>	97,67%	1,39%	98,61%	57,94%
<i>M-RSD</i>	97,67%	11,11%	89,35%	57,34%
<i>S-FDE</i>	97,69%	6,94%	95,07%	56,50%
<i>M-CWW</i>	97,67%	11,11%	86,81%	53,77%
<i>T-TEB</i>	97,67%	13,89%	86,84%	52,98%
<i>I-GEN</i>	97,67%	0,00%	99,54%	50,00%
<i>S-GEN</i>	97,67%	0,00%	98,41%	50,00%
<i>S-INM</i>	97,69%	0,00%	98,75%	50,00%

A.3.3 Low-Level Feature Performances of Iso-Dense Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>M-GLS</i>	70,93%	63,41%	64,92%	72,21%
<i>T-HOT</i>	69,38%	58,64%	64,30%	68,76%
<i>I-GEN</i>	65,50%	52,58%	68,07%	65,61%
<i>T-HIS</i>	65,12%	47,12%	68,24%	64,69%
<i>M-CMD</i>	64,34%	43,71%	71,62%	62,78%
<i>T-LBP</i>	62,02%	39,09%	68,47%	62,11%
<i>T-GLD</i>	62,02%	40,15%	66,89%	62,11%
<i>M-CWM</i>	62,40%	32,50%	75,84%	61,04%
<i>T-GLC</i>	61,24%	37,42%	70,61%	60,50%
<i>S-RBS</i>	62,02%	35,08%	69,80%	60,24%
<i>S-DFD</i>	62,02%	32,75%	69,71%	60,12%
<i>M-RMD</i>	59,30%	30,38%	72,58%	59,86%
<i>S-INM</i>	63,08%	31,54%	72,04%	59,86%
<i>S-RDD</i>	61,24%	43,82%	62,41%	58,98%
<i>S-ZER</i>	59,23%	38,53%	66,50%	58,65%
<i>S-GEN</i>	58,14%	33,64%	69,54%	57,79%
<i>S-RDF</i>	57,75%	36,73%	65,32%	57,57%
<i>T-EDH</i>	56,98%	33,79%	71,73%	56,21%
<i>M-CSD</i>	57,36%	30,08%	72,30%	55,90%
<i>M-CWK</i>	58,53%	24,32%	74,21%	55,39%
<i>M-CWS</i>	58,53%	28,56%	73,37%	54,42%
<i>M-CWW</i>	58,53%	24,47%	75,56%	54,36%
<i>S-CFD</i>	57,36%	26,08%	73,25%	53,86%
<i>S-FDE</i>	56,20%	30,75%	68,81%	53,29%
<i>M-RSD</i>	57,36%	24,92%	74,32%	52,70%
<i>T-TEB</i>	57,36%	26,21%	74,10%	51,56%

A.3.4 Low-Level Feature Performances of High-Dense Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>M-GLS</i>	77,91%	69,70%	71,37%	77,38%
<i>I-GEN</i>	74,03%	74,22%	59,64%	74,21%
<i>T-HIS</i>	73,64%	69,49%	62,98%	73,72%
<i>T-HOT</i>	74,03%	58,33%	73,69%	73,61%
<i>M-CMD</i>	67,05%	73,09%	48,27%	65,98%
<i>T-LBP</i>	64,34%	66,53%	46,07%	64,28%
<i>M-CWM</i>	62,40%	73,59%	36,43%	63,09%
<i>T-GLC</i>	63,18%	67,51%	45,36%	62,94%
<i>M-RMD</i>	62,40%	71,26%	36,13%	62,29%
<i>T-GLD</i>	62,40%	60,45%	48,45%	61,96%
<i>S-RDF</i>	61,63%	62,53%	46,44%	61,91%
<i>S-INM</i>	60,85%	71,51%	38,15%	61,46%
<i>S-GEN</i>	62,02%	65,75%	43,81%	61,27%
<i>S-ZER</i>	62,02%	52,67%	54,39%	60,94%
<i>S-RDD</i>	60,77%	54,15%	53,50%	60,81%
<i>S-DFD</i>	59,30%	63,39%	44,54%	60,30%
<i>S-RBS</i>	59,23%	64,06%	44,48%	59,20%
<i>S-FDE</i>	59,23%	60,71%	42,26%	57,76%
<i>M-CSD</i>	57,75%	55,37%	48,63%	57,14%
<i>S-CFD</i>	56,54%	53,38%	46,42%	56,32%
<i>M-CWS</i>	55,04%	70,34%	34,52%	56,15%
<i>T-EDH</i>	55,43%	54,66%	49,64%	55,20%
<i>T-TEB</i>	54,65%	53,95%	46,67%	53,88%
<i>M-CWK</i>	53,88%	70,13%	30,60%	53,64%
<i>M-RSD</i>	55,04%	55,51%	45,18%	52,96%
<i>M-CWW</i>	55,43%	57,34%	42,32%	52,54%

A.4 Low-Level Feature Performance Results of BI-RADS Property

A.4.1 Low-Level Feature Performances of BI-RADS 0 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-GEN</i>	93,30%	14,44%	95,49%	78,98%
<i>S-RBS</i>	93,30%	14,42%	93,50%	74,13%
<i>S-DFD</i>	93,25%	20,63%	92,31%	73,59%
<i>T-LBP</i>	93,56%	30,05%	92,08%	72,05%
<i>T-GLC</i>	94,00%	23,13%	96,62%	70,92%
<i>S-FDE</i>	93,30%	10,88%	93,38%	67,13%
<i>S-INM</i>	93,30%	10,55%	93,93%	66,86%
<i>T-HOT</i>	93,49%	27,41%	95,59%	66,67%
<i>T-GLD</i>	93,30%	27,87%	92,86%	66,55%
<i>S-ZER</i>	93,30%	13,03%	93,63%	63,05%
<i>S-CFD</i>	93,30%	4,53%	97,40%	60,92%
<i>T-HIS</i>	93,45%	13,21%	98,10%	60,73%
<i>T-EDH</i>	93,30%	11,93%	95,24%	59,81%
<i>S-RDD</i>	93,30%	16,84%	86,29%	59,03%
<i>I-GEN</i>	93,30%	3,01%	98,28%	55,72%
<i>S-RDF</i>	93,32%	4,43%	97,77%	54,93%
<i>M-CWS</i>	93,30%	5,60%	97,28%	54,71%
<i>M-GLS</i>	93,30%	5,28%	98,06%	54,09%
<i>M-CMD</i>	93,30%	6,28%	94,81%	53,52%
<i>M-CWK</i>	93,30%	3,32%	97,93%	53,29%
<i>M-CSD</i>	93,30%	9,47%	91,62%	52,83%
<i>M-RMD</i>	93,30%	10,11%	90,48%	51,95%
<i>M-CWM</i>	93,30%	1,46%	98,54%	51,38%
<i>M-CWW</i>	93,30%	4,96%	95,46%	51,18%
<i>M-RSD</i>	93,30%	5,37%	94,42%	51,13%
<i>T-TEB</i>	93,30%	3,19%	97,22%	50,98%

A.4.2 Low-Level Feature Performances of BI-RADS 1 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-TEB</i>	99,89%	11,11%	96,73%	75,70%
<i>S-RDD</i>	99,93%	2,78%	99,34%	66,67%
<i>M-CWW</i>	99,89%	2,78%	92,48%	62,31%
<i>T-HOT</i>	99,89%	5,56%	97,80%	60,35%
<i>I-GEN</i>	99,89%	2,78%	97,10%	59,78%
<i>T-LBP</i>	99,89%	5,56%	96,12%	55,29%
<i>M-CSD</i>	99,89%	8,33%	89,99%	52,07%
<i>S-CFD</i>	99,89%	8,33%	89,20%	50,00%
<i>M-RSD</i>	99,89%	2,78%	93,80%	50,00%
<i>M-CWM</i>	99,89%	0,00%	98,76%	50,00%
<i>T-GLC</i>	99,89%	0,00%	97,31%	50,00%
<i>S-DFD</i>	99,89%	0,00%	99,77%	50,00%
<i>T-HIS</i>	99,89%	0,00%	97,66%	50,00%
<i>T-GLD</i>	99,89%	0,00%	99,66%	50,00%
<i>S-RBS</i>	99,89%	0,00%	99,91%	50,00%
<i>M-CMD</i>	99,89%	0,00%	97,40%	50,00%
<i>T-EDH</i>	99,89%	0,00%	99,96%	50,00%
<i>M-CWK</i>	99,89%	0,00%	99,66%	50,00%
<i>M-CWS</i>	99,89%	0,00%	99,66%	50,00%
<i>S-ZER</i>	99,89%	0,00%	97,32%	50,00%
<i>M-GLS</i>	99,89%	0,00%	99,16%	50,00%
<i>S-RDF</i>	99,89%	0,00%	99,87%	50,00%
<i>M-RMD</i>	99,89%	0,00%	92,16%	50,00%
<i>S-FDE</i>	99,89%	0,00%	94,22%	50,00%
<i>S-GEN</i>	99,89%	0,00%	98,66%	50,00%
<i>S-INM</i>	99,89%	0,00%	99,68%	50,00%

A.4.2 Low-Level Feature Performances of BI-RADS 2 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-RDF</i>	95,22%	16,27%	92,14%	75,81%
<i>T-HIS</i>	95,21%	16,32%	95,56%	71,47%
<i>S-RBS</i>	95,39%	12,78%	92,28%	68,14%
<i>T-HOT</i>	95,21%	15,63%	94,52%	67,79%
<i>T-GLC</i>	95,21%	19,71%	92,71%	67,61%
<i>S-RDD</i>	95,26%	18,67%	93,00%	67,05%
<i>T-LBP</i>	95,21%	19,03%	92,90%	66,90%
<i>T-EDH</i>	95,35%	14,78%	92,55%	66,60%
<i>S-DFD</i>	95,22%	17,13%	89,03%	65,67%
<i>S-FDE</i>	95,22%	13,47%	91,87%	63,95%
<i>S-ZER</i>	95,43%	9,18%	96,00%	62,95%
<i>I-GEN</i>	95,21%	10,32%	95,01%	61,84%
<i>M-GLS</i>	95,21%	17,35%	90,01%	61,31%
<i>T-GLD</i>	95,21%	16,97%	93,44%	61,31%
<i>S-CFD</i>	95,22%	9,27%	92,39%	60,57%
<i>S-GEN</i>	95,21%	9,57%	94,51%	59,51%
<i>M-CWM</i>	95,39%	9,20%	94,35%	58,83%
<i>M-CWS</i>	95,39%	8,80%	94,42%	56,79%
<i>M-CWK</i>	95,21%	6,00%	94,30%	56,38%
<i>M-RMD</i>	95,21%	14,48%	87,17%	56,35%
<i>M-RSD</i>	95,21%	14,37%	87,47%	56,07%
<i>M-CMD</i>	95,21%	8,63%	91,41%	56,02%
<i>M-CWW</i>	95,21%	10,29%	90,55%	55,59%
<i>T-TEB</i>	95,21%	9,05%	93,67%	55,23%
<i>S-INM</i>	95,37%	10,84%	92,61%	54,90%
<i>M-CSD</i>	95,21%	6,96%	92,92%	54,19%

A.4.3 Low-Level Feature Performances of BI-RADS 3 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-FDE</i>	80,46%	25,36%	84,32%	75,08%
<i>T-HIS</i>	80,46%	35,24%	83,08%	74,03%
<i>S-GEN</i>	80,46%	28,58%	85,48%	73,82%
<i>S-ZER</i>	80,46%	22,66%	85,97%	71,00%
<i>T-GLD</i>	80,62%	26,27%	86,75%	70,55%
<i>S-RBS</i>	80,46%	19,21%	88,12%	69,38%
<i>S-RDF</i>	80,46%	22,79%	86,90%	67,47%
<i>S-INM</i>	80,46%	17,12%	87,70%	67,41%
<i>S-RDD</i>	80,46%	15,70%	92,22%	65,68%
<i>S-DFD</i>	80,41%	22,95%	83,22%	65,30%
<i>T-EDH</i>	80,46%	14,08%	91,10%	64,52%
<i>S-CFD</i>	80,46%	15,05%	87,97%	62,46%
<i>I-GEN</i>	80,46%	19,72%	85,60%	61,87%
<i>M-RSD</i>	80,46%	11,06%	90,96%	61,83%
<i>M-GLS</i>	80,46%	17,29%	86,98%	61,13%
<i>T-LBP</i>	80,46%	14,85%	90,36%	60,94%
<i>M-CWS</i>	80,46%	10,26%	90,90%	60,55%
<i>T-GLC</i>	80,46%	18,00%	88,48%	60,22%
<i>T-HOT</i>	80,46%	23,18%	85,52%	60,15%
<i>M-CMD</i>	80,46%	9,77%	91,37%	57,76%
<i>M-RMD</i>	80,46%	9,49%	92,22%	56,34%
<i>M-CSD</i>	80,53%	11,17%	89,35%	56,17%
<i>T-TEB</i>	80,46%	7,01%	94,44%	55,24%
<i>M-CWM</i>	80,46%	16,11%	85,45%	54,45%
<i>M-CWW</i>	80,46%	10,13%	90,93%	54,16%
<i>M-CWK</i>	80,46%	8,81%	89,75%	51,87%

A.4.4 Low-Level Feature Performances of BI-RADS 4 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-HIS</i>	57,30%	71,87%	36,82%	60,13%
<i>S-GEN</i>	59,46%	64,73%	42,10%	58,98%
<i>T-GLC</i>	57,41%	64,01%	44,22%	58,85%
<i>T-HOT</i>	56,42%	62,52%	46,88%	58,62%
<i>S-ZER</i>	56,06%	72,84%	33,35%	58,10%
<i>S-DFD</i>	57,68%	68,86%	36,04%	57,94%
<i>M-GLS</i>	56,17%	71,69%	34,14%	57,61%
<i>T-GLD</i>	58,18%	65,49%	41,93%	57,48%
<i>S-RBS</i>	58,41%	68,17%	36,36%	55,81%
<i>T-EDH</i>	54,59%	70,26%	34,07%	55,42%
<i>M-CWK</i>	52,84%	80,03%	23,83%	55,36%
<i>S-INM</i>	56,84%	81,06%	21,54%	54,77%
<i>T-LBP</i>	55,95%	65,27%	40,11%	54,74%
<i>S-FDE</i>	58,98%	62,64%	41,65%	54,72%
<i>S-RDD</i>	59,02%	71,48%	31,27%	54,68%
<i>S-RDF</i>	54,94%	78,82%	24,64%	54,63%
<i>S-CFD</i>	56,95%	59,80%	40,59%	53,96%
<i>M-CWS</i>	53,20%	82,77%	19,47%	53,03%
<i>M-CWM</i>	55,91%	72,51%	29,43%	52,56%
<i>M-RSD</i>	56,64%	59,66%	40,57%	51,97%
<i>M-CSD</i>	56,97%	59,23%	41,08%	51,70%
<i>M-CMD</i>	56,17%	65,68%	34,79%	51,26%
<i>I-GEN</i>	54,04%	64,11%	35,91%	50,93%
<i>T-TEB</i>	53,46%	67,33%	32,63%	50,75%
<i>M-CWW</i>	56,82%	60,62%	39,08%	50,58%
<i>M-RMD</i>	56,53%	57,39%	42,55%	50,47%

A.4.5 Low-Level Feature Performances of BI-RADS 4A Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-GEN</i>	93,02%	4,63%	98,47%	63,75%
<i>M-CWM</i>	93,02%	9,26%	92,43%	62,29%
<i>M-CWS</i>	93,02%	6,02%	96,63%	60,97%
<i>M-RSD</i>	93,02%	9,72%	91,56%	60,63%
<i>M-RMD</i>	93,02%	9,72%	93,19%	60,14%
<i>M-CWW</i>	93,02%	8,33%	96,42%	59,03%
<i>M-CSD</i>	93,02%	5,09%	96,25%	57,78%
<i>T-HIS</i>	93,41%	5,56%	98,33%	57,57%
<i>S-DFD</i>	93,08%	3,70%	97,41%	56,74%
<i>M-CMD</i>	93,02%	2,78%	97,22%	56,32%
<i>S-CFD</i>	93,08%	3,24%	97,44%	56,04%
<i>S-RBS</i>	93,08%	4,17%	96,44%	55,44%
<i>T-LBP</i>	93,41%	10,19%	95,90%	55,35%
<i>S-ZER</i>	93,08%	4,17%	95,68%	54,82%
<i>M-CWK</i>	93,02%	4,63%	95,73%	54,38%
<i>T-GLC</i>	93,02%	3,70%	96,67%	54,31%
<i>S-RDF</i>	93,08%	3,24%	97,92%	54,17%
<i>M-GLS</i>	93,02%	3,70%	96,98%	53,68%
<i>T-GLD</i>	93,02%	4,63%	95,56%	52,92%
<i>S-RDD</i>	93,08%	0,93%	98,10%	52,66%
<i>S-FDE</i>	93,08%	2,78%	97,09%	50,90%
<i>T-EDH</i>	93,02%	2,78%	97,08%	50,56%
<i>I-GEN</i>	93,02%	0,46%	98,68%	50,07%
<i>T-HOT</i>	93,02%	1,85%	95,63%	50,00%
<i>T-TEB</i>	93,02%	1,85%	97,85%	50,00%
<i>S-INM</i>	93,08%	0,00%	97,58%	50,00%

A.4.6 Low-Level Feature Performances of BI-RADS 4B Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-EDH</i>	95,35%	8,33%	96,61%	61,69%
<i>S-CFD</i>	95,38%	8,33%	92,93%	61,49%
<i>M-RSD</i>	95,35%	7,64%	93,22%	61,18%
<i>I-GEN</i>	95,35%	1,39%	99,22%	56,30%
<i>T-LBP</i>	95,35%	4,17%	98,37%	56,30%
<i>M-CMD</i>	95,35%	2,08%	98,34%	56,10%
<i>S-RBS</i>	95,38%	2,78%	95,75%	56,10%
<i>T-GLC</i>	95,35%	2,78%	96,07%	55,89%
<i>T-HIS</i>	95,35%	6,94%	97,09%	55,69%
<i>S-INM</i>	95,38%	3,47%	98,28%	55,28%
<i>M-GLS</i>	95,35%	3,47%	96,14%	54,78%
<i>S-RDD</i>	95,38%	2,78%	97,71%	54,50%
<i>M-CSD</i>	95,74%	7,64%	91,23%	54,17%
<i>S-RDF</i>	95,38%	1,39%	98,25%	53,56%
<i>M-CWS</i>	95,35%	1,39%	97,93%	53,56%
<i>S-ZER</i>	95,38%	3,47%	96,66%	53,46%
<i>M-CWK</i>	95,35%	0,69%	98,10%	53,15%
<i>S-DFD</i>	95,38%	3,47%	96,56%	52,76%
<i>M-CWW</i>	95,35%	8,33%	91,97%	52,74%
<i>S-GEN</i>	95,35%	0,69%	98,75%	52,34%
<i>S-FDE</i>	95,38%	2,78%	97,53%	52,15%
<i>T-HOT</i>	95,35%	2,78%	96,27%	51,93%
<i>T-TEB</i>	95,35%	1,39%	97,97%	50,00%
<i>M-RMD</i>	95,35%	0,69%	96,58%	50,00%
<i>T-GLD</i>	95,35%	0,00%	97,49%	50,00%
<i>M-CWM</i>	95,35%	0,00%	97,83%	50,00%

A.4.7 Low-Level Feature Performances of BI-RADS 4C Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-HOT</i>	92,64%	19,30%	90,38%	67,68%
<i>S-RDD</i>	92,69%	8,33%	97,12%	67,29%
<i>S-ZER</i>	92,69%	17,11%	88,01%	64,50%
<i>S-FDE</i>	92,69%	11,84%	91,21%	63,95%
<i>S-GEN</i>	92,64%	13,16%	95,61%	63,10%
<i>T-HIS</i>	93,02%	17,54%	96,03%	62,32%
<i>S-INM</i>	92,69%	12,72%	91,25%	59,51%
<i>M-CWS</i>	92,64%	3,95%	96,48%	58,02%
<i>M-GLS</i>	92,64%	8,33%	95,68%	57,18%
<i>T-LBP</i>	92,64%	4,39%	96,79%	56,85%
<i>M-RMD</i>	92,64%	8,77%	92,89%	56,41%
<i>S-DFD</i>	92,69%	4,82%	96,15%	54,99%
<i>T-GLD</i>	92,64%	4,39%	95,68%	54,88%
<i>I-GEN</i>	92,64%	2,19%	98,43%	54,34%
<i>S-RBS</i>	92,69%	3,95%	96,42%	54,23%
<i>M-CWM</i>	92,64%	4,39%	95,57%	54,13%
<i>M-CWW</i>	92,64%	4,39%	94,77%	54,13%
<i>T-TEB</i>	92,64%	3,51%	97,14%	54,04%
<i>M-CWK</i>	92,64%	4,39%	94,39%	53,80%
<i>S-RDF</i>	92,69%	3,51%	97,64%	53,59%
<i>S-CFD</i>	92,69%	3,07%	95,45%	53,33%
<i>M-CMD</i>	92,64%	2,19%	95,96%	52,21%
<i>T-GLC</i>	92,64%	7,02%	92,19%	51,79%
<i>M-CSD</i>	92,64%	7,89%	91,00%	51,79%
<i>T-EDH</i>	92,64%	13,16%	85,91%	51,29%
<i>M-RSD</i>	92,64%	2,19%	96,30%	51,17%

A.4.8 Low-Level Feature Performances of BI-RADS 5 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>T-EDH</i>	80,23%	43,35%	72,40%	77,80%
<i>S-GEN</i>	77,13%	48,33%	70,08%	76,64%
<i>S-FDE</i>	75,00%	38,49%	77,25%	75,86%
<i>S-INM</i>	75,77%	33,48%	77,37%	75,44%
<i>S-DFD</i>	79,23%	30,30%	78,66%	73,95%
<i>S-RDD</i>	81,92%	40,60%	73,59%	73,42%
<i>S-RBS</i>	77,31%	42,84%	71,36%	72,95%
<i>S-ZER</i>	77,69%	36,88%	79,09%	72,83%
<i>S-RDF</i>	74,28%	38,38%	74,00%	72,34%
<i>T-HIS</i>	76,25%	43,03%	78,96%	72,25%
<i>T-HOT</i>	76,74%	44,26%	75,47%	71,02%
<i>T-GLC</i>	74,61%	41,31%	70,57%	67,64%
<i>M-GLS</i>	74,53%	40,02%	73,37%	66,22%
<i>T-LBP</i>	74,28%	38,52%	70,86%	64,87%
<i>T-GLD</i>	74,28%	38,39%	74,43%	64,08%
<i>S-CFD</i>	74,28%	36,53%	67,52%	63,82%
<i>I-GEN</i>	74,46%	44,85%	60,26%	61,06%
<i>M-CWS</i>	74,28%	35,22%	67,60%	59,84%
<i>M-CWM</i>	74,28%	35,11%	69,82%	59,42%
<i>M-RSD</i>	74,28%	30,90%	71,44%	58,72%
<i>M-RMD</i>	74,28%	29,54%	71,39%	57,08%
<i>T-TEB</i>	74,28%	42,01%	60,09%	55,64%
<i>M-CWW</i>	74,28%	30,79%	69,72%	55,04%
<i>M-CWK</i>	74,28%	34,22%	66,85%	54,66%
<i>M-CMD</i>	74,28%	36,84%	63,93%	53,17%
<i>M-CSD</i>	74,28%	30,37%	69,65%	52,58%

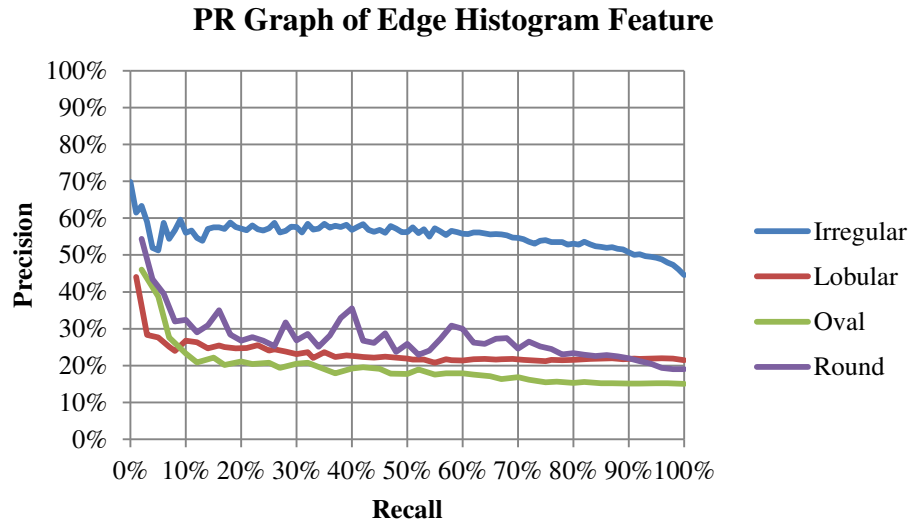
A.4.9 Low-Level Feature Performances of BI-RADS 6 Class

Low-Level Feature	Max. Accuracy	Avg. Sensitivity	Avg. Specificity	Max. B.Accuracy
<i>S-GEN</i>	91,47%	27,01%	95,16%	71,10%
<i>S-RBS</i>	88,76%	19,81%	94,83%	69,33%
<i>T-GLC</i>	88,76%	22,41%	91,19%	67,83%
<i>M-CWS</i>	88,76%	18,39%	92,36%	67,54%
<i>S-ZER</i>	89,92%	23,29%	92,69%	67,22%
<i>T-HIS</i>	89,92%	31,61%	93,89%	67,22%
<i>M-CMD</i>	88,76%	15,52%	89,45%	65,71%
<i>T-HOT</i>	88,76%	24,14%	91,70%	64,60%
<i>M-CWM</i>	88,76%	11,78%	94,21%	62,44%
<i>S-RDD</i>	88,76%	17,38%	89,99%	62,26%
<i>T-GLD</i>	88,76%	23,56%	89,16%	61,78%
<i>T-LBP</i>	88,76%	22,41%	89,08%	60,77%
<i>T-EDH</i>	88,76%	10,34%	94,40%	60,25%
<i>M-CSD</i>	88,76%	8,05%	95,60%	59,64%
<i>S-RDF</i>	89,15%	12,66%	94,40%	59,63%
<i>S-FDE</i>	88,76%	10,62%	95,92%	59,45%
<i>S-INM</i>	88,76%	9,97%	93,41%	57,75%
<i>S-DFD</i>	88,76%	10,09%	93,12%	57,36%
<i>I-GEN</i>	88,76%	2,30%	97,89%	56,41%
<i>M-RSD</i>	88,76%	8,05%	93,52%	56,20%
<i>M-CWW</i>	88,76%	14,08%	88,86%	56,15%
<i>M-GLS</i>	88,76%	9,20%	94,32%	55,35%
<i>S-CFD</i>	88,76%	5,02%	95,74%	54,25%
<i>M-RMD</i>	88,76%	6,61%	92,39%	53,42%
<i>M-CWK</i>	88,76%	10,63%	90,17%	53,16%
<i>T-TEB</i>	88,76%	8,62%	85,66%	50,00%

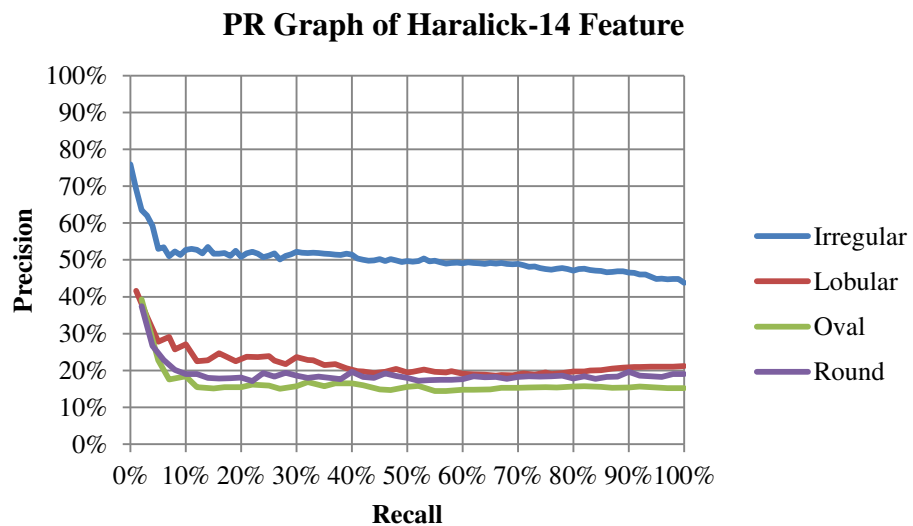
B Precision-Recall Graphs of Low-Level Features

B.1 Precision-Recall Graphs of Shape Property Queries

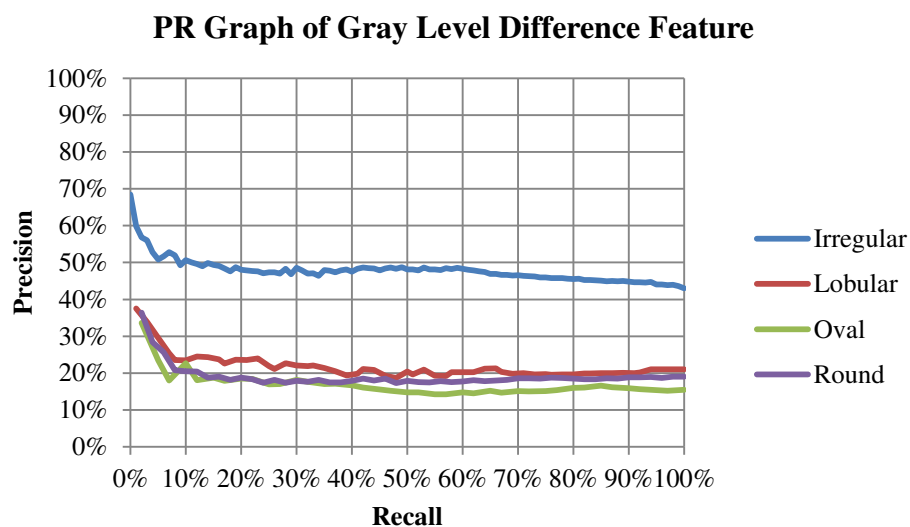
B.1.1 Precision-Recall Graph of Edge Histogram Feature



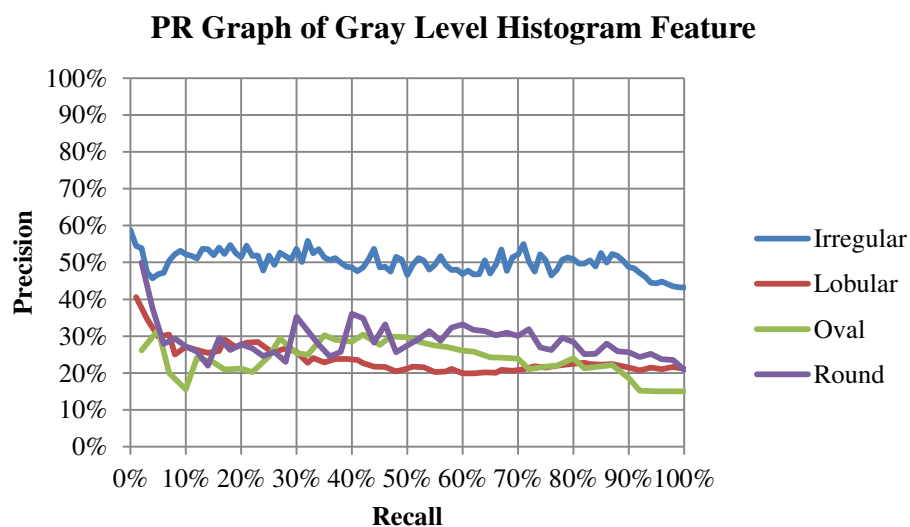
B.1.2 Precision-Recall Graph of Haralick-14 Feature



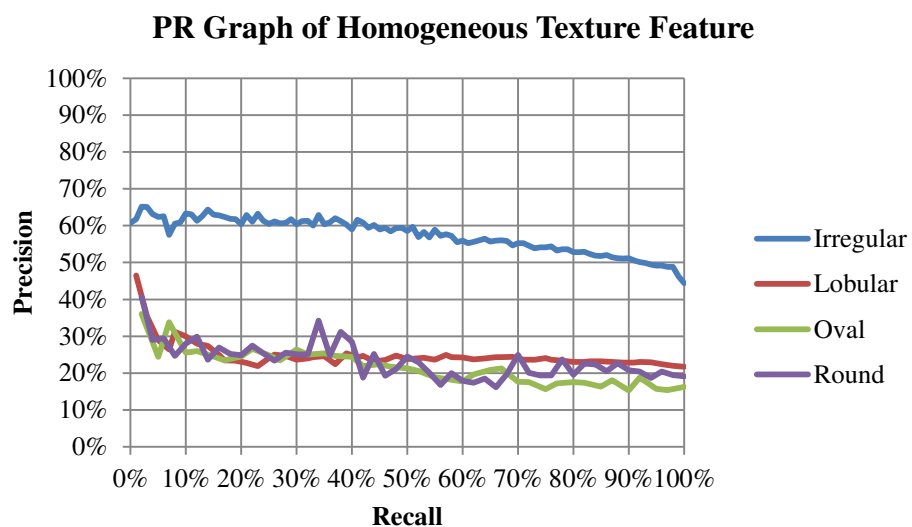
B.1.3 Precision-Recall Graph of Gray Level Difference Feature



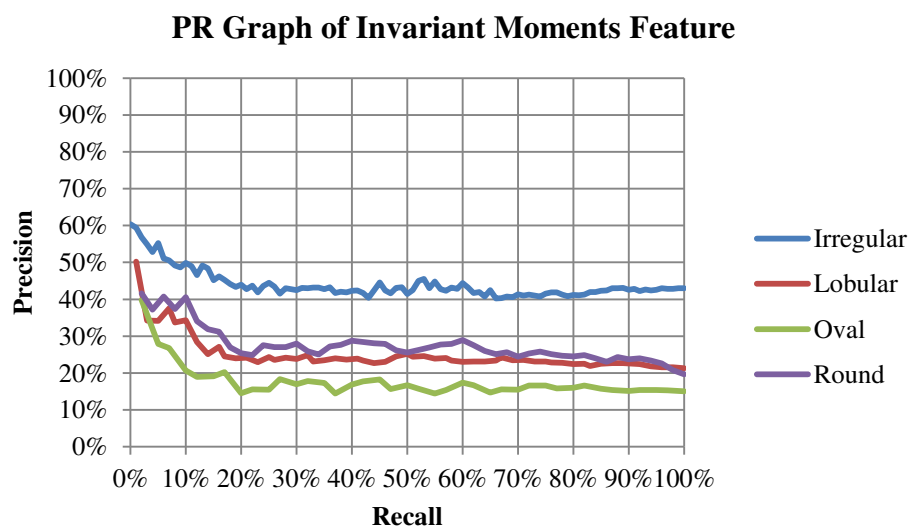
B.1.4 Precision-Recall Graph of Gray Level Histogram Feature



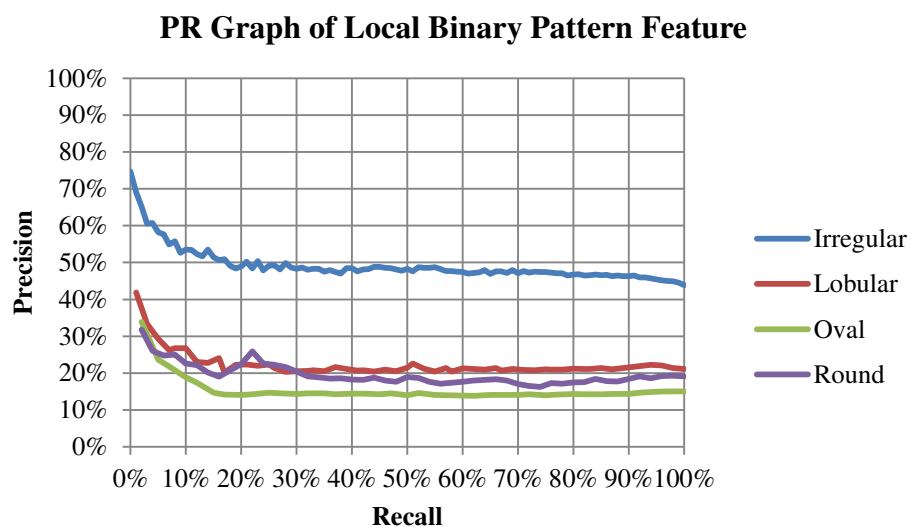
B.1.5 Precision-Recall Graph of Homogeneous Texture Feature



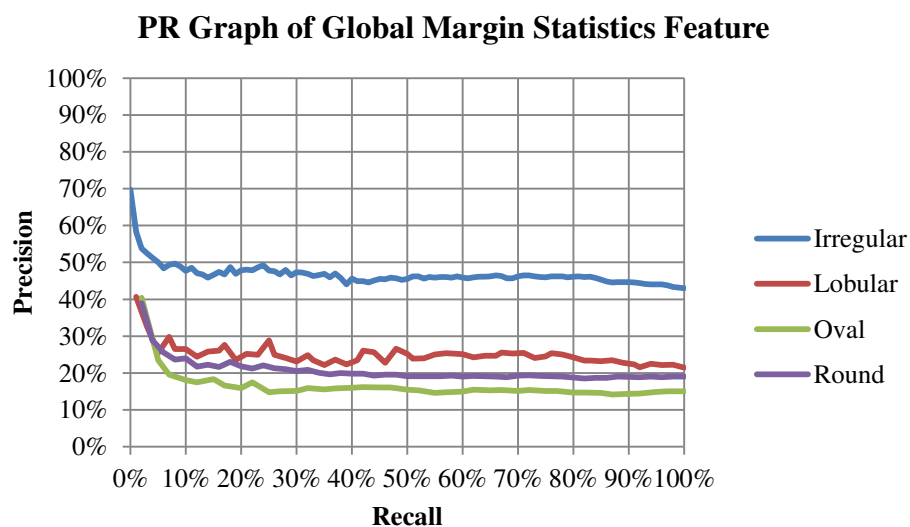
B.1.6 Precision-Recall Graph of Invariant Moments Feature



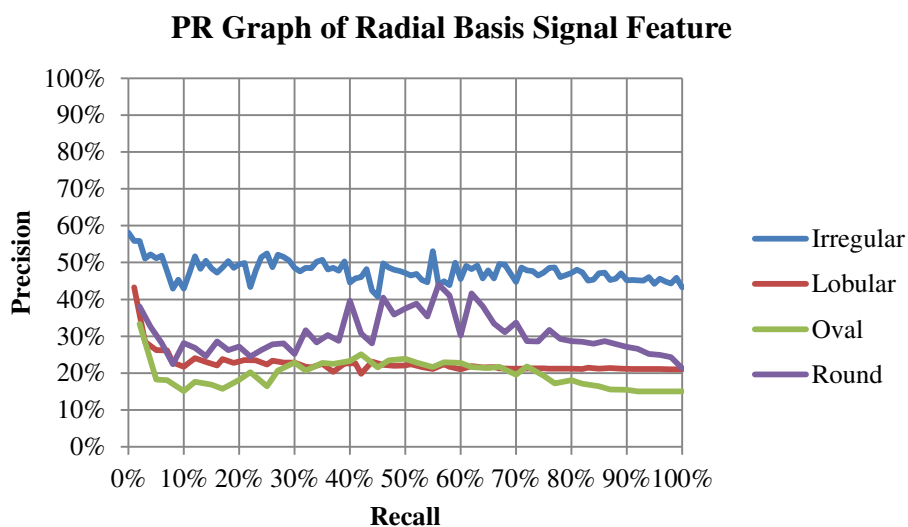
B.1.7 Precision-Recall Graph of Local Binary Pattern Feature



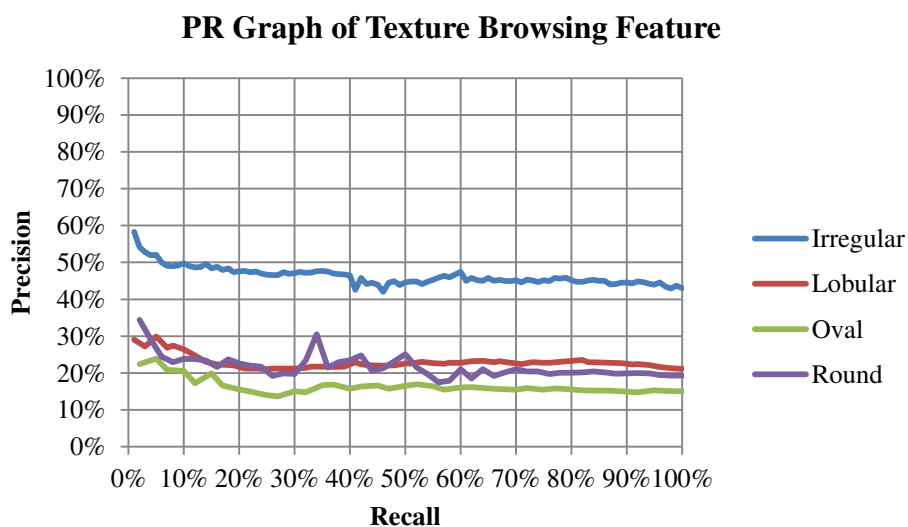
B.1.8 Precision-Recall Graph of Global Margin Statistics Feature

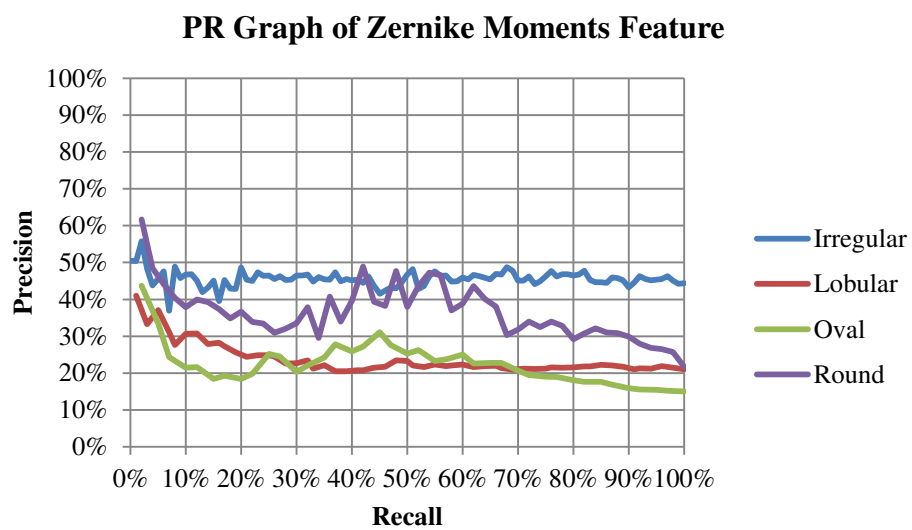


B.1.9 Precision-Recall Graph of Radial Basis Signal Feature



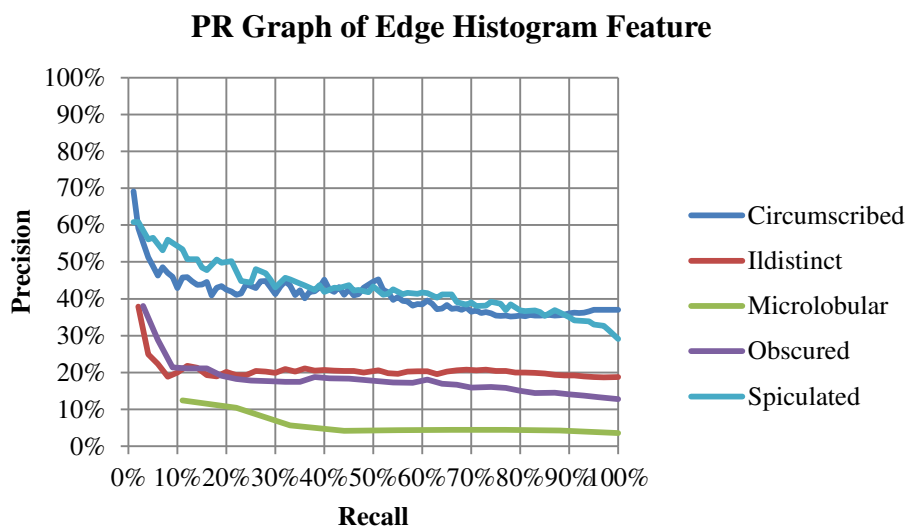
B.1.10 Precision-Recall Graph of Texture Browsing Feature



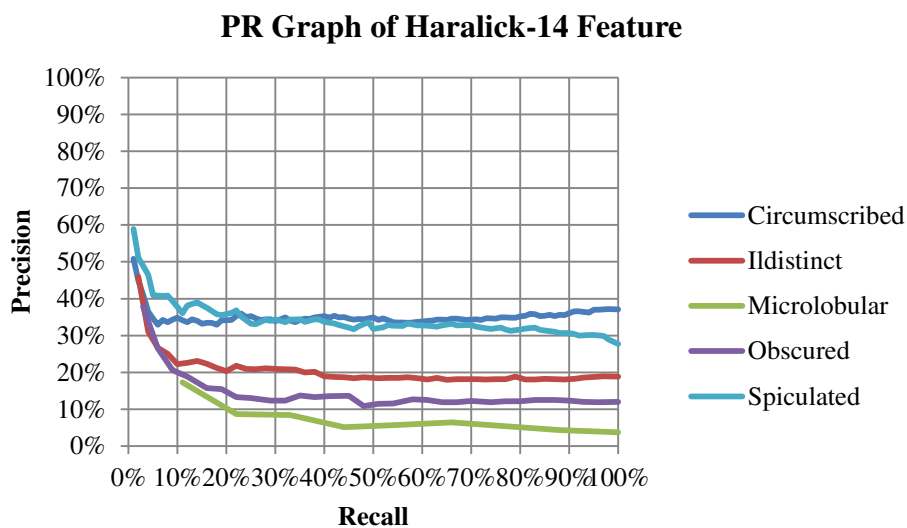
B.1.11 Precision-Recall Graph of Zernike Moments Feature

B.2 Precision-Recall Graphs of Margin Property Queries

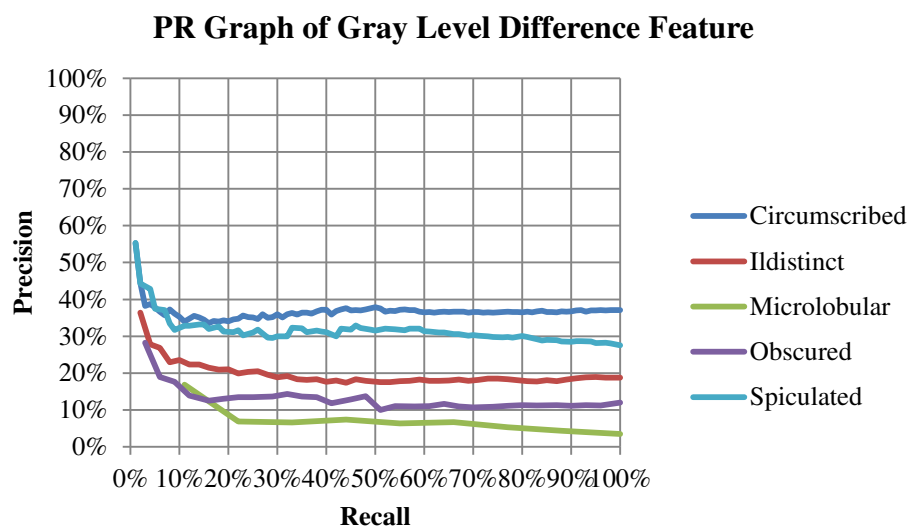
B.2.1 Precision-Recall Graph of Edge Histogram Feature



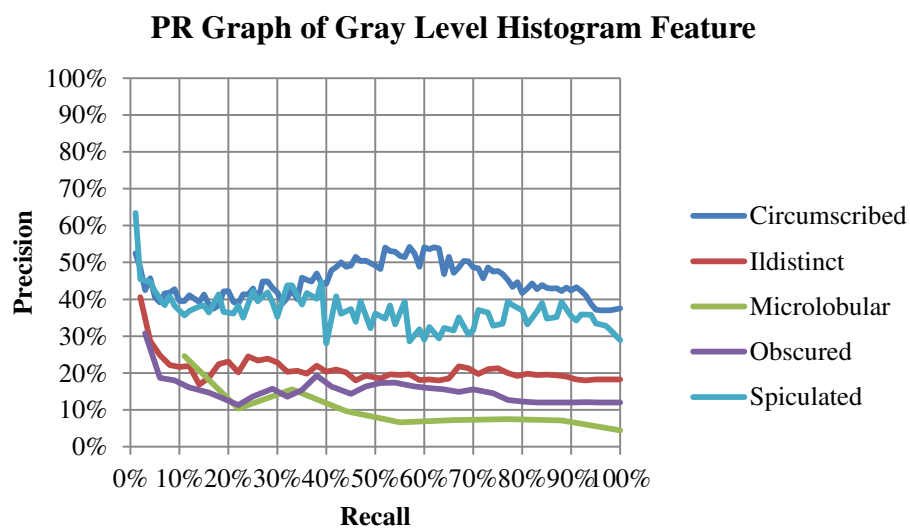
B.2.2 Precision-Recall Graph of Haralick-14 Feature



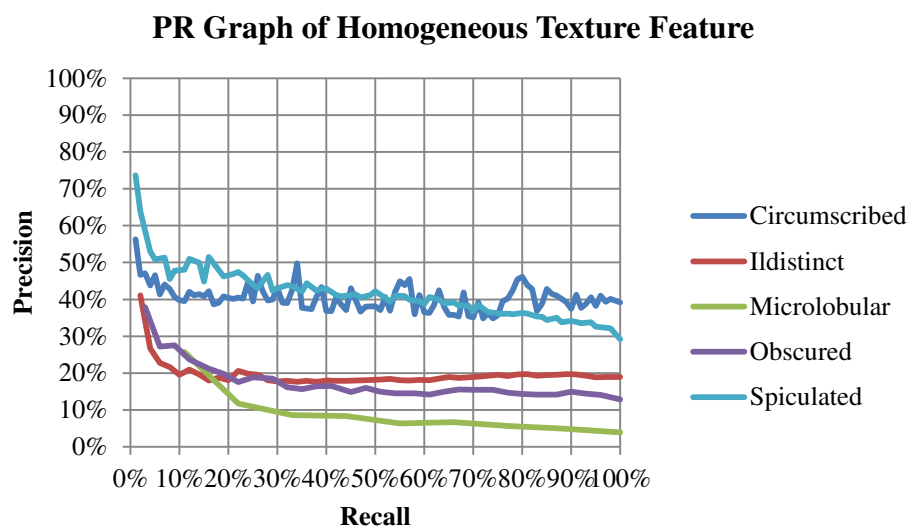
B.2.3 Precision-Recall Graph of Gray Level Difference Feature



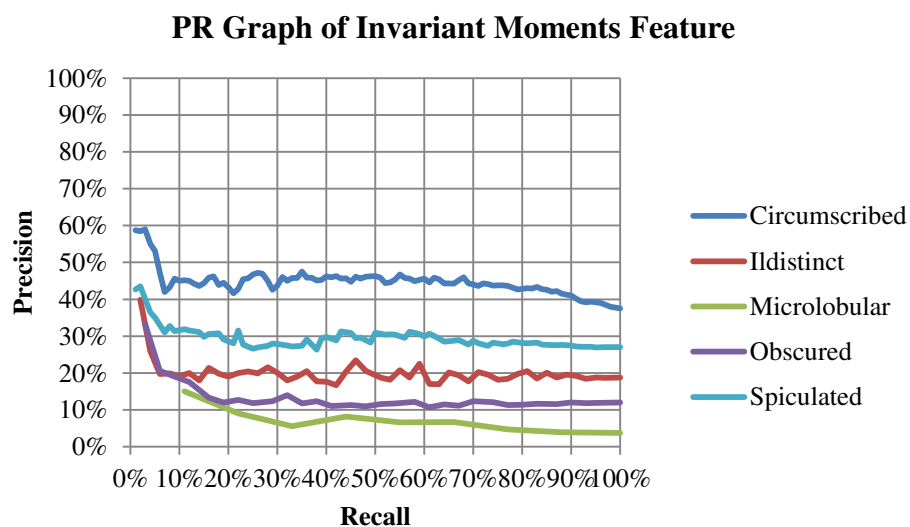
B.2.4 Precision-Recall Graph of Gray Level Histogram Feature



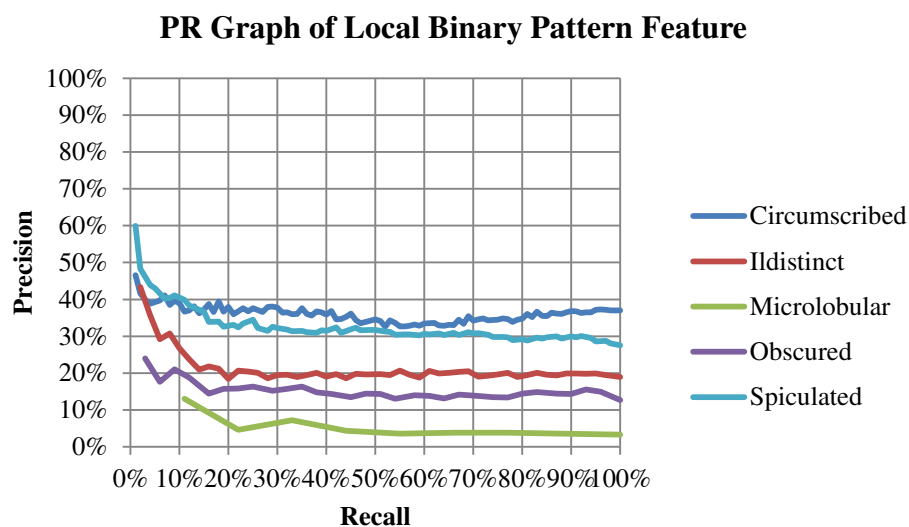
B.2.5 Precision-Recall Graph of Homogeneous Texture Feature



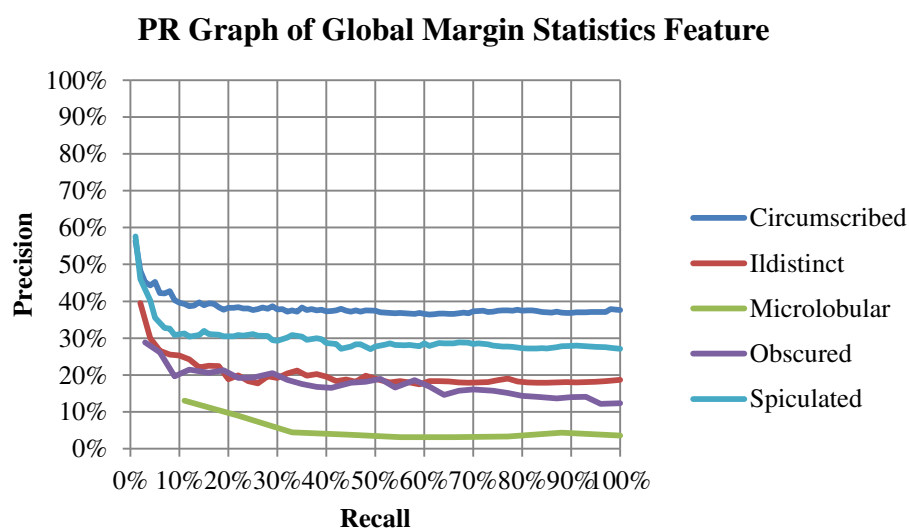
B.2.6 Precision-Recall Graph of Invariant Moments Feature



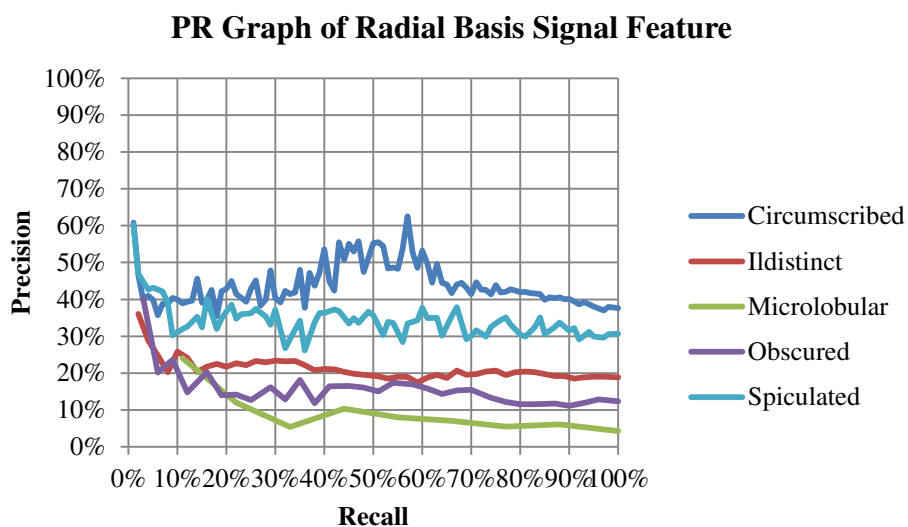
B.2.7 Precision-Recall Graph of Local Binary Pattern Feature



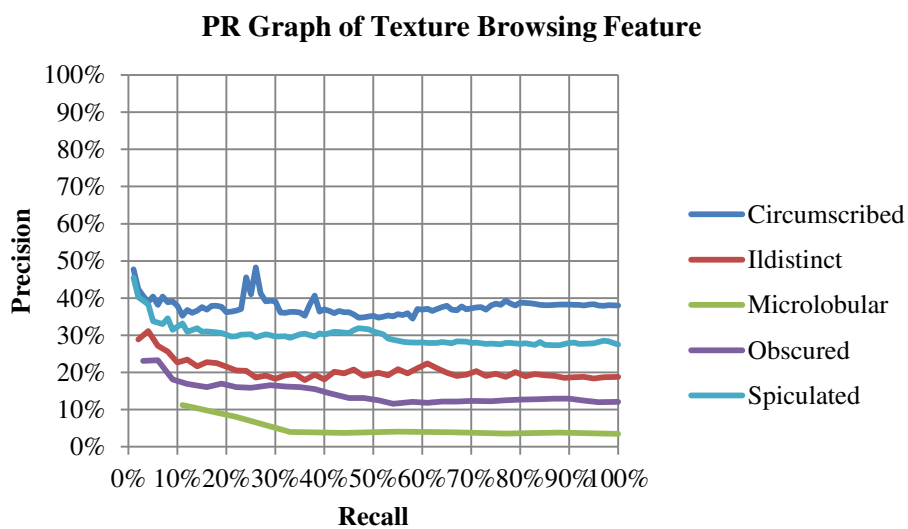
B.2.8 Precision-Recall Graph of Global Margin Statistics Feature



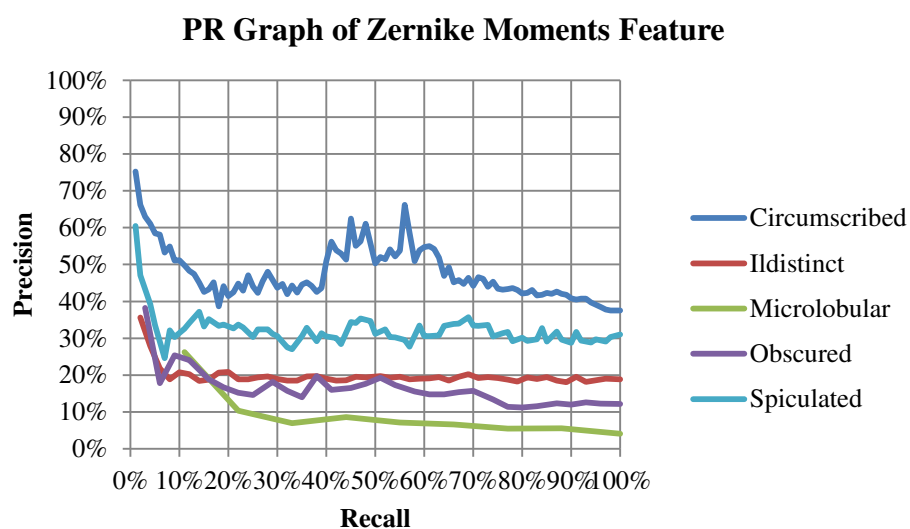
B.2.9 Precision-Recall Graph of Radial Basis Signal Feature



B.2.10 Precision-Recall Graph of Texture Browsing Feature

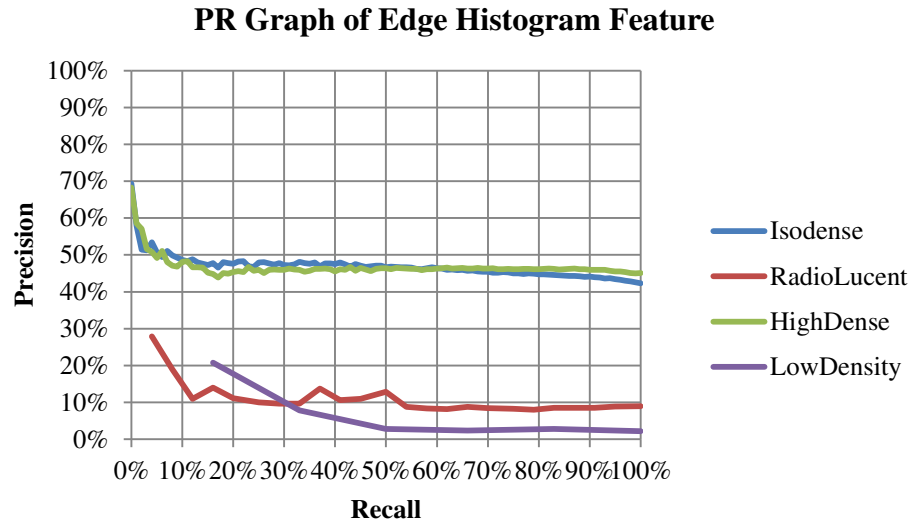


B.2.11 Precision-Recall Graph of Zernike Moments Feature

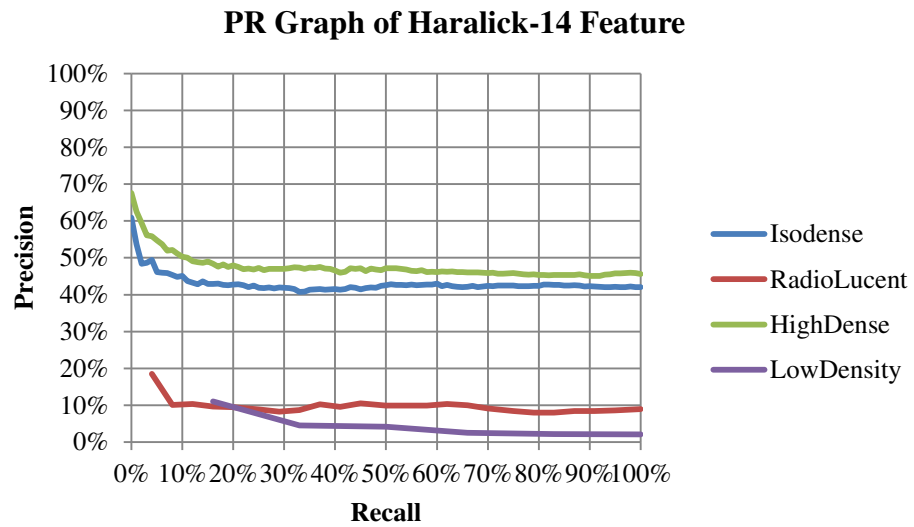


B.3 Precision-Recall Graphs of Density Property Queries

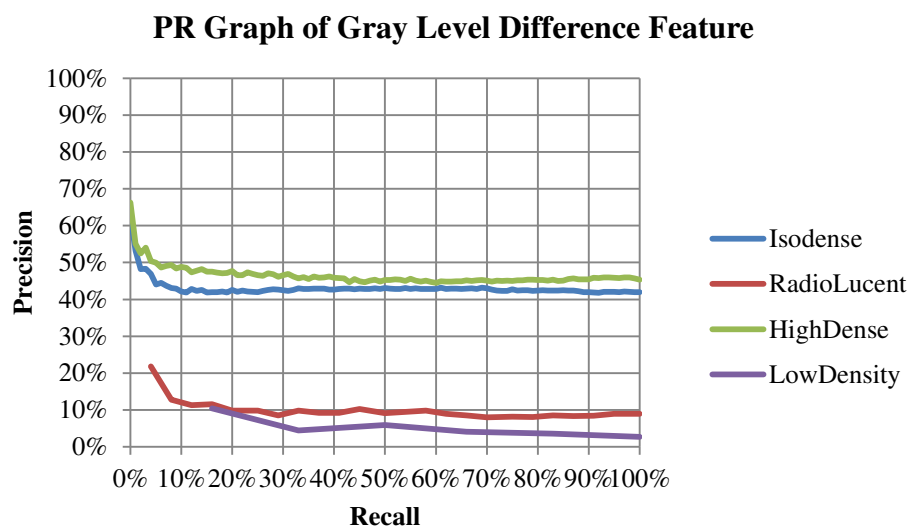
B.3.1 Precision-Recall Graph of Edge Histogram Feature



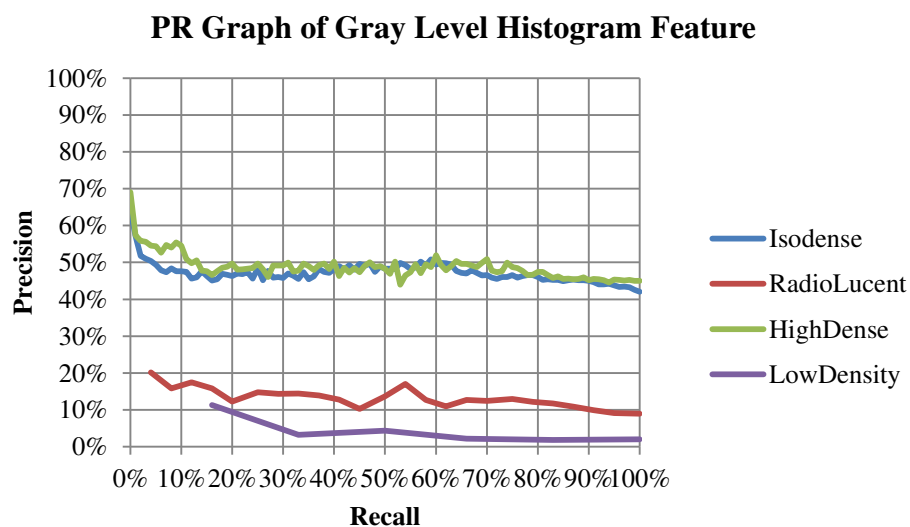
B.3.2 Precision-Recall Graph of Haralick-14 Feature



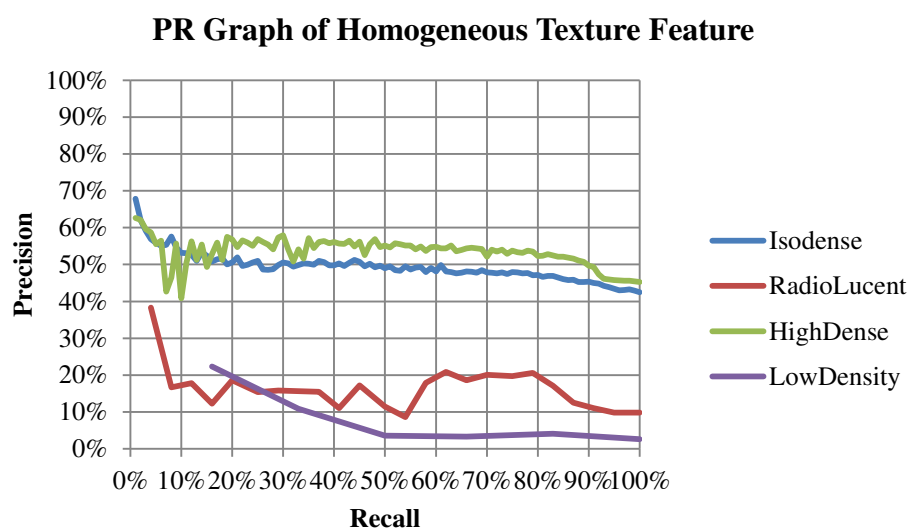
B.3.3 Precision-Recall Graph of Gray Level Difference Feature



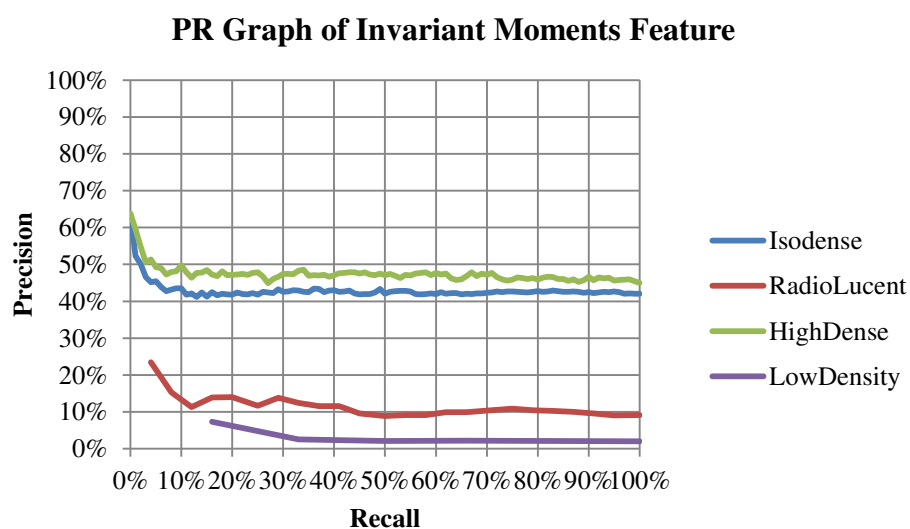
B.3.4 Precision-Recall Graph of Gray Level Histogram Feature



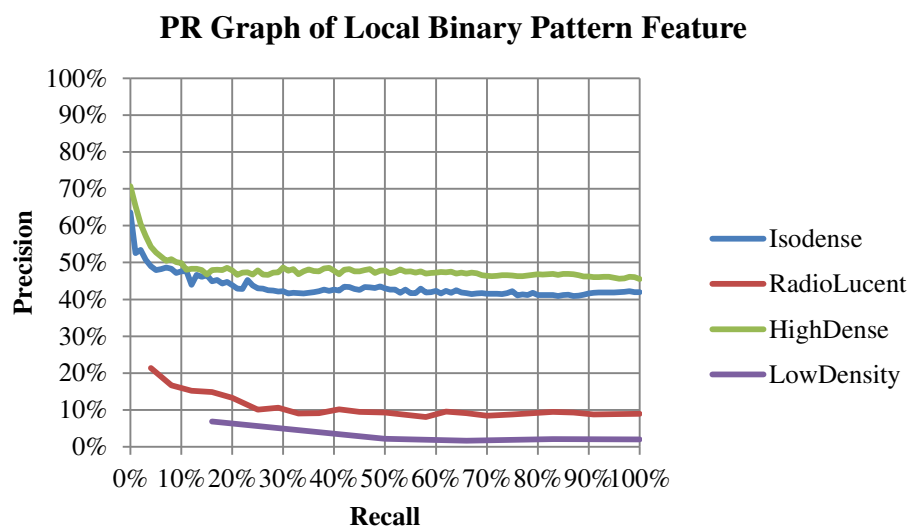
B.3.5 Precision-Recall Graph of Homogeneous Texture Feature



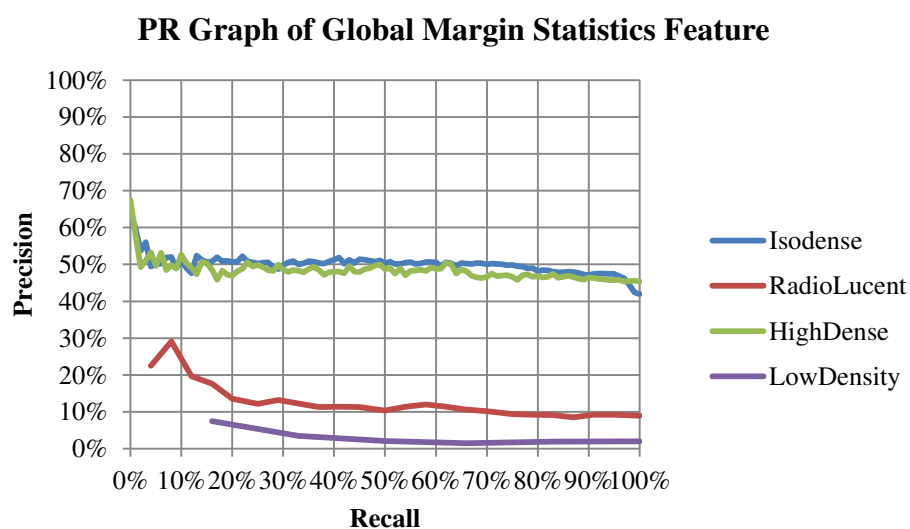
B.3.6 Precision-Recall Graph of Invariant Moments Feature



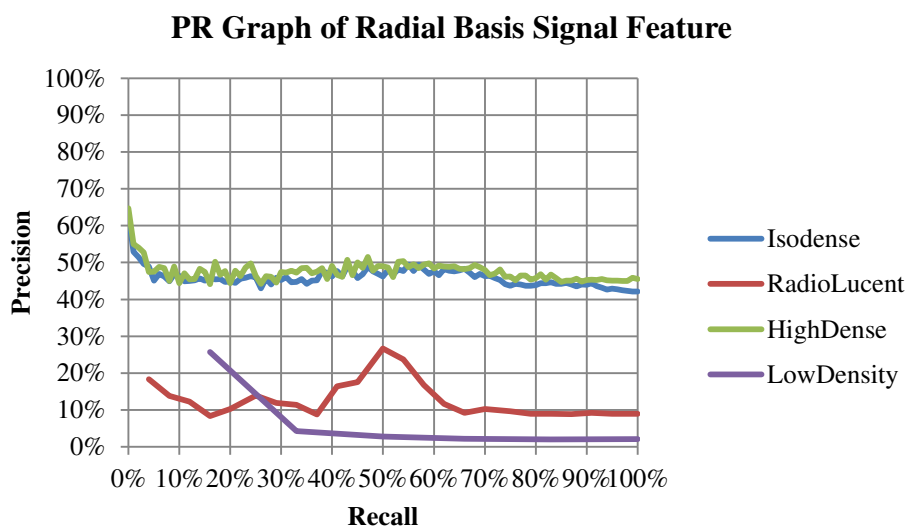
B.3.7 Precision-Recall Graph of Local Binary Pattern Feature



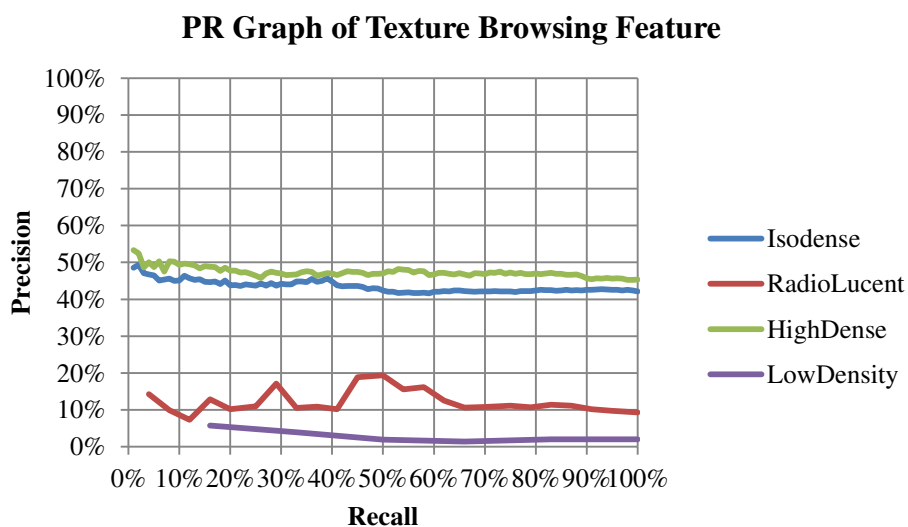
B.3.8 Precision-Recall Graph of Global Margin Statistics Feature



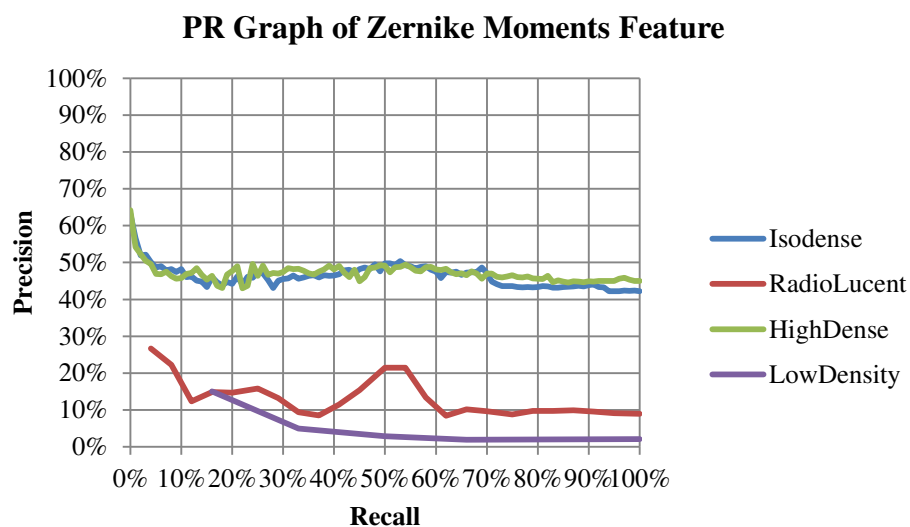
B.3.9 Precision-Recall Graph of Radial Basis Signal Feature



B.3.10 Precision-Recall Graph of Texture Browsing Feature

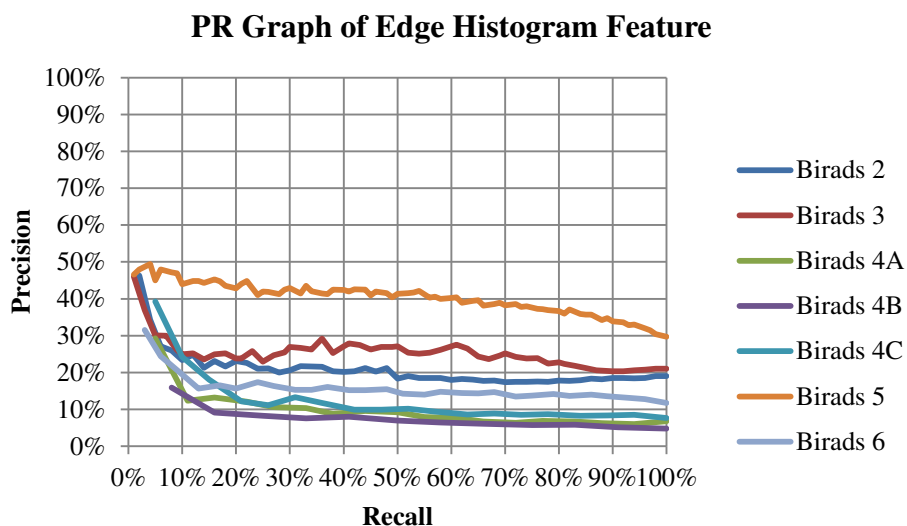


B.3.11 Precision-Recall Graph of Zernike Moments Feature

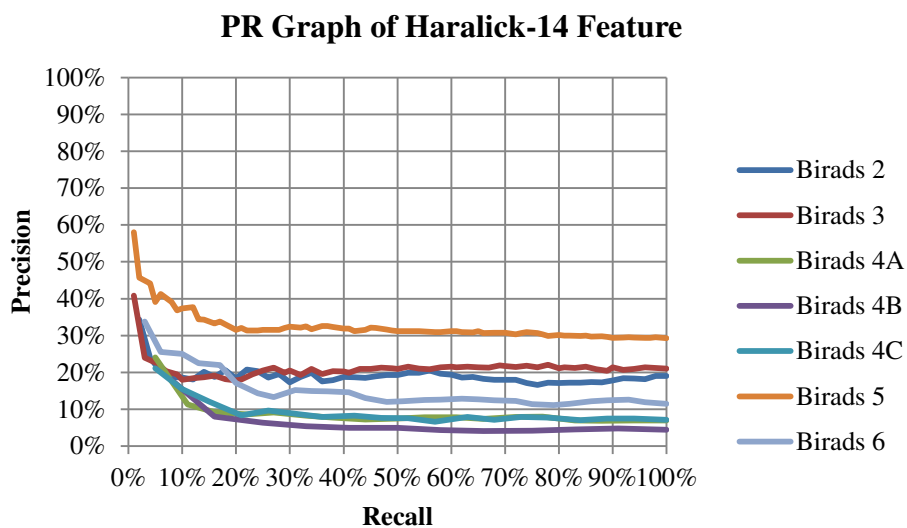


B.4 Information Retrieval Performance of BI-RADS Property Queries

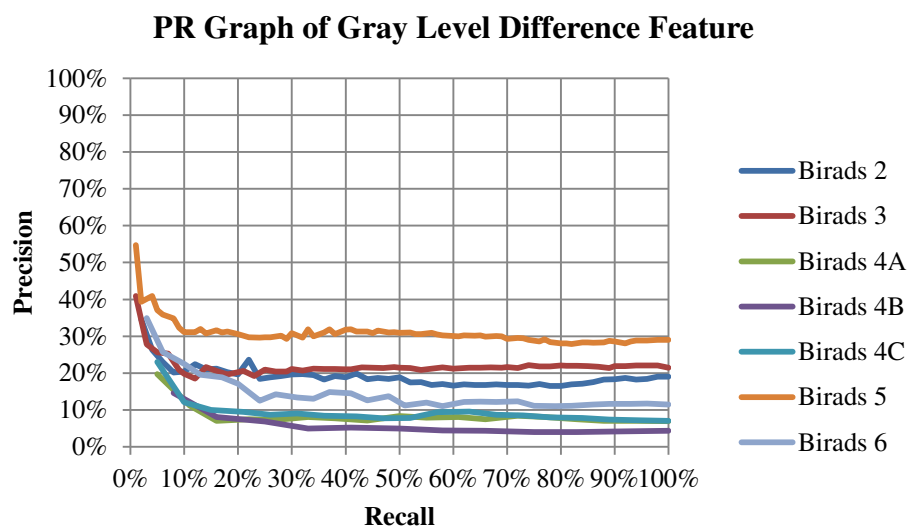
B.4.1 Precision-Recall Graph of Edge Histogram Feature



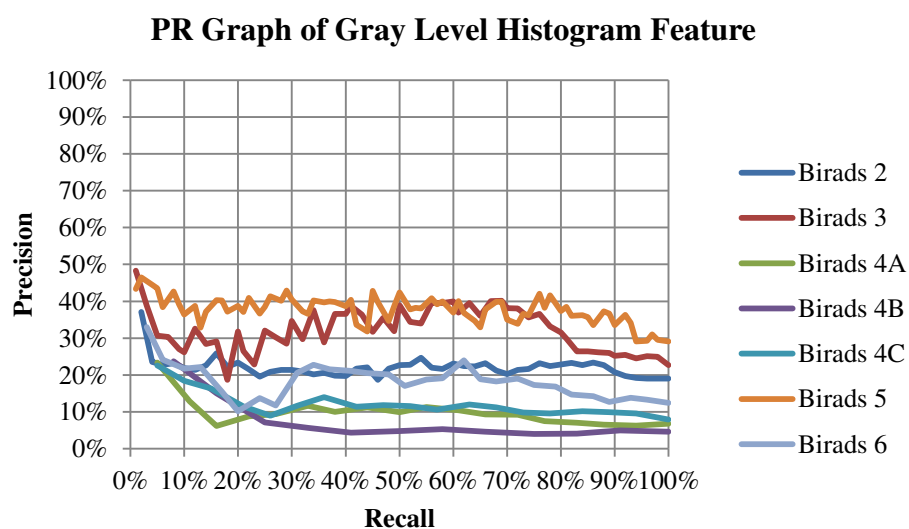
B.4.2 Precision-Recall Graph of Haralick-14 Feature



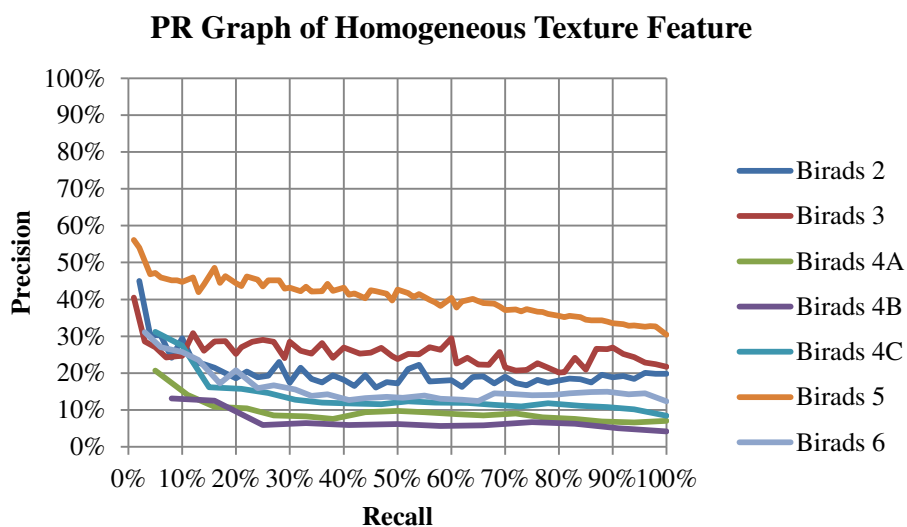
B.4.3 Precision-Recall Graph of Gray Level Difference Feature



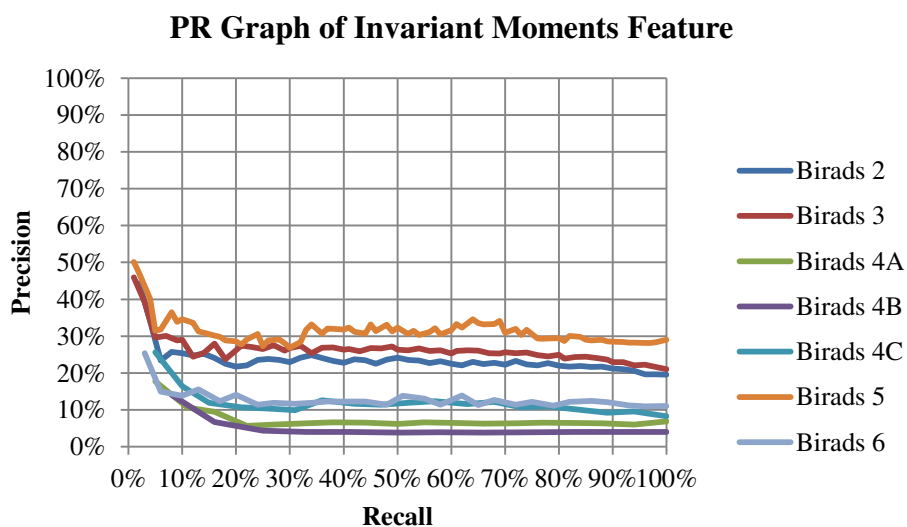
B.4.4 Precision-Recall Graph of Gray Level Histogram Feature



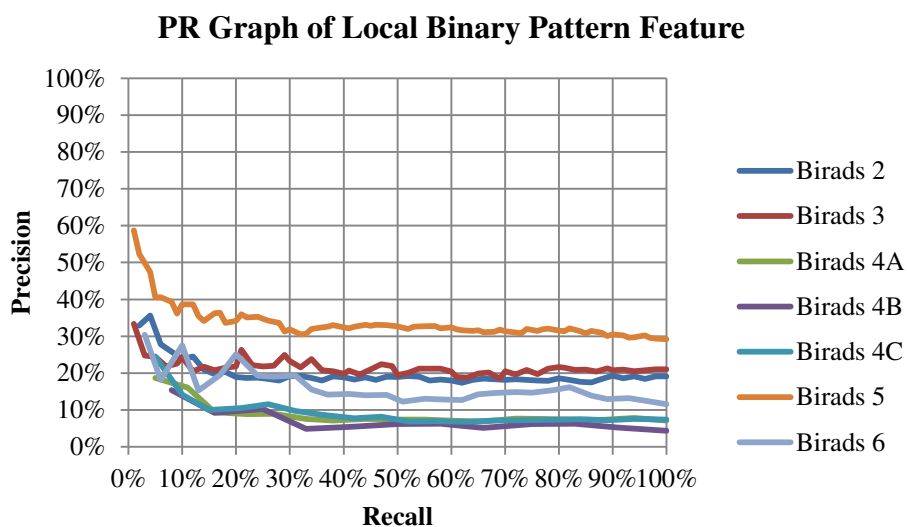
B.4.5 Precision-Recall Graph of Homogeneous Texture Feature



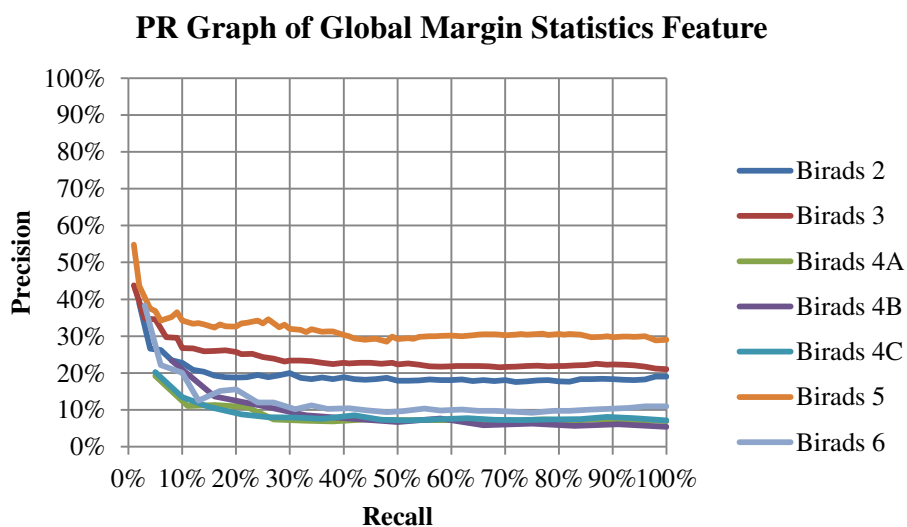
B.4.6 Precision-Recall Graph of Invariant Moments Feature



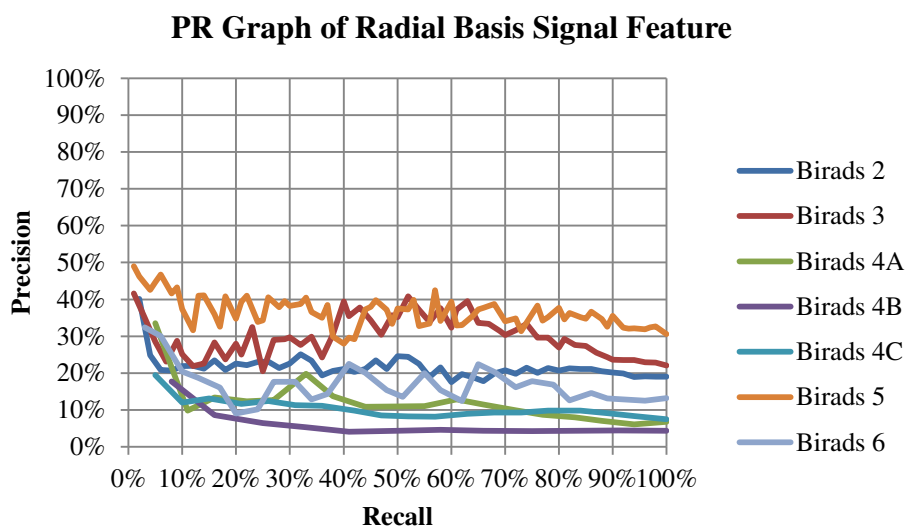
B.4.7 Precision-Recall Graph of Local Binary Pattern Feature



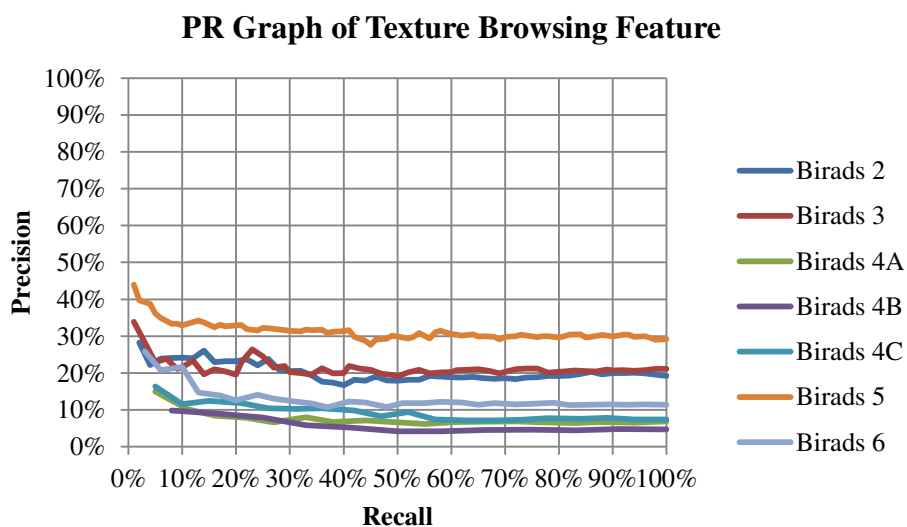
B.4.8 Precision-Recall Graph of Global Margin Statistics Feature



B.4.9 Precision-Recall Graph of Radial Basis Signal Feature



B.4.10 Precision-Recall Graph of Texture Browsing Feature



B.4.11 Precision-Recall Graph of Zernike Moments Feature