ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE

A CONTRIBUTION TOWARDS CONTENT-BASED IMAGE RETRIEVAL ISSUE : COMPARISON BETWEEN GLCM, 2DPCA, AND SURF IN CBIR FOR SCREENING MAMMOGRAPHY ANALYSIS

M.Sc. THESIS

Nabila Sabatini PURWADI

Department of Electronics & Communication Engineering

Biomedical Engineering Program

JUNE 2016

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE

A CONTRIBUTION TOWARDS CONTENT-BASED IMAGE RETRIEVAL ISSUE : COMPARISON BETWEEN GLCM, 2DPCA, AND SURF IN CBIR FOR SCREENING MAMMOGRAPHY ANALYSIS

M.Sc. THESIS

Nabila Sabatini PURWADI
(504131413)

Department of Electronics & Communication Engineering

Biomedical Engineering Program

Thesis Advisor: Dr. Serkan Türkeli

JUNE 2016

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

İÇERİK TABANLI GÖRÜNTÜ ERİŞİM SİSTEMİ KONUSUNA YÖNELİK KATKI: TARAMA MAMOGRAFİ ANALİZİ İÇİN CBIR GLCM, 2DPCA, VE SURF KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

Nabila Sabatini PURWADI
(504131413)

Tez Danışmanı: Dr. Serkan Türkeli

HAZİRAN 2016

Nabila Sabatini PURWADI, a M.Sc. student of ITU Graduate School of ScienceEngineering and Technology 504131413 successfully defended the thesis entitled "A CONTRIBUTION TOWARDS CONTENT-BASED IMAGE RETRIEVAL ISSUE : COMPARISON BETWEEN GLCM, 2DPCA, AND SURF IN CBIR FOR SCREENING MAMMOGRAPHY ANALYSIS", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Dr. Serkan Türkeli** Istanbul Technical University	.....
Jury Members :	**Doç. Dr. İlker Bayram** Istanbul Technical University	.....
	Doç. Dr. Serkat Özekes Üsküdar University	.....
		.....

Date of Submission : **29 April 2016**
Date of Defense : **8 June 2016**

To my dear brother Feaza; stay huer.

FOREWORD

I would like to offer my deepest gratitude to my advisor, Dr. Serkan Türkeli for his great guidance and help, not just in this thesis, but in my study in Turkey as well.

I would also like to thank Araş. Gör. Bilge Akkoca for her help in improving the result of this thesis.

Never ending gratitude to my soul-soothing brother Feaza, my extraordinary parents, my significant other, and all my friends in Istanbul. The special ones who tremendously helped me directly while making this thesis; Majd, Audry, Yaya, Tysha, Muhammad, Vera, and Dila. Staff from Indonesian General Consulate who gave me encouragements and help; Mbak Maya and Mbak Ida.

June 2016 Nabila Sabatini PURWADI

TABLE OF CONTENTS

Page

ABBREVIATIONS

CBIR	**:** Content-based Image Retrieval
CAD	**:** Computer-aided Diagnosis
GLCM	**:** Grey-level Co-occurrence Matrix
2DPCA	**:** Two-dimensional Principal Component Analysis
SURF	**:** Speeded-up Robust Feature
SVM	**:** Support Vector Machine
MI	**:** Mutual Information
SIFT	**:** Scale-invariant Feature Transform
CLAHE	**:** Contrast-limited Adaptive Histogram Equalization

LIST OF TABLES

Page

LIST OF FIGURES

Page

A CONTRIBUTION TOWARDS CONTENT-BASED IMAGE RETRIEVAL ISSUE : COMPARISON BETWEEN GLCM, 2DPCA, AND SURF IN CBIR FOR SCREENING MAMMOGRAPHY ANALYSIS

SUMMARY

Breast cancer is one of the leading cause of cancer deaths in women worldwide. One of the prevention methods for breast cancer death is making use of screening mammography. Screening mammography is used to detect anomaly in the breast so those with abnormal breasts can get earlier treatment, because some abnormality is not palpable, thus might go undetected by the patient or doctor.

There have been active research areas surrounding computer-aided diagnosis (CAD) to aid radiologists on making diagnostics based on mammography images. The aim of these systems is not to replace doctors, but to aid them during their diagnosis. Some CAD is aiding doctors by providing them with previous similar cases, or case-based reasoning. In mammography, such system is studied in the field of Content-Based Image Retrieval (CBIR).

There are several issues in the field of CBIR. Some studies are researching about how the semantic layer is going to be used, some are about which low-level feature and similarity metric is used, and some are researching about the data management part of CBIR.

This thesis is trying to contribute towards the second problem of CBIR, in which we are trying to compare Grey-level Co-occurrence Matrix (GLCM), Two-dimensional Principal Component Analysis (2DPCA), and Speeded-Up Robust Features (SURF) as low level features for mammography in CBIR system.

First, we are trying to find the best parameter for each methods by comparing them in classification of normal and abnormal class of mammography images from MIAS database. After that, with the best parameter taken from the previous step, we are trying to see the performance of each methods when the database is divided into 9 class; 3 subclasses of normal class and 6 subclasses of abnormal class. Lastly, we are simulating the CBIR environment by doing the image retrieval stage while comparing euclidean distance and MI for similarity measure.

The result of this thesis shows that SURF performed the best compared with GLCM and 2DPCA in differentiating normal and abnormal, differentiating breast density and abnormality, and for image retrieval.

İÇERİK TABANLI GÖRÜNTÜ ERİŞİM SİSTEMİ KONUSUNA YÖNELİK KATKI: TARAMA MAMOGRAFİ ANALİZİ İÇİN CBIR GLCM, 2DPCA, VE SURF KARŞILAŞTIRILMASI

ÖZET

Meme kanseri tüm dünyada kadınların kansere bağlı ölümlerin başında gelen nedenlerden biridir. Amerika Birleşik Devletleri'nin kadınları bir yılda en az 200.000 kadınlar meme kanseri olmaktadır. Her 8 kadından biri hayatın boyunca meme kanserine tanımlanmaktadır. Meme dokusunun herhangi bir meme yerinden gelebilir. En çok ortaya çıkan tipi; "duktal" kanser ve "lobüler" kanserdir. Duktal kanser, memelerin kanallarından gelen bir kanser türüdür. Lobüler kanser ise süt üreten bezlerden bir kanser türüdür. Diğer dokulardan gelen, az bulunan kanserler medüller, tübüler, müsinöz gibi türleri de vardır. Meme kanseri oluşumunda mühim olan genetik değişikliklerdir. Genetik yapıda çeşitli faktörlerin ve normal yaşlanmanın kanser etkilerin en sık görülen bozukluklar kanserin sebebidir. Ama meme kanserlerinin sadece %7-9'luk bir kısmı ailesel kalıtsaldır. Özellikle anne tarafında genç yaşta meme kanseri ve erkek meme kanseri oluşması ailesel kaynaklanan bir kanseri işaret edebilir.

Meme kanserini teşhis edebilmek için tarama mamografi yöntemi kullanılıyor. Tarama mamografi yöntemiyle hastalık anormal göğüslü insanlarda erken tedavi için kullanılıyor. Çünkü bazı durumlarda anormallik tespit edilemiyor ve doktorlar tarafından görülemiyor. Hiç bir şikâyeti olmayan kadınlarda, meme kanserlerini erken aşamada öğrenmeleri için Tarama mamografi kullanılır. Eğer kontrol amaçlı mamografi düzenli olarak yapılınca erken tanı olasılığını arttırarak tedavinin başarı ihtimalı önemli bir oranda arttırır. 40 yaşından fazla her kadınlar yılda bir kez kontrol için mamografi yaptırması önerilmektedir. Ayrıca daha yüksek risk taşıdığı var sayılan kadınlar doktor önerisi ile daha erken yaşlarda kontrola gitmeleri önerilmektedir.

Radyologların mamografi görüntülerine dayalı teşhislerini kolaylaştırmak için Bilgisayarlı Tanı (Computer-aided Diagnosis, CAD) sisteminde aktif araştırmalar yapılıyor. Bu sistemin hedefi doktorların yerini almak değil, aksine doktorlara teşhisleri sırasında yardımcı olmaktır. Bazı CAD sistemleri önceki benzer vakalarda ya da durum tabanlı çıkarsamalar sunarak doktorlara yardım ediyor. Mamografi alanında bu tür sistemler İçerik Tabanlı Görüntü Erişim Sistemi (Content-based Image Retrieval, CBIR) alanında incelenmistir.

Genelde, CBIR sisteminin iki aşamadan oluştuğu tahmin edilebilir. İlk aşamada, görüntü veri tabanlarındaki her mamografi görüntüden öznitelikler çıkarılır. İkinci aşamada, bir sorgu görüntüsünün öznitelik hesaplanır ve daha sonra en benzer mamografi görüntüleri bulmak için veri tabanlarındaki görüntü öznitelik vektörleri ile karşılaştırılır. Sorgu görüntüsüyle veritabanlarındaki görüntüler arasındaki özniteliklerin karşılaştırılması, benzerlik ölçüsü ile yapılmaktadır. Görüntüye ya da veri tabanındaki en benzer görüntüye erişmek için en benzer görüntülerin indeksi verilmiştir.

CBIR alanında çeşitli sorunlar vardır. Bazı çalışmalar anlamsal katmanın nasıl kullanılacağı hakkında araştırmalar yapıyor, bazılarıysa hangi düşük seviye özelliği ve benzerlik metriği kullanıldığını araştırıyor, ve bazıları da veri yönetimi bölümü hakkında araştırmalar yapıyor.

Bu tez CBIR sisteminin ikinci sorununa katkıda bulunmaya çalışıyor. Biz CBIR mamografi alanında Gri seviye Eş-oluşum Matrisi (Grey-level Co-occurrence Matrix, GLCM), İki boyutlu Temel Bileşen Analizi (Two-dimensional Principal Component Analysis, 2DPCA) ve Hızlandırılmış Sağlam Özellikleri (Speeded-up Robust Feature, SURF) yöntemlerini karşılaştırmak için çalışıyoruz.

GLCM, dokusal öznitellik çıkarılması için bir yöntemdir. GLCM M. Haralick tarafından ortaya atılmış bir özellik çıkarma yöntemi olup, gri tonlu bir görüntünün özniteliğini çıkarmaya yaramaktadır. GLCM, iki komşu piksel arasındaki ilişkiyi tanımlar; birinci referans pikseli, ikincisi de komşu piksel olarak bilinir. Matristeki dağılım, pikseller arasındaki mesafe ve açıya göre ayarlanır. Bu matris, X boyutlu bir kare matris olup, matrisin her bir elemanı d mesafesindeki i ve j piksel değerlikli çiftin oluşum sayısını belirtir. Görüntü arşivindeki görüntülerin kontrast bileşeni diğer doku özellik vektörü bileşenlerinden değer olarak önemli ölçüde büyük olduğundan benzerlik ölçümünde diğer bileşenlerin etkisini azaltmaktadır. Bu nedenle görüntü arşivindeki bütün görüntülerin kontrast parametresi normalleştirilmiştir.

PCA, orijinal p değişkeninin varyans yapısını daha az sayıda ve bu değişkenlerin doğrusal bileşenleri olan yeni değişkenlerle ifade etme için bir yöntemidir. PCA ve 2DPCA arasındaki fark; 2DPCA temel bileşen bir vektördür ise PCA, temel bileşen bir skalerdir. Aralarında korelasyon bulunan p sayıda değişkeni açıkladığı yapıyı, aralarında korelasyon bulunmayan ve sayıca orijinal değişken sayısından daha az sayıda (p>k) orijinal değişkenlerin doğrusal bileşenleri olan değişkenlerle ifade etme yöntemidir. Veri matrisinde yer alan p değişkenin doğrusal bileşenlerini kovaryans matrisinin ya da korelasyon matrisinin kullanılabilir. Eğer değişkenler aynı birim veya karşılaştırılabilir birimlerdeyse, değişken varyansları aynı boyuttaysa varyans-kovaryans matrisi kullanılır. Bu durumlar sağlanmadığında varyans-kovaryansmatrisi yerine korelasyon matrisi kullanılır.

Son kullandığımız yöntemi SURF'dur. SURF algoritmasıyla bir görüntüden çok sayıda öznitelik çıkarılabilir. Çıkarılan öznitelik vektörlerinin her biri 64 boyutludur. Öznitelik çıkarımı için görüntü farklı ölçeklerde incelenerek önemli noktalar bulunur. Görüntünün farklı ölçeklerde incelenmesi için taklit Gauss'ların farkı yöntemi kullanılır. Bu yöntemde görüntüye farklı varyanslı Gauss filtreleri uygulanır. Bu filtrelerin uygulanmasıyla görüntü farklı miktarlarda bulanıklaştırılmış olur. Bu görüntülerin farkı alınarak görüntüdeki kenarlar ve köşeler elde edilir. Bu farkların bazı yöntemlerle elenmesiyle önemli noktalar bulunur. Önemli noktalarda ve komşularında gradyanlar hesaplanır. Her önemli nokta için bir genlik ve yön bilgisi hesaplanır. Bir Gauss penceresi kullanılarak önemli noktaya yakın olan noktaların etkisi arttırılırken, uzak olanlarınki azaltılır. Hesaplanan yön histogramları kullanılarak öznitelik vektörleri elde edilir.

Görüntülerden çıkarılan özniteliklerin sınıflandırılması için Destek Vektör Makinesi (Support Vector Machine, SVM) kullanılmıştır. SVM yöntemi iki sınıftaki öznitelik vektörlerinin ortasındaki optimal hiperdüzlemi bulmaya çalışır. Bu yöntem temelde doğrusal olarak ayrıştırılabilen iki sınıfın sınıflandırılması için önerilmiştir; ancak bazı ek işlemlerle daha çok sınıf için, çekirdek fonksiyonlarının kullanımıyla da

doğrusal olarak ayrıştırılamayan veri kümelerinde kullanılabilmektedir. Bunun için, doğrusal olarak sınıflandırılamayan veri bir çekirdek fonksiyonuyla doğrusal olarak sınıflandırılabileceği bir uzaya taşınır.

İlk olarak, MIAS veri tabanında normal ve anormal sınıf sınıflandırılmasında karşılaştırarak her yöntem için en iyi parametreyi bulmaya çalışıyoruz. Sonra, 3 alt sınıf ve 6 anormal alt sınıf olmak üzere 9 sınıfa ayrılan veritabanı bize önceki adımla alınan en iyi parametreyle her yöntemin performansını gösteriyor. Son olarak, görüntü alma adımını kullanarak CBIR sistemine benzer bir sistem oluşturuyoruz. Bu durumda Öklit Distans ve MI benzerlik ölçüsünü karşılaştırmış oluyoruz.

Bu tezin sonucu bize GLCM, 2DPCA, ve SURF yöntemlerinin normal ve anormal, meme yoğunluğu ve anormallık, ve görüntü alma adımı alanlarında karşılaştırıldığında SURF yönteminin en iyi yöntem olduğunu gösteriyor.

1. INTRODUCTION

1.1 Breast Cancer

Breast cancer is the 2nd most common cancer for both sexes worldwide in 2012 [5]. The mortality rate for breast cancer, especially for 40-55 age range, takes the first place in women. In 2012, 522.000 people died because of breast cancer, approximately 6% of all cancer deaths [5]. In the United States, the incidence and mortality rate of breast cancer generally raise in parallel with age as shown in Figure 1.1.

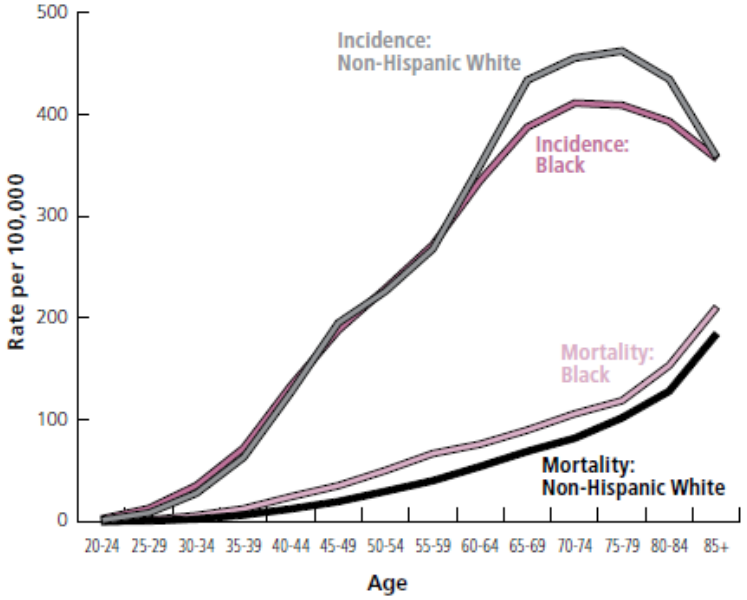


Figure 1.1 : Age-specific female breast cancer incidence and mortality rates in the US, 2008-2012. Sources: Incidence: North American Association of Central Cancer Registries (NAACR), 2015. Mortality: US mortality data, National Center for Health Statistics, Centers for Disease Control and Prevention [2].

One of the methods to reduce the mortality rate of breast cancer is to employ mammography in the pre-diagnosis stage. It has been proven that screening mammography has been able to reduce death caused by breast cancer, even as high as 28% in a population study in Norway [6]. Mammography is claimed to be the

most effective method for early detection of lesions and microcalcifications, which are important signs of early stage of breast cancer [7] [8] [9].

1.2 Mammography

Mammography is one of the x-ray based imaging modality used to obtain detailed internal structure human breasts. It is specially designed for breast imaging with high-contrast and high-resolution using low-dose x-ray. Typically, they are in order of 0.7 mSv.

Mammography can determine changes in breast even before they are detected by the patient or doctor. This is because some mass could be identified by mammography even two years before it becomes palpable [10]. The high sensitivity of mammography is highly dependent on the patient's age, the size and conspicuity of the lesion, the hormone status of the tumor, the density of the breast, and the overall image quality and the skills and experience of the radiologist [11].

One of the main problems is that it is very difficult to distinguish between benign and malignant lesions just from mammogram images [12]. If an abnormality is present, additional imaging methods like ultrasound or biopsy is usually performed to confirm its malignancy. Mammography and ultrasound are usually paired to determine whether biopsy is needed for the patient. There are two kinds of mammography scan; screening mammography and diagnostic mammography.

Screening mammography is done for women without symptoms. The main purpose of screening mammography is to detect abnormalities in their early stages [13]. Since early diagnosis of breast abnormalities that could possibly lead to breast cancer greatly improves chances of successful treatment, every women older than 40 is suggested to perform screening mammography every year [10].

Diagnostic mammography is more detailed and is intended for women who already discover abnormalities with their breasts. The main purpose of diagnostic mammography is to determine the exact location of the mass or abnormalities and most of the time it is used to predict the malignancy before ultrasound or biopsy is taken [10].

1.3 Content-Based Image Retrieval

Content-based Image Retrieval (CBIR) is an approach to retrieve digital images from a data repository by the contents of the images as opposed to the description of the images. CBIR is also referred to as Query by Image Contents (QBIC).It retrieves images by the low-level features of the image such as color, texture, shape, salient point, and other features that give most information about the image such as statistical measures [14].

In medical images, the main goal of CBIR is to provide radiologists with examples of labelled images with known pathology that are similar to the case being evaluated to boost their diagnostic accuracy. In order for a CBIR system to be beneficial to aid radiologist in their diagnosis, the retrieved images must be relevant to the query image as perceived by the radiologists. The importance of this lies in the fact that physicians usually try to recall similar cases by seeking images with similar pathological features to a given image [15].

Although CBIR has been used widely in several areas, there are several issues in the development of this technology. These issues can be grouped into three major subjects [16]:

1. What kind and how semantic layer is used?
2. Which low-level features is used and which similarity metrics is used?
3. How to make data management and organization more effective?

Medical image archives such as Picture Archiving and Communication System (PACS) and Computer-Aided Diagnostics (CAD) are some of the areas that CBIR is most needed. CBIR system can be viewed as a CAD tool to provide evidence for case-based reasoning.

1.4 Recent Studies

In recent years, there is a growing number studies that focused on CBIR on mammography with several different approaches and methods.

El Naqa *et.al* have studied the similarity learning approach to CBIR with application to digital mammography [17]. They were using 9 shapes of microcalcifications (MCC) in mammographic images, namely cross-sectional area, density, invariant moment, compactness, scatteredness, moment signature, eccentricity, solidity, and normalized Fourier descriptor, as features. Human feedback was used as first step in classifying whether the mammographic image is relevant enough to be classified by the system. In their similar previous study [18], they compared artificial neural network (ANN) and support vector machines (SVM) for the image retrieval part and found that SVM outperformed ANN in this case. Thus SVM was used in this study as classifier and the whole system achieved 76.7% highest average matching fraction.

Chia-hung Wei *et.al* were using MIAS database to study general framework for CBIR to differentiate "normal" and "abnormal" ROIs in mammographic images [19]. They were using 11 textural features based on grey-level co-occurrence matrix (GLCM). In the retrieval part, *L2* norm distance was used and it managed to achieve 51% maximum precision and 19% recall rate.

A validation study by Yihua Lan *et.al* studied CBIR-CAD approaches, specifically to compare ROI and similarity measures for their system [20]. They were using DDSM as their database and *k*-nearest neighbor (KNN) and mutual information (MI) as similarity measurement methods. Areas under ROC curve was used as evaluation parameter and it achieved 0.851 at its highest.

Jinchang Ren presented a paper [21] to compare the performance of artificial neural network (ANN) and SVM as classification method of MCCs in mammographic images taken from DDSM database. With both unbalanced and balanced data, SVM outperformed ANN with 0.671 and 0.797 precision rate respectively.

CADx of mammographic masses was studied by Shaoting Zang *et.al* with scale-invariant feature transform (SIFT) as feature extraction method [22]. SIFT was chosen because most CBIR methods fall short of scalability in the retrieval stage [23] which makes their diagnostic accuracy restricted. The features extracted with SIFT then searched in a vocabulary tree that stores quantized features of previously diagnosed ROIs. DDSM database was used in this study and it achieved 88.4% highest

precision rate. The [22] was using *k*-means clustering with the same visual vocabulary tree and achieved areas under ROC of 0.91.

Wiesmüller *et.al* presented a paper which used a generalized version of GLCM called grey-level aura matrix (GLAM) [24] with DDSM database. The advantage of GLAM is its features are completely independent from shape and size of ROI. KNN was used as a classification method and the whole system achieved 82.2% average precision rate.

Honda *et.al* was investigating textural features in their paper about CBIR of breast lesions classified according to BI-RADS [25]. They were using Haralick's textural features [1], which is based on GLCM to automatically extract features from ROIs previously established in digitized mammograms. Principal component analysis (PCA) was used to simplify indexing and retrieving process. The precision values between 22% and 100% using private database.

Novel approach for feature extraction called fuzzy texton was introduced by Putheli *et.al* in their study of classification of normal, benign, and malignant tissues in mammographic images [26]. Using MIAS database and SVM in retrieval process, they managed to achieve 90% mean success rate.

Kinoshita *et.al* presented a CBIR approach using visual features related to breast density patterns [27]. With private database, they are using mainly grey-level-based features such as Haralick's texture features with addition to Radon domain and granulometric measure. Kohonen's self organizing map (SOM) neural network is used as classifier in retrieval stage and they achieved 79%-83% precision rate.

Zheng *et.al* was trying to improve performance of CBIR schemes in searching for similar breast mass regions using commonly used retrieval methods, namely MI, Pearson's correlation, and KNN [28]. The feature used was 14 features based on pixel values in ROI. Under these circumstances, KNN managed to perform the best with areas under curve valued at 0.893.

There are also several papers which were using IRMA database. Oliviera *et.al* were trying to make a CBIR system using singular value decomposition (SVD) as feature extraction method [29]. SVM was used in retrieval stage and it achieved 90% precision rate on the first 10% of retrieved images. 2D-PCA was used by Deserno *et.al* in [30]

and [31] in feature extraction stage. Also with SVM, they managed to achieve average precision rate of 83%-97%.

1.5 Purpose of Thesis

This study is trying to contribute to one issue in CBIR as mentioned in section 1.3, which is; *Which low-level features is used and which similarity metrics is used?* The detailed purposes of this thesis can be summarized as follows:

1. To compare GLCM, 2DPCA, and SURF for feature extraction methods with SVM to differentiate between normal and abnormal case of mammogram images.
2. To compare GLCM, 2DPCA, and SURF for feature extraction methods with SVM to differentiate between breast density subclasses in normal breast and abnormality subclasses for abnormal breast.
3. To compare GLCM, 2DPCA, and SURF with euclidean distance and mutual information for CBIR with breast density subclasses for normal breasts and abnormality subclasses for abnormal breasts.

2. MATERIAL AND METHODS

2.1 Software and Hardware Specifications

The main software used in this thesis is Matlab R2015a running in Microsoft Windows 7 64 bit operating system. LibSVM [32] is used for the SVM calculation and cross-validation testing is done with Waikato Environment for Knowledge Analysis (WEKA) version 3.6.13 software. GLCM toolbox for Matlab is made by Patrik Brynolfsson [33]. Mutual Information toolbox is made by Hanchuan Peng [34]. Manual segmentation of each of the images are done using GNU Image Manipulation Program (GIMP) version 2.8.16.

The processing hardware used Intel Core i3 CPU with 2.40 GHz each and 4 GB RAM.

2.2 Dataset Explanation

Mammography dataset used in this thesis is taken from the Mammographic Image Analysis Society (MIAS) database [35] consisted of 322 images, with 209 images of normal class and 113 images abnormal class. The dataset does not differentiate between craniocaudal (CC) or mediolateral oblique (MLO) in terms of mammographic views.

To avoid bias, we took 113 normal images randomly to build a balanced database, thus making the final dataset of 226 images of normal and abnormal images with 113 images in each class.

Each images has properties as shown in Table 2.1 to Table 2.3.

The malignancy property has two options; B (Benign) and M (Malignant), which is not used in this thesis. The x and y coordinate correspond to the central point of the abnormality. The approximate size property corresponds to the radius of abnormality.

The distribution of breast density subclass in normal class and abnormality subclass in abnormal class are shown on Table 2.4 and Table 2.5 respectively.

Table 2.1 : Example of properties labeled in one image in MIAS database.

Properties	Code
Filename	mdb001
Breast density	G
Abnormality class	CIRC
Malignancy	B
x-coordinate	535
y-coordinate	425
Approximate size	197

Table 2.2 : Breast density explanation from MIAS properties.

Breast density code	Type
F	Fatty
D	Dense-glandular
G	Fatty-glandular

The original, unprocessed images sample in normal and abnormal classes are shown in figure 2.1 and 2.2.

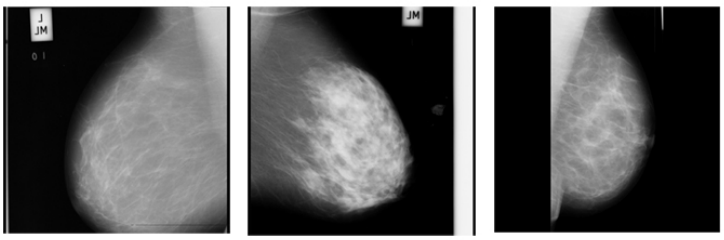


Figure 2.1 : Samples of normal breasts in MIAS database. Left to right : Fatty tissue, Dense tissue, Fatty-Glandular tissue.

2.3 Contrast-limited Adaptive Histogram Equalization

Contrast-limited Adaptive Histogram (CLAHE) is an adaptive contrast enhancement method based on adaptive histogram equalization (AHE), where a local histogram is calculated and a different grayscale transform is computed at each image location based on the neighbourhood. The pixel's intensity transformed to a value between the display range, proportional to the pixel intensity's rank in the local intensity histogram.

The basis of CLAHE is histogram equalization equation as shown in equation 2.1. Variable r represents the grey levels of the image, and equation 2.1 produces a level s for each pixel with grey level r in the original image. $P(r)$ corresponds to the

Table 2.3 : Abnormality explanation from MIAS properties.

Abnormality code	Type
CALC	Calcification
CIRC	Circumscribed
SPIC	Spiculated mass
MISC	Other, ill-defined mass
ARCH	Architectural distortion
ASYM	Asymmetry

Table 2.4 : Distribution of breast density subclass in normal class.

Normal Class	
Breast density	Total
Fatty	33
Dense-glandular	37
Fatty-glandular	43

Table 2.5 : Distribution of abnormality subclass in abnormal class.

Abnormal class	
Abnormality	Total
Calcification	23
Circumscribed	23
Spiculated mass	19
Other, ill-defined mass	14
Architectural distortion	19
Asymmetry	15

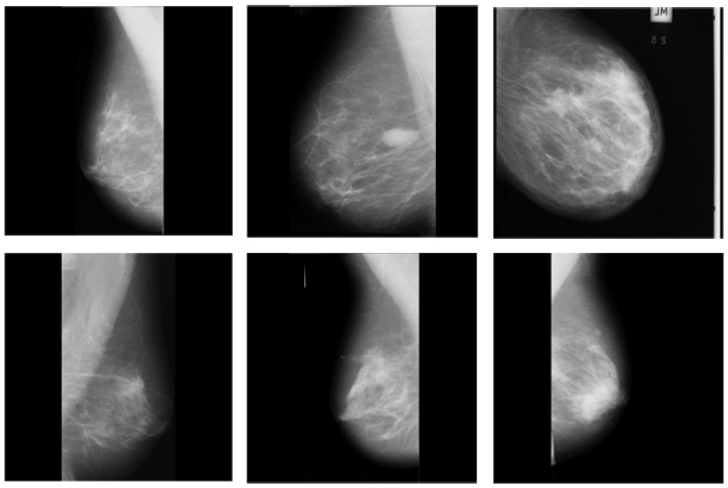


Figure 2.2 : Samples of abnormal breasts in MIAS database. Top row, left to right : Calcification, Circumscribed, Spiculated mass. Bottom row, left to right: Ill-defined mass, Architectural distortion, Asymmetry.

probability of occurrence of grey level *rk* in the image, which can be determined by the histogram of the image [36]. Therefore, *T(r)* equalizes the histogram of the given image, resulted in a uniform histogram.

$$s = T(r) = \sum_{j=0}^{k} p_r(r_j) = \sum_{j=0}^{k} \frac{n_j}{n}, k = 0, 1, 2, ...L \tag{2.1}$$

CLAHE is considered an enhancement of AHE, because a user-specified maximum value is imposed to the height of the local histogram in enhancement calculation. The enhancement is therefore reduced in uniform areas of the image, which prevents over-enhancement of noise and reduce edge-shadowing effect found in AHE [37].

CLAHE procedure is as follow [38] :

1. obtain a local histogram with the neighbours of every pixel,
2. clip the histogram with user-specified maximum value,
3. modify the histogram by redistributing pixels by equation 2.1,
4. and integrate the histogram up to the value of the pixel to obtain the final value.

Previous studies have already tested the effectiveness of CLAHE for processing mammography images. [39] was trying to improve detection of simulated spiculated mass in dense mammograms using CLAHE. Twenty student observers were asked to detect the orientation of phantom spiculated mass after it was embedded into dense mammographic images. The result showed that CLAHE improves the rate of detection observers greatly compared to raw images. [40] was trying to detect calcifications in mammographic images using modified CLAHE. The result showed that the method has increased detectability of calcifications compared to the raw images.

2.4 Feature Extraction Methods

2.4.1 Grey-level co-occurrence matrix

Grey-Level Co-occurrence Matrix (GLCM) is a square matrix whose rows and columns containing grey level intensity values. Because of that, the dimension of

the GLCM matrix depends on the number of grey levels in the image. GLCM matrix is calculated using the formula in equation 2.2.

$$C_{\Delta x,\Delta y(i,j)} = \sum_{p=1}^{m}\sum_{q=1}^{n}\begin{cases}1, & I(p,q)=i\ and; I(p+\Delta x, q+\Delta y)=j\\ 0, & otherwise\end{cases} \tag{2.2}$$

I(x,y) is the image function, m and n are image width and height respectively, Δx and Δy are distance parameters defined by d and θ respectively. Figure 2.3 shows neighbouring pixels at d=1 and θ=(0°, 45°, 90°, 135°).

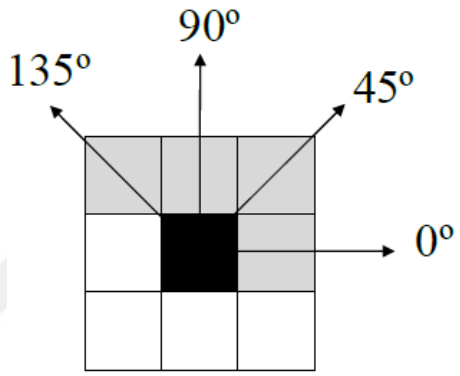


Figure 2.3 : Example of a single pixel with 8 neighbours at d=1 and θ=(0°, 45°, 90°, 135°).

After the co-occurrence matrix is calculated, each element of the matrix is normalized by using equation 2.3 to build a normalized co-occurrence matrix *(p(i,j))* as shown in equation 2.3.

$$p(i,j) = \frac{C(i,j)}{\sum_{k=1}^{N_g}\sum_{l=1}^{N_g} C(k,l)} \tag{2.3}$$

Statistical properties of GLCM are intensively used in texture related studies, as proposed by [1]. The features aim to descrive texture of an image by using spatial intensity dependencies. Table 2.6 shows the list of statistical features extracted from GLCM.

$p_x(i)$ in Table 2.6 refers to *i*th entry in the marginal-probability matrix obtained by summing the rows of *p(i,j)*. N_g in Table 2.6 and 2.3 refers to number of distinct grey levels in the quantized image.

GLCM has been used in a number of studies of CAD and CBIR for mammography. [19] was using GLCM to differentiate between normal and abnormal mammography

Table 2.6 : Statistical features extracted from GLCM [1].

Feature Description	Formal Definition
Angular Second Moment / Energy	$\sum_{i=1}^{N_g}\sum_{i=1}^{N_g} p(i,j)^2$
Contrast	$\sum_{n=0}^{N_g-1} n^2 \sum_{i=1}^{N_g}\sum_{i=j}^{N_g} p(i,j) \quad where\ n = \lvert i-j\rvert$
Correlation	$\frac{\sum_{i=1}^{N_g}\sum_{i=1}^{N_g} ijp(i,j)-\mu_x\mu_y}{\rho_x\rho_y}$
Sum of Squares : Variance	$\sum_{i=1}^{N_g}\sum_{i=1}^{N_g}(i-\mu_x)^2 p(i,j)$
Inverse Difference Moment / Homogenity	$\sum_{i=1}^{N_g}\sum_{i=1}^{N_g}\frac{1}{1+(i-j)^2}p(i,j)$
Sum Average	$\sum_{i=2}^{2N_g} ip_{x+y}(i)$
Sum Variance	$\sum_{i=2}^{2N_g}(i-SumEntropy)^2 p_{x+y}(i)$
Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i)log(p_{x+y}(i))$
Entropy	$-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)log(p(i,j))$
Difference Variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$
Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i)log(p_{x-y}(i))$
Information Measures of Correlation 1	$\frac{HXY-HXY1}{max\{HX,HY\}}$
Information Measures of Correlation 2	$(1-exp[-2.0(HXY2-HXY)])^{1/2}$
	$HXY = -\sum_i\sum_j p(i,j)log(p(i,j))$
	$HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)log(p_x(i)p_y(i))$
	$HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)p_y(j)log(p_x(i)p_y(i))$

images taken from MIAS database. With euclidean distance as similarity measure, it has managed to achieve maximum precision of 51%. [41] was trying to differentiate abnormalities from MIAS database for CBIR system. It has managed to achieve maximum mean precision of 76.9% with GLCM of distance 5. [42] was trying to differentiate breast density patterns in mammography images with GLCM textural features and self-organizing map neural network for retrieval method. It has managed to achieve precision rate in the range from 78%-83%.

2.4.2 Two dimensional principal component analysis

Two Dimensional Principal Component Analysis (2DPCA) is a feature extraction and data representation technique based on the classic Principal Component Analysis (PCA). PCA can be described as a method to build a less redundant and more compact representation of data, in which a reduced number of components can be independently responsible for data variation [43].

PCA transforms p-dimensional vector in another q-dimensional vector, by projecting the original vector into a lower-dimensional basis composed of q orthogonal axes of maximum variance, the principal components. Rather than 1D vectors basis used in PCA, the 2DPCA is using 2D matrices.

2DPCA is trying to project image A, with a matrix size of *m x n* pixels, onto X by linear transformation :

$$Y = AX \tag{2.4}$$

Y, a projected m-dimensional vector, is obtained and defined as the projected feature vector of image A.

To get a good projection vector X, the trace of the covariance matrix of the projected feature vectors is obtained through the following criterion:

$$J(X) = tr(S_x) \tag{2.5}$$

where S_x denotes the covariance matrix of the projected feature vectors of the training examples and $tr(S_x)$ denoted the trace of S_x :

$$tr(S_x) = X^T \left[E(A - EA)^T (A - EA) \right] X \tag{2.6}$$

The image covariance matrix G of image A can be defined as:

$$G = E \left[(A - EA)^T (A - EA) \right] \tag{2.7}$$

Then, the criterion in 2.5 can be expressed as:

$$J(X) = X^T GX \tag{2.8}$$

Where X is a single-column vector. The optimal projection axis X_{opt} is a vector that maximizes J(X), or can be described as the eigenvector of G corresponding to the largest eigenvalue. These optimal projections vectors of 2DPCA, $X_1, X_2, ..., X_d$, where d is the number of selected eigenvalues, are used for feature extraction.

For image A, let:

$$Y_k = AX_k, \quad k = 1, 2, 3, ...d \tag{2.9}$$

Projected features $Y_1, Y_2, Y_3, ..., Y_k$ is called principal components of image A. Unlike PCA where the principal components is a scalar, in 2DPCA each principal component is a vector. The principal component vectors are used to form an m x d matrix $L = \left[Y_1^T, T_2^T, ..., Y_k^T\right]$ which is called the feature matrix of image A.

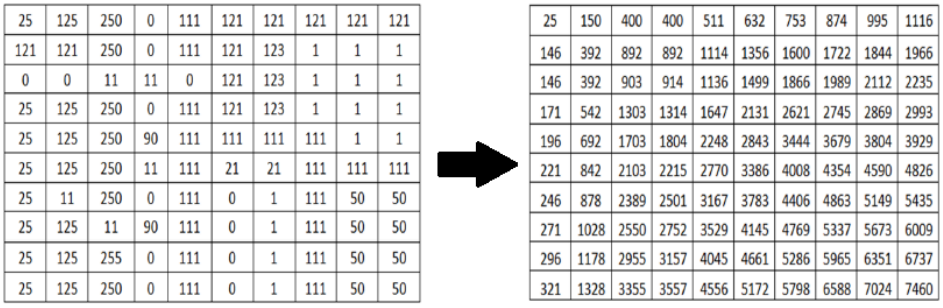
Several studies has been using 2DPCA for processing mammography images in CAD and CBIR system. [30] was trying to develop a CBIR system to categorise mammography images based on breast density classes with 2DPCA. It achieved average precision rates in the range from 83%-97%. [44] was trying trying to use 2DPCA to develop a CAD system to differentiate 12 and 20 classes of breast lesions. It managed to reach 61.6% highest accuracy of raw classifications.

2.4.3 Speeded-up robust features

Speeded-Up Robust Features (SURF) is a method for key-point feature detector and descriptor which is highly similar to Shift-Invariant Feature Transform (SIFT). Proposed by [3], this method is using box-filter response approach, computed using *Integral Images*. From the results in [3], it is shown that for a given image, the number of feature points extracted using both Difference of Gaussian (DoG) approach used in SIFT and simple box-filter used in SURF are comparable.

Integral Image, which is also known as *Summed Area Tables*, that has the same dimensions with the given image. In the *Integral Image*, a value at any position (x,y) is

the sum of all the pixel values above and to its left.The computation of *Integral Image* is illustrated in figure 2.4.



25	125	250	0	111	121	121	121	121	121
121	121	250	0	111	121	123	1	1	1
0	0	11	11	0	121	123	1	1	1
25	125	250	0	111	121	123	1	1	1
25	125	250	90	111	111	111	111	1	1
25	125	250	11	111	21	21	111	111	111
25	11	250	0	111	0	1	111	50	50
25	125	11	90	111	0	1	111	50	50
25	125	255	0	111	0	1	111	50	50
25	125	250	0	111	0	1	111	50	50

25	150	400	400	511	632	753	874	995	1116
146	392	892	892	1114	1356	1600	1722	1844	1966
146	392	903	914	1136	1499	1866	1989	2112	2235
171	542	1303	1314	1647	2131	2621	2745	2869	2993
196	692	1703	1804	2248	2843	3444	3679	3804	3929
221	842	2103	2215	2770	3386	4008	4354	4590	4826
246	878	2389	2501	3167	3783	4406	4863	5149	5435
271	1028	2550	2752	3529	4145	4769	5337	5673	6009
296	1178	2955	3157	4045	4661	5286	5965	6351	6737
321	1328	3355	3557	4556	5172	5798	6588	7024	7460

Figure 2.4 : *Left:* the given image with size 10 x 10 pixels. *Right :* Integral Image formed from the given image.

The Integral Images provide a way to calculate responses for box-type filters. After the structure is built, the response for a box-filter of any size inside the image can be built by only four operations inside any image. SURF is using an approximation of the Hessian matrix for detection (Fast-Hessian Detector).

$$det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \tag{2.10}$$

D_{xx}, D_{yy}, D_{xy} in 2.10 are approximations for Gaussian second order derivatives with the lowest scale (the same as SIFT). The Gaussian derivatives are approximated by box-filters as illustrated in Figure 2.5.

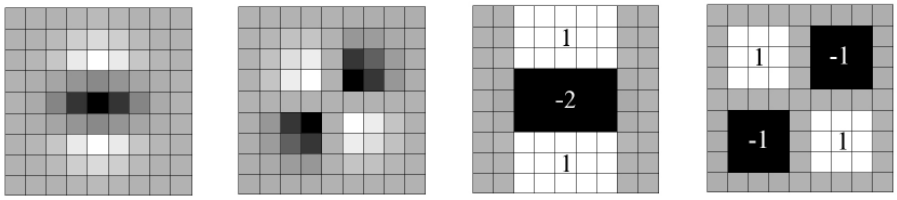


Figure 2.5 : From left to right : the discretised and cropped Gaussian second order partial derivative in y and xy direction in SIFT, and SURF approximation for the Gaussian second order partial derivative in y (D_{yy}) and xy (D_{xy}) direction. Grey regions are equal to zero. [3].

SURF detector is also utilizing scale space presentation to provide scale invariance. Same with DoG detector in SIFT, the octaves are images with increasing Gaussian kernel size. This is done so the filter is able to calculate faster as the box is scaled

instead of the image. Finally, the local features are selected as the local maxima in the $3x3x3$ neighbourhood in the scale-space.

The SURF descriptors is a local feature descriptors that is trying to achieve rotation invariance by finding reproducible orientation for the local neighbourhood of a keypoint. When the scale of a detected keypoint is s, the Haar wavelet responses in both x and y direction are calculated in circular neighbourhood of size $6s$.

After calculating the filter responses, the local neighbourhood is weighted with a Gaussian with $\sigma = 2$ to make the nearest intensity changes significantly. A sliding window of size $\pi/3$ is used surrounding the keypoint to calculate the sum of horizontal and vertical responses. Sum of those responses are then used to calculate the local orientation vector for each direction. The longest of those vectors is finally selected to represent the dominant orientation of a descriptor.

The responses of Haar wavelets are used in building the local feature descriptor. The first step is select an area surrounding the keypoint (detected in scale s) of size $20s$. The region is separated into $4x4$ sub-regions and in each sub-regions, there are $2x2$ smaller regions where response strength are summed. A feature vector v is calculated from these response strength, as follows:

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \tag{2.11}$$

Where d_x, d_y are the wavelet responses in horizontal and vertical directions, respectively. 64-dimensional vector is combined from the 16 vectors from the sub-regions and 4 smaller regions inside it ($16x4 = 64$), as opposed to SIFT with 128-dimensional feature vector.

The original paper [3] reports results that show SURF descriptors is $5x5$ faster to calculate and provides more than 10% better recognition rate than the SIFT descriptor.

Previous studies had been using SURF for processing mammography images. [45] was proposing an unsupervised method to automatically detect suspicious region in mammography images using SURF. It managed to achieve almost 90% sensitivity which indicated a very low false negative rate.

2.5 Support Vector Machine

Support Vector Machine (SVM) is a binary learning machine that tries to classify an input vector X by using the decision function in equation 2.12. SVMs work by determining which side of the hyperplane X lies, and in 2.12, the hyperplane is perpendicular to W and at a distance $\frac{b}{\|W\|}$ from the origin. SVMs maximize the margin around the separating hyperplane. The decision function is specified by a subset of training samples, and they are called the support vectors.

$$f(X) =< X, W > +b \tag{2.12}$$

When the training set is linearly separable, it's called hard margin classification. On the other hand, when it is not linearly separable, it's called soft margin classification which is using slack variables ξ_i to allow some misclassification of noisy or difficult samples where $\xi_i > 0, i = 1, 2..., n$. Figure 2.6 shows examples of 2 slack variables $\xi 1$ and $\xi 2$.

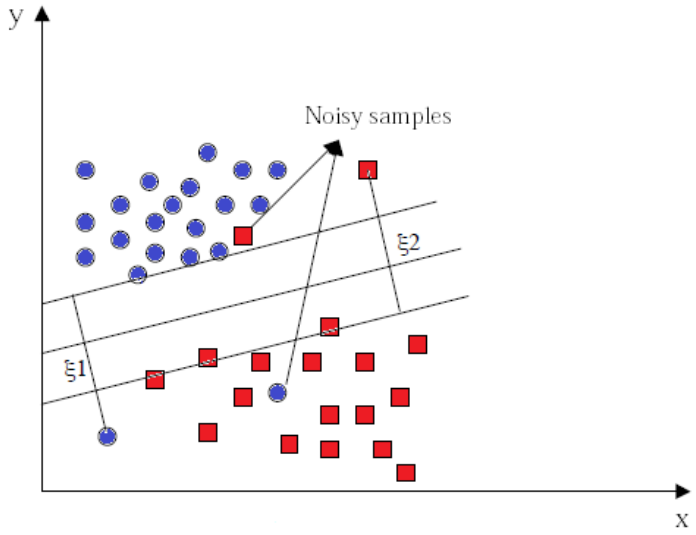


Figure 2.6 : Example of noisy samples.

The slack variable approach is not very efficient for separating various non-separable class [46]. More refined technique for separating non-separable class required the use of non-linear classifier rather than a linear hyperplane. Non-linear classifiers require a mapping function, which can be in form of Φ or kernel functions. Φ is a function that

maps the input data patterns into a higher dimensional space, and kernel functions are used for mapping the input space to a feature space.

The reason of choosing kernel functions over Φ is because they reduce the complexity of calculating Φ [47]. In kernel-based methods, input vectors are classified using the maximum-margin hyperplane in a specific feature space. Figure 2.7 illustrates two-dimensional feature set of non-linearly separable class. The figure 2.7 shows how they are mapped to a three-dimensional feature space with the Radial Basis Function and then the maximum-margin hyperplane in the projected space is mapped back to the original input space where it serves as a non-linear hyperplane.

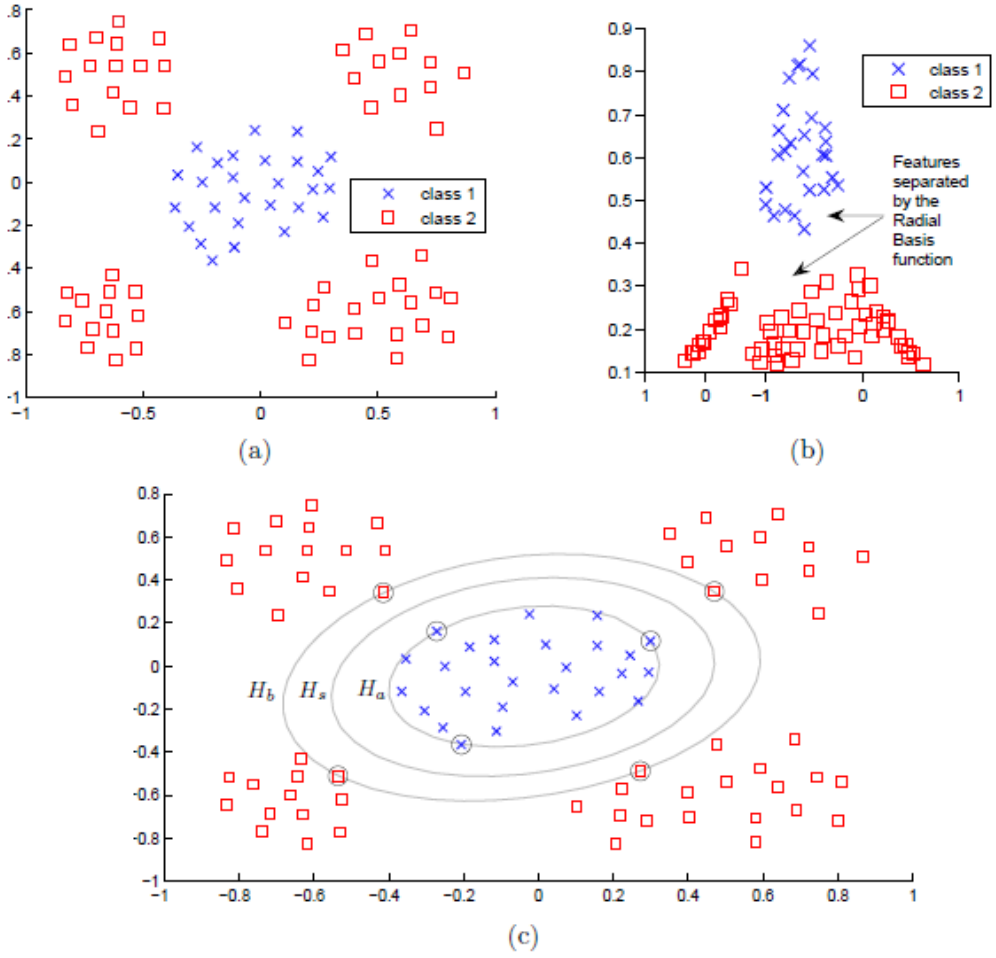


Figure 2.7 : (a) Two-class, two-dimensional feature set that cannot be linerarly separated. (b) Mapped to three-dimensional feature space with Radial Basis Function. (c) The hyperplane from feature space is mapped back to the original input space. (from [4], pg. 40).

There are many kernel functions available and some of the popular ones are:

- Radial Basis Function : $k(x_j, x_i) = e^{-\gamma \|x_j - x_i\|^2}, for \gamma > 0$

- Polynomial : $k(x_j, x_i) = (x_j.x_i)^d$
- Sigmoidal : $k(x_j, x_i) = tanh(ax_j.x_i - b)$

To use a non-linear SVM, first each feature vector x is projected to new feature space with the kernel mapping $x \mapsto \phi(x)$, where $\phi(x)$ represents the kernel function. Using Soft Margin SVM, the optimization problem becomes

$$maximize(-0.5 \sum_{j=0}^{n-1} \sum_{i=0}^{n-1} \alpha_j C_j C_i k(x_j, x_i) + \sum_{i=0}^{n-1} \alpha_i) \tag{2.13}$$

such that $0 \leq \alpha_i \leq C, \sum_{i=0}^{n-1} \alpha_i C_i = 0$ and $k(x_j, x_i) = \phi(x_j).\phi(x_i)$ where C is the cost function. This can be solved using QP-solver [48]. After finding α, w can be calculated using equation 2.14, and b is found with equation 2.15 where n_{sv} is the number of support vectors.

$$w = \sum_{i=0}^{n-1} \alpha_i C_i \phi(x_i) \tag{2.14}$$

$$b = \frac{1}{n_{sv}} \sum_{(x_i, c_i) \in S_{sv}} (c_i - \sum_{(x_i, c_i) \in S_{sv}} \alpha_j C_j \phi(x_j).\phi(x_i)) \tag{2.15}$$

A new input vector z is then classified with the signed distance function $y = f(z) = w.\phi(z) + b$.

Since SVM is designed for binary classification, there are several approaches to adapt it into multi-class problems. The most widely used ones are One-Against-All (OAA) classification and Pairwise classification (also known as One-Against-One).

OAA SVM was introduced by [49]. The formulation of OAA method required unanimity among all SVMs: a data would be classified under a certainty that if and only if that class's SVM accepted it and all other classes' SVMs rejected it. On the other hand, in pairwise SVMs, a classifier is trained for every combination of classes. The input is classified with each classifier, and the class obtaining the most assignments is chosen as the overall assigned class.

Studies of CAD and CBIR in mammography had been using SVM recurrently. [26], [29], [50], [15], [30], [44], [31] have all been using SVM as their method of choice for classification purposes. [21] was comparing SVM with Artificial Neural Network

(ANN) to classify microcalcifications in mammography images. Their result shows that SVM outperformed ANN when balanced database is present. It also showed that SVM has better generalization capability, less prone to overfitting, and less computational complexity than ANN.

2.6 Similarity Measures

2.6.1 Euclidean distance

Euclidean distance is the basis of many measures of similarity between data points or vectors. The distance between vectors X and Y is defined as follows:

$$d(x,y) = \sqrt{\sum_{i}^{n}(x_i - y_i)^2} \tag{2.16}$$

where n denotes the total number of data. Euclidean distance is also sometimes called L_2 norm or L_2 distance.

Euclidean distance is the square root of the sum of squared differences between corresponding elements of the two vectors. Since no adjustment is made for differences in scale, euclidean distance is only suitable for data measured in the same scale.

2.6.2 Mutual information

Mutual information (MI) is the description of the relatedness of two sources S_i and S_j in information theory. Equation 2.17 shows the calculation of MI.

$$I(S_i;S_j) = H(S_i) - H(Si|S_j) = H(S_j) - H(S_j|S_i) \tag{2.17}$$

where H(S) is the entropy rate of S. Mutual information is an absolute measure of information common to both sources. For independent sources, it will converge to zero.

However, MI itself cannot be useful to measure distance, because it's not a distance function and not bounded. A study done by [51] was trying to make MI a bounded distance by normalizing it and subtracting it from one. One way to normalize it is to

normalize by the maximum possible possible mutual information the two sources can share, as follows:

$$d_{CR}(S_i, S_j) = 1 - \frac{I(S_i; S_j)}{min(H(S_i).H(S_j))} \tag{2.18}$$

The other way to normalize it is by the maximum entropu of both sources resulting in:

$$d_{CL}(S_i, S_j) = 1 - \frac{I(S_i; S_j)}{max(H(S_i).H(S_j))} \tag{2.19}$$

The distance in 2.19 evaluates to zero only for identical spurces $S_i = S_j$, because only identical sources share the maximum possible information and have identical entropies at the same time. This distance reflects whether the sources are identical or not.

[51] already done the comparison on d_{CR} and d_{CL}, and it resulted in d_{CL} is the better choice for classification purposes while d_{CR} is better for content recognition purposes.

3. EXPERIMENTAL SETUPS

3.1 Experiment Stages

The experiment in this thesis is divided into four stages; image preprocessing, feature extraction methods and SVM parameter selection, feature extraction method and SVM testing, and image retrieval stage.

Flowcharts of these stages are shown in figure 3.1, 3.5, 3.6, and 3.7 respectively.

3.1.1 Image preprocessing stage

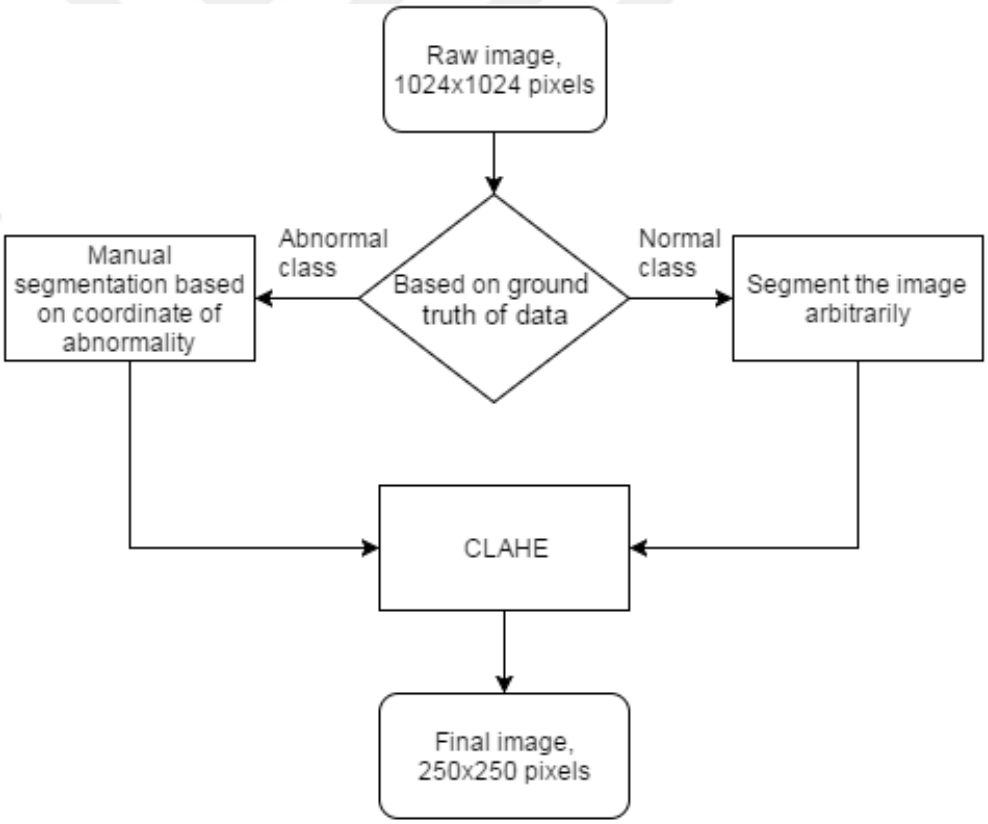


Figure 3.1 : Flowchart of the preprocessing stage.

This stage serves the purpose of segmenting and enhancing the images. There are no parameter changes in this stage because this is beyond the scope of this thesis.

The preprocessing was done in these steps:

1. Segment each of the images from 1024x1024 pixels to 250x250 pixels. The size is chosen arbitrarily based on the average size of abnormalities. The segmented image will try to centralize the abnormality on the images, if present and possible (i.e. the abnormality is not on the edge). The central coordinate of the abnormality was given from the database. For the normal class, in which there is no abnormality present, the segmentation is done arbitrarily, excluding the nipple area. This step is done manually. Examples of the segmented images for each class is shown on figure 3.2 and 3.3.

Figure 3.2 : Samples of segmented normal breast tissue images. From left to right: Fatty subclass, Dense subclass, and Fatty-glandular subclass.

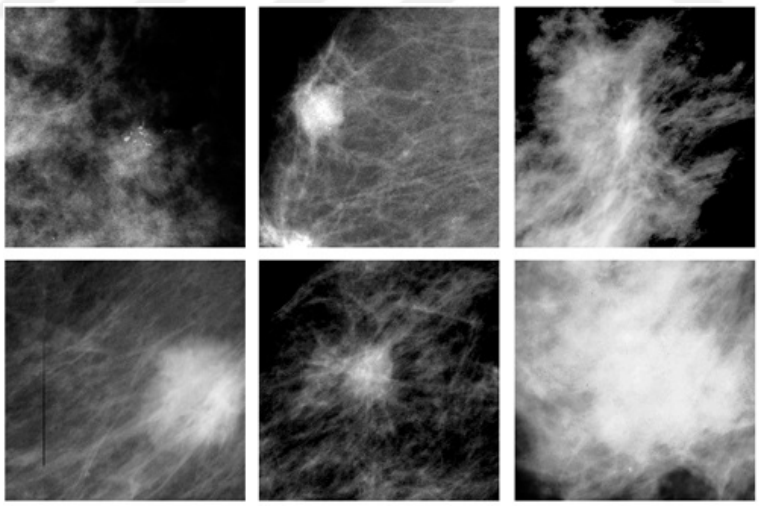


Figure 3.3 : Samples of segmented abnormal breast tissue images. Top row, left to right: Calcification subclass, Circumscribed mass subclass, and Spiculated mass subclass. Bottom row, left to right: Ill-defined mass subclass, Architectural distortion subclass, and Asymmetry class. The images were enhanced manually with GIMP to superimpose the differences of each class.

2. The decision to do the segmentation manually is because each mammography image must be annotated by the radiologist beforehand. Since any CAD or CBIR

system for radiology case is intended to be used by radiologists, it is fair to say that the manual segmentation done in this thesis is trying to simulate how the radiologists annotate the images before entering them into the system.

3. Enhance the contrast of each of the images with CLAHE. The parameter used is the default parameter set my Matlab, and is shown in table 3.1. The sample image before and after CLAHE application is shown on figure 3.4.

Table 3.1 : CLAHE default parameter.

Parameter Name	Value
Size of tiles	8x8
Clip limit	0.01
Number of histogram bins	256
Histogram distribution	Uniform

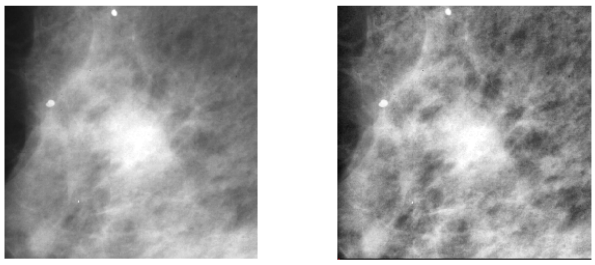


Figure 3.4 : Sample of an image before (left) and after CLAHE (right).

3.1.2 Two-class problem: feature extraction methods and SVM parameter selection

This stage serves the purpose of selecting the best parameter for each feature extraction method; GLCM, 2DPCA, and SURF, and also best kernel for SVM classification.

Two-class problem is used to measure the performance of each parameter pairs; 113 normal and 113 abnormal class. The 10 groups partition is used to do the 10-fold cross-validation to make sure all data have been used as test and training data at least once.

This stage is done in these steps:

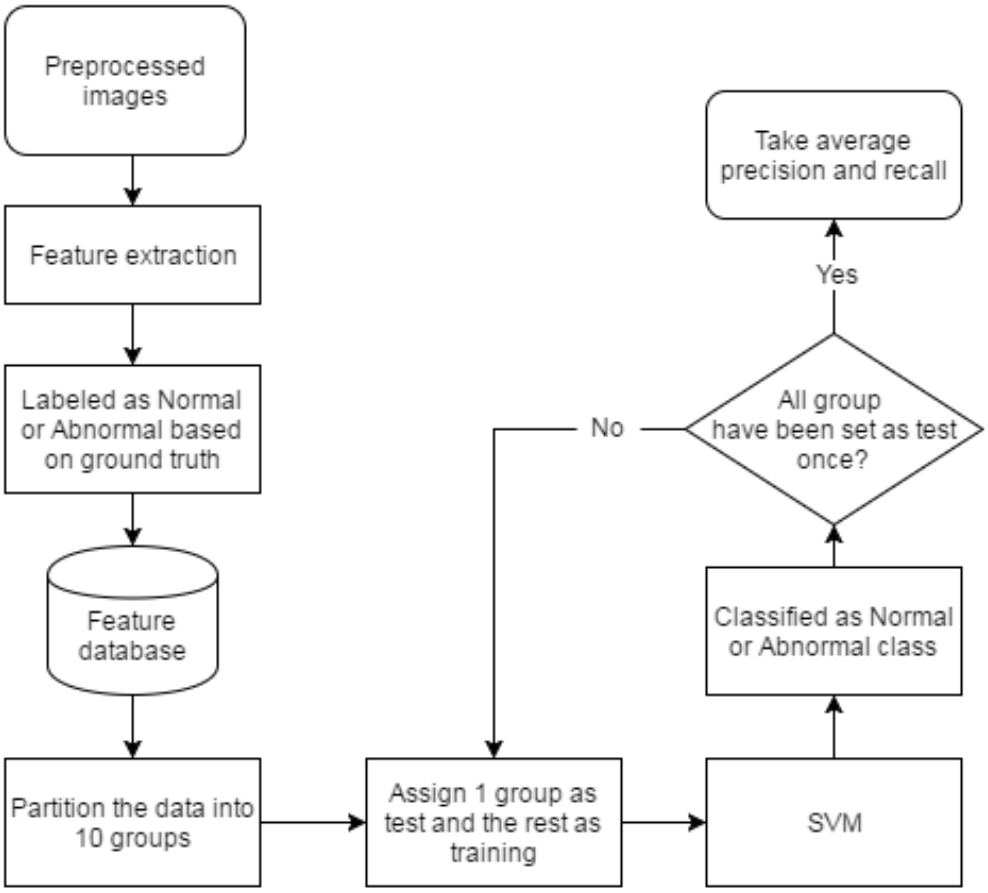


Figure 3.5 : Flowchart of the feature extraction methods and SVM parameter selection stage.

1. Preprocessed image is feed into feature extraction method; GLCM, 2DPCA, or SURF.

2. The feature vectors are stored in the database, labelled with either normal or abnormal with regards of the ground truth in the database.

3. Once all 226 feature vectors are stored in the database, they are partitioned into 10 groups. 10-fold cross validation commenced to measure their performance with SVM.

4. Precision and recall rate are averaged from the folds.

5. These steps is done for each of the parameters assigned for GLCM, 2DPCA, SURF, and SVM.

3.1.3 Nine-class problem: feature extraction methods and SVM testing

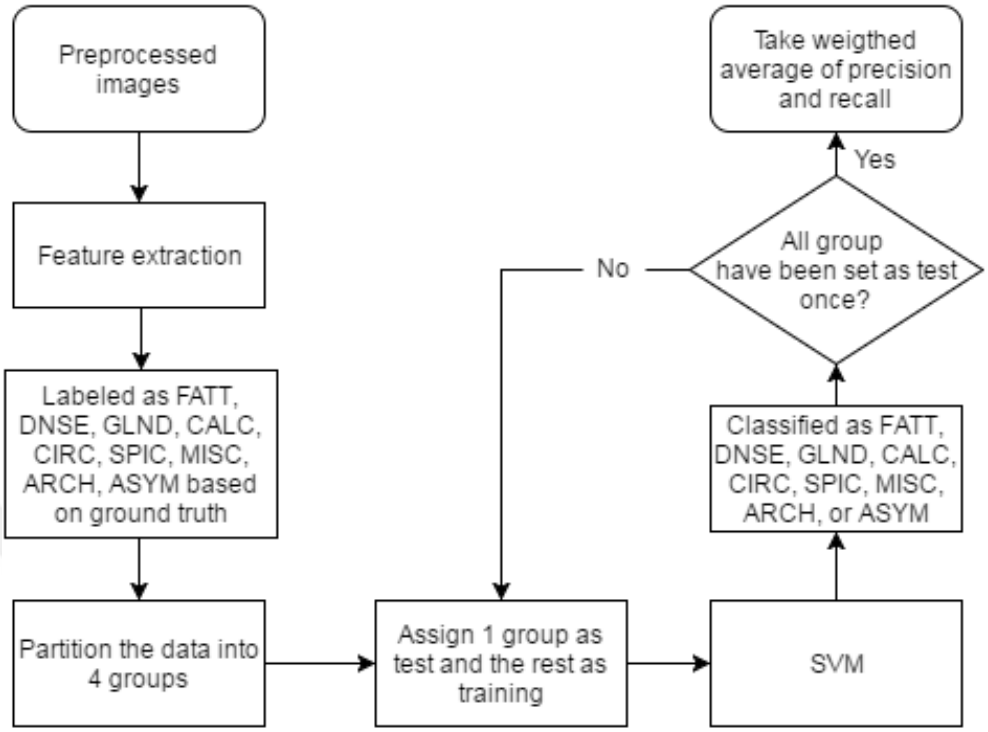


Figure 3.6 : Flowchart of the feature extraction and SVM testing stage. FATT: Normal, fatty tissue, DNSE: Normal, dense tissue. GLND: Normal, fatty-glandular tissue. CALC: Abnormal, calcification. CIRC: Abnormal, circumscribed mass. SPIC: Abnormal, spiculated mass. MISC: Abnormall, ill-defined mass. ARCH: Abnormal, architectural distortion. ASYM: Abnormal, asymmetry.

This stage serves the purpose of seeing how well each feature extraction method and SVM, with their respective best parameter taken from the previous stage, to do the classification.

Nine-class problem is used in this stage, named Normal fatty tissue, Normal dense tissue, Normal fatty-glandular tissue, Abnormal calcification, Abnormal circumscribed mass, Abnormal spiculated mass, Abnormal ill-defined mass, Abnormal architectural distortion, and Abnormal asymmetry.

The partitioning step is used to do the 4-fold cross validation to measure the performance of each methods. The choice of 4 is to simulate the separation of 75% training and 25% test data.

This stage is done in these steps:

1. Preprocessed image is feed into feature extraction method; GLCM, 2DPCA, or SURF with their best parameter taken from previous stage.

2. The feature vectors are stored in the database, labelled with either Normal fatty tissue, Normal dense tissue, Normal fatty-glandular tissue, Abnormal calcification, Abnormal circumscribed mass, Abnormal spiculated mass, Abnormal ill-defined mass, Abnormal architectural distortion, and Abnormal asymmetry with regards of the ground truth in the database.

3. Once all 226 feature vectors are stored in the database, they are partitioned into 4 groups. 4-fold cross validation commenced to measure their performance with SVM.

4. Precision and recall rate are averaged from the folds.

3.1.4 Image retrieval stage

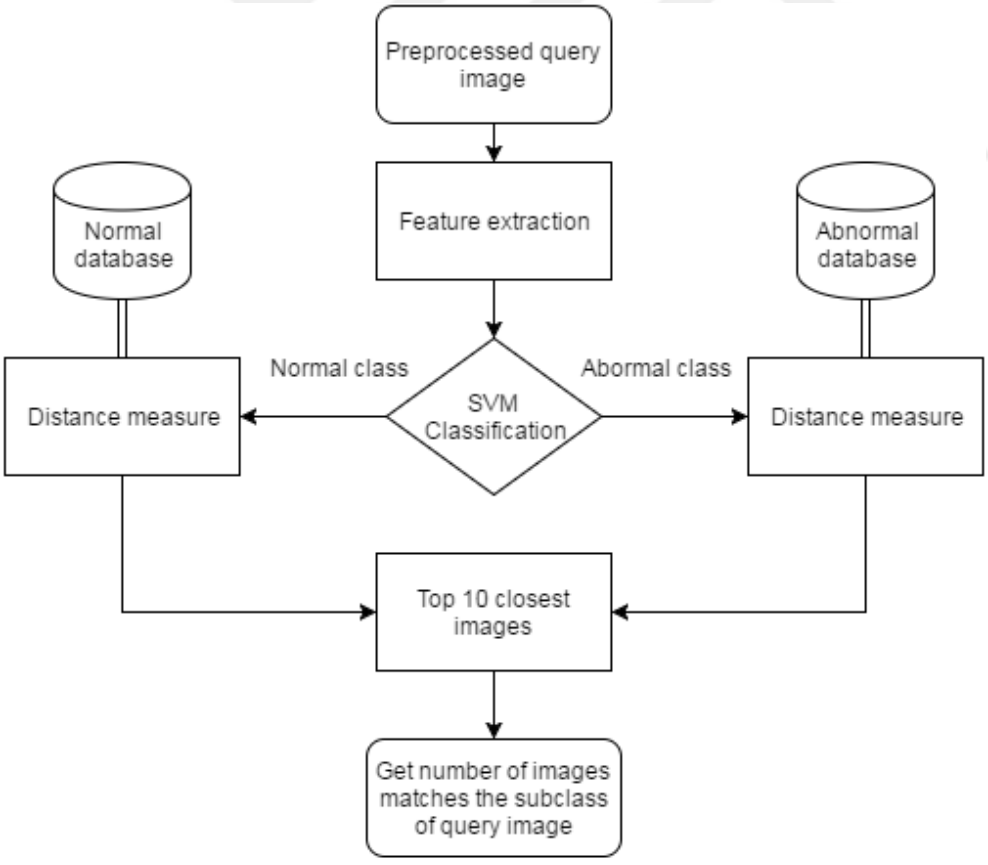


Figure 3.7 : Flowchart of the image retrieval stage.

This stage serves the purpose of simulating CBIR with the feature extraction methods, classification, and distance measure previously studied in the prior stage.

Depending on the classification, each query image will either be compared by the distance measure method with 3 class-problem (Fatty, Dense, or Fatty-glandular) when it's classified as normal, or with 6-class problem (Calcification, Circumscribed mass, Spiculated mass, Ill-defined mass, Architectural distortion, or Asymmetry).

This stage is done in these steps:

1. Preprocessed image is fed into feature extraction method; GLCM, 2DPCA, or SURF with their best parameter taken from previous stage.

2. The feature vectors are stored in two databases; data labelled with Normal fatty tissue, Normal dense tissue, and Normal fatty-glandular tissue are stored in Normal database, Abnormal calcification, Abnormal circumscribed mass, Abnormal spiculated mass, Abnormal ill-defined mass, Abnormal architectural distortion, and Abnormal asymmetry are stored in Abnormal database.

3. Once all 226 feature vectors are stored in their respective database, each of the subclasses are partitioned into group of training and test set as shown in table 3.2. The partition is based on the 70:30 ratio of training and test images.

4. An image in a test group is classified by SVM, whether it is belong to normal or abnormal class.

5. Distance measure is calculated with euclidean distance or MI within the query feature vector and each of feature vectors stored in their respective classified database. Ten of the closest images in the database is selected.

6. The number of correctly selected subclass out of 10 is counted and score is taken. *Score = number of correct subclass x 10*. Maximum possible score is 100.

7. Step 3-6 is repeated 4 times until all the data have been assigned in the test set at least once. Final score is taken as aggregated score calculated in step 6.

8. These steps are done for both euclidean distance and MI for comparing distance measure methods.

Table 3.2 : Partitioned data for the retrieval stage.

Subclass	Training	Test
Fatty	24	9
Dense	26	11
Fatty-glandular	30	13
Calcification	16	7
Circumscribed mass	15	8
Spiculated mass	14	5
Ill-defined mass	10	4
Architectural distortion	14	5
Asymmetry	11	4

3.2 Feature Extraction Setups

3.2.1 GLCM setup

The preprocessed image is converted into feature vector by GLCM method as follows:

1. GLCM is created with *graycomatrix* function in Matlab with grey level number of 8 and offset/distance of 1, 3, 5, or 7.
2. Prior step resulted in 8x8 GLCM.
3. Statistical measures are taken from table 2.6.
4. Prior step resulted in 13x1 vector which is the feature vector.

The main parameter for experiment with GLCM is the distance. Distance of 1, 3, 5, and 7 were tested with SVM to differentiate between normal and abnormal class from the database.

3.2.2 2DPCA setup

The preprocessed image is converted into feature vector by 2DPCA method as follows:

1. The image format will be converted from 8-bit unsigned-integer to double for calculation.
2. Take the average matrix from all images.

3. Find the image covariance matrix G by 2.6.

4. Take the eigenvector of the largest eigenvalue of G as the optimal projection axis X_{opt}.

5. Principal components of each image is formed by 2.4.

6. Each column from the prior step corresponds to principal component vectors. 1, 2, 3, 4, 5, or 6 columns will be taken as final feature vector.

The main parameter for experiment with 2DPCA is the number of principal component. Principal component of 1, 2, 3, 4, 5, and 6 were tested with SVM to differentiate between normal and abnormal class from the database.

3.2.3 SURF setup

The preprocessed image is converted into feature vector by SURF method as follows:

1. Take detected features from the image with *detectSURFFeatures* function in Matlab.

2. Take the 10, 15, 20, 25, or 30 strongest features using *selectStrongest* function in Matlab.

3. If the number of detected features are smaller than the number of strongest features taken, append the feature vectors with zeros.

4. Reshape the vectors into single-column vector and use as feature vector.

The main parameter for experiment with SURF is the number of strongest features detected. 10, 15, 20, 25, and 30 features were tested with SVM to differentiate between normal and abnormal class from the database.

3.3 SVM Parameter

Parameter used in SVM classification is based on values given in LibSVM [32] and shown in table 3.3.

The main parameter for SVM is kernel type and normalization. The kernels tested were Linear, Radial Basis Function, and Polynomial. Each kernel is tested with and without normalization.

Table 3.3 : LibSVM parameter values.

Parameter	Value
SVM type	nu-SVC
Kernel type	linear, radial basis function, or polynomial
Degree	3
Gamma	1/number of feature
nu	0.5
Cache size	100 MB
Normalization	yes or no
Weight	1

4. FINDINGS AND RESULTS

4.1 Two-class Problem: Feature Extraction Methods and SVM Parameter Selection Stage

4.1.1 GLCM experiments

The result of experiment stage 2 with GLCM is given in table 4.1 and 4.2.

Table 4.1 : Result of GLCM distances and SVM kernels with normalization.

GLCM distance	Linear		RBF		Polynomial	
	Precision	Recall	Precision	Recall	Precision	Recall
1	0.879	0.867	0.873	0.863	0.863	0.845
3	0.871	0.863	0.873	0.863	0.894	0.872
5	0.868	0.863	0.86	0.854	0.894	0.872
7	0.877	0.872	0.878	0.872	0.909	0.889

Table 4.1 shows the result of classification with SVM with normalization. Distance of 7 achieved overall best performance in all SVM kernels except for linear kernel. Polynomial kernel achieved overall best performance in all distances except for distance of 1.

Table 4.2 : Result of GLCM distances and SVM kernels without normalization.

GLCM distance	Linear		RBF		Polynomial	
	Precision	Recall	Precision	Recall	Precision	Recall
1	0.897	0.881	0.888	0.876	0.894	0.881
3	0.894	0.876	0.892	0.881	0.894	0.881
5	0.897	0.881	0.895	0.885	0.892	0.881
7	0.897	0.881	0.892	0.881	0.894	0.881

Table 4.2 shows the result of classification with SVM without normalization. Distance of 1, 5, and 7 achieved very similar performance in linear SVM kernel, which is also the best performance achieved in this case.

Overall, the best parameter for GLCM case is distance of 7 with 0.909 precision and 0.889 recall rate with normalization and polynomial SVM kernel.

4.1.2 2DPCA experiments

The result of experiment stage 2 with 2DPCA is given in table 4.3 and 4.4.

Table 4.3 : Result of 2DPCA principal components number and SVM kernels with normalization.

Prin. Comp. No.	Linear		RBF		Polynomial	
	Precision	Recall	Precision	Recall	Precision	Recall
1	0.553	0.553	0.567	0.566	0.572	0.571
2	0.675	0.673	0.688	0.686	0.718	0.708
3	0.747	0.743	0.742	0.739	0.773	0.761
4	0.710	0.708	0.742	0.739	0.735	0.730
5	0.769	0.765	0.773	0.770	0.804	0.796
6	0.766	0.761	0.763	0.757	0.782	0.774

Table 4.3 shows the result of classification with SVM with normalization. Principal components of 5 achieved overall best performance in all SVM kernels. Polynomial kernel achieved overall best performance in all number of principal components.

Table 4.4 : Result of 2DPCA principal components number and SVM kernels without normalization.

Prin. Comp. No.	Linear		RBF		Polynomial	
	Precision	Recall	Precision	Recall	Precision	Recall
1	0.562	0.562	0.477	0.478	0.751	0.504
2	0.685	0.681	0.477	0.478	0.753	0.513
3	0.735	0.730	0.482	0.482	0.653	0.513
4	0.710	0.708	0.482	0.482	0.416	0.496
5	0.783	0.779	0.482	0.482	0.500	0.500
6	0.736	0.735	0.477	0.478	0.248	0.491

Table 4.4 shows the result of classification with SVM without normalization. Principal components of 5 achieved highest precision rate with linear SVM.

Overall, the best parameter for 2DPCA case is principal components of 5 with 0.804 precision and 0.796 recall rate with normalization and polynomial SVM kernel.

4.1.3 SURF experiments

The result of experiment stage 2 with SURF is given in table 4.5 and 4.6.

Table 4.5 : Result of SURF strongest features numbers and SVM kernels with normalization.

Strongest features	Linear		RBF		Polynomial	
	Precision	Recall	Precision	Recall	Precision	Recall
10	0.851	0.850	0.859	0.859	0.865	0.864
15	0.878	0.877	0.860	0.859	0.891	0.890
20	0.908	0.908	0.908	0.908	0.908	0.908
25	0.921	0.921	0.917	0.917	0.921	0.921
30	0.914	0.912	0.914	0.912	0.943	0.938

Table 4.5 shows the result of classification with SVM with normalization. For linear and RBF kernels, 25 strongest features yielded the best precision and recall rate. However, for polynomial kernel, 30 strongest features yield the best result. Polynomial kernel achieved overall best performance in all number of strongest features.

Table 4.6 : Result of SURF strongest features numbers and SVM kernels without normalization.

Strongest features	Linear		RBF		Polynomial	
	Precision	Recall	Precision	Recall	Precision	Recall
10	0.859	0.859	0.855	0.855	0.550	0.549
15	0.922	0.921	0.917	0.917	0.577	0.575
20	0.895	0.895	0.908	0.908	0.553	0.553
25	0.917	0.917	0.912	0.912	0.545	0.544
30	0.904	0.904	0.913	0.912	0.586	0.584

Table 4.6 shows the result of classification with SVM without normalization. 15 strongest features achieved highest precision rate with linear SVM. Linear kernel achieved overall best performance in all number of strongest features.

Overall, the best parameter for SURF case is 30 strongest features with 0.943 precision and 0.938 recall rate with normalization and polynomial SVM kernel.

4.1.4 Confusion matrices with selected best parameters

Table 4.7, 4.8, and 4.9 showed the confusion matrices when the experiment is done with the selected best parameters; distance of 7 for GLCM, 5 principal components for 2DPCA, 30 strongest features for SURF, and polynomial kernel with normalization for SVM.

Table 4.7 : GLCM confusion matrix with distance 7.

Normal	Abnormal	<= classified as
113	0	Normal
25	88	Abnormal

Table 4.8 : 2DPCA confusion matrix with 5 principal components.

Normal	Abnormal	<= classified as
81	32	Normal
14	99	Abnormal

Table 4.9 : SURF confusion matrix with 30 strongest features.

Normal	Abnormal	<= classified as
112	1	Normal
13	100	Abnormal

From the overall experiments in this stage, SURF with 30 strongest feature performed the best as feature extraction method in this stage.

4.2 Nine-class Problem: Feature Extraction Methods and SVM Testing Stage

4.2.1 GLCM experiment

Confusion matrix for experiment in stage 3 with GLCM is shown on table 4.10. The measurement of class-wise and weighted average of true positive rate, false positive, precision, and recall is shown on table 4.11.

Table 4.10 : GLCM with distance 7 with polynomial SVM confusion matrix.

FATT	DNSE	GLND	CALC	CIRC	SPIC	MISC	ARCH	ASYM	
18	7	8	0	0	0	0	0	0	FATT
2	21	14	0	0	0	0	0	0	DNSE
4	6	33	0	0	0	0	0	0	GLND
1	3	0	9	2	3	3	0	2	CALC
1	2	4	1	6	2	2	2	3	CIRC
2	4	0	1	3	2	4	1	2	SPIC
0	1	0	2	1	3	4	1	2	MISC
0	2	2	1	0	4	1	6	3	ARCH
0	3	0	2	1	2	4	1	2	ASYM

From the confusion matrix, it is clear that GLCM was able to avoid classifying normal classes into abnormal class with 100% accuracy. However, to classify them with breast

Table 4.11 : Class-wise measurement of table 4.10.

Class	True Positive	False Positive	Precision	Recall
FATT	0.545	0.052	0.643	0.545
DNSE	0.568	0.148	0.429	0.568
GLND	0.767	0.153	0.541	0.767
CALC	0.391	0.034	0.563	0.391
CIRC	0.261	0.034	0.462	0.261
SPIC	0.105	0.068	0.125	0.105
MISC	0.286	0.066	0.222	0.286
ARCH	0.316	0.024	0.545	0.316
ASYM	0.133	0.057	0.143	0.133
Weighted average	0.447	0.084	0.451	0.447

density subclass, the performance is not as high. A particular problem was found in dense subclass, where 14 out of 37 images were misclassified as fatty-glandular subclass. Fatty-glandular subclass achieved best performance with 0.767 true positive rate. Although, it also has the biggest false positive rate, which could indicate that the method is more likely to classify images into fatty-glandular subclass.

In abnormal subclasses case, circumscribed mass subclass was the most problematic because it failed to classify 7 out of 23 images into abnormal class. Spiculated mass and asymmetry performed the worse with less than 0.2 true positive rate. The highest true positive rate is achieved by calcification subclass with 0.391.

4.2.2 2DPCA experiment

Confusion matrix for experiment in stage 3 with 2DPCA is shown on table 4.12. The measurement of class-wise and weighted average of true positive rate, false positive, precision, and recall is shown on table 4.13.

From the confusion matrix, 2DPCA appeared to fail to classify normal subclasses with high accuracy as GLCM did. Dense subclass performed the best in normal class with 0.595 true positive rate. Fatty-glandular class performed the worst because it failed to avoid classifying 19 out of 43 images to abnormal subclasses.

However, for abnormal subclasses, 2DPCA performed better than GLCM generally. This is because no abnormal subclass in 2DPCA result got true positive rate below 0.2. It appeared to perform best on classifying circumscribed mass subclass with 0.478 true positive rate. However, this is also a drawback since circumscribed mass subclass has

Table 4.12 : 2DPCA with 5 principal components and polynomial SVM confusion matrix.

FATT	DNSE	GLND	CALC	CIRC	SPIC	MISC	ARCH	ASYM	
14	7	4	1	3	3	0	0	1	FATT
4	22	6	0	1	2	0	2	0	DNSE
5	7	12	1	5	2	5	3	3	GLND
1	0	1	9	5	2	2	3	0	CALC
2	0	2	4	11	1	1	2	0	CIRC
0	0	2	4	5	6	2	0	0	SPIC
2	0	0	2	3	0	4	3	0	MISC
0	3	2	5	4	1	0	4	0	ARCH
0	3	2	1	3	2	0	1	3	ASYM

Table 4.13 : Class-wise measurement of 4.12.

Class	True Positive	False Positive	Precision	Recall
FATT	0.424	0.073	0.500	0.424
DNSE	0.595	0.106	0.524	0.595
GLND	0.279	0.104	0.387	0.279
CALC	0.391	0.089	0.333	0.391
CIRC	0.478	0.143	0.275	0.478
SPIC	0.316	0.063	0.316	0.316
MISC	0.286	0.047	0.286	0.286
ARCH	0.211	0.068	0.222	0.211
ASYM	0.200	0.019	0.429	0.200
Weighted average	0.376	0.086	0.386	0.376

the biggest false positive rate which could indicate that the method is more likely to classify images into circumscribed mass subclass.

4.2.3 SURF experiment

Confusion matrix for experiment in stage 3 with SURF is shown on table 4.14. The measurement of class-wise and weighted average of true positive rate, false positive, precision, and recall is shown on table 4.15.

From the confusion matrix, SURF was able to avoid classifying normal classes into abnormal class with almost 100% accuracy. However, to classify them with breast density subclass, the performance is not as high. Similar to GLCM case, dense subclass is the most problematic in normal class and fatty-glandular subclass performed the best with 0.721 true positive rate.

Table 4.14 : SURF with 30 strongest features and polynomial SVM confusion matrix.

FATT	DNSE	GLND	CALC	CIRC	SPIC	MISC	ARCH	ASYM	
19	4	10	0	0	0	0	0	0	FATT
5	19	13	0	0	0	0	0	0	DNSE
5	6	31	0	1	0	0	0	0	GLND
1	0	1	14	3	3	0	1	0	CALC
1	2	2	4	9	1	2	2	2	CIRC
0	0	2	2	3	6	1	2	3	SPIC
0	1	2	0	3	0	5	1	2	MISC
0	0	0	2	4	3	1	8	1	ARCH
0	1	0	2	2	4	2	3	1	ASYM

Table 4.15 : Class-wise measurement of 4.14.

Class	True Positive	False Positive	Precision	Recall
FATT	0.576	0.062	0.613	0.576
DNSE	0.514	0.074	0.576	0.514
GLND	0.721	0.164	0.508	0.721
CALC	0.609	0.049	0.583	0.609
CIRC	0.391	0.079	0.360	0.391
SPIC	0.316	0.053	0.353	0.316
MISC	0.357	0.028	0.455	0.357
ARCH	0.421	0.043	0.471	0.421
ASYM	0.067	0.028	0.143	0.067
Weighted average	0.496	0.077	0.483	0.496

In abnormal cases, calcification subclass performed significantly better than other abnormal subclass, with 0.609 true positive rate. Other abnormal subclass also gained better result compared to GLCM and 2DPCA, with exception of circumscribed mass and asymmetry.

Overall, from the weighted averages of each methods in this stage, SURF performed the best with 0.496 true positive rate, 0.077 false positive rate, 0.483 precision rate, and 0.496 recall rate.

4.3 Image Retrieval Stage

4.3.1 Normal class retrieval final score

Table 4.16 shows the result of aggregated final score for retrieval of normal class with euclidean distance.

Table 4.16 : Final score for normal class retrieval with euclidean distance measure.

Method	Average score per-query
GLCM	61.3
2DPCA	46.1
SURF	66.7

From table 4.16, with euclidean distance SURF is performed the best to retrieve correct normal subclasses with the average of 66.7 score per-query.

Table 4.17 shows the result of aggregated final score for retrieval of normal class with MI.

Table 4.17 : Final score for normal class retrieval with MI.

Method	Average score per-query
GLCM	58.2
2DPCA	38.7
SURF	60

From table 4.17, with MI SURF is performed the best to retrieve correct normal subclasses with the average of 60 score per-query.

Overall, euclidean distance achieved better performance as a distance measure method for normal class retrieval.

4.3.2 Abnormal class retrieval final score

Table 4.18 shows the result of aggregated final score for retrieval of abnormal class with euclidean distance.

From table 4.18, with euclidean distance SURF is performed the best to retrieve correct abnormal subclasses with the average of 47.6 score per-query.

Table 4.18 : Final score for abnormal class retrieval with euclidean distance.

Method	Average score per-query
GLCM	26.2
2DPCA	38.3
SURF	47.6

Table 4.19 shows the result of aggregated final score for retrieval of abnormal class with MI.

Table 4.19 : Final score for abnormal class retrieval with MI.

Method	Average score per-query
GLCM	29.8
2DPCA	38
SURF	44.2

From table 4.19, with MI SURF is performed the best to retrieve correct abnormal subclasses with the average of 44.2 score per-query.

Overall, euclidean distance achieved better performance as a distance measure method for both normal and abnormal class retrieval.

5. CONCLUSION AND FUTURE WORKS

5.1 Conclusion

From all the experiment stages done in this thesis, there are several conclusions than can be derived.

From the Two-class Problem experiment stage, we can see that to differentiate between normal and abnormal class, GLCM with distance 7 yielded the best result compared to distance of 1, 3, and 5. It managed to achieve average precision rate of 0.909 and average recall rate of 0.889. When 2DPCA is used in the same experiment stage, 5 principal components yielded the best result compared to 1, 2, 3, 4, and 6 principal components. It managed to achieve average precision rate of 0.804 and average recall rate of 0.796. Lastly, in the same experiment stage when SURF is used, SURF with 30 strongest features yielded the best result compared to 10, 15, 20, and 25 strongest features. It managed to achieve average precision rate of 0.943 and average recall rate of 0.938. For the classification in the this stage, Polynomial kernel of SVM with normalization performed best.

Considering all the parameters that have already been compared in the Two-class Problem experiment stage, we can conclude that SURF with 30 strongest feature and SVM with Polynomial kernel is the best parameter. SURF can be said to be suitable to differentiate between normal and abnormal because of its ability to detect edges and blob features. This is straightforward because for the normal class, it should be expected that SURF will fail to detect more blobs or features, thus images with less features will belong to normal class.

In the Nine-class Problem experiment stage, GLCM achieved weighted average precision of 0.451 and recall of 0.447. The highest class-wise precision for normal class is achieved by Fatty subclass with 0.643, while for abnormal class is achieved by Calcification subclass with 0.563.

In the same stage with 2DPCA, it managed to achieve weighted average precision of 0.386 and recall of 0.376. The highest class-wise precision for normal class is achieved by Dense subclass with 0.524, while for abnormal class is achieved by Asymmetry with 0.429.

With SURF in the same stage, it managed to achieve weighted average precision of 0.483 and recall of 0.496. The highest class-wise precision for normal class is achieved by Fatty subclass with 0.613, while for abnormal class is achieved by Calcification subclass with 0.583.

Considering all the parameters, we can conclude that SURF is the best feature extraction method in the Nine-class Problem experiment stage. From the result of abnormal subclass classification, we can see that SURF is more suitable to detect calcification than other abnormalities. This is most probably because calcification is a single distinct abnormality unlike the other 5 mass-based abnormalities. Other abnormalities might look really similar to each other, such as circumscribed mass and spiculated mass, and also to the breast tissue surrounding the mass itself. Since SURF can be said as a blob-feature detection method, mass-based abnormalities might be prone to be having similar area of detected features.

The last stage, which is the Image Retrieval experiment stage to simulate a CBIR environment, we found that GLCM achieved highest average score of 61.3 for normal class and 26.2 for abnormal class. Both scores were achieved with euclidean distance measurement. From these scores we can derive conclusion that GLCM is not suitable to differentiate abnormality, but performed adequately well to distinguish breast density subclasses in normal class.

2DPCA in the same stage managed to achieve highest average score of 46.1 for normal class and 38.3 for abnormal class. Both scores were achieved with euclidean distance measurement. From these scores we can see that 2DPCA performed less effective than GLCM to differentiate breast density subclasses in normal class, but performed better to distinguish abnormality subclasses compared to GLCM.

In the same stage, SURF managed to achieve highest average score of 66.7 for normal class and 47.6 for abnormal class. Both scores were achieved with euclidean distance

measurement. From these scores we can conclude that SURF performed better overall compared to both GLCM and 2DPCA.

The tables containing retrieval results for each methods to show the retrieval and scoring schemes are shown on Appendix A.1.

As the final conclusion, this thesis has concluded that SURF performed better for CBIR in screening mammography analysis, compared to GLCM and 2DPCA.

5.2 Suggestions for Future Works

The classification between normal and abnormal class in this study has yielded much better result than classification of abnormal subclasses. This might be the result of uncorrelated breast tissue being detected as feature and interfere the classification and retrieval processes. To overcome this, a more specialized segmentation of abnormality is deemed necessary. For example, all uncorrelated breast tissue could be filled with black colour as background area.

Specialized feature extraction for calcification-based abnormality and mass-based abnormality is also suggested, since each of them has different distinct characteristic and might help the classifier to classify them to suitable category.

Bigger database is also suggested for better retrieval result. In this thesis, in the retrieval stage in each abnormal subclasses there are less than 10 test images, which might be affecting the performance of the system.

REFERENCES

[1] **Haralick, R.M.**, **Shanmugam, K. and Dinstein, I.** (1973). Textural Features for Image Classification, *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610–621.

[2] **Society, A.C.** (2015). *Breast Cancer Facts & Figures 2015-2016*, American Cancer Society, Inc., Atlanta.

[3] **Bay, H.**, **Tuytelaars, T. and Van Gool, L.**, (2006). Surf: Speeded up robust features, Computer vision–ECCV 2006, Springer, pp.404–417.

[4] **Lodder, S.** (2009). Single-trial classification of an EEG-based brain computer interface using the wavelet packet decomposition and cepstral analysis, *Ph.D. thesis*, Department of Electrical and Electronic Engineering, University of Stellenbosch.

[5] **Cdc.gov**, (2016), CDC - Global Cancer Statistics, `http://www.cdc.gov/cancer/international/statistics.htm`.

[6] **Weedon-Fekjær, H.**, **Romundstad, P.R. and Vatten, L.J.** (2014). Modern mammography screening and breast cancer mortality: population study, *BmJ*, *348*, g3701.

[7] **Cheng, H.D.**, **Cai, X.**, **Chen, X.**, **Hu, L. and Lou, X.** (2003). Computer-aided detection and classification of microcalcifications in mammograms: a survey, *Pattern recognition*, *36*(12), 2967–2991.

[8] **Cheng, H.**, **Shi, X.**, **Min, R.**, **Hu, L.**, **Cai, X. and Du, H.** (2006). Approaches for automated detection and classification of masses in mammograms, *Pattern recognition*, *39*(4), 646–668.

[9] **Jian, W.**, **Sun, X. and Luo, S.** (2012). Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform, *Biomedical engineering online*, *11*(1), 96.

[10] **Smolkin, J.C. and M Margaret Kemeny, M.** (2001). *Reduce Your Breast Cancer Risks*, iUniverse.

[11] **Mushlin, A.I.**, **Kouides, R.W. and Shapiro, D.E.** (1998). Estimating the accuracy of screening mammography: a meta-analysis, *American journal of preventive medicine*, *14*(2), 143–153.

[12] **Zhang, P.**, **Verma, B. and Kumar, K.** (2004). A neural-genetic algorithm for feature selection and breast abnormality classification in digital mammography, *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 3, IEEE, pp.2303–2308.

[13] **Wells, J.**, **Ong, G.J. and Bastian, H.** (1998). Mammography and the politics of randomised controlled trials/Commentaries, *British Medical Journal*, *317*(7167), 1224.

[14] **Lee, R. and Kim, H.** (2008). *Computer and Information Science*, Studies in Computational Intelligence, Springer, `https://books.google.com.tr/books?id=7g3hoQAIUR8C`.

[15] **Wei, C.H.**, **Li, Y. and Huang, P.J.** (2011). Mammogram retrieval through machine learning within BI-RADS standards, *Journal of biomedical informatics*, *44*(4), 607–614.

[16] **Smeulders, A.W.**, **Worring, M.**, **Santini, S.**, **Gupta, A. and Jain, R.** (2000). Content-based image retrieval at the end of the early years, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(12), 1349–1380.

[17] **El-Naqa, I.**, **Yang, Y.**, **Galatsanos, N.**, **Nishikawa, R. and 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.

[18] **El-Naqa, I.**, **Yang, Y.**, **Galatsanos, N. and Wernick, M.** Content-based image retrieval for digital mammography, *Proceedings. International Conference on Image Processing*.

[19] **Wei, C.H. and Li, C.T.** (2005). A general framework for content-based medical image retrieval with its application to mammograms, *Proc. SPIE 5748, Medical Imaging 2005: PACS and Imaging Informatics*.

[20] **Lan, Y.**, **Ren, H.**, **Zhang, Y. and Yu, H.** (2012). Similarity in Mammography CAD Using CBIR Approach: A Validation Study, *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*.

[21] **Ren, J.** (2012). ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging, *Knowledge-Based Systems*, *26*, 144–153.

[22] **Jiang, M.**, **Zhang, S. and Metaxas, D.N.** (2014). Detection of Mammographic Masses by Content-Based Image Retrieval, *Machine Learning in Medical Imaging*, 33–41.

[23] **Jiang, M.**, **Zhang, S.**, **Li, H. and Metaxas, D.N.** (2015). Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval, *IEEE Transactions on Biomedical Engineering*, *62*(2), 783–792.

[24] **Wiesmüller, S. and Chandy, A.** (2009). Content based mammogram retrieval using Gray Level Aura Matrix, *Singaporean Journal Scientific Research*.

[25] **Honda, M.O.**, **de Azevedo Marques, P.M. and Rodrigues, J.A.H.** (2003). Content-based image retrieval in mammography: using texture features for correlation with BI-RADS categories, *Digital Mammography*, 231–233.

[26] **RaghaDeepthiLoka, V. and Putheti, S.** (2013). Classification of Normal, Benign and Malignant Tissues using Fuzzy Texton and Support Vector Machine in Mammographic Images, *International Journal of Computer Applications*, *82*(15), 36–39.

[27] **Kinoshita, S.K.**, **de Azevedo-Marques, P.M.**, **Pereira, R.R.**, **Rodrigues, J.A.H. and Rangayyan, R.M.** (2007). Content-based Retrieval of Mammograms Using Visual Features Related to Breast Density Patterns, *Journal of Digital Imaging*, *20*(2), 172–190.

[28] **Zheng, B.** (2009). Computer-Aided Diagnosis in Mammography Using Content-Based Image Retrieval Approaches: Current Status and Future Perspectives, *Algorithms*, *2*(2), 828–849.

[29] **Oliveira, J.E.**, **B.Lopes, A.P. and Deserno, T.M.** (2009). MammoSVD: a Content-Based Image Retrieval System Using a Reference Database of Mammographies.

[30] **de Oliveira, J.E.**, **Machado, A.M.**, **Chavez, G.C.**, **Lopes, A.P.B.**, **Deserno, T.M. and Araujo, A.d.A.** (2010). MammoSys: A content-based image retrieval system using breast density patterns, *Computer Methods and Programs in Biomedicine*, *99*(3), 289–297.

[31] **Deserno, T.**, **Soiron, M.**, **Oliveira, J. and Araujo, A.** (2011). Towards Computer-Aided Diagnostics of Screening Mammography Using Content-Based Image Retrieval, *2011 24th SIBGRAPI Conference on Graphics, Patterns and Images*.

[32] **Chang, C.C. and Lin, C.J.** (2011). LIBSVM: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

[33] **Brynolfsson, P.**, (2016), `http://www.mathworks.com/matlabcentral/fileexchange/55034-glcmfeatures-glcm-`.

[34] **Peng, H.**, (2016), Mutual Information computation - File Exchange - MATLAB Central, `http://www.mathworks.com/matlabcentral/fileexchange/14888-mutual-information-computation`.

[35] **Suckling, J.**, **Parker, J.**, **Dance, D.**, **Astley, S.**, **Hutt, I.**, **Boggis, C.**, **Ricketts, I.**, **Stamatakis, E.**, **Cerneaz, N.**, **Kok, S.** *et al.* (1994). The mammographic image analysis society digital mammogram database, *Exerpta Medica. International Congress Series*, volume1069, pp.375–378.

[36] **Gonzales, R.C.W.** (2002). Digital image processing 2nd edition, *New York: Addison*.

[37] **Pizer, S.M.**, (1986). Psychovisual issues in the display of medical images, Pictorial information systems in medicine, Springer, pp.211–233.

[38] **Sivaramakrishna, R.**, **Obuchowski, N.A.**, **Chilcote, W.A.**, **Cardenosa, G. and Powell, K.A.** (2000). Comparing the performance of mammographic enhancement algorithms: a preference study, *American Journal of Roentgenology*, *175*(1), 45–51.

[39] **Pisano, E.D.**, **Zong, S.**, **Hemminger, B.M.**, **DeLuca, M.**, **Johnston, R.E.**, **Muller, K.**, **Braeuning, M.P. and Pizer, S.M.** (1998). Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms, *Journal of Digital Imaging*, *11*(4), 193–200.

[40] **Sundaram, M.**, **Ramar, K.**, **Arumugam, N. and Prabin, G.** (2011). Histogram Modified Local Contrast Enhancement for mammogram images, *Applied Soft Computing*, *11*(8), 5809–5816.

[41] **Chandy, D.A.**, **Johnson, J.S. and Selvan, S.E.** (2014). Texture feature extraction using gray level statistical matrix for content-based mammogram retrieval, *Multimedia tools and applications*, *72*(2), 2011–2024.

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

[43] **Duda, R.O.**, **Hart, P.E. and Stork, D.G.** (2012). *Pattern classification*, John Wiley & Sons.

[44] **Deserno, T.M.**, **Soiron, M.**, **de Oliveira, J.E. and Araújo, A.d.A.** (2012). Computer-aided diagnostics of screening mammography using content-based image retrieval, *SPIE Medical Imaging*, International Society for Optics and Photonics, pp.831527–831527.

[45] **Insalaco, M.**, **Bruno, A.**, **Farruggia, A.**, **Vitabile, S. and Ardizzone, E.** An unsupervised method for suspicious regions detection in mammogram images.

[46] **Gunn, S.R.** *et al.* (1998). Support vector machines for classification and regression, *ISIS technical report*, *14*.

[47] **Shawe-Taylor, J. and Cristianini, N.** (2004). *Kernel methods for pattern analysis*, Cambridge university press.

[48] **Nocedal, J. and Wright, S.** (2006). *Numerical optimization*, Springer Science & Business Media.

[49] **Vladimir, V.N. and Vapnik, V.**, (1995), The nature of statistical learning theory.

[50] **Tsochatzidis, L.**, **Zagoris, K.**, **Savelonas, M.**, **Papamarkos, N.**, **Pratikakis, I.**, **Arikidis, N. and Costaridou, L.** (2014). Microcalcification oriented content-based mammogram retrieval for breast cancer diagnosis, *Imaging Systems and Techniques (IST), 2014 IEEE International Conference on*, IEEE, pp.257–262.

[51] **Dawy, Z.**, **Hagenauer, J.**, **Hanus, P. and Mueller, J.C.** (2005). Mutual information based distance measures for classification and content recognition with applications to genetics, *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, volume 2, IEEE, pp.820–824.

APPENDICES

APPENDIX A : Scoring table for image retrieval stage with GLCM
APPENDIX B : Scoring table for image retrieval stage with 2DPCA
APPENDIX C : Scoring table for image retrieval stage with SURF

APPENDIX A

Table A.1 : Scoring table for retrieval of Normal, Fatty subclass with GLCM.

NORMAL FATTY TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	4	6	40
2	Normal	4	6	40
3	Normal	7	3	70
4	Normal	5	5	50
5	Normal	3	7	30
6	Normal	6	4	60
7	Normal	7	3	70
8	Normal	7	3	70
9	Normal	6	4	60
Average		5.44	4.56	54.44
Std.dev		1.51	1.51	15.09

Table A.2 : Scoring table for retrieval of Normal, Dense subclass with GLCM.

NORMAL DENSE TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	6	4	60
2	Normal	6	4	60
3	Normal	7	3	70
4	Normal	5	5	50
5	Normal	7	3	70
6	Normal	8	2	80
7	Normal	3	7	30
8	Normal	4	6	40
9	Normal	10	0	100
10	Normal	4	6	40
11	Normal	5	5	50
Average		5.91	4.09	59.09
Std.dev		2.02	2.02	20.20

Table A.3 : Scoring table for retrieval of Normal, Fatty-glandular subclass with GLCM.

NORMAL GLANDULAR TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	7	3	70
2	Normal	10	0	100
3	Normal	8	2	80
4	Normal	6	4	60
5	Normal	7	3	70
6	Normal	8	2	80
7	Normal	8	2	80
8	Normal	6	4	60
9	Normal	6	4	60
10	Normal	4	6	40
11	Normal	6	4	60
12	Normal	8	2	80
13	Normal	8	2	80
Average		7.08	2.92	70.77
Std.dev		1.50	1.50	15.78

Table A.4 : Scoring table for retrieval of Abnormal, Calcification subclass with GLCM.

ABNORMAL CALC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	4	6	40
2	Abnormal	4	6	40
3	Abnormal	5	5	50
4	Abnormal	5	5	50
5	Normal	0	10	0
6	Normal	0	10	0
7	Abnormal	6	4	60
Average		3.43	6.57	34.29
Std.dev		2.44	2.44	24.40

Table A.5 : Scoring table for retrieval of Abnormal, Circumscribed subclass with GLCM.

ABNORMAL CIRC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	3	7	30
2	Normal	0	10	0
3	Normal	0	10	0
4	Abnormal	6	4	60
5	Abnormal	4	6	40
6	Abnormal	4	6	40
7	Normal	0	10	0
8	Abnormal	3	7	30
Average		2.50	7.50	25.00
Std.dev		2.27	2.27	22.68

Table A.6 : Scoring table for retrieval of Abnormal, Spiculated subclass with GLCM.

ABNORMAL SPIC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Normal	0	10	0
2	Normal	0	10	0
3	Abnormal	0	10	0
4	Abnormal	4	6	40
5	Abnormal	4	6	40
Average		1.60	8.40	16.00
Std.dev		2.19	2.19	21.91

Table A.7 : Scoring table for retrieval of Abnormal, Ill-defined subclass with GLCM

ABNORMAL MISC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	3	7	30
2	Abnormal	1	9	10
3	Abnormal	3	7	30
4	Abnormal	2	8	20
Average		2.25	7.75	22.50
Std.dev		0.96	0.96	9.57

Table A.8 : Scoring table for retrieval of Abnormal, Architectural distortion subclass with GLCM.

ABNORMAL ARCH TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Normal	0	10	0
2	Abnormal	4	6	40
3	Abnormal	4	6	40
4	Normal	0	10	0
5	Abnormal	5	5	50
Average		2.60	7.40	26.00
Std.dev		2.41	2.41	24.08

Table A.9 : Scoring table for retrieval of Abnormal, Asymmetry subclass with GLCM.

ABNORMAL ASYM TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	3	7	30
2	Abnormal	2	8	20
3	Abnormal	2	8	20
4	Normal	0	10	0
Average		1.75	8.25	17.50
Std.dev		1.26	1.26	12.58

APPENDIX B

Table B.1 : Scoring table for retrieval of Normal, Fatty subclass with 2DPCA.

NORMAL FATTY TEST	**Classified as**	**Top 10 Result**		**Score**
Query		Same Density	Different Density	
1	Normal	3	7	30
2	Normal	6	4	60
3	Abnormal	0	10	0
4	Normal	5	5	50
5	Normal	4	6	40
6	Normal	6	4	60
7	Abnormal	0	10	0
8	Abnormal	0	10	0
9	Normal	7	3	70
Average		3.44	6.56	34.44
Std.dev		2.83	2.83	28.33

Table B.2 : Scoring table for retrieval of Normal, Dense subclass with 2DPCA.

NORMAL DENSE TEST	**Classified as**	**Top 10 Result**		**Score**
Query		Same Density	Different Density	
1	Normal	6	4	60
2	Normal	4	6	40
3	Normal	4	6	40
4	Normal	4	6	40
5	Abnormal	0	10	0
6	Normal	5	5	50
7	Normal	3	7	30
8	Normal	6	4	60
9	Normal	3	7	30
10	Normal	7	3	70
11	Normal	7	3	70
Average		4.45	5.55	44.54
Std.dev		2.07	2.07	20.70

Table B.3 : Scoring table for retrieval of Normal, Fatty-glandular subclass with 2DPCA.

NORMAL GLANDULAR TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	4	6	40
2	Normal	3	7	30
3	Normal	3	7	30
4	Normal	5	5	50
5	Normal	2	8	20
6	Normal	4	6	40
7	Normal	4	6	40
8	Abnormal	0	10	0
9	Normal	3	7	30
10	Normal	4	6	40
11	Abnormal	0	10	0
12	Normal	5	5	50
13	Normal	4	6	40
Average		3.15	6.85	31.54
Std.dev		1.63	1.63	16.40

Table B.4 : Scoring table for retrieval of Abnormal, Calcification subclass with 2DPCA.

ABNORMAL CALC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	5	5	50
2	Abnormal	3	7	30
3	Normal	0	10	0
4	Abnormal	5	5	50
5	Abnormal	5	5	50
6	Abnormal	4	6	40
7	Abnormal	5	5	50
Average		3.86	6.14	38.57
Std.dev		1.86	1.86	18.64

Table B.5 : Scoring table for retrieval of Abnormal, Circumscribed subclass with 2DPCA.

ABNORMAL CIRC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	3	7	30
2	Normal	0	10	0
3	Abnormal	6	4	60
4	Abnormal	6	4	60
5	Abnormal	6	4	60
6	Abnormal	4	6	40
7	Normal	0	10	0
8	Abnormal	5	5	50
Average		3.75	6.25	37.50
Std.dev		2.55	2.55	25.50

Table B.6 : Scoring table for retrieval of Abnormal, Spiculated subclass with 2DPCA.

ABNORMAL SPIC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	5	5	50
2	Normal	0	10	0
3	Abnormal	0	10	0
4	Abnormal	4	6	40
5	Abnormal	4	6	40
Average		2.60	7.40	26.00
Std.dev		2.41	2.41	24.08

Table B.7 : Scoring table for retrieval of Abnormal, Ill-defined subclass with 2DPCA.

ABNORMAL MISC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Normal	0	10	0
2	Abnormal	3	7	30
3	Abnormal	1	9	10
4	Abnormal	4	6	40
Average		2.00	8.00	20.00
Std.dev		1.83	1.83	18.26

Table B.8 : Scoring table for retrieval of Abnormal, Architectural distortion subclass with 2DPCA.

ABNORMAL ARCH TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	2	8	20
2	Abnormal	5	5	50
3	Abnormal	4	6	40
4	Abnormal	5	5	50
5	Abnormal	5	5	50
Average		4.20	5.80	42.00
Std.dev		1.30	1.30	13.04

Table B.9 : Scoring table for retrieval of Abnormal, Asymmetry subclass with 2DPCA.

ABNORMAL ASYM TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	4	6	40
2	Normal	0	10	0
3	Abnormal	2	8	20
4	Abnormal	4	6	40
Average		2.50	7.50	25.00
Std.dev		1.91	1.91	19.15

APPENDIX C

Table C.1 : Scoring table for retrieval of Normal, Fatty subclass with SURF.

NORMAL FATTY TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	4	6	40
2	Normal	7	3	70
3	Normal	7	3	70
4	Normal	5	5	50
5	Normal	4	6	40
6	Normal	10	0	100
7	Normal	3	7	30
8	Normal	7	3	70
9	Normal	7	3	70
Average		6.00	4.00	60.00
Std.dev		2.18	2.18	21.79

Table C.2 : Scoring table for retrieval of Normal, Dense subclass with SURF.

NORMAL DENSE TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	5	5	50
2	Normal	6	4	60
3	Normal	4	6	40
4	Normal	5	5	50
5	Normal	5	5	50
6	Normal	4	6	40
7	Normal	6	4	60
8	Normal	4	6	40
9	Normal	5	5	50
10	Normal	5	5	50
11	Normal	7	3	70
Average		5.09	4.91	50.90
Std.dev		0.94	0.94	9.40

Table C.3 : Scoring table for retrieval of Normal, Fatty-glandular subclass with SURF.

NORMAL GLANDULAR TEST	Classified as	Top 10 Result		Score
Query		Same Density	Different Density	
1	Normal	8	2	80
2	Normal	7	3	70
3	Normal	8	2	80
4	Normal	8	2	80
5	Normal	10	0	100
6	Normal	7	3	70
7	Normal	5	5	50
8	Normal	5	5	50
9	Normal	10	0	100
10	Normal	7	3	70
11	Normal	7	3	70
12	Normal	7	3	70
13	Normal	7	3	70
Average		7.38	2.62	73.85
Std.dev		1.50	1.50	16.35

Table C.4 : Scoring table for retrieval of Abnormal, Calcification subclass with SURF.

ABNORMAL CALC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	6	4	60
2	Abnormal	6	4	60
3	Abnormal	8	2	80
4	Abnormal	5	5	50
5	Abnormal	8	2	80
6	Abnormal	10	0	100
7	Abnormal	4	6	40
Average		6.71	3.29	67.14
Std.dev		2.06	2.06	20.59

Table C.5 : Scoring table for retrieval of Abnormal, Circumscribed subclass with SURF.

ABNORMAL CIRC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	5	5	50
2	Abnormal	5	5	50
3	Abnormal	9	1	90
4	Abnormal	4	6	40
5	Abnormal	6	4	60
6	Normal	0	10	0
7	Normal	0	10	0
8	Abnormal	6	4	60
Average		4.38	5.63	43.75
Std.dev		3.07	3.07	30.68

Table C.6 : Scoring table for retrieval of Abnormal, Spiculated subclass with SURF.

ABNORMAL SPIC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	5	5	50
2	Abnormal	4	6	40
3	Abnormal	4	6	40
4	Abnormal	4	6	40
5	Abnormal	4	6	40
Average		4.20	5.80	42.00
Std.dev		0.45	0.45	4.47

Table C.7 : Scoring table for retrieval of Abnormal, Ill-defined subclass with SURF.

ABNORMAL MISC TEST	Classified as	Top 10 Result		Score
Query		Same Abnormality	Different Abnormality	
1	Abnormal	3	7	30
2	Abnormal	3	7	30
3	Abnormal	5	5	50
4	Normal	0	10	0
Average		2.75	7.25	27.50
Std.dev		2.06	2.06	20.62

Table C.8 : Scoring table for retrieval of Abnormal, Ill-defined subclass with SURF.

ABNORMAL ARCH TEST	**Classified as**	**Top 10 Result**		**Score**
Query		Same Abnormality	Different Abnormality	
1	Abnormal	7	3	70
2	Abnormal	5	5	50
3	Abnormal	7	3	70
4	Abnormal	7	3	70
5	Abnormal	6	4	60
Average		6.40	3.60	64.00
Std.dev		0.89	0.89	8.94

Table C.9 : Scoring table for retrieval of Abnormal, Asymmetry subclass with SURF.

ABNORMAL ASYM TEST	**Classified as**	**Top 10 Result**		**Score**
Query		Same Abnormality	Different Abnormality	
1	Abnormal	2	8	20
2	Normal	0	10	0
3	Abnormal	0	10	0
4	Abnormal	3	7	30
Average		1.25	8.75	12.50
Std.dev		1.50	1.50	15.00

CURRICULUM VITAE

Name Surname: Nabila Sabatini Purwadi

Place and Date of Birth: Wonogiri, 5 December 1990

Adress: Bekasi, Indonesia

E-Mail: nabilasabatini@gmail.com

B.Sc.: Telecommunication Engineering from Telkom Institute of Technology, Indonesia

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- Purwadi N.S., Atay H.T., Türkeli S., 2016: Assessment of Content-Based Image Retrieval Approaches for Mammography Based on Breast Density Patterns. *Health-Exploring Complexity: An Interdisciplinary Systems Approach, Medical Informatics Europe*, 28 August-2 September 2016 Munich, Germany.

- Purwadi N.S., Akkoca B.S., Türkeli S., 2016: Tarama Mamografi Analizi için İçerik Tabanlı Görüntü Erişim Sistemi Yaklaşımları Değerlendirme ve İyileştirme. *24. IEEE Sinyal Işleme ve Iletişim Uygulamaları Kutultayı.*, 16-19 May 2016 Zonguldak, Turkey.